Stephen N. Lyle

Acceleration in Relativity

A Critical Introduction

April 19, 2012

Springer

Preface

An inertially moving observer in the flat spacetime of special relativity has a natural choice of coordinates adapted to her worldline which makes things look simpler than they would for any other choice. This is ultimately because all our field theories of physics which govern whatever is being measured and whatever is being used to measure them have the Lorentz symmetry.

Likewise for a freely falling observer in any curved spacetime of general relativity, certain coordinates adapted to that worldline will make things look simpler than any other coordinates, at least to a good approximation, this time because, when we carry our field theories of physics over from their flat spacetime formulation to the curved spacetime, we do this *in principle* in such a way that those theories have local Lorentz symmetry, and precisely so that freely falling observers will be privileged in this way.

All this is based on the assumption that the relativity theories are good theories, and that we should build our field theories of matter in this way. But what happens for an accelerating observer in a flat spacetime, or an observer who is not freely falling in a curved spacetime? Is there any natural choice of coordinate frame for such a person, relative to which things look simpler?

It has always surprised me the confidence with which quantities expressed relative to non-inertial coordinate frames are interpreted physically as though we were just trying to understand those quantities using inertial coordinate systems. And the way certain frames are clearly presented as though they were somehow the 'natural' frame to use, but with no attempt to justify this presumption, and in particular without any reference to the operational procedures that might be adopted by accelerating observers to carry out their measurements.

After all, general relativity (GR) reminds us that coordinates are just coordinates. Of course, that does not mean that it is not useful to describe coordinate frames that are somehow adapted to the motion of an accelerating observer. And indeed this book is intended as a didactic introduction to such coordinate systems, not to mention non-holonomic frames such as tetrad fields and a lot of related mathematical machinery for describing the motion of observers or continuous fluids in the context of the relativity theories.

But the tone is critical throughout, particularly with regard to the ubiquitous notion of observers. The real target of the book, particularly in the later chapters which exemplify some of the problems, is the interface between the mathematical theory and the real world of measurements and what we like to think of as *physical* understanding.

This is a book for the undergraduate, or indeed anyone, who has been through a first course on general relativity and found some of the claims hard to integrate into their world picture. It is for someone who is left wondering sometimes whether there is not perhaps some problem of careless wording or even a real problem of physical understanding, since the two often go together. And it is for the kind of person whose interest in physics does not stop at the elegant formulation but really wants to know what the underlying logic might be, or why it might work in practice. Only a superficial acquaintance with general relativity is assumed.

Chapter 11 summarises the long-discussed question of whether stationary electric charges in static spacetimes can radiate EM energy, the subject of a book published in 2008 [30]. The chapter itself is basically a transcript of a talk given in Bad Honnef (Germany) at the conference *Problems and Developments of Classical Electrodynamics*, sponsored by the Heraeus Foundation in 2011. It was preparation for this talk and associated investigation of the Unruh effect (discussed in Chap. 14) which inspired the present book.

The two problems are related. They both consider Rindler observers in flat spacetime. These are people with eternal uniform acceleration in a straight line, who happen to be mathematically identical to freely falling people in what is usually considered to be a static homogeneous gravitational field. And it was the apparently unquestioning assumption of what they would take to be a natural coordinate frame, and equally uncritical assumption of the way they would interpret mathematical objects expressed relative to this frame, that spurred the investigations described here.

The book discusses uniform acceleration and incorporates a recent generalisation by Friedman (who was at the Bad Honnef meeting) and Scarr, so there are still developments in this area. These authors address the matter of relating theory to reality through their weak locality hypothesis, which was in turn inspired by Mashhoon's locality hypothesis, and all these things are part of an ongoing investigation. The locality hypothesis encompasses what are sometimes known as the clock and ruler hypotheses, also discussed at length in this book.

Chapter 6 attempts to extend Bell's courageous paper *How to teach special relativity* [2] to the curved spacetime context, explaining the link with the clock and ruler hypotheses and the idea recently promoted by Brown [7] that the metric field in general relativity ultimately gets its meaning from dynamical considerations regarding the kind of equipment we use to measure lengths and times. The view taken in this book is that the dynamical perspective on relativity theory championed by Brown and Pooley [6–9] is a useful addition to our understanding of these theories. This view is taken further in Chap. 9.

The ideas in [32], which are old ideas but throw some light on modern physics, in particular particle physics, are reviewed in Chap. 8. The content of this chapter was presented at a meeting in Munich entitled *Physics, Maths, and Philosophy of*

Nature, from 28 to 30 June 2011, to celebrate the sixtieth birthday of Detlef Dürr. That prepares the ground for Chap. 9, which discusses dynamical explanations for relativistic effects such as the velocity dependence of a particle's resistance to acceleration, as gauged by its inertial mass, and the contribution of binding energy to inertial effects in bound state particles, just as Brown and Pooley advocate dynamical explanations for relativistic length contraction and time dilation. However, the position in this book is not the constructivist one criticised by Norton [44]. The above arguments are just considered to provide further insight.

Acceleration is really the underlying theme of the book. Hence, the discussion in Chaps. 8–10 which describe the link between the self-force ideas advocated in [32] and the ongoing problem of mass renormalisation in quantum field theory (QFT), a problem intimately related to the phenomenon of acceleration. At the same time, Chap. 10 illustrates the mathematical machinery introduced earlier on.

The point of all this is just to show that these issues are still evolving and still relevant to teaching and research in physics.

Acknowledgements

I am indebted first and foremost to all those whose work is cited in the book, whether criticised or not, but more specifically to Domenico Giulini and Volker Perlick who organised the meeting No. 475, *Problems and Developments of Classical Electrodynamics*, WE-Heraeus Seminar, Bad Honnef in 2011, and of course to the Heraeus Foundation who sponsored it. Thanks to Robert Low and Lewis Ryder for discussions there and afterwards which have become important for this book. I am grateful to Yaakov Friedman and Tzvi Scarr for sharing their idea for generalising the notion of uniform acceleration.

Thanks also to Harvey Brown and Oliver Pooley for discussions regarding their recent resuscitation and development of the Bell approach to length contraction and time dilation, and Graham Nerlich and Vesselin Petkov for patient attempts to dissuade me of those developments. The most useful exchanges are often with those who disagree.

I am especially grateful to Detlef Dürr for taking an interest in the matters described in this book, and others presented at a conference in honour of his sixtieth birthday at the Ludwigs-Maximilians-Universität in Munich from 28 to 30 June 2011, some of which are mentioned here. Thanks therefore to the organisers of that meeting: Anna Dürr, Peter Pickl, Herbert Spohn, Stefan Teufel, and Roderich Tumulka, but also to Florian Hoffmann for sharing with me his thesis on the Unruh effect, prepared at LMU Munich. I am particularly grateful to Detlef for his staunch reminders that we need to know what we are talking about.

Finally, the most important, without whom nothing would be possible: Thérèse, Paul, and Martin.

Alzen, France *Stephen N. Lyle*
December 2011

Contents

1 Introduction and Guide 1

2 Adapted Frames in Special Relativity 11
2.1 Inertial Reference Frames 11
2.2 Non-Inertial Reference Frames 12
2.2.1 Construction 12
2.2.2 General Considerations 14
2.2.3 SE Connection Coefficients 17
2.2.4 Viewing Free Particles 19
2.2.5 Conclusion So Far 20
2.3 Toward Explicit Construction of SE Coordinates 20
2.3.1 Proper Metric and Rigid Motions 23
2.3.2 Imposing Rigidity 24
2.3.3 Fermi–Walker Transport 25
2.3.4 Semi-Euclidean Coordinates Rediscovered 27
2.3.5 SE Metric for FW Transported Coordinate Axes 28
2.3.6 SE Connection for FW Transported Coordinate Axes 30
2.3.7 SE Coordinates with Rotating Axes 32
2.3.8 SE Metric for Rotating Coordinate Axes 33
2.3.9 SE Connection for Rotating Coordinate Axes 34
2.4 Generalised Uniform Acceleration 36
2.4.1 Definition 37
2.4.2 Tensorial Nature of A and $\bar{A}$ 38
2.4.3 Nature of Generalisation 38
2.4.4 Coordinate Frame for Generalised Uniform Acceleration . 40
2.4.5 Rigidity 44
2.4.6 Reduction to FW Transport 48
2.4.7 Extent of Generalisation 53
2.4.8 Metric for Friedman–Scarr Coordinates 55
2.4.9 More about Observers at Fixed Space Coordinates 57
2.5 Velocity Transformations 62

2.6 Four-Velocity of an Object at Fixed Space Coordinates in the Accelerating Frame . . . 66
2.7 Acceleration Transformations for GUA Observers . . . 67
2.8 Summary . . . 73
2.9 Translational Uniform Acceleration . . . 76
2.10 Uniform Circular Motion as Generalised Uniform Acceleration . . . 83
2.11 General Circular Motion . . . 89
2.11.1 A First Tetrad for an Adapted Coordinate System . . . 90
2.11.2 Acceleration Matrix for the Non-Precessing Tetrad . . . 91
2.11.3 Fermi–Walker Transported Tetrad . . . 94
2.11.4 Understanding Acceleration Matrices . . . 98
2.11.5 Geodesic Coordinates for Rotating Tetrad . . . 102
2.11.6 Metric for Rotating Tetrad . . . 104
2.11.7 Metric for FW Transported Tetrad . . . 108
2.12 Range of Validity of SE Coordinates . . . 109
2.12.1 Range of Validity for General Circular Motion and Rotating Tetrad . . . 111
2.12.2 Light Cylinder for Uniform Circular Motion and Rotating Tetrad . . . 112
2.12.3 Intersecting Hyperplanes of Simultaneity for Circular Motion . . . 114
2.12.4 A Brief Conclusion for Uniform Rotation . . . 121
2.13 General Analysis of the Surface $g_{00} = 0$ . . . 122
2.13.1 Eigenvalues and Determinant of M . . . 123
2.13.2 Simplifying the Condition $g_{00} = 0$ . . . 126
2.13.3 The Case $\det M \neq 0$ . . . 127
2.13.4 The Case $\det M = 0$ . . . 131
2.13.5 The Case of Rotating Tetrad Coordinates . . . 135

3 Adapted Frames in General Relativity . . . 141
3.1 Normal Coordinates . . . 142
3.2 Normal Frames . . . 146
3.2.1 Construction . . . 146
3.2.2 Checking Properties . . . 149
3.2.3 Connection Coefficients . . . 151
3.2.4 Eliminating the Rotation Matrix . . . 152
3.2.5 Viewing Free Particles . . . 153
3.3 Locally Inertial Frames . . . 154

4 Holonomic and Non-Holonomic Frames in General Relativity . . . 155
4.1 Lie Bracket and Structure Coefficients . . . 156
4.2 Connection and Torsion . . . 158
4.3 Timelike Congruences and the Tetrad Formalism . . . 161
4.3.1 Tetrad . . . 163
4.3.2 Form Tetrad . . . 164

4.3.3 Tetrad and Projection Operators 165
4.3.4 Tetrad and Determinant of the Metric 165
4.3.5 Tetrad Components of a Tensor 165
4.3.6 Some Other Relations 166
4.3.7 Tetrad as Coordinate Frame 166
4.3.8 Propagation of Tetrads 166
4.3.9 Fermi Rotation Coefficients 168
4.3.10 Ricci Rotation Coefficients 169
4.3.11 Torsion-Free Connection 170
4.3.12 Fermi Rotation Coefficients as Ricci Rotation Coefficients 171
4.3.13 Expansion and Vorticity 171
4.3.14 Vorticity-Free Congruence 175
4.3.15 Rigid Motions in Curved Spacetimes 177
4.3.16 Rate of Strain Tensor 179
4.3.17 Stationary and Static Spacetimes 187
4.3.18 Transformation Properties of the Tetrad Quantities 189
4.4 Energy–Momentum Tensor of a Conservative Continuous Medium 193
4.4.1 Three Reference Frames 195
4.4.2 Useful Identities 197
4.4.3 Energy Density 200
4.4.4 Stress Tensor 202
4.4.5 Dynamics. Accounting for the Energy Balance 206
4.4.6 Dynamics. Accounting for the Momentum Balance 212
4.4.7 Energy–Momentum–Stress Density 215
4.4.8 Summary 216

5 Weak Locality Hypothesis 219

6 Extending Bell's Approach to General Relativity 229
6.1 De Sitter Empty Universe 233
6.2 Schwarzschild Spacetime 237
6.3 Static Homogeneous Gravitational Field 246
6.4 Taking Stock 253
6.5 Linking Theory to Measurement. Interpretation of the Metric 256

7 Mashhoon's Locality Hypothesis 263
7.1 Mashhoon on Length Measurement by Rotating Observers 268
7.2 Calculating a Length in the Rotating Tetrad Coordinate System . . . 270
7.2.1 Calculating the Length in a Constant T Hyperplane 271
7.2.2 Calculating the Length in a Constant t Hyperplane 276
7.3 Conclusion 279

8 Acceleration, Self-Force, and Inertia 281
8.1 Electromagnetic Mass and Newton's Laws 281
8.2 Classical Mass Renormalisation 286
8.3 Self-Force Calculations 287
8.4 Rewriting Newton's Second Law 293
8.5 Self-Force Effects in Modern Particle Physics 293
8.6 An Interim Conclusion 296
8.7 Passive Gravitational Mass and the Geodesic Principle 297
8.7.1 Equality of Inertial Mass and Passive Gravitational Mass . 297
8.7.2 Geodesic Principle 300
8.7.3 Equality of Inertial and Passive Gravitational Mass Revisited 302
8.7.4 Do Einstein's Equations Explain Inertia? 304
8.7.5 Geodesic Principle from Einstein's Equations: Another 'Proof' 310
8.7.6 A Brief Conclusion 317

9 Dynamical Explanations for Relativistic Effects 319
9.1 Introduction 319
9.2 FitzGerald Contraction and Time Dilation 320
9.3 Velocity Dependence of Particle Mass 325
9.4 Inertial Consequences of Binding Energy 331
9.5 Conclusion 333

10 Non-Renormalisability of Electromagnetic Self-Force. A Classical Picture 335
10.1 Overview 335
10.2 EM Fields Close to the Worldline of a Point Charge 337
10.3 Charge Dumbbell and Rigidity Assumption 341
10.4 Leading Order Term in EM Self-Force 346
10.4.1 Transverse Linear Acceleration 348
10.4.2 Longitudinal Linear Acceleration 348
10.5 Spherical Symmetry 349
10.6 Non-Relativistic Limit 351
10.7 Conclusion 352

11 Electromagnetic Radiation and the Coming of Age of the Equivalence Principle 353
11.1 The Scenario 353
11.2 The Fields 355
11.3 The Problem 356
11.4 Equivalence Principles 356
11.5 Stationary Charge in a Static Spacetime 358
11.6 Killing Vector Fields Revisited 360
11.7 Equivalence Principles Revisited 362

11.8 Which Frame? 364
11.9 Stationary Charge in a Static Spacetime Revisited 364
11.10 Interpreting Physical Quantities in Non-Inertial Frames 366
11.11 Conclusion 369

12 KVFs, SHGFs, and Uniform Gravitational Fields 371

13 Using Tetrad Fields as Accelerating Frames 381
13.1 Tetrad for Semi-Euclidean Coordinate Frame 382
13.1.1 Uniformly Accelerating Worldline at $y^1 = 0$ 383
13.1.2 Uniformly Accelerating Worldline at $y^1 = \kappa$ 384
13.1.3 SE Coordinate Frame 386
13.2 Tetrad for Another Field of Uniformly Accelerating Observers 389
13.3 Tetrads Adapted to a Given Worldline 390
13.4 Faraday Tensor Components Relative to SE Tetrad 394
13.5 A Brief Conclusion 396

14 Unruh Effect 399
14.1 Observer with Translational Uniform Acceleration 399
14.2 Observer with Circular Motion 403
14.2.1 Cylindrical Coordinates and Killing Vector Field 404
14.2.2 Metric in Cylindrical Coordinates 404
14.2.3 Connection in Cylindrical Coordinates 404
14.2.4 Killing Vector Field 405
14.2.5 Rotating Cylindrical Coordinates 407
14.2.6 Metric in Rotating Cylindrical Coordinates 407
14.2.7 Connection in Rotating Cylindrical Coordinates 408
14.2.8 Killing Vector Field Revisited 409
14.2.9 Detectors 411
14.3 A Conclusion of Sorts 415

References 417

Index 421

Acronyms

Here are some of the abbreviations used throughout the book:

AO	accelerating observer
EM	electromagnetic
FS	Friedman–Scarr
FW	Fermi–Walker
GR	general relativity
HOS	hyperplane or hypersurface of simultaneity
ICIF	instantaneously comoving inertial frame
ICIO	instantaneously comoving inertial observer
LH	locality hypothesis
LIF	locally inertial frame
MEME	minimal extension of Maxwell's equations
PGM	passive gravitational mass
QFT	quantum field theory
SE	semi-Euclidean
SEP	strong equivalence principle
SHGF	static homogeneous gravitational field
SR	special relativity
WLH	weak locality hypothesis
WEP	weak equivalence principle

Chapter 1
Introduction and Guide

There are two aspects to this book. One is a rather general description of frames of reference that might be adopted by observers with various kinds of worldline. This part is intended to be didactic, although not in a moralizing sense. The other exemplifies and discusses the physical interpretation of quantities expressed relative to such frames, and the interface between the theories of relativity and the real world of observation. This part is intended to be critical, even moralizing.

Both flat and curved spacetimes are discussed, but a lot of the book is about frames that might be used by accelerating observers of various kinds in flat spacetime. It is sometimes said that special relativity (SR) has nothing to say about acceleration because it deals only in Lorentz transformations between inertial frames, while accelerating observers are not supposed to use inertial frames to describe their environment. These more general observers are supposed to use what are often called accelerating frames.

As a matter of fact, something happens in the transition from a Newtonian to an Einsteinian world view that messes up this notion of accelerating frame. In the Newtonian world, time was the same for everyone, and any observer, however he was moving, could carry along a rigid piece of spacetime that did not deform in any way, and was not subject to any kind of mixing between the space and time dimensions. Things were very simple in the Newtonian world, and this simple world view continues to serves as an approximation under the right circumstances.

In special relativity, there are still natural frames of reference for observers moving at constant velocities relative to any inertial frame, usually called inertial observers. (One should perhaps say inertially moving observers, or perhaps non-accelerating observers.) These frames exist because all our field theories of physical phenomena have a symmetry associated with them. In fact, they are Lorentz symmetric, i.e., they look the same in some sense when formulated relative to any inertial frame. Put another way, inertial frames are the ones relative to which our field theories take on their simplest form.

But in special relativity, as in general relativity (GR), there are no acceleration symmetries. It may be that our field theories should be acceleration symmetric in some sense, but if so, we have not discovered that yet. As a consequence, there are no *natural* frames for accelerating observers in SR, or indeed in GR. When an observer is accelerating and wants to identify a set of coordinates relative to which her worldline is just the time coordinate axis, there is no general choice that will make our field theories look especially simple.

Acceleration seems to raise another problem for people who are actually accelerating when they make measurements, due to the very same fact that our fundamental theories of matter have no acceleration symmetry in the way they have constant velocity (Lorentz) symmetry. The point is that two different detectors designed to measure the same physical quantity will always deliver the same value of that physical quantity when moving inertially under the same physical conditions, but they are unlikely to do so when accelerating in any way under the same physical conditions.

These issues are carried over wholesale to the curved spacetime of general relativity. In the usual formulation of GR, for any selected event, one can always find a locally inertial frame in which the metric takes the Minkowski form at that event and changes only very slowly as one moves away from that event, a situation referred to in this book as the weak equivalence principle (WEP). Our theories of non-gravitational physical phenomena are then shipped into the curved spacetime context by saying that they must look roughly as they do in flat spacetime when expressed relative to such locally inertial frames. We refer to this ploy as the strong equivalence principle (SEP). This idea can be made precise in different ways, although one stands out for its simplicity and is generally the one chosen. SEP should be thought of as a bold hypothesis about how to do non-gravitational physics in the curved spacetime context.

So for inertially moving, i.e., non-accelerating observers, there is a natural choice of frame virtually by decree, the decree being SEP, which makes our field theories look the same, and indeed makes them look their simplest, at least to a very good approximation. These observers are identified with ones who are freely falling, i.e., not themselves subject to any non-gravitational effects. They follow geodesics of the spacetime to a good approximation. They benefit from the fact that our field theories of non-gravitational physics are locally Lorentz symmetric in GR.

But when the observer in the curved spacetime is accelerating in some way, which now means that she is being pushed off her free fall geodesic by some non-gravitational effect, there is no longer any natural choice of frame for this person, for exactly the same reason as in flat spacetime. Just as the naturalness of the locally inertial frame is shipped into GR by SEP for freely falling observers, so the lack of any natural frame is shipped into GR by SEP for accelerating observers, by decree as it were. Our field theories of physics as they stand have no acceleration symmetry.

And once again, acceleration raises another problem for people who are actually accelerating when they make measurements, due to the very same fact that our fundamental theories of matter have no acceleration symmetry in the way they have local Lorentz symmetry. The point is the same: if our theories are right, two different detectors designed to measure the same physical quantity will always deliver the same value of that physical quantity when in free fall under the same physical conditions, but they are unlikely to do so when accelerating in any way under the same physical conditions.

So does special relativity have anything to say about acceleration? And does general relativity have anything more to say about acceleration than special relativity? As we shall see there is nothing to stop us adapting coordinates in some highly convenient ways to the worldlines of accelerating observers in flat spacetime. But some would say, on the basis of a vaguely stated, or simply unstated, version of 'the' equivalence principle, that passage to an accelerating frame is somehow equivalent to introducing a gravitational field and hence brings us under the jurisdiction of general relativity.

Such claims are never spelt out very clearly because they are unfounded. We shall argue that this is a vestige of the Newtonian world view that is completely superseded by the Einsteinian one in GR. Further, we shall suggest that nothing more is gained in our understanding of accelerating observers in GR. All the problems of the accelerating observer and how she should describe and understand physical phenomena are carried over unchanged from flat to curved spacetimes. In this book we shall consider special relativity as a special case of general relativity, viewing special relativity as general relativity with no gravitational effects, i.e., with no matter or energy anywhere, and saying of course that special relativity treats gravity very differently when there is any gravity.

There are always two aspects of acceleration:

- Accelerating objects.
- Accelerating observers who would like to describe what they measure.

So far we have been referring rather to the second of these. What about the first? Of course, one can perfectly well consider accelerating test particles in SR, as in GR, but if some process is occurring in the particle, e.g., an electron orbiting a central nucleus, we do not know a priori whether that process is going just as it would for an instantaneously comoving inertial particle of the same kind, insofar as the two processes could be compared.

It seems unlikely, but a detailed calculation with the relevant theories would allow one to estimate the discrepancy, assuming those theories to be correct, or good approximations. This will be one of the themes later in the book. Presumably such calculations would show the discrepancy to be very small for most devices we use as clocks, for example, and the scale of accuracy on which no physical process fits with the theoretical proper time associated with the given worldline would be the one where we would have to admit that our relativity theory was beginning to fail.

This also brings us back to the accelerating observer who is using the clock just mentioned. If an observer is accelerating in a flat spacetime, what coordinates would this observer set up? When the acceleration is uniform, everyone seems to use the semi-Euclidean (or Rindler) coordinates described in Sects. 2.2 and 2.3, as though there were something special about them. Of course, the observer remains at the spatial origin of those coordinates, the time coordinate is the proper time of the observer, and other obvious things like that. But are those the coordinates the observer would *naturally* set up? If we are thinking about using clocks and rulers to set them up in a real world, it would seem that we do not actually know. The clock and ruler hypotheses described in Chaps. 5 and 6 merely assert that they would be in that context. Whether our actual physical clocks and rulers would fit the bill is another matter.

But would it change anything if we were working with GR in a curved spacetime here? Of course, there are nice coordinates for any timelike worldline, in which the worldline remains at the spatial origin and the time coordinate is the proper time, and so on, as explained in Chap. 3. But are those the coordinates that an observer following that worldline would *naturally* set up using clocks and rulers, or light signals? It would seem that we are in exactly the same situation as in the last paragraph.

The whole of Chap. 2 is about coordinate frames adapted to the motion of observers with various timelike worldlines in a flat spacetime, i.e., without gravity. They are *adapted* to these worldlines in the sense that the worldline $\sigma(\tau)$ of the observer satisfies the equations

$$x^0 = \tau\,, \qquad x^i = 0\,, \quad i = 1, 2, 3\,,$$

in the associated coordinate system, where τ is the proper time along the worldline. Thus a particle is at rest at the origin of its associated frame, and the coordinate time $x^0 = t$ of the associated coordinate system agrees with the proper time τ of the particle along its worldline.

We begin by considering inertial worldlines, then an arbitrary accelerating worldline. We use the convention that a Greek index runs over $\{0,1,2,3\}$, while a Latin index runs over $\{1,2,3\}$, and try to stick to this throughout the book. Likewise, in Chap. 2, we adopt the convention that timelike vectors have positive length, but elsewhere these conventions are necessarily altered when the discussion aims specifically at contributions to the literature. Since we must all consult the literature, it is good practice to be flexible about the conventions.

Section 2.2 is a somewhat qualitative introduction to the construction of semi-Euclidean coordinate frames adapted to not necessarily inertial worldlines. Then Sect. 2.3 provides a much more explicit mathematical construction based on the unusual account due to DeWitt in [14]. We describe in detail the notions of proper metric, rigidity, and Fermi–Walker transport which are relevant to these frames and obtain the corresponding form of the Minkowski metric and flat spacetime connection, noting that the latter encodes the acceleration of the observer worldline.

Section 2.4 is about a specific kind of accelerating worldline which is said to be uniformly accelerating. This idea has recently been generalised in a rather elegant way from uniform acceleration in a straight line (translational uniform acceleration TUA) to a whole class of other interesting kinds of motion, including circular motion at constant angular velocity [23]. Once again we set up semi-Euclidean coordinate systems and discuss rigidity.

Section 2.5 discusses velocity transformations for observed objects in an arbitrary semi-Euclidean frame, and Sect. 2.6 the four-velocity of objects sitting at fixed space coordinates in such frames. Section 2.7 does the same for accelerations of observed objects but only when the observer worldline is uniformly accelerating in the generalised sense.

As mentioned above, the semi-Euclidean coordinate frames are more difficult to understand physically than accelerating frames in a Newtonian world. Indeed, the very name of accelerating frame loses something of its attraction because, even though the geometry induced on the hyperplanes of simultaneity of these frames from the Minkowski metric is in fact Euclidean, justifying the name semi-Euclidean frame, objects sitting at fixed space coordinates all have different accelerations to the observer who sets those coordinates up. Unless of course the latter is moving inertially.

After a short summary of all the results concerning SE coordinate frames in Sect. 2.8, we pay specific attention to the case of translational uniform acceleration in Sect. 2.9. This is quite often mentioned in the literature, because the SE coordinates and associated form of the Minkowski metric are often taken to model a static homogeneous gravitational field (SHGF). This kind of frame is also crucial in discussions of the radiation problem, which asks whether uniformly accelerating charged particles radiate electromagnetic energy, and the Unruh effect, which claims that the quantum vacuum will feel warm if one accelerates uniformly through it. Section 2.10 exemplifies generalised uniform acceleration with the simple example of uniform circular motion.

Section 2.11 treats the case of an observer with arbitrary circular motion, i.e., moving in a circle but with possibly changing angular speed. We consider two different SE coordinate systems adapted to this motion, one that is not rigid unless the angular speed is constant in Sects. 2.11.1, 2.11.2, and 2.11.5, and a rigid one in Sect. 2.11.3. The role of the acceleration matrix is reviewed in Sect. 2.11.4 and the component forms of the Minkowski metric for the two coordinate systems are obtained in Sects. 2.11.6 and 2.11.7.

The tone is critical. It is very easy to generate a lot of mathematics in relativity theories, even when the spacetime is flat! So for the sake of form, we do indeed generate plenty of mathematics. The reader should keep asking what it is all about, especially when faced with claims that this or that choice of coordinates is somehow natural. It should be borne in mind that there are no natural choices of coordinates for accelerating observers, unless we somehow know what they would naturally measure. But we do not.

In order to know what they would naturally measure, we would need to spell out operational procedures for measuring lengths and times, say. But it is quite obvious that the results will depend on the acceleration of the observer and on the selected operational procedure, precisely because there are no acceleration symmetries in our field theories of matter. The situation is quite different for inertially moving observers, precisely because our field theories of matter are Lorentz symmetric. Of course, it could be that our field theories of matter should actually have some kind of acceleration symmetry that we have not yet discovered, although that seems unlikely.

Apart from illustrating the different possibilities that exist for setting up SE coordinate systems even in a relatively simple situation such as circular motion, the results of this section are used later to discuss Mashhoon's locality hypothesis (see Chap. 7). He illustrates his idea by describing a length assessment that might be made by an observer in circular motion. We describe another way the length might be assessed and compare the 'naturalness' of the two. More about that in a moment.

Section 2.12 discusses some rather tedious problems that arise with all SE coordinate systems, unless the observer who sets them up is moving inertially. These problems are illustrated for the case of circular motion. The most interesting of them is the intersection of hyperplanes of simultaneity (HOS) of the observer when one moves too far from the observer worldline. This kind of problem is a pure product of relativity theory. The problem is that the observer may consider herself to be simultaneous with a given event at two different values of her proper time.

Here again a critical attitude is worthwhile. Simultaneity is a somewhat arbitrary matter for accelerating observers. In the standard construction of SE coordinate systems, the observer borrows the hyperplane of simultaneity of an instantaneously comoving inertial observer. In curved spacetime, things become even more arbitrary as the physical notion of simultaneity is phased out altogether. Observers can at best take spacelike hypersurfaces orthogonal to their worldline to specify simultaneous events, but there will be lots of these hypersurfaces.

Another problem, or pseudoproblem, is that the (00) component of the semi-Euclidean form of the Minkowski metric may go to zero in some regions off the observer worldline. A completely general analysis of the region where $g_{00}=0$ is given in Sect. 2.13. Outside the surface $g_{00}=0$, this metric component may also go negative, suggesting that what was previously a time coordinate may have become a space coordinate. There is no mathematical problem with that, because coordinates are just coordinates, but it does mean that the physical interpretation must change. In particular, it means that observers can no longer remain at fixed values of what were intended to be the space coordinates. The situation is exemplified for an observer with circular motion.

As mentioned above, Chap. 3 discusses the possibilities for adapted coordinate frames in general relativity, dealing first with normal coordinates at an event (Sect. 3.1). Here we introduce the exponential map from the tangent space at an event to the manifold and basically show that a suitable choice of coordinates can reduce the metric to the Minkowski form at the chosen event and at the same time make all the connection coefficients zero there (but only there, in general), provided that the spacetime is torsion free. This bears out the mathematical counterpart of the 'principle' referred to as the weak equivalence principle in this book, showing how it is in fact built into the very fabric of the theory. It only remains to identify what such coordinates would refer to in the real world (see Sect. 6.5).

Section 3.2 shows how far one can generalise the notion of semi-Euclidean frame to curved spacetimes. Starting with an arbitrary worldline, the result is a normal frame adapted to that worldline with a whole string of nice properties, although not quite as many as could be obtained in a flat spacetime. As in the flat spacetime case, the connection coefficients cannot all be made to vanish on the observer worldline (Sect. 3.2.3). They encode its four-acceleration and also the rotation of the space triad along the worldline, which can be chosen rather arbitrarily for the purposes of the construction. The notion of Fermi–Walker (FW) transport is introduced to reduce that rotation to a minimum and hence get more of the connection coefficients equal to zero along the worldline (Sect. 3.2.4).

Section 3.3 shows what happens when the observer is in free fall, whence her worldline is a geodesic. The connection coefficients can then (and only then) be made to vanish right along the worldline. This kind of frame is called a locally inertial frame. Its theoretical existence is built into standard GR in the case where the spacetime is torsion free. What it corresponds to physically is another matter. How would we set up such coordinates in the real world?

Not all frames in relativity theory need be coordinate (or holonomic) frames. Chapter 4 is about more general frames, introducing the Lie bracket and structure coefficients for such a frame (Sect. 4.1), then discussing the connection and the torsion tensor (Sect. 4.2).

Section 4.3 discusses the tetrad formalism and congruences of timelike worldlines in a general curved spacetime, introducing the notions of expansion and vorticity, vorticity-free congruences, and stationary and static spacetimes. The link between expansion and rigid motions is established with reference to the rate of strain tensor for the congruence. Physically, this section is about the motion of continuous media with arbitrary motion. Section 4.4 applies some of the ideas to the case of a conservative continuous medium and derivation of its energy–momentum tensor, through which it will source its own gravitational effects, in particular using a label coordinate system introduced by DeWitt [14].

Chapter 5 discusses what Friedman and Scarr refer to as the weak locality hypothesis (WLH) [23]. These authors were concerned with uniformly accelerating observers in flat spacetime, in their generalised sense, and WLH refers to the relationship between the SE coordinate frames such observers might set up and lengths and times that they would measure physically. The prospects for this 'hypothesis' are discussed critically and the connection is made with the clock and ruler 'hypotheses' encountered in the literature. Scare quotes here warn the reader that there is some debate over the status of these statements. This section contains a brief discussion of how infinitesimal proper distances might be measured using light signals.

Chapter 6 reconsiders the clock and ruler hypotheses from another angle, with reference to the famous, and for some notorious, paper entitled *How to Teach Special Relativity* by Bell [2]. As part of this section, we introduce the notion of a static homogeneous gravitational field, which is usually taken to have the same metric components as would be adopted by a translationally uniformly accelerating observer using the standard SE coordinate frame in a flat spacetime (see Sect. 6.3).

The considerations in Chap. 6 lead back, in Sect. 6.5, to the problem of linking theory to measurement and interpreting the metric field in general relativity. Chapter 7 considers Mashhoon's locality hypothesis (LH), which inspired WLH, considering length measurements that might be made by rotating observers. As usual, the tone is critical. How should proper length be defined? What would naturally be measured? Do we actually know?

Although there seems to be a suggestion sometimes that LH is a fundamental principle of some kind, a standpoint that is opposed here, Mashhoon defends the following idea that is strongly supported here: if one does have an operational description of the way some measurements are being made by an accelerating observer in any specific situation, our theories of relativity together with any relevant theories of non-gravitational physics that are shipped into the curved spacetime context using the strong equivalence principle will be able to estimate the error in the observer's assuming that she is using an inertial coordinate frame. That this is indeed Mashhoon's main agenda is borne out by the fact that he cites examples of specific calculations to estimate such errors. Naturally, this does assume that all theory brought in here, such as the relativity theory itself, is good theory.

Chapter 8 introduces a general theme that is closely related to the phenomenon of acceleration and prepares the ground for Chaps. 9 and 10: spatially extended charge distributions exert forces on themselves when accelerated. This is a synopsis of [32]. It ends with a circular but nevertheless instructive proof of the equality of inertial mass and passive gravitational mass in general relativity theory, and an in-depth investigation of what is usually referred to as the geodesic principle, which states that freely falling test particles will follow geodesics. The global aim here is to investigate the nature of inertia itself, since some have claimed [6] that it might be a straight consequence of Einstein's equations for the curvature of spacetime, a thesis that is opposed here.

Chapter 9 is concerned with the undying debate about dynamical explanations for relativistic effects such as length contraction, time dilation, the velocity dependence of a particle's resistance to acceleration, and the contribution of binding energy to inertial effects in bound state particles. The claim here is that such explanations do give some physical insight, and are of course logically consistent with the usual geometrical or other explanations.

Chapter 10 has two aims. One is to illustrate a practical application of rigidity assumptions in electromagnetic self-force calculations, in particular using the earlier discussion of label coordinate systems and Fermi–Walker transport. The other is to show the relevance of acceleration to the problem of classical mass renormalisation, as first revealed by Dirac in 1938 [16]. This section extends what was said in Chap. 8, but a full discussion of this issue can be found in [32].

Chapter 11 is concerned with the radiation problem, discussed in the literature for over 100 years now. Several problems are tangled up together, but they all concern the way we interpret physical quantities expressed relative to non-inertial reference frames: Do uniformly accelerating charged particles in flat spacetime radiate EM energy? Do 'stationary' charges in static homogeneous gravitational fields, or more generally in static gravitational fields, radiate EM energy? The scare quotes warn the reader that stationarity can be interpreted in different ways. Regarding the second question, we are forced to ask what we mean by 'the' equivalence principle.

The notion of Killing vector field comes to the fore here (Sect. 11.6). It is common practice to use Killing vector fields to construct conserved vector fields in conjunction with an energy–momentum tensor, e.g., the energy–momentum tensor for some EM fields. This provides a redefinition of energy for accelerating observers. Of course, energy is a frame-dependent concept in the flat spacetime of special relativity. We know how to transform the energy of a thing from inertial frame to another by carrying out a Lorentz transformation of an energy–momentum four-vector, and this delivers a different energy for the thing in each inertial frame. But how should we transform it, or redefine it, in non-inertial frames? How should an accelerating observer define energy? What energy would be measured by an accelerating observer using standard techniques? As a vector, an energy–momentum four-vector can be represented relative to any coordinate or other frame, but here we are suggesting a different definition which favours the idea that the relevant quantity should be a conserved quantity. What is then the physical significance of such a pseudo-energy? This question never seems to be actively addressed.

Chapter 12 picks up on the notion of static homogeneous gravitational field (SHGF) and the idea that the flow of a timelike Killing vector field somehow specifies a state of rest in any curved spacetime that happens to have such a symmetry of the metric. The claim here is always that the metric would 'look the same' to someone following such a flow curve, because the Lie derivative of the metric is zero along such a curve. This provides a kind of geometric notion of being at rest which is questioned here.

Chapter 13 examines the possibilities for describing the electromagnetic fields generated by a charged particle with various kinds of motion in flat spacetime, but using tetrad frames rather than coordinate frames. The subject is a recent suggestion by Maluf and Ulhoa [34] whereby they claim to show that a uniformly accelerating charged particle will radiate EM energy by expressing the Faraday tensor relative to such a tetrad field. The problem is that one can adapt many different tetrad fields to such a worldline and the EM fields will look different mathematically depending on which is chosen. This does not help in any way with the problem of physical interpretation because it does not try to relate operational descriptions of actual measurements in the real world to the mathematical machinery.

The last chapter is about the Unruh effect, giving an overview of the aspects relevant to this book, viz., those aspects depending on an interpretation of some physical quantities when they are expressed relative to a non-inertial coordinate frame. The view taken here is that these investigations reveal a very interesting aspect of the quantum field theoretical (QFT) vacuum, but that they teach us something even more interesting about the phenomenon of acceleration. And this QFT vacuum result so much discussed under the heading of the Unruh effect may well raise other problems for understanding a physical scenario if we allow accelerating observers to redefine things like energy. The Unruh–DeWitt detector (pointlike, with two energy levels, and interacting linearly with the field) will not excite when it moves at constant velocity through the usual QFT vacuum (in flat spacetime), but it will whenever it accelerates. The question there is whether it is useful to try to set up an explanatory particle picture of that for an observer moving with the detector.

A very large part of the discussion about the Unruh effect aims to do just that. However, it is only possible for a very small subset of accelerations, viz., eternal uniform acceleration and circular motion at constant angular speed, precisely the ones where the observer is following the flow line of a KVF. A lot of time and effort is being expended on this kind of particle interpretation, so we ought to make sure it is really worthwhile and not just a mathematician's play area.

In summary, a large part of the present book is intended to be didactic, introducing the mathematical machinery used to describe spacetime in the relativity theories. So it should serve as a good mathematical introduction to all that. But it is all too easy to generate mathematics in this context, and the other aim of the book is to convince the reader that we need to take more care over interpreting that mathematics, i.e., over relating the mathematics to real world observation, at least on the didactic or metaphysical level, of saying what we are talking about.

So the mathematical theory given here comes with a warning: whenever an observer is accelerating, and the fact is that we are always accelerating to some extent when we make measurements, we should begin with an operational description of the way we make those measurements and then see what approximations are appropriate. The fact that we can define, or redefine, well known physical quantities within the theory means nothing in itself.

Chapter 2
Adapted Frames in Special Relativity

We work here in the context of special relativity, which is to say, in a flat Minkowski space. In such a space, there are inertial coordinate systems, defined by the vanishing of the connection everywhere, and the diagonalisation of the metric into the form

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{2.1}$$

2.1 Inertial Reference Frames

An inertial frame is an inertial coordinate system adapted to a moving particle in the sense that the particle worldline is the time axis and the time coordinate its proper time. We now claim that there exists an inertial coordinate system adapted to the particle with worldline $\sigma(\tau)$, parametrised by its proper time τ, if and only if $\sigma(\tau)$ is a timelike geodesic. To see this from left to right, suppose there is such an associated coordinate system. Then $\sigma(\tau)$ satisfies the equations

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = 0 ,$$

and we know that the connection coefficients are all zero in an inertial coordinate system, so $\sigma(\tau)$ is a geodesic. The tangent to $\sigma(\tau)$ has the form $(1, 0, 0, 0)$, so $\sigma(\tau)$ is timelike.

To see the converse, suppose $\sigma(\tau)$ is a timelike geodesic. Then in an arbitrary inertial system $\{y^\mu\}$, $\sigma(\tau)$ satisfies

$$\frac{\mathrm{d}^2 y^\mu}{\mathrm{d}\tau^2} = 0 ,$$

and so is described by an equation

$$y^\mu(\tau) = a^\mu \tau + b^\mu ,$$

for some a^μ and b^μ. A linear transformation, in fact, a Poincaré transformation, gives us an inertial system $\{x^\mu\}$ in which $\sigma(\tau)$ satisfies

$$x^0 = \tau , \qquad x^i = 0 , \quad i = 1, 2, 3 .$$

Hence inertial frames can be associated only with particles performing inertial motion, that is, particles satisfying the equation of motion

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\nu\sigma} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} = 0 ,$$

in a general coordinate system, where Γ is the connection. Note that we adopt the view that Minkowski space is a genuine special case of general relativity, in the sense that its generally covariant formulation is valid. In other words, we are not working only in inertial frames, where the connection is zero. Indeed, we are about to consider what an accelerating frame might look like.

2.2 Non-Inertial Reference Frames

In the four dimensional, generally covariant Newtonian spacetime described in [22], it is shown how rigid Euclidean frames can be associated with non-inertial trajectories. This is a well-known result, except that Friedman shows how Newtonian spacetime can be described as a 4D space, and in a generally covariant way, which indeed reveals the main difference with the special relativistic case, viz., the connection and spatiotemporal metric can vary independently in the Newtonian world. This is not so in Minkowski spacetime, and it is a consequence that, if a Euclidean frame can be associated with a worldline, then that worldline must be inertial.

2.2.1 Construction

The best that can be done in Minkowski spacetime is to associate what we shall call a *semi-Euclidean (SE) frame* with a given timelike, not necessarily geodesic, worldline $\sigma(\tau)$. This is a coordinate system with the following properties:

1. Curves $y^i = \text{constant}$ $(i = 1, 2, 3)$ are timelike, and any curve with $y^0 = \text{constant}$ is spacelike.
2. $y^0 = \tau$ along $\sigma(\tau)$.
3. $(g_{\mu\nu}) = \mathrm{diag}\,(1, -1, -1, -1)$ along σ.

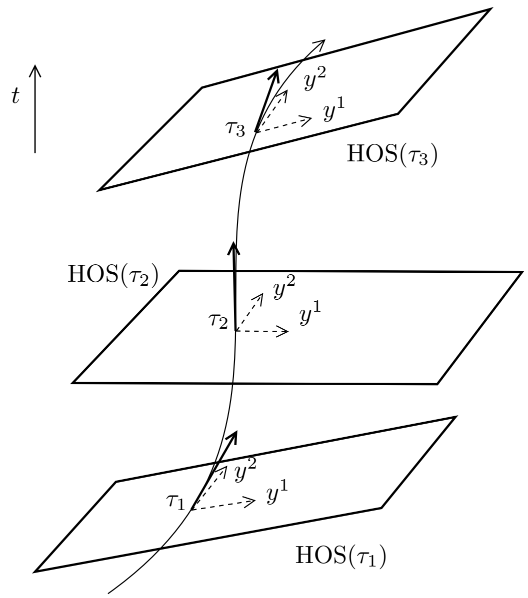

Fig. 2.1 Constructing a semi-Euclidean (SE) frame for an accelerating observer. View from an inertial frame with time coordinate t. The curve is the observer worldline. Three hyperplanes of simultaneity (HOS) are shown at three successive proper times τ_1, τ_2, and τ_3 of the observer. These hyperplanes of simultaneity are borrowed from the instantaneously comoving inertial observer, as are the coordinates y^1, y^2, and y^3 used to coordinatise them. Only two of the latter coordinates can be shown in the spacetime diagram

4. y^1, y^2, y^3 are Cartesian on every hypersurface $y^0 =$ constant, meaning that the spatial distance between (y^0, y^1, y^2, y^3) and $(y^0, y^{1\prime}, y^{2\prime}, y^{3\prime})$ is

$$\sqrt{(y^1 - y^{1\prime})^2 + (y^2 - y^{2\prime})^2 + (y^3 - y^{3\prime})^2}\ .$$

5. $\sigma(\tau)$ satisfies $y^i = 0$ for $i = 1, 2, 3$.

Such a coordinate system can be constructed by choosing, for each value of τ along σ, an inertial system $\{x^\mu\}$ whose origin coincides with $\sigma(\tau)$ and whose x^0 axis is tangent to σ at $\sigma(\tau)$ (see Fig. 2.1). We then assign each point on the spatial hypersurface $x^0 = 0$ the coordinates

$$y^0 = \tau\ , \quad y^i = x^i\ , \quad i = 1, 2, 3\ .$$

Of course, the choice of successive inertial systems must be sufficiently smooth, so that the curves $y^i =$ constant are suitably continuous and differentiable. More will be said about this later. In the meantime, note how we have used the instantaneously comoving inertial frames, and how the coordinates we adopt differ from the inertial coordinates that these would give.

We can see that property (3) on the above list, viz.,

$$(g_{\mu\nu}) = \mathrm{diag}\,(1, -1, -1, -1) \quad \text{along } \sigma \; ,$$

follows necessarily from the other properties. Firstly, it is clear that we have

$$(g_{\mu\nu}) = \begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & -1 & 0 & 0 \\ g_{20} & 0 & -1 & 0 \\ g_{30} & 0 & 0 & -1 \end{pmatrix} \; .$$

This is because, for any neighbouring points in the same plane of simultaneity as some point on the worldline, their separation is just given by the metric in the inertial coordinates of the instantaneously comoving frame for that point on the worldline. But events which are also separated by some small difference in the proper times of their associated planes of simultaneity are quite a different matter! The problem here is that values of the metric on σ are also supposed to give lengths to neighbouring points which may not be on σ.

However, we can see that, for each value of τ, any curve with $y^0 =$ constant intersecting the worldline at $\sigma(\tau)$ will be orthogonal to σ there. The point is that any such curve lies entirely within the hyperplane of simultaneity of the instantaneously comoving inertial frame at $\sigma(\tau)$ by construction. We can see the orthogonality in this frame, because g has the usual Minkowski form in this inertial frame, and its time axis is tangent to σ at $\sigma(\tau)$. Finally, orthogonality is a property which is independent of coordinates, so we have orthogonality in semi-Euclidean coordinates and $g_{0j} = 0$ for $j = 1, 2, 3$.

It remains only to show that $g_{00} = 1$ in the semi-Euclidean coordinates, but this follows because

$$g_{00} = g(\partial_\tau, \partial_\tau) = g(u, u) = 1 \; .$$

These arguments do of course use the other properties listed above for the SE coordinate system.

2.2.2 General Considerations

For a non-inertial observer, the above semi-Euclidean frame can be well-defined only on a finite neighbourhood of σ, for at finite distances from σ, the hypersurfaces $y^0 =$ constant may well intersect one another. This situation is well illustrated by the case of translational uniform acceleration discussed in detail in Sect. 2.9.

Apart from highlighting a major difference with the rigid Euclidean frames which exist in Newtonian spacetime, this is a rather intriguing idea. It means that, following the procedure described above, the accelerating observer will declare herself to be simultaneous with a certain event at two distinct proper times along her worldline. This illustrates the limited utility of the notion of simultaneity in relativistic theories.

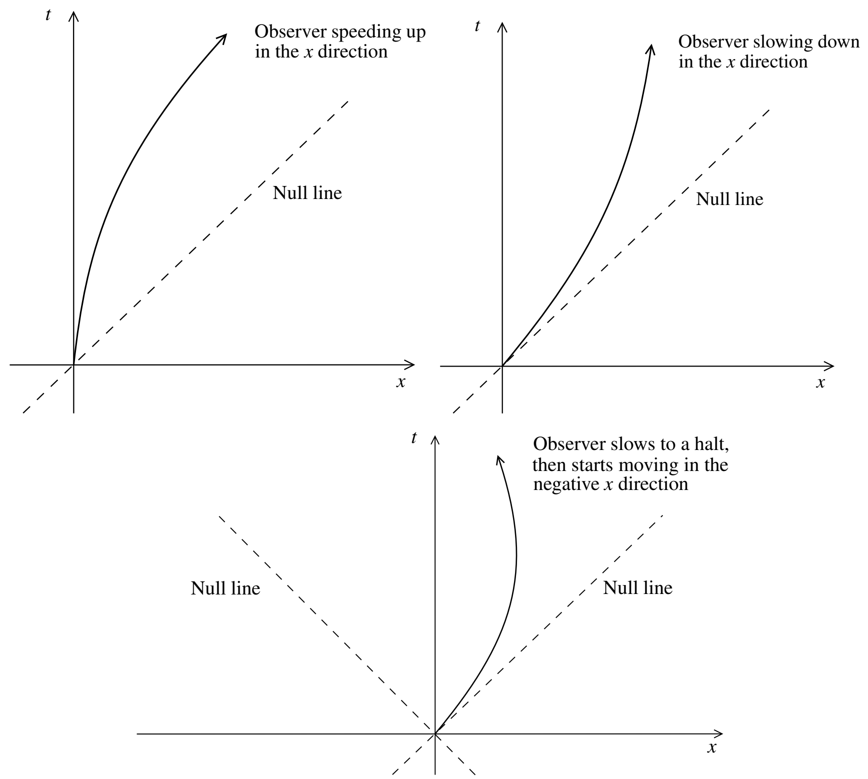

Fig. 2.2 Three different worldlines in a 2D spacetime

Note how the surfaces of simultaneity are defined in the semi-Euclidean frame by the relations $\tau =$ constant, for various constants. This is not an operational definition for the accelerating observer, for she merely borrows the findings of the simultaneously comoving inertial observer. The everyday notion of simultaneity is simply replaced by a temporal coordinate that happens to take equal values for two different events.

Let us briefly consider regions where the SE coordinates become problematic, because the construction process breaks down, viz., regions where there is no unique point on the worldline of the observer such that this observer considers a given point in those regions to be simultaneous. The existence of a point and its uniqueness are both crucial, but let us just consider here exactly how the non-uniqueness problem arises in three typical, indeed generic situations that refer to the simple case of one-dimensional motion:

- A worldline curves down toward the null line in the right quadrant of the 2D spacetime diagram, i.e., the observer is speeding up in the x direction (Fig. 2.2 top left).

- A worldline curves up and away from the null line in the 2D spacetime diagram, i.e., the observer is slowing down in the x direction (Fig. 2.2 top right).
- The worldline in the last example curves right over and curves down toward the null line in the left quadrant of the 2D spacetime diagram, i.e., the observer slows to a stop and then speeds up in the negative x direction (Fig. 2.2 bottom).

The hyperplane of simultaneity (HOS) of an observer at a given event on her worldline is found as follows (see Fig. 2.3). We draw the relevant null line at the event, i.e., the null line making the smallest angle with the worldline, and the tangent to the worldline at the event. The latter is the instantaneous time axis for that observer. Her HOS lies at the same angle from the null line but below it. We now have the following observations:

- As the worldline curves down in the first case above, the HOS swings up, so neighbouring HOSs on such a worldline would meet to the left of the worldline in the spacetime diagram (see Fig. 2.3).
- As the worldline steepens in the second case above, the HOS swings down, so neighbouring HOSs on such a worldline would meet to the right of the worldline in the spacetime diagram.
- As the worldline curves right over and begins to curve down on the left in the third case above, the intersections of HOSs will all occur to the right of the worldline.

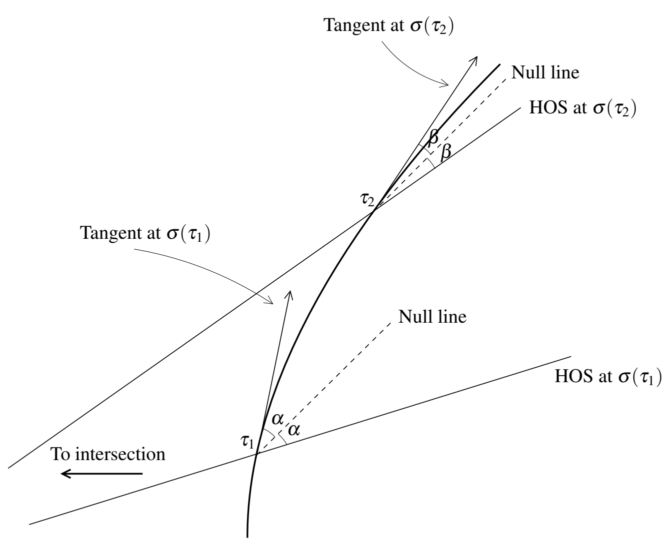

Fig. 2.3 Construction of two hyperplanes of simultaneity (HOS) at different proper times τ_1 and τ_2 along the observer worldline in a 2D spacetime. At each proper time, the tangent to the worldline lies at the same angle (α at τ_1 and β at τ_2) above the null line as the hyperplane of simultaneity below it

Note that, when there is some acceleration in the worldline, there will always be intersection of HOSs somewhere, and it turns out that this intersection is always on the opposite side of the worldline to its centre of curvature.

In the third case we consider an observer who changes direction. When this observer has the same speed but in the opposite direction at two events on her worldline, her HOSs will meet somewhere to the right if the change of speed is to the left, and vice versa. In the case of motion in a circle of radius r, for example, as considered in Sect. 2.10, this happens all the time, although here we involve a second space dimension, because the observer will have moved a distance $2r$ in that other space direction between two velocity reversals. The consequences for the HOSs are discussed in some detail there.

Another consideration here is clock synchronisation. Spatially separated clocks at rest in a semi-Euclidean frame and synchronised at one time τ_1, will not in general remain synchronised at a second time τ_2. The proper length of the worldline of a clock at rest at one point in the frame between τ_1 and τ_2 will differ from the proper length of the worldline of a clock at rest at a second point between those same two SE coordinate times. Note again that we understand 'at rest' here to mean having the same spatial coordinates in a succession of the simultaneously comoving inertial frames. This idea will be made much more precise later on.

There is sometimes a suggestion or implication that clocks sitting at fixed space coordinates in a semi-Euclidean frame are somehow accelerating with the observer, but this is a presumption, for we do not know a priori what a physical object would have to do to occupy such a sequence of events in spacetime. In any case, the difference with the rigid Euclidean frame of Newtonian spacetime is once again highlighted. This, too, will be investigated in what follows.

Another thing can go wrong when constructing the SE coordinate system in the way described. It may be that for some events E there is no event on the observer worldline such that the observer considers E to be simultaneous. This situation is illustrated in Sect. 2.9.

2.2.3 SE Connection Coefficients

A general analysis of the connection coefficients can be made along σ. They will not be zero unless σ is an inertial worldline. Recall how the four-velocity and four-acceleration are defined in the general context of 4D spacetime manifolds. The four-velocity is

$$u^{\mu} := \frac{\mathrm{d}y^{\mu}}{\mathrm{d}\tau} ,$$

the components of the tangent vector T_{σ} to the worldline, if it is parametrised by the proper time, and the four-acceleration is $\mathrm{D}_{T_{\sigma}} T_{\sigma} = \mathrm{D}_u u$, with components

$$a^{\mu} := \frac{\mathrm{d}^2 y^{\mu}}{\mathrm{d}\tau^2} + \Gamma^{\mu}_{\nu\sigma} \frac{\mathrm{d}y^{\nu}}{\mathrm{d}\tau} \frac{\mathrm{d}y^{\sigma}}{\mathrm{d}\tau} . \tag{2.2}$$

Note that $y^0 = \tau$ and $y^i = 0$ for $i = 1, 2, 3$, describes the worldline in the semi-Euclidean coordinates, so the four-velocity has components $u^\mu = (1, 0, 0, 0)$ in the $\{y^\mu\}$ coordinate system.

The four-acceleration is $a = (a^0, a^1, a^2, a^3)$ with $a^0 = 0$, since $u^2 = 1$ is constant and this implies that $u \cdot a = 0$. However, the definition (2.2) gives

$$a^\mu = \Gamma^\mu_{00} \; ,$$

so we deduce immediately that

$$\boxed{\Gamma^0_{00} = 0 \; , \qquad \Gamma^i_{00} = a^i \; , \quad i = 1, 2, 3} \; , \tag{2.3}$$

this being true right along the observer worldline.

What can be said about the other connection coefficients along σ? If we consider any spacelike geodesic $\rho(s)$, with affine parameter s, lying within a hyperplane of simultaneity $y^0 =$ constant, the curve $\rho(s)$ must satisfy

$$\frac{\mathrm{d}^2 y^\mu}{\mathrm{d}s^2} = 0 \; , \tag{2.4}$$

since y^1, y^2, y^3 are supposed to be Cartesian. Actually, some care is needed to sustain the logic of our arguments here. We have specified a way of constructing these coordinates in Fig. 2.1. Showing that they are Cartesian is equivalent to showing that the above formula holds! So, what we are about to deduce concerning the connection coefficients is exactly what we would need to show if we were to show that the coefficients are Cartesian. We may either appeal to intuition, which is clearly risky in this domain, or else provide some explicit formulas for the semi-Euclidean coordinates. In fact we shall do the latter shortly. Returning to (2.4), the consequence for the connection is that

$$\boxed{\Gamma^\mu_{ij} = 0 \; , \quad \mu = 0, 1, 2, 3, \;\; i = 1, 2, 3} \; . \tag{2.5}$$

In contrast to (2.3), this result holds everywhere, for (2.4) can be applied to any geodesic lying entirely within some plane of simultaneity $y^0 =$ constant, whether it intersects σ or not.

Therefore the only nonzero connection components are $\Gamma^i_{00} = a^i$, $\Gamma^i_{0j} = \Gamma^i_{j0}$ for $i, j = 1, 2, 3$, and $\Gamma^0_{i0} = \Gamma^0_{0i}$ for $i, j = 1, 2, 3$. To see what the remainder of these represent, note that

$$\mathrm{D}_u g = 0 \; , \tag{2.6}$$

since this is a coordinate-independent relation for any vector u, once we have started out with a metric connection [see (4.26) on p. 160]. We apply it to the unit tangent vector u, i.e., the four-velocity, at points along σ. Since

$$\mathrm{D}_u g = u^\mu g_{\nu\sigma,\mu} - u^\mu \Gamma^\alpha_{\nu\mu} g_{\alpha\sigma} - u^\mu \Gamma^\alpha_{\sigma\mu} g_{\nu\alpha} \; ,$$

and since $g_{\nu\sigma}$ is constant along the worldline, whence

$$u^{\mu} g_{\nu\sigma,\mu} = 0 \,,$$

the relation (2.6) amounts to

$$\Gamma^{\alpha}_{\nu 0} g_{\alpha\sigma} + \Gamma^{\alpha}_{\sigma 0} g_{\nu\alpha} = 0 \,.$$

It then follows that

$$\boxed{\Gamma^{j}_{0i} = -\Gamma^{i}_{0j} \,, \quad i, j = 1, 2, 3} \,,$$

and

$$\boxed{\Gamma^{i}_{00} = \Gamma^{0}_{0i} \,, \quad i = 1, 2, 3} \,,$$

these results applying only on σ.

We can summarise our results for the connection on σ by

$$\boxed{\Gamma^{i}_{00} = \Gamma^{0}_{0i} = \Gamma^{0}_{i0} = a^{i} \,, \quad i = 1, 2, 3} \,, \tag{2.7}$$

$$\boxed{\Gamma^{\mu}_{ij} = 0 \,, \quad \mu = 0, 1, 2, 3, \;\; i = 1, 2, 3} \,, \tag{2.8}$$

$$\boxed{\Gamma^{i}_{0j} = \Gamma^{i}_{j0} = \Omega^{i}{}_{j} \,, \quad i, j = 1, 2, 3} \,, \tag{2.9}$$

where $\Omega^{i}{}_{j}$ is an antisymmetric rotation matrix. We shall see later that this matrix is related to the choice of instantaneously comoving inertial frames used to construct the SE coordinate system as we move along the worldline. If we obtain these instantaneously comoving inertial frames by Fermi–Walker transporting the coordinate axes in the instantaneously comoving inertial frame chosen at some initial point on the worldline, we can make this matrix equal to zero.

2.2.4 Viewing Free Particles

We can write down the equation of motion of a free particle on σ now, assuming that it intersects the observer's worldline, and only at the unique event at which it coincides with the observer. We obtain

$$\boxed{\frac{\mathrm{d}^2 y^i}{\mathrm{d}y^{02}} + a^i + 2\Omega^{i}{}_{j} \frac{\mathrm{d}y^j}{\mathrm{d}y^0} - 2a^j \frac{\mathrm{d}y^j}{\mathrm{d}y^0} \frac{\mathrm{d}y^i}{\mathrm{d}y^0} = 0} \,, \tag{2.10}$$

where indices run over values 1, 2, 3. Before describing how this is derived, observe that it contains a term a^i which could be referred to as an inertial force, a term proportional to the semi-Euclidean 3-velocity which corresponds to a Coriolis force, and finally a relativistic correction. As mentioned at the end of the last section, we shall show that the Coriolis term can be removed by Fermi–Walker transporting the coordinate axes along the worldline.

To establish (2.10), we start with the geodesic equation for $i = 1, 2, 3$,

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}\tau'^2} + a^i \left(\frac{\mathrm{d}y^0}{\mathrm{d}\tau'}\right)^2 + 2\Omega^i{}_j \frac{\mathrm{d}y^j}{\mathrm{d}\tau'} \frac{\mathrm{d}y^0}{\mathrm{d}\tau'} = 0 ,$$

where τ' is proper time along this geodesic. If we change parameter from the affine parameter τ' to the non-affine parameter y^0, we simplify the left-hand side here, but the right-hand side is now

$$-\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2} \frac{\mathrm{d}y^i}{\mathrm{d}y^0} \Big/ \left(\frac{\mathrm{d}y^0}{\mathrm{d}\tau'}\right)^2 .$$

The zero component of the geodesic equation is

$$\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2} + \Gamma^0_{\mu\nu} \frac{\mathrm{d}y^\mu}{\mathrm{d}\tau'} \frac{\mathrm{d}y^\nu}{\mathrm{d}\tau'} = 0 ,$$

which gives, since $\Gamma^0_{00} = 0 = \Gamma^0_{ij}$, for all $i, j = 1,2,3$,

$$\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2} = -2a^j \frac{\mathrm{d}y^j}{\mathrm{d}\tau'} \frac{\mathrm{d}y^0}{\mathrm{d}\tau'} ,$$

where j is summed from 1 to 3. The result follows.

Note finally that if the acceleration a^i and the rotation $\Omega^i{}_j$ are both zero, then the semi-Euclidean frame becomes an inertial frame, and the law of motion takes the familiar form

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}y^{02}} = 0 .$$

2.2.5 Conclusion So Far

The above discussion is still rather qualitative. Despite the cogency of the arguments, it would be satisfying to obtain more explicit expressions for the SE coordinates, SE metric, and SE connection. This is what we shall now do.

2.3 Toward Explicit Construction of SE Coordinates

The discussion here was inspired by [14]. Imagine a situation with a main observer, whose worldline is given, and other observers moving rigidly with this one in a sense to be defined. The latter will eventually be sitting at fixed SE space coordinates, once we have set up this frame. We shall use the notation $x^\mu(\xi,\tau)$ for the worldlines of all these observers, where τ is their proper time and $\xi = (\xi^1,\xi^2,\xi^3) \in \mathbb{R}^3$ are labels for the observers which will eventually become SE space coordinates.

The worldline of the main observer is thus denoted by $x^\mu(0,\tau)$, where τ is her proper time and the ξ^i are all zero, so that this observer will sit at the space origin of the system, as required for an adapted coordinate system. The general idea in this construction is to extend the four-velocity of the main observer to a tetrad along this observer's worldline by finding a spacelike triad $n_i(\tau)$, $i = 1,2,3$, along the worldline with the properties

$$n_i \cdot n_j = -\delta_{ij} \;, \qquad n_i \cdot u_0 = 0 \;, \qquad u_0^2 = 1 \;, \tag{2.11}$$

where we use u_0 to denote the observer 4-velocity here.

We now decree that the worldlines of all the other observers will be given by a relation of the form

$$x^\mu(\xi,\tau) = x^\mu(0,\sigma) + \xi^i n_i{}^\mu(\sigma) \;, \tag{2.12}$$

where σ is a certain function of the ξ^i and τ to be determined. It is important to remember that, on the left of (2.12), τ is the proper time of the observer labelled by ξ. Furthermore, as mentioned above, we shall aim to arrange things so that the observers form a rigid assembly, in a sense to be specified shortly. This will be achieved by laying down a requirement on the triad $\{n_i\}_{\{i=1,2,3\}}$.

To achieve a relation like (2.12), given τ and ξ, we must find the proper time σ of the main observer $\xi = 0$ such that the point $x^\mu(\xi,\tau)$ is simultaneous with the event $x^\mu(0,\sigma)$ in the instantaneous rest frame of the observer labelled by $\xi = 0$. Assuming that a given event off the main observer worldline is simultaneous with just one event on that worldline, the latter event will be defined by orthogonality of the tangent to the main observer worldline there, i.e., its four-velocity there, and the geodesic from that event on the main observer worldline to the given event off the worldline. But that means that $x^\mu(\xi,\tau) - x^\mu(0,\sigma)$ must be a linear combination of the n_i, and that is precisely what (2.12) states to be the case.

Note that problems begin to occur in the construction if several events on the main observer worldline are simultaneous with the chosen $x^\mu(\xi,\tau)$. This becomes more probable as one moves away from the worldline of the main observer. In general, this construction will only be possible over some neighbourhood of that worldline. Yet another thing can go wrong: it may be that no event on the main observer worldline is simultaneous with the chosen $x^\mu(\xi,\tau)$. Both these things are exemplified in the case where the main observer has translational uniform acceleration, discussed in detail in Sect. 2.9.

To determine the function $\sigma(\xi^i,\tau)$, write

$$u^\mu = \dot{x}^\mu(\xi,\tau) = (u_0{}^\mu + \xi^i \dot{n}_i{}^\mu)\dot{\sigma} \;,$$

all arguments being suppressed in the final expression. Here and in what follows, it is to be understood that dots over u_0 and the n_i denote differentiation with respect to σ, while the dot over σ denotes differentiation with respect to τ.

In order to proceed further, we must expand $\dot{n}_i$ in terms of the orthonormal tetrad u_0, n_i:

$$\boxed{\dot{n}_i{}^\mu = a_{0i} u_0{}^\mu + \Omega_{ij} n_j{}^\mu} \,. \tag{2.13}$$

The coefficients a_{0i} are determined, from the identity

$$\dot{n}_i \cdot u_0 + n_i \cdot \dot{u}_0 = 0 \,,$$

to be just the components of the absolute acceleration of the observer $\xi = 0$ in her local rest frame:

$$a_{0i} = -n_i \cdot \dot{u}_0 \,, \tag{2.14}$$

and the identity

$$\dot{n}_i \cdot n_j + n_i \cdot \dot{n}_j = 0$$

tells us that Ω_{ij} is antisymmetric:

$$\Omega_{ij} = -\Omega_{ji} \,.$$

We now have

$$\boxed{u^\mu = \Big[\big(1 + \xi^i a_{0i}\big) u_0{}^\mu + \xi^i \Omega_{ij} n_j{}^\mu\Big] \dot{\sigma}} \,. \tag{2.15}$$

But

$$1 = u^2 = \Big[\big(1 + \xi^i a_{0i}\big)^2 - \xi^i \xi^j \Omega_{ik} \Omega_{jk}\Big] \dot{\sigma}^2 \,,$$

whence

$$\boxed{\dot{\sigma} = \Big[\big(1 + \xi^i a_{0i}\big)^2 - \xi^i \xi^j \Omega_{ik} \Omega_{jk}\Big]^{-1/2}} \,. \tag{2.16}$$

The right-hand side of this equation is a function solely of σ and the ξ^i. Therefore the equation may be integrated along each worldline $\xi =$ constant, subject to the boundary condition

$$\sigma(\xi, 0) = 0 \,,$$

which ensures that zero proper time $\tau = 0$ for any observer is simultaneous for the main observer with zero proper time $\sigma = 0$ for the main observer. We shall, in particular, have the necessary condition

$$\sigma(0, \tau) = \tau \,.$$

Note that the rigid set of observers must be confined to regions where

$$\big(1 + \xi^i a_{0i}\big)^2 > \xi^i \Omega_{ik} \xi^j \Omega_{jk} \quad (\geq 0) \,, \tag{2.17}$$

otherwise some of its component observers will be moving faster than light. The parameters (ξ, σ) are of course semi-Euclidean coordinates, a point to be developed below (see Sect. 2.3.4).

2.3.1 Proper Metric and Rigid Motions

We can now calculate a *proper metric* of the set of observers. Rather than the somewhat absurd idea of a rigidly moving set of observers, let us speak here of a rigidly moving medium. The component particles of the medium are labeled by three parameters ξ^i, $i = 1,2,3$, and the worldline of particle ξ is given by four functions $x^\mu(\xi,\tau)$, $\mu = 0,1,2,3$, where τ is its proper time. In general relativity, the x^μ may be arbitrary coordinates in curved spacetime, but here we assume them to be standard coordinates of some inertial frame.

If $\xi^i + \delta\xi^i$ are the labels of a neighbouring particle, its worldline is given by the functions

$$x^\mu(\xi+\delta\xi,\tau) = x^\mu(\xi,\tau) + x^\mu{}_{,i}(\xi,\tau)\delta\xi^i , \tag{2.18}$$

where the comma followed by a Latin index denotes partial differentiation with respect to the corresponding ξ. Note that the quantity $x^\mu{}_{,i}(\xi,\tau)\delta\xi^i$, representing the difference between the two sets of worldline functions, is formally a 4-vector, being basically an infinitesimal coordinate difference. However, it is not generally orthogonal to the worldline of ξ. In other words, it does not lie in the hyperplane of simultaneity of either particle.

To get such a vector one applies the projection tensor onto the instantaneous hyperplane of simultaneity:

$$P^{\mu\nu} := \eta^{\mu\nu} - \dot{x}^\mu\dot{x}^\nu ,$$

where the dot denotes partial differentiation with respect to τ. The result is

$$\delta x^\mu := P^\mu{}_\nu x^\nu{}_{,i}(\xi,\tau)\delta\xi^i = x^\mu{}_{,i}\delta\xi^i - \dot{x}^\mu\dot{x}_\nu x^\nu{}_{,i}\delta\xi^i . \tag{2.19}$$

We find that application of the projection tensor corresponds to a simple proper time shift of amount

$$\delta\tau = -\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu{}_{,i}\delta\xi^i , \tag{2.20}$$

so that

$$\delta x^\mu = x^\mu(\xi+\delta\xi,\tau+\delta\tau) - x^\mu(\xi,\tau) .$$

Indeed,

$$x^\mu(\xi+\delta\xi,\tau+\delta\tau) = x^\mu(\xi,\tau) + x^\mu{}_{,i}\delta\xi^i + \dot{x}^\mu\delta\tau ,$$

and feeding in the proposed expression for $\delta\tau$, we obtain

$$\begin{aligned}\delta x^\mu &= x^\mu(\xi,\tau) + x^\mu{}_{,i}\delta\xi^i + \dot{x}^\mu\delta\tau - x^\mu(\xi,\tau)\\ &= x^\mu{}_{,i}\delta\xi^i - \eta_{\nu\sigma}\dot{x}^\nu\dot{x}^\sigma{}_{,i}\delta\xi^i\dot{x}^\mu ,\end{aligned}$$

which is precisely δx^μ as defined in (2.19).

What can we conclude from this analysis? The two particles ξ and $\xi+\delta\xi$ appear, in the instantaneous rest frame of either, to be separated by a distance δs given by

$$(\delta s)^2 = -\delta x \cdot \delta x = -\gamma_{ij}\delta\xi^i\delta\xi^j \;, \tag{2.21}$$

where

$$\boxed{\gamma_{ij} := P_{\mu\nu}x^\mu{}_{,i}x^\nu{}_{,j}} \;. \tag{2.22}$$

This follows because

$$\delta x \cdot \delta x = P_{\mu\sigma}x^\sigma{}_{,i}\delta\xi^i P^\mu{}_\nu x^\nu{}_{,j}\delta\xi^j \;,$$

and

$$\begin{aligned} P_{\mu\sigma}P^\mu{}_\nu &= (\eta_{\mu\sigma} - \dot{x}_\mu\dot{x}_\sigma)(\delta^\mu{}_\nu - \dot{x}^\mu\dot{x}_\nu) \\ &= \eta_{\nu\sigma} - \dot{x}_\nu\dot{x}_\sigma - \dot{x}_\sigma\dot{x}_\nu + (\dot{x}_\mu\dot{x}^\mu)\dot{x}_\sigma\dot{x}_\nu \\ &= \eta_{\nu\sigma} - \dot{x}_\nu\dot{x}_\sigma = P_{\nu\sigma} \;, \end{aligned}$$

whence

$$\delta x \cdot \delta x = P_{\nu\sigma}x^\sigma{}_{,i}x^\nu{}_{,j}\delta\xi^i\delta\xi^j = \gamma_{ij}\delta\xi^i\delta\xi^j \;,$$

for the given γ_{ij}, as claimed. We shall call the quantity γ_{ij} the proper metric of the medium.

The point about γ_{ij} is that the two particles or observers labelled by ξ and $\xi + \delta\xi$ appear in the instantaneous rest frame of either to be separated by a proper distance δs as they would measure it given by

$$\boxed{(\delta s)^2 = -\gamma_{ij}\delta\xi^i\delta\xi^j} \;. \tag{2.23}$$

We shall say that the set of particles or observers undergoes rigid motion if and only if the proper metric is everywhere independent of τ. This is expressed by

$$\boxed{\dot{\gamma}_{ij} = 0} \;. \tag{2.24}$$

Under rigid motion, the instantaneous separation distance between any pair of neighbouring observers is constant in time as they would see it in an instantaneously comoving inertial frame. This kind of relativistic rigidity is also known as Born rigidity.

2.3.2 Imposing Rigidity

Let us now return to the problem of a rigidly moving set of observers associated with a principal observer with arbitrary motion. We have

$$n_i \cdot u = \Omega_{ij}\xi^j\dot{\sigma} \;, \tag{2.25}$$

by (2.15), and also

$$x^{\mu}{}_{,i} = n_i{}^{\mu} + (u_0{}^{\mu} + \xi^j \dot{n}_j{}^{\mu})\sigma_{,i} = n_i{}^{\mu} + u^{\mu}\dot{\sigma}^{-1}\sigma_{,i} \,, \tag{2.26}$$

$$u_{\mu}x^{\mu}{}_{,i} = \Omega_{ij}\xi^j\dot{\sigma} + \dot{\sigma}^{-1}\sigma_{,i} \,,$$

and then the proper metric γ_{ij} is given by

$$\begin{aligned}
\gamma_{ij} &= P_{\mu\nu}x^{\mu}{}_{,i}x^{\nu}{}_{,j} \\
&= -\delta_{ij} + \Omega_{ik}\xi^k\sigma_{,j} + \Omega_{jk}\xi^k\sigma_{,i} + \dot{\sigma}^{-2}\sigma_{,i}\sigma_{,j} \\
&\quad -\big(\Omega_{ik}\xi^k\dot{\sigma} + \dot{\sigma}^{-1}\sigma_{,i}\big)\big(\Omega_{jl}\xi^l\dot{\sigma} + \dot{\sigma}^{-1}\sigma_{,j}\big) \\
&= -\delta_{ij} - \dot{\sigma}^2\Omega_{ik}\Omega_{jl}\xi^k\xi^l \\
&= -\delta_{ij} - \frac{\Omega_{ik}\Omega_{jl}\xi^k\xi^l}{\big(1+\xi^m a_{0m}\big)^2 - \xi^n\xi^r\Omega_{ns}\Omega_{rs}} \,,
\end{aligned} \tag{2.27}$$

using the expression (2.16) for $\dot{\sigma}$.

From this expression we see that there are two ways in which the motion of the set of observers can be rigid:

- All the Ω_{ij} are zero.
- All the Ω_{ij} and all the a_{0i} are constants, independent of σ.

We shall see that case one leads to a semi-Euclidean coordinate system for an arbitrary timelike worldline in Minkowski spacetime. In the second case the motion is one of a six-parameter family, with the Ω_{ij} and the a_{0i} as fixed parameters. These special motions are sometimes called superhelical motions. They will be considered in Sect. 2.4, and in particular Sect. 2.4.5.

But first, briefly, what is the point in requiring rigidity for the observer ensemble? It just means that any given observer can use a rigid ruler to indicate spatial lengths, i.e., any ruler that satisfies what is sometimes called the *ruler hypothesis*. Whatever its motion, such a ruler is always instantaneously ready to measure length as specified in an instantaneously comoving inertial frame.

Geometry is of course Euclidean in any hyperplane of simultaneity of any observer, by construction, but here we shall find that different observers share hyperplanes of simultaneity. However, we shall find that they all have different rest frame accelerations, whence the coordinate time that the main observer attributes to a given event for another observer will not generally coincide with the latter's proper time.

2.3.3 Fermi–Walker Transport

Saying that the Ω_{ij} are all zero amounts to saying that the triad n_i^{μ} is Fermi–Walker transported along the worldline of the particle $\xi = 0$. Let us see briefly what this means. If $u_0(\sigma)$ is the 4-velocity of the worldline, the equation for Fermi–Walker transport of a contravector A^{μ} along the worldline is

$$\boxed{\dot{A} = -(A \cdot \dot{u}_0)u_0 + (A \cdot u_0)\dot{u}_0} \ . \tag{2.28}$$

This preserves inner products, i.e., if A and B are FW transported along the worldline, then $A \cdot B$ is constant along the worldline:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}(A \cdot B) &= \dot{A} \cdot B + A \cdot \dot{B} \\ &= \Big[-(A \cdot \dot{u}_0)u_0 + (A \cdot u_0)\dot{u}_0\Big] \cdot B + A \cdot \Big[-(B \cdot \dot{u}_0)u_0 + (B \cdot u_0)\dot{u}_0\Big] \\ &= 0 \ .\end{aligned}$$

Furthermore, the tangent vector u_0 to the worldline is itself FW transported along the worldline, and if the worldline is a spacetime geodesic (a straight line in Minkowski coordinates), then FW transport is the same as parallel transport.

Now recall that the Ω_{ij} were defined by

$$\dot{n}_i^{\ \mu} = a_{0i}u_0^{\ \mu} + \Omega_{ij}n_j^{\ \mu} \ . \tag{2.29}$$

When $\Omega_{ij} = 0$, this becomes

$$\dot{n}_i^{\ \mu} = a_{0i}u_0^{\ \mu} \ , \tag{2.30}$$

and this is indeed the Fermi–Walker transport equation for $n_i^{\ \mu}$, found by inserting $A = n_i$ into (2.28), because we insist on $n_i \cdot u_0 = 0$ and we have $a_{0i} = -n_i \cdot \dot{u}_0$ [see (2.14) on p. 22].

In fact, the orientation in spacetime of the local rest frame triad $n_i^{\ \mu}$ is not kept constant along a worldline here unless that worldline is straight (we are referring to flat spacetimes here). Under Fermi–Walker transport, however, the triad remains as constantly oriented, or as rotationless, as possible, in a certain sense: at each instant of time σ, the triad is subjected to a pure Lorentz boost without rotation in the instantaneous hyperplane of simultaneity. On a closed orbit, this process can still lead to spatial rotation of axes upon return to the same space coordinates, an effect known as Thomas precession. For a general non-Fermi–Walker transported triad, the Ω_{ij} are the components of the angular velocity tensor that describes the instantaneous rate of rotation of the triad in the instantaneous hyperplane of simultaneity.

Of course, given any triad $n_i^{\ \mu}$ at one point on the worldline, it is always possible to Fermi–Walker transport it to other points by solving (2.28). We are then saying that motions that can be given by (2.12) on p. 21, viz.,

$$x^\mu(\xi, \tau) = x^\mu(0, \sigma) + \xi^i n_i^{\ \mu}(\sigma) \ , \tag{2.31}$$

where the ξ^i are assumed to label observers in our set of observers, are rigid in the sense of the criterion given above. Furthermore, the proper geometry of the set of observers given by the proper metric γ_{ij} in (2.22) on p. 24 is then flat, i.e.,

$$\gamma_{ij} = -\delta_{ij} \ . \tag{2.32}$$

Note finally that, in the special case where we have arranged for $\Omega_{ij} = 0$, the expression for the 4-velocity in (2.15) on p. 22 becomes

$$u = \left(u_0^\mu + \xi^i \dot{n}_i{}^\mu\right)\dot{\sigma} = u_0\left(1 + a_{0i}\xi^i\right)\dot{\sigma} \ ,$$

and since (2.16) implies that

$$\dot{\sigma} = (1 + a_{0i}\xi^i)^{-1} \ ,$$

it follows that

$$\boxed{u(\xi,\tau) = u_0\big(0,\sigma(\xi,\tau)\big)} \ . \tag{2.33}$$

2.3.4 Semi-Euclidean Coordinates Rediscovered

We now note that (σ,ξ^i) are indeed semi-Euclidean coordinates for an observer with worldline $x^\mu(0,\sigma)$, moving with the base particle $\xi = 0$, regardless of whether the space triad n_i, $i = 1,2,3$, is FW transported along that worldline. What we are doing here is to label the observer ξ^i by her spatial coordinates ξ^i in the semi-Euclidean system moving with the observer $\xi = 0$. Geometrically, we have the worldline of the arbitrarily chosen observer O at the origin, viz., $x^\mu(0,\sigma)$, with σ her proper time. We have another worldline $x^\mu(\xi^i,\tau)$ of an observer P labelled by ξ, with proper time τ. For given τ, we seek σ such that $x^\mu(\xi^i,\tau)$ is in the hyperplane of simultaneity (HOS) of O at her proper time σ. Then (ξ^i) is the position of P in the tetrad moving with O. Indeed, $\{\xi^i\}$ are the space coordinates of P relative to O in that frame.

As attested by (2.25) on p. 24, or directly from (2.33) above, in the specific case of an FW transported space triad n_i, $i = 1,2,3$, we also have

$$n_i \cdot u = 0 \ , \tag{2.34}$$

so that the instantaneous hyperplane of simultaneity of the observer at $\xi = 0$ is an instantaneous hyperplane of simultaneity for all the other observers in the rigid set as well, and the triad n_i^μ serves to define a rotationless rest frame for the whole set of observers. In other words, the coordinate system defined by the observer labels ξ^i may itself be regarded as being Fermi–Walker transported, and all the observers in the rigid set have a common designator of simultaneity in the parameter σ. In this semi-Euclidean system, associated with observer O, σ is taken to be the time coordinate.

Put another way, (2.34) says that the $n_i(\sigma)$ are in fact orthogonal to the worldline of the observer labelled by ξ^i at the value of τ corresponding to σ. This happens because $u(\xi,\tau) = u_0(0,\sigma)$. In words, the 4-velocity of observer ξ at her proper time τ is the same as the 4-velocity of the base observer when she is simultaneous with the latter in the reckoning of the base observer, quite a remarkable thing, which is in fact a direct consequence of the rigidity assumption.

For more about the HOS sharing effect, see also (2.149) on p. 60 and the discussion thereafter. In the case of generalised uniform acceleration discussed in Sect. 2.4, and for the frame construction due to Friedman and Scarr, HOS sharing only occurs when the motion of the main observer is purely translational. That is a subcase of the one discussed here, where the motion is purely translational but does not need to be uniform, i.e., the Ω_{ij} are all zero but the a_{0i} do not need to be constants.

Because σ is not generally equal to τ, it is not usually possible for the observers to have a common synchronisation of standard clocks. The relation between σ and τ is given by (2.16) on p. 22 as

$$\dot{\sigma} = \left(1 + \xi^i a_{0i}\right)^{-1} . \tag{2.35}$$

We can thus find the absolute acceleration a_i of an arbitrary observer (or particle, if we are considering a medium) in terms of a_{0i} and the ξ^i:

$$\begin{aligned} a_i &= -n_i \cdot \dot{u} = -n_i \cdot \frac{\partial u}{\partial \sigma}\dot{\sigma} = -\dot{\sigma} n_i \cdot \frac{\partial}{\partial \sigma}\left[\left(1 + \xi^j a_{0j}\right) u_0 \dot{\sigma}\right] \\ &= -\dot{\sigma} n_i \cdot \dot{u}_0 \\ &= \frac{a_{0i}}{1 + \xi^j a_{0j}} . \end{aligned} \tag{2.36}$$

Here we have used the fact mentioned above that $u = \left(1 + \xi^j a_{0j}\right) u_0 \dot{\sigma} = u_0$. We see that, although the motion is rigid and rotationless in the sense described above, not all the observers (or parts of the medium) are subject to the same acceleration.

It is important to note that, when we find ξ^i and σ, they constitute semi-Euclidean coordinates (adapted to $\xi = 0$) for the point $x^\mu(\xi, \tau)$ whether or not that point follows an observer for fixed ξ. What we are imagining here are material observers that follow all these points with fixed ξ, for a whole 3D range of values of ξ. We conclude that this rigid motion possesses only the three degrees of freedom that the observer $\xi = 0$ herself possesses. The base observer $\xi = 0$ can move any way she wants, but the rest of the observers must then follow in a well defined way.

2.3.5 SE Metric for FW Transported Coordinate Axes

In these coordinates, the metric tensor takes the form

$$g_{00} = \left.\frac{\partial x^\mu}{\partial \sigma}\right|_\xi \left.\frac{\partial x^\nu}{\partial \sigma}\right|_\xi \eta_{\mu\nu} = u^2 \dot{\sigma}^{-2} = (1 + \xi^i a_{0i})^2 , \tag{2.37}$$

$$g_{i0} = g_{0i} = \left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma \left.\frac{\partial x^\nu}{\partial \sigma}\right|_\xi \eta_{\mu\nu} = (n_i \cdot u)\dot{\sigma}^{-1} = 0 , \tag{2.38}$$

$$g_{ij} = \left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma \left.\frac{\partial x^\nu}{\partial \xi^j}\right|_\sigma \eta_{\mu\nu} = n_i \cdot n_j = -\delta_{ij} , \tag{2.39}$$

which has a simple diagonal structure, in particular because $n_i \cdot u = n_i \cdot u_0 = 0$ in this case. In these calculations, we need to be a little careful about the partial derivatives.

For example, note that, using

$$x^\mu(\xi,\tau) = x^\mu\big(0,\sigma(\xi,\tau)\big) + \xi^i n_i{}^\mu\big(\sigma(\xi,\tau)\big) \ , \tag{2.40}$$

we have

$$\begin{aligned} x^\mu{}_{,i} = \left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\tau &= \left.\frac{\partial x^\mu(0,\sigma)}{\partial \xi^i}\right|_\tau + n_i{}^\mu + \dot{n}_j{}^\mu \xi^j \sigma_{,j} \\ &= u_0^\mu \left.\frac{\partial \sigma}{\partial \xi^i}\right|_\tau + n_i{}^\mu + \dot{n}_j{}^\mu \xi^j \sigma_{,j} \\ &= n_i{}^\mu + \big(u_0^\mu + \xi^j \dot{n}_j{}^\mu\big)\sigma_{,i} \ , \end{aligned}$$

as claimed in (2.26). But we are now considering something much simpler to calculate! Still using (2.40),

$$\left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma = n_i{}^\mu \ , \tag{2.41}$$

and in a similar vein,

$$\left.\frac{\partial x^\mu}{\partial \sigma}\right|_\xi = u_0^\mu + \dot{n}_i{}^\mu \xi^i = u^\mu \dot{\sigma}^{-1} \ , \tag{2.42}$$

which justifies the above calculations for the metric.

We note that this metric becomes static, i.e., time-independent, with the parameter σ playing the role of time, in the special case in which the absolute acceleration of each observer is constant. Such a motion is called *uniform acceleration*. The notion of uniform acceleration is discussed further in Sect. 2.4.

Note also that the matrix of metric components, viz.,

$$(g_{\mu\nu}) = \begin{pmatrix} (1+\xi^i a_{0i})^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \ , \tag{2.43}$$

can become singular if $1+\xi^i a_{0i} = 0$. Indeed,

$$\det(g_{\mu\nu}) = -(1+\xi^i a_{0i})^2 \ . \tag{2.44}$$

This does not mean that the Minkowski metric is singular, only that the transformation to SE coordinates is singular at such events. Clearly, the SE coordinates cannot be extended to these events. This situation is discussed further in Sect. 2.9.

2.3.6 SE Connection for FW Transported Coordinate Axes

Note that the above relations give the Minkowski metric components in terms of the SE coordinate system, not only on the worldline $\xi = 0$ where we see immediately that they reduce to the form $\eta_{\mu\nu}$ given in (2.1), but also throughout the region where the SE coordinates are well defined. This allows us to calculate the connection coefficients using the standard relation

$$\Gamma^{\mu}_{\nu\sigma} = \frac{1}{2} g^{\mu\tau}(g_{\tau\sigma,\nu} + g_{\nu\tau,\sigma} - g_{\nu\sigma,\tau}) , \tag{2.45}$$

which follows from the metric condition in the torsion free case, i.e., when the connection coefficients are assumed symmetric in the lower two indices.

For this, we note that

$$\left(g^{\mu\tau}\right) = \begin{pmatrix} \left(1+\xi^j a_{0j}\right)^{-2} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} .$$

This is diagonal, which facilitates the task. For example, we obtain

$$\begin{aligned} \Gamma^0_{00} &= \frac{1}{2}\left(1+\xi^j a_{0j}\right)^{-2}(g_{00,0} + g_{00,0} - g_{00,0}) \\ &= \left(1+\xi^j a_{0j}\right)^{-2} \partial_\sigma \left[\frac{1}{2}\left(1+\xi^j a_{0j}\right)^2\right] , \end{aligned}$$

whence

$$\boxed{\Gamma^0_{00} = \frac{\xi^i \dot{a}_{0i}}{1+\xi^j a_{0j}}} . \tag{2.46}$$

This gives this connection coefficient throughout the region in which the SE coordinates are defined. On the worldline at the space origin of the SE coordinate system, i.e., when $\xi = 0$, we find

$$\boxed{\Gamma^0_{00}\big|_{\xi=0} = 0} . \tag{2.47}$$

Consider now

$$\Gamma^{\mu}_{ij} = \frac{1}{2} g^{\mu\mu}(g_{\mu j,i} + g_{i\mu,j} - g_{ij,\mu}) \qquad \text{(no sum over } \mu) ,$$

which is clearly zero, since only g_{00} has any spacetime coordinate dependence. This is also true throughout the region where the SE coordinates are defined, so that we may write

$$\boxed{\Gamma^{\mu}_{ij} = 0 , \qquad i,j = 1,2,3, \;\; \mu = 0,1,2,3} , \tag{2.48}$$

and hence a fortiori on $\xi = 0$, whence

$$\boxed{\Gamma^{\mu}_{ij}\Big|_{\xi=0} = 0\,, \qquad i,j = 1,2,3,\;\; \mu = 0,1,2,3}\,. \tag{2.49}$$

We also have

$$\Gamma^{i}_{00} = \frac{1}{2}g^{ii}(g_{i0,0} + g_{0i,0} - g_{00,i}) = \frac{1}{2}g_{00,i} = \frac{1}{2}\partial_{\xi^i}\Big[\big(1+\xi^j a_{0j}\big)^2\Big]\,,$$

with no sum over i, whence

$$\boxed{\Gamma^{i}_{00} = \big(1+\xi^j a_{0j}\big)a_{0i}\,, \qquad i = 1,2,3}\,. \tag{2.50}$$

Then on $\xi = 0$, we obtain

$$\boxed{\Gamma^{i}_{00}\big|_{\xi=0} = a_{0i}\,, \qquad i = 1,2,3}\,. \tag{2.51}$$

Likewise,

$$\begin{aligned}\Gamma^{0}_{0i} &= \frac{1}{2}g^{00}(g_{0i,0} + g_{00,i} - g_{0i,0})\\ &= \frac{1}{2}g^{00}g_{00,i} = \big(1+\xi^j a_{0j}\big)^{-2}\partial_{\xi^i}\Big[\frac{1}{2}\big(1+\xi^j a_{0j}\big)^2\Big]\,,\end{aligned}$$

whence

$$\boxed{\Gamma^{0}_{0i} = \frac{a_{0i}}{1+\xi^j a_{0j}}\,, \qquad i = 1,2,3}\,. \tag{2.52}$$

Then on $\xi = 0$, we obtain

$$\boxed{\Gamma^{0}_{0i}\big|_{\xi=0} = a_{0i}\,, \qquad i = 1,2,3}\,. \tag{2.53}$$

Finally,

$$\Gamma^{i}_{0j} = \frac{1}{2}g^{ii}(g_{ij,0} + g_{0i,j} - g_{0j,i})\,,$$

with no sum over i, whence

$$\boxed{\Gamma^{i}_{0j} = 0\,, \qquad i = 1,2,3}\,. \tag{2.54}$$

Then on $\xi = 0$, we obtain

$$\boxed{\Gamma^{i}_{0j}\big|_{\xi=0} = 0\,, \qquad i = 1,2,3}\,. \tag{2.55}$$

In this way, we completely recover the relations (2.7)–(2.9) on p. 19, for the case ${\Omega^i}_j = 0$. The time has come to understand precisely what led to this rotation matrix in our earlier considerations.

2.3.7 SE Coordinates with Rotating Axes

Clearly, the simpler result in the latest calculations has something to do with the rigidity assumption, which was implemented by assuming that the spacelike triad $\{n_i\}_{i=1,2,3}$ was FW transported along the observer worldline, whence the coefficients Ω_{ij} were set to zero in (2.13) on p. 22. When we drop this assumption, we may no longer have a rigid coordinate system in general (unless all the a_{0i} and Ω_{ij} are constant, as mentioned on p. 25 and discussed further in Sects. 2.4 and 2.4.4), but we still obtain one that is semi-Euclidean in the sense of the five conditions back on p. 12.

So we shall now retain the more general relations

$$x^\mu(\xi,\tau) = x^\mu(0,\sigma) + \xi^i {n_i}^\mu(\sigma) \,, \tag{2.56}$$

with

$$u^\mu = \frac{\partial x^\mu(\xi,\tau)}{\partial \tau} = \left(u_0^\mu + \xi^i \dot{n}_i{}^\mu\right)\dot{\sigma} \tag{2.57}$$

and

$$\dot{n}_i{}^\mu = a_{0i}(\sigma) u_0^\mu + \Omega_{ij}(\sigma) {n_j}^\mu \,. \tag{2.58}$$

This leads to (2.16) on p. 22, viz.,

$$\dot{\sigma} = \left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]^{-1/2} . \tag{2.59}$$

The relation (2.56) now implies

$$\left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma = {n_i}^\mu \,, \tag{2.60}$$

exactly as in (2.41), and

$$\left.\frac{\partial x^\mu}{\partial \sigma}\right|_\xi = u_0^\mu + \dot{n}_i{}^\mu \xi^i = u^\mu \dot{\sigma}^{-1} \,, \tag{2.61}$$

exactly as in (2.42).

2.3.8 SE Metric for Rotating Coordinate Axes

It is instructive to calculate the components of the Minkowski metric in these more general SE coordinates. This time we have

$$g_{00} = \left.\frac{\partial x^\mu}{\partial \sigma}\right|_\xi \left.\frac{\partial x^\nu}{\partial \sigma}\right|_\xi \eta_{\mu\nu} = u^2\dot{\sigma}^{-2} = \left(1+\xi^i a_{0i}\right)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk} \,, \tag{2.62}$$

$$g_{i0} = g_{0i} = \left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma \left.\frac{\partial x^\nu}{\partial \sigma}\right|_\xi \eta_{\mu\nu} = (n_i \cdot u)\dot{\sigma}^{-1} \,, \tag{2.63}$$

$$g_{ij} = \left.\frac{\partial x^\mu}{\partial \xi^i}\right|_\sigma \left.\frac{\partial x^\nu}{\partial \xi^j}\right|_\sigma \eta_{\mu\nu} = n_i \cdot n_j = -\delta_{ij} \,, \tag{2.64}$$

which no longer has the diagonal structure since we do not now have $n_i \cdot u = 0$ in general. In the rigid case, we had $u = u_0$ to ensure diagonality of the metric everywhere.

If we set $\xi = 0$, we do indeed obtain the form $\eta_{\mu\nu}$ of (2.1), noting for example that $n_i \cdot u|_{\xi=0} = n_i \cdot u_0 = 0$. If we set $\Omega_{ij} = 0$, we retrieve the metric components in Sect. 2.3.5, as expected (see p. 28). When $\Omega_{ij} \neq 0$, we have

$$\begin{aligned}(n_i \cdot u)\dot{\sigma}^{-1} &= n_i \cdot \left(u_0 + \dot{n}_j \xi^j\right)\\ &= n_i \cdot \dot{n}_j \xi^j\\ &= \xi^j n_i \cdot (a_{0j}u_0 + \Omega_{jk}n_k)\\ &= \Omega_{ji}\xi^j \,,\end{aligned}$$

whence

$$g_{i0} = \xi^j \Omega_{ji} = g_{0i} \,. \tag{2.65}$$

The matrix of metric components, viz.,

$$(g_{\mu\nu}) = \begin{pmatrix} \left(1+\xi^i a_{0i}\right)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk} & \xi^j\Omega_{j1} & \xi^j\Omega_{j2} & \xi^j\Omega_{j3} \\ \xi^j\Omega_{j1} & -1 & 0 & 0 \\ \xi^j\Omega_{j2} & 0 & -1 & 0 \\ \xi^j\Omega_{j3} & 0 & 0 & -1 \end{pmatrix} , \tag{2.66}$$

can become singular if $1+\xi^i a_{0i} = 0$. Indeed, for any a, b, c, and d,

$$\begin{aligned}\det\begin{pmatrix} a & b & c & d \\ b & -1 & 0 & 0 \\ c & 0 & -1 & 0 \\ d & 0 & 0 & -1 \end{pmatrix} &= -a - b\begin{vmatrix} b & 0 & 0 \\ c & -1 & 0 \\ d & 0 & -1 \end{vmatrix} + c\begin{vmatrix} b & -1 & 0 \\ c & 0 & 0 \\ d & 0 & -1 \end{vmatrix} + d\begin{vmatrix} b & -1 & 0 \\ c & 0 & -1 \\ d & 0 & 0 \end{vmatrix}\\ &= -a - b^2 - c^2 - d^2 \,.\end{aligned}$$

So for the most general SE coordinate system, the matrix of components of the Minkowski metric has determinant

$$\boxed{\det g_{\mathrm{SE}}^{\mathrm{Mink}} = -\left(1+\xi^i a_{0i}\right)^2} \, . \tag{2.67}$$

Interestingly, this is always independent of the rotation chosen for the space triad $\{n_i\}_{i=1,2,3}$, as specified by Ω_i, $i=1,2,3$, but it does depend on the acceleration of the worldline as specified by its absolute components a_{0i}, $i=1,2,3$.

Furthermore, as just mentioned, it can be zero. We have already seen how this happens for translational uniform acceleration in Sect. 2.3.5. In this more general case, we see that it occurs for all ξ^i satisfying

$$\xi^i a_{0i}(\sigma) = -1 \, ,$$

for some value of the proper time σ of the observer. This specifies a 2-plane of the 3D space of ξ^i for each proper time σ. Note also that

$$g_{00} \leq 0 \, , \quad \text{when} \quad \det g_{\mathrm{SE}}^{\mathrm{Mink}} = 0 \, . \tag{2.68}$$

This issue is discussed further in Sect. 2.12.

We can now assess the validity of the first property on the list at the beginning of Sect. 2.2.1 back on p. 12. This claimed that we could arrange the coordinates so that the curves $\xi^i = \text{constant}$ $(i=1,\,2,\,3)$ would be timelike, and also so that any curve with $\sigma = \text{constant}$ would be spacelike:

- When a curve has $\xi^i = \text{constant}$ $(i=1,\,2,\,3)$, only the zeroth component of its tangent vector is nonzero. Its Lorentzian pseudolength is then a positive multiple of g_{00} in (2.66). So this property of our SE frame is only satisfied in regions where $g_{00} > 0$. Hence the further discussion in Sect. 2.12.
- When a curve has $\sigma = \text{constant}$, this implies that the zeroth component of its tangent vector is zero, so its Lorentzian pseudolength will always be negative, according to (2.66).

2.3.9 SE Connection for Rotating Coordinate Axes

It is interesting here to see how we obtain (2.9) on p. 19 with $\Omega^i{}_j \neq 0$. Once again, we use the standard formula

$$\Gamma^{\mu}_{\nu\sigma} = \frac{1}{2} g^{\mu\tau} \left(g_{\tau\sigma,\nu} + g_{\nu\tau,\sigma} - g_{\nu\sigma,\tau} \right) \, , \tag{2.69}$$

which expresses the fact that we impose the metric condition on the connection and assume that the latter is symmetric in its two lower indices. To do the full calculation for the connection components throughout the region in which the SE coordinates are defined, we would have to invert the metric matrix to obtain $g^{\mu\tau}$. This is quite feasible, but let us focus on the values of the connection coefficients along the worldline $\xi = 0$. Then we only require the contravariant metric components on the worldline and we have

$$g^{\mu\nu}|_{\xi=0} = \eta^{\mu\nu} .$$

Thus,

$$\Gamma^0_{00}|_{\xi=0} = \frac{1}{2} g_{00,0}\Big|_{\xi=0} = \frac{1}{2} \frac{\partial}{\partial\sigma}\Big|_{\xi} \Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk}\Big]\Big|_{\xi=0} ,$$

and a little thought shows that this will be zero.

Then

$$\Gamma^i_{00}|_{\xi=0} = -\frac{1}{2}(g_{i0,0}+g_{0i,0}-g_{00,i})\Big|_{\xi=0} = -g_{i0,0}|_{\xi=0} + \frac{1}{2} g_{00,i}\Big|_{\xi=0} ,$$

with

$$g_{i0,0}|_{\xi=0} = \frac{\partial}{\partial\sigma}\Big|_{\xi} (\xi^j\Omega_{ji})\Big|_{\xi=0} = 0$$

and

$$\begin{aligned}
\frac{1}{2} g_{00,i}\Big|_{\xi=0} &= \frac{1}{2} \frac{\partial}{\partial\xi^i}\Big|_{\sigma} \Big[\big(1+\xi^j a_{0j}\big)^2 - \xi^j\xi^k\Omega_{jl}\Omega_{kl}\Big]\Big|_{\xi=0} \\
&= \frac{1}{2} \frac{\partial}{\partial\xi^i}\Big|_{\sigma} \Big[\big(1+\xi^j a_{0j}\big)^2\Big]\Big|_{\xi=0} \\
&= a_{0i} ,
\end{aligned}$$

as expected.

Next we have

$$\Gamma^0_{0i}|_{\xi=0} = \frac{1}{2}(g_{0i,0}+g_{00,i}-g_{0i,0})\Big|_{\xi=0} = \frac{1}{2} g_{00,i}\Big|_{\xi=0} ,$$

and we have just seen that this gives a_{0i}.

Then

$$\Gamma^\mu_{ij}|_{\xi=0} = \frac{1}{2}\eta^{\mu\mu}(g_{\mu j,i}+g_{i\mu,j}-g_{ij,\mu})\Big|_{\xi=0} ,$$

with no sum over μ. If $\mu = 1,2,3$, we get zero since $g_{ij} = \delta_{ij}$. When $\mu = 0$, we get

$$\Gamma^0_{ij}|_{\xi=0} = \frac{1}{2}(g_{0j,i}+g_{i0,j})\Big|_{\xi=0} ,$$

but

$$g_{0j,i} = \frac{\partial}{\partial \xi^i}\bigg|_\sigma (\xi^k \Omega_{kj}) = \Omega_{ij} ,$$

which is antisymmetric in i and j, whence we obtain zero for all coefficients $\Gamma^\mu_{ij}|_{\xi=0}$, as before.

As a matter of fact, we can do slightly better here. We can show that the connection coefficients Γ^μ_{ij} are actually zero wherever the SE coordinates are defined, and not just on the observer worldline. To see this, begin with

$$\Gamma^\mu_{ij}|_{\xi=0} = \frac{1}{2} g^{\mu\nu}(g_{\nu j,i} + g_{i\nu,j} - g_{ij,\nu}) .$$

If $\nu = 0$, we obtain a term proportional to

$$g_{0j,i} + g_{i0,j} - g_{ij,0} = \Omega_{ij} + \Omega_{ji} = 0 ,$$

and if $\nu = k$, we obtain a term proportional to

$$g_{kj,i} + g_{ik,j} - g_{ij,k} = 0 ,$$

so we always get zero. This result means that any geodesic lying in one of the spacelike hypersurfaces $\tau =$ constant is a straight line when expressed relative to the SE coordinates, even if it does not intersect the observer worldline.

So everything is so far as before. The most interesting coefficients are

$$\begin{aligned}
\Gamma^i_{0j}|_{\xi=0} &= \frac{1}{2}\eta^{ii}(g_{ij,0} + g_{0i,j} - g_{0j,i})\bigg|_{\xi=0} \\
&= -\frac{1}{2}\Big[\partial_j(\xi^k\Omega_{ki}) - \partial_i(\xi^k\Omega_{kj})\Big]\bigg|_{\xi=0} \\
&= \Omega_{ij} .
\end{aligned}$$

We conclude that the Ω_{ij} introduced in the expansion (2.58) is in fact the same as ${\Omega^i}_j$ in (2.9) on p. 19.

2.4 Generalised Uniform Acceleration

At the end of Sect. 2.3.5 (see p. 29), we defined uniform acceleration as being a motion in which the absolute acceleration, i.e., the acceleration expressed in a suitably chosen continuous sequence of instantaneously comoving inertial frames, is constant. To be more precise, it was motion in which the acceleration $\dot{u}_0$ of the observer had the form

$$\dot{u}_0(\sigma) = a_{0i}(\sigma) n_i(\sigma) ,$$

with $\{n_i(\sigma)\}$ a Fermi–Walker transported spatial triad along the observer worldline, and with the a_{0i} actually independent of the proper time σ of the observer.

This notion has been generalised by Friedman and Scarr [23] in a way we shall now describe, since it ties up a loose end from Sect. 2.3.2. There we noted that there was a second way to achieve the rigid motion of a set of observers, viz., superhelical motion, in which all the functions $a_{0i}(\sigma)$ and $\Omega_{ij}(\sigma)$ in (2.13) on p. 22 are actually independent of σ. Let us now examine how this works.

2.4.1 Definition

Here we closely follow the discussion by Friedman and Scarr [23]. We work in an inertial (laboratory) frame denoted by K. For any timelike worldline, we take the 4-velocity to be the dimensionless unit 4-vector

$$u = (u^0, u^1, u^2, u^3) := \frac{1}{c}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} ,$$

where τ is the proper time, and hence define the 4-acceleration to be

$$a^\mu := c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} ,$$

which has units of acceleration.

We define a uniformly accelerating worldline to be one that satisfies

$$\boxed{c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = A^\mu{}_\nu u^\nu} , \tag{2.70}$$

with some specified initial value $u(0) = u_0$, where $A^\mu{}_\nu$ is a tensor under Lorentz transformations and independent of τ. We also require that the type (2,0) form $\bar{A}$ of this tensor, with components $A_{\mu\nu} := \eta_{\mu\sigma}A^\sigma{}_\nu$, should be antisymmetric, for the following reason. Since $u^2 = 1$ is constant, we must have $a \cdot u = 0$, whence we require

$$0 = \eta_{\mu\nu}a^\mu u^\nu = \eta_{\mu\nu}A^\mu{}_\sigma u^\sigma u^\nu = u^\nu A_{\nu\sigma}u^\sigma ,$$

a sufficient condition for which is that the type (2,0) tensor $A_{\mu\nu}$ should be antisymmetric, i.e.,

$$A_{\mu\nu} = -A_{\nu\mu} . \tag{2.71}$$

Equation (2.70) has the unique solution

$$\boxed{u(\tau) = \exp(A\tau/c)u_0 = \left(\sum_{n=0}^{\infty}\frac{A^n}{n!c^n}\tau^n\right)u_0} , \tag{2.72}$$

where A is the type (1,1) tensor. A key motivation for the above definition is that this kind of motion is covariant in the sense that uniformly accelerated motion in one inertial frame is uniformly accelerated motion in every inertial frame. This in turn follows straight from the definition because proper time is an invariant, u is a four-vector, $A^\mu{}_\nu$ is a tensor (see below), and $A_{\mu\nu}$ will be antisymmetric in every inertial frame if it is so in one.

2.4.2 Tensorial Nature of A and $\bar{A}$

The equation of motion (2.70) has been expressed relative to some arbitrarily chosen inertial frame K. But how would we transform this equation of motion in order to describe the worldline relative to a new inertial frame?

The answer is that we will get *the same equation* expressed relative to the new frame if we transform the object so suggestively written as $A^\mu{}_\nu$ as a type (1,1) tensor. Indeed, if we are to obtain the same equation expressed relative to the new frame, *it has to transform like this* because the left-hand side of (2.70) transforms as a contravector, and so does u.

Relative to any other choice of laboratory inertial frame K_1 related to K by a homogeneous Lorentz transformation, the acceleration matrix will have the form

$$A_1 = L^{-1} A L \,, \tag{2.73}$$

where L is the homogeneous Lorentz transformation from K to K_1.

Naturally then, the object $\bar{A}$ with components $A_{\mu\nu} := \eta_{\mu\sigma} A^\sigma{}_\nu$ must transform as a type (2,0) tensor when we rewrite the equation of motion relative to some other inertial frame. Relative to any other choice of laboratory inertial frame K_1 related to K by a homogeneous Lorentz transformation L, the type (2,0) acceleration matrix $\bar{A}$ will have the form

$$\bar{A}_1 = L^{\mathrm{T}} \bar{A} L \,. \tag{2.74}$$

Note that $\bar{A}$ is antisymmetric if and only if $L^{\mathrm{T}}\bar{A}L$ is antisymmetric.

We shall see concrete examples of such transformations shortly.

2.4.3 Nature of Generalisation

Let us see how the above extends the usual definition of uniform acceleration. The first thing is to write down the most general possible matrix versions of $A_{\mu\nu}$ and $A^\mu{}_\nu$ in the chosen laboratory inertial frame K, bearing in mind the antisymmetry of the former:

$$A_{\mu\nu}(\mathbf{g},\omega) = \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ -\mathbf{g} & -c\pi(\omega) \end{pmatrix} \,, \qquad A^\mu{}_\nu(\mathbf{g},\omega) = \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & c\pi(\omega) \end{pmatrix} \,. \tag{2.75}$$

Here we have used the notation introduced in [23]: $\mathbf{g}$ is a 3-component object with units of acceleration and transpose $\mathbf{g}^{\mathrm{T}}$, and $\omega = (\omega^1, \omega^2, \omega^3)$ is another 3-component object but this time with units of 1/time, and

$$\pi(\omega) := \varepsilon_{ijk}\omega^k \;, \tag{2.76}$$

with ε_{ijk} the completely antisymmetric Levi-Civita symbol. The factor of c with $\pi(\omega)$ just ensures that this entry has units of acceleration. Since $A_{\mu\nu}$ is independent of τ, the same goes for $\mathbf{g}$ and ω.

Now when $\omega = 0$, the above definition of uniform acceleration reduces to the usual definition of uniform acceleration in a straight line. Indeed, we then have

$$c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = A^\mu{}_\nu u^\nu = \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & 0 \end{pmatrix} \begin{pmatrix} u^0 \\ \mathbf{u}/c \end{pmatrix} \;, \tag{2.77}$$

so that

$$c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = (c^{-1}\mathbf{g}\cdot\mathbf{u}, \mathbf{g}u^0) \;, \tag{2.78}$$

since we are taking

$$(u^\mu) = (u^0, \mathbf{u}/c) \;, \qquad u^0 = \gamma(v) \;, \qquad \mathbf{u} = \gamma(v)\mathbf{v} \;.$$

We note also that

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma \;,$$

whence

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t} = \mathbf{g}u^0\gamma^{-1} = \mathbf{g} \;. \tag{2.79}$$

Since $\mathbf{g}$ is independent of time, this is indeed the usual definition for translational uniform acceleration. It has solution

$$\mathbf{u} = \mathbf{g}t + \mathbf{u}_0 \;, \tag{2.80}$$

where $\mathbf{u}_0$ is the value of $\mathbf{u}$ at time $t = 0$. We shall see shortly how this accords with the definition on p. 29.

Interestingly, this is precisely the motion we would obtain from a naive special relativistic model of gravity in which gravity is just a force, for the case where that force is given by $m\mathbf{g}$ for some constant acceleration $\mathbf{g}$ due to gravity and m is the passive gravitational mass of the particle, assuming of course that passive gravitational mass and inertial mass are exactly equal, as seems to be suggested by experiment. Purely translational uniform acceleration (TUA) will be examined explicitly in Sect. 2.9.

Note that the definition of purely translational uniform acceleration does not give rise to a covariant notion of uniform acceleration, since it depends on having $\omega = 0$. This standard notion would thus only be covariant under transformations that preserve this condition, viz., Lorentz boosts in the direction of $\mathbf{g}$ and space rotations about the direction of **g**. Rather than being the whole homogeneous Lorentz group, as for the new definition of uniform acceleration, the covariance group would be the little group fixing the space axis along **g**. In fact, the generalised form of uniform acceleration is even covariant under the transformations of the inhomogeneous Lorentz group (Poincaré group), including spacetime translations.

A further motivation for the extended definition is that the action of the Lorentz force on a charged particle can be represented by the action of the antisymmetric electromagnetic tensor $F_{\mu\nu}$ on the four-velocity of the particle:

$$m\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = qF^\mu{}_\nu u^\nu \,, \tag{2.81}$$

where m is the particle rest mass and q its charge. Then any constant and uniform electromagnetic field will lead to uniformly accelerated motion.

The type (1,1) tensor A and the associated antisymmetric type (2,0) tensor $\bar{A}$ are both referred to as the acceleration tensor. A uniformly accelerated motion is uniquely defined by its acceleration tensor A and its initial four-velocity u_0. The associated worldline $\hat{x}(\tau)$ can be found if we know the initial position $\hat{x}(0)$. The basic equation (2.70) along with an initial value $u(0) = u_0$ can be solved exactly to obtain $u(\tau)$, and the resulting expression is easily integrated to obtain $\hat{x}(\tau)$ if we have the initial value $\hat{x}(0)$. The details can be found in [23], but also in [17], where geometric algebra techniques are applied to the case of a constant electromagnetic field tensor.

2.4.4 Coordinate Frame for Generalised Uniform Acceleration

Once again, we follow Friedman and Scarr for this construction [23]. The aim will be to construct a coordinate frame K' such that the observer with uniform acceleration and proper time τ has worldline $(c\tau, 0, 0, 0)$. We shall do this in a mathematically natural way and we shall find that the result is a rigid frame in the sense that two nearby particles sitting at fixed space coordinates in K' are always the same distance apart as measured in the instantaneously comoving inertial frame of either.

Note that we already have a rigid frame, obtained by FW transporting a space triad along the observer worldline and using the construction described in detail in Sects. 2.3.4–2.3.6. In the present case, we shall obtain a generally different rigid coordinate frame that depends crucially on the observer worldline being generated as in (2.70) by a constant acceleration tensor. We shall once again transport a space triad along the observer worldline, but the transport will be a generalisation of FW transport, although it will be a generalisation *only* for the case of generalised uniform acceleration, and only reduce to it in the special case where we have the purely translational form of uniform acceleration in some inertial frame in which the observer comes to rest at some event.

Consider the worldline $\hat{x}(\tau)$ of a uniformly accelerating observer in the sense of (2.70), with motion determined by a constant acceleration matrix $\bar{A} = (A_{\mu\nu})$, initial 4-velocity $u(0)$, and initial position $\hat{x}(0)$. The first step is to define an instantaneously comoving inertial frame (ICIF) K_τ at each proper time τ, specified by a tetrad $\lambda(\tau) = \{\lambda_{(\kappa)}(\tau)\}_{\kappa=0,1,2,3}$. To do so we choose an initial ICIF K_0 with origin at $\hat{x}(0)$, specified by a tetrad $\hat{\lambda} = \{\hat{\lambda}_{(\kappa)}\}_{\kappa=0,1,2,3}$, where as usual we take $\hat{\lambda}_{(0)} = u(0)$ and $\{\hat{\lambda}_{(i)}\}_{i=1,2,3}$ can be chosen arbitrarily to complete the tetrad.

We must now transport this initial tetrad along the worldline. Previously we used FW transport and obtained a perfectly good rigid coordinate frame in that way. However, there is a mathematically more natural way to transport our space triad in the present case. For $\tau > 0$, we define K_τ by requiring the origin of K_τ at time τ to be $\hat{x}(\tau)$ and requiring the basis of K_τ to be the unique solution $\lambda(\tau) = \{\lambda_{(\kappa)}(\tau)\}_{\kappa=0,1,2,3}$ of the initial value problem

$$\boxed{c\frac{\mathrm{d}\lambda^\mu_{(\kappa)}}{\mathrm{d}\tau} = A^\mu{}_\nu \lambda^\nu_{(\kappa)}\ , \qquad \lambda_{(\kappa)}(0) = \hat{\lambda}_{(\kappa)}}\ . \tag{2.82}$$

Since $\hat{\lambda}_{(0)} = u(0)$ and $u(\tau)$ satisfies this differential equation according to (2.70), we deduce that $\lambda_{(0)}(\tau) = u(\tau)$ for all values of τ. Furthermore, the type (1,1) tensor A is a constant matrix, we can immediately solve the system (2.82) to obtain

$$\lambda(\tau) = \exp(A\tau/c)\hat{\lambda}\ . \tag{2.83}$$

Note also that this kind of transport is an isometry, i.e., it preserves the Lorentzian scalar product. To see this, suppose that v and w are any two 4-vectors at $\hat{x}(0)$ and solve (2.82) to obtain vector fields $v(\tau)$ and $w(\tau)$ along $\hat{x}(\tau)$. Then consider

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\big[v(\tau)\cdot w(\tau)\big] &= \big[Av(\tau)\big]\cdot\big[w(\tau)\big] + \big[v(\tau)\big]\cdot\big[Aw(\tau)\big]\\ &= \eta_{\mu\sigma}A^\mu{}_\nu v^\nu w^\sigma + \eta_{\mu\sigma}v^\mu A^\sigma{}_\nu w^\nu\\ &= A_{\sigma\nu}v^\nu w^\sigma + A_{\mu\nu}v^\mu w^\nu\\ &= A_{\mu\nu}(v^\nu w^\mu + v^\mu w^\nu) = 0\ ,\end{aligned} \tag{2.84}$$

due to the antisymmetry of the type (2,0) tensor $\bar{A}$. Interestingly, we do not use the constancy of the matrix A in this proof, only the differential relations that v and w must satisfy, so this kind of transport is isometric for quite general, possibly time-varying acceleration matrices A, provided that the associated matrix $\bar{A}$ is antisymmetric.

The fact that this transport is isometric is important, because it means that the solution to (2.82) will be orthonormal right along the observer worldline, i.e., it will be a tetrad. We shall examine the resulting coordinate system in a moment. Before doing so, it is interesting to rewrite (2.82) in a slightly different way. To begin with, let us think of our initial frame $\hat{\lambda}$ as a matrix with columns

$$\hat{\lambda} = (\hat{\lambda}_{(0)}\ \hat{\lambda}_{(1)}\ \hat{\lambda}_{(2)}\ \hat{\lambda}_{(3)})\ , \tag{2.85}$$

where each column comprises the components $\hat{\lambda}_{(\kappa)}^{\mu}$ of the given tetrad 4-vector expressed relative to the inertial (laboratory) frame K. This matrix maps the basis

$$e_0 := \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} , \quad e_1 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} , \quad e_2 := \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} , \quad e_3 := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} , \tag{2.86}$$

of K to the basis we have chosen for the initial ICIF K_0. Now define the new matrix

$$\tilde{A} := \hat{\lambda}^{-1} A \hat{\lambda} , \tag{2.87}$$

the representation of the type (1,1) tensor A relative to the basis of the initial ICIF K_0.

We check formally that this is the right transformation. If $\{x^{\mu}\}$ are the coordinates in K and $\{\tilde{x}^{(\mu)}\}$ are the coordinates in K_0, then the correct transformed components of A relative to the frame K_0 should be

$$\tilde{A}^{(\mu)}{}_{(\nu)} = \frac{\mathrm{d}\tilde{x}^{(\mu)}}{\mathrm{d}x^{\sigma}} \frac{\mathrm{d}x^{\tau}}{\mathrm{d}\tilde{x}^{(\nu)}} A^{\sigma}{}_{\tau} \tag{2.88}$$

Let us check that the matrix $\hat{\lambda}$ is

$$\hat{\lambda} = \left(\mathrm{d}x^{\tau}/\mathrm{d}\tilde{x}^{(\nu)}\right) \tag{2.89}$$

in this notation, with τ labelling rows and ν labelling columns. This has inverse

$$\left(\mathrm{d}x^{\tau}/\mathrm{d}\tilde{x}^{(\nu)}\right)^{-1} = \left(\mathrm{d}\tilde{x}^{(\mu)}/\mathrm{d}x^{\sigma}\right) ,$$

with μ labelling rows and σ labelling columns, so (2.88) would then be the matrix relation $\tilde{A} = \hat{\lambda}^{-1} A \hat{\lambda}$. Now (2.89) amounts to the relations

$$\hat{\lambda}_{(\nu)}^{\tau} = \frac{\mathrm{d}x^{\tau}}{\mathrm{d}\tilde{x}^{(\nu)}} . \tag{2.90}$$

But we have

$$x^{\tau} e_{\tau} = \tilde{x}^{(\nu)} \hat{\lambda}_{(\nu)} , \tag{2.91}$$

where the e_{τ} are defined in (2.86). This in turn implies that

$$x^{\tau} = \tilde{x}^{(\nu)} \hat{\lambda}_{(\nu)}^{\tau} ,$$

whence we obtain (2.90).

Put another way, we now have

$$\hat{\lambda} \tilde{A} = A \hat{\lambda} , \tag{2.92}$$

or in component form

$$\hat{\lambda}^{\nu}_{(\gamma)}\tilde{A}^{(\gamma)}{}_{(\kappa)} = A^{\nu}{}_{\sigma}\hat{\lambda}^{\sigma}_{(\kappa)} \ . \tag{2.93}$$

The point is that we can now write

$$\begin{aligned} c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} &= \big[\exp(A\tau/c)\big]^{\mu}{}_{\nu}A^{\nu}{}_{\sigma}\hat{\lambda}^{\sigma}_{(\kappa)} \\ &= \big[\exp(A\tau/c)\big]^{\mu}{}_{\nu}\hat{\lambda}^{\nu}_{(\gamma)}\tilde{A}^{(\gamma)}{}_{(\kappa)} \ , \end{aligned}$$

whence

$$\boxed{c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\gamma)}\tilde{A}^{(\gamma)}{}_{(\kappa)}} \ . \tag{2.94}$$

This is just a slightly different, but equivalent version of (2.82). It will be useful for drawing the parallel with previous constructions, and also for showing that this kind of transport generalises FW transport in the case of purely translational uniform acceleration in the initial instantaneously comoving inertial frame $K_0 = \hat{\lambda}$.

These considerations generalise easily to the case of acceleration matrices that are not constant along the observer worldline. This is discussed in greater detail in Sect. 2.11.4 [see in particular (2.321) on p. 101 and the following]. It is shown there that we can always find a matrix $A^{\mu}{}_{\nu}$ such

$$c\frac{\mathrm{d}\lambda^{\mu}{}_{(0)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}{}_{(0)} \ , \tag{2.95}$$

although it will not generally be constant, with the further property that

$$c\frac{\mathrm{d}\lambda^{\mu}{}_{(i)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}{}_{(i)} \ , \quad i = 1,2,3 \ , \tag{2.96}$$

whatever smooth choice of space tetrad $\{\lambda^{\mu}{}_{(i)}\}_{i=1,2,3}$ has been made along the worldline (but noting that the matrix A then depends on that choice).

Let us now see how to set up coordinates $\{y^{(\mu)}\}$ adapted to the generalised uniformly accelerating observer worldline. For any event X with coordinates x^{μ} in K, we find a proper time τ for the observer for which $\hat{x}(\tau)$ is simultaneous with X in the ICIF K_{τ}. We then define the zero (or time) coordinate of X in the proposed accelerating frame K' to be $y^{(0)} = c\tau$. Note that all events in this particular hyperplane of simultaneity for K_{τ} will be attributed the same time coordinate $c\tau$.

Put another way, K' is borrowing the hyperplanes of simultaneity of an instantaneously comoving inertial observer, so given that we obtain the tetrad field along the observer worldline by the isometric propagation (2.82) [or (2.94)], we have a standard construction of semi-Euclidean coordinates as described previously. And as in our earlier constructions, we still have the problem that such hyperplanes can intersect off the observer worldline. These coordinates will generally only be valid on some neighbourhood of the worldline, and not throughout the whole of spacetime.

Now the displacement $\bar{y}$ of X relative to the observer at proper time τ can be expressed in terms of the space triad $\{\lambda_{(i)}(\tau)\}_{i=1,2,3}$, since

$$\bar{y} \cdot u(\tau) = \left[x - \hat{x}(\tau)\right] \cdot u(\tau) = 0 \, ,$$

by the specific choice of τ. Hence, there are $y^{(i)} \in \mathbb{R}$, $i = 1,2,3$, such that

$$\bar{y} = y^{(i)} \lambda_{(i)}(\tau) \, .$$

In short, we have found τ such that

$$\boxed{x^\mu = \hat{x}^\mu(\tau) + y^{(i)} \lambda_{(i)}(\tau)} \, . \tag{2.97}$$

The coordinates of event X in the coordinate frame K' will then be defined as $(c\tau, y^{(1)}, y^{(2)}, y^{(3)})$, and the relation (2.97) tells us how to convert from these coordinates to the original laboratory coordinate system K.

2.4.5 Rigidity

We shall show that this is a rigid coordinate system in the sense of Sect. 2.3.1. It is important to understand that this notion of rigidity is *not* the same as saying that the geometry of the hyperplanes of simultaneity is Euclidean, which is true by construction.

Imagine two particles A and B at rest relative to the space coordinates of the proposed accelerating frame K', with worldlines of the form $\{(c\tau, \mathbf{y}_A) : \tau \in \mathbb{R}\}$ and $\{(c\tau, \mathbf{y}_B) : \tau \in \mathbb{R}\}$, respectively. At a given τ, the particles lie in the same hyperplane of simultaneity of the inertial frame K_τ, and since this is also the hyperplane of simultaneity adopted in K', the proper distance between A and B at coordinate time τ in the K' system is just the length of the vector $\left[y_B^{(i)} - y_A^{(i)}\right] \lambda_{(i)}(\tau_0)$, viz.,

$$\sqrt{\delta_{ij} \left[y_B^{(i)} - y_A^{(i)}\right] \left[y_B^{(j)} - y_A^{(j)}\right]} \, .$$

The fact that this is independent of τ does not prove Born rigidity.

Rigidity according to Sect. 2.3.1 means that neighbouring worldlines with fixed space coordinates are always the same proper distance apart *as measured in the instantaneous rest frame of either*. We need therefore to examine the proper distance between A and B as measured in the instantaneous rest frame of A, for example, which depends on the motion of A. The frame K' is nevertheless rigid in this sense, ultimately because the acceleration matrix is constant, but this is non-obvious and requires more work.

We have already done this work in Sect. 2.3.2, however. The key will be (2.94). Recall that A is a constant matrix, i.e., independent of proper time τ, and so of course is the matrix $\hat{\lambda}$ of (2.85). This means that the matrix $\tilde{A}$ in (2.94) is also independent of τ. But (2.94) corresponds exactly to the key relation (2.13) in Sect. 2.3 (see p. 22), viz.,

$$c\dot{n}_i{}^\mu = a_{0i} u^\mu + c\Omega_{ij} n_j{}^\mu \, , \tag{2.98}$$

where we have reinstated c and replaced the notation u_0 for the 4-velocity of the observer by the present notation u, recalling that we made the latter dimensionless.

Now we have the correspondence $n_i \leftrightarrow \lambda_{(i)}$, $i = 1,2,3$, while $u \leftrightarrow \lambda_{(0)}$. By (2.14) on p. 22, we also have

$$a_{0i} = -cn_i \cdot \dot{u} \ , \tag{2.99}$$

which means that

$$c\dot{u} = a_{0i} n_i \ , \tag{2.100}$$

since $\dot{u}$ is orthogonal to u.

So, in the notation of Sect. 2.3, which was a completely general construction using any smoothly chosen tetrad along the worldline, and for an arbitrary smooth timelike worldline, the relation

$$c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\nu)} \tilde{A}^{(\nu)}{}_{(\kappa)} \tag{2.101}$$

is replaced by

$$\begin{cases} c\dot{\lambda}_{(i)} = a_{0i}\lambda_{(0)} + c\Omega_{ij}\lambda_{(j)} \ , \\ c\dot{\lambda}_{(0)} = a_{0i}\lambda_{(i)} \ . \end{cases} \tag{2.102}$$

We can now read off the matrix $\tilde{A}$, obtaining

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \ , \tag{2.103}$$

with ν specifying the row and κ the column.

It looks at first glance as though there may be a proviso here: it looks as though $\tilde{A}$ may be a snapshot of the matrix on the right-hand side of (2.103), taken in the initial ICIF $\hat{\lambda}$, whereas the matrix on the right-hand side looks as though it may be taken relative to ICIF(τ) for each τ. But note that we have $\lambda^{\mu}_{(\nu)}(\tau)$ on the right-hand side of (2.101), so it really is another version of the relations (2.102).

In any case, when $\lambda(\tau)$ is obtained by isometric transport, as in (2.83) on p. 41, viz.,

$$\lambda(\tau) = \exp(A\tau/c)\hat{\lambda} \ , \tag{2.104}$$

we get the same result for $\tilde{A}$ no matter what ICIF$(\tau) =: K_\tau$ is used to reexpress A, since

$$\begin{aligned} \lambda^{-1}(\tau)A\lambda(\tau) &= \hat{\lambda}^{-1}\exp(-A\tau/c)A\exp(A\tau/c)\hat{\lambda} \\ &= \hat{\lambda}^{-1}A\hat{\lambda} = \tilde{A} \ . \end{aligned} \tag{2.105}$$

Returning to the above identification of the matrix $\tilde{A}$ with the matrix on the right-hand side of (2.103), we can immediately deduce what we need to know here in order to prove that we have another rigid frame by this construction, despite the evident fact that the initial tetrad need not be FW transported along the worldline, since we are not assuming $\Omega_{ij} = 0$ for all $i, j \in \{1,2,3\}$. The point is that A is a constant matrix if and only if $\tilde{A}$ is a constant matrix, and this is true if and only if a_{0i} and Ω_{ij} are constant for all $i, j \in \{1,2,3\}$. This corresponds exactly to superhelical motion as introduced on p. 25.

At least, we have shown that the theory in [23] always leads to cases of superhelical motion, but it is not yet entirely clear that superhelical motion always corresponds to a case of GUA with isometrically transported triad. After all, if we begin with the relations (2.102), we obtain a relation like (2.101), viz.,

$$c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\nu)}\underline{A}^{(\nu)}{}_{(\kappa)} \,, \tag{2.106}$$

where $\underline{A}$ is the constant matrix

$$\underline{A}^{(\nu)}{}_{(\kappa)} := \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \,, \tag{2.107}$$

with ν specifying the row and κ the column, but we have not said anything about how the space triad should be propagated along the worldline. Superhelical motion occurs when the Ω_{ij} are not necessarily zero, but all the a_{0i} and Ω_{ij} are constant, but to show that we have GUA according to the definition, we need to show that we have

$$c\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}u^{\nu} \,, \tag{2.108}$$

for some constant matrix A and for some choice of inertial frame, and we also need to know that the space triad $\{\hat{\lambda}_{(i)}\}_{i=1,2,3}$ has been transported isometrically according to the rule

$$c\frac{\mathrm{d}\lambda^{\mu}_{(i)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(i)} \,, \quad i = 1,2,3 \,. \tag{2.109}$$

This needs to be carefully considered if we are to claim that superhelical motion corresponds precisely to the general GUA construction of Friedman and Scarr.

We can see how to carry out this construction. Starting with (2.106) and (2.107), we have the solution

$$\lambda^{\mu}_{(\kappa)}(\tau) = \lambda^{\mu}_{(\nu)}(0)\big[\exp(\underline{A}\tau/c)\big]^{\nu}_{\;(\kappa)} \,, \tag{2.110}$$

and we define $\hat{\lambda}^{\mu}_{(\nu)} := \lambda^{\mu}_{(\nu)}(0)$, which is basically the initial ICIF, whence

$$\lambda^{\mu}_{(\kappa)}(\tau) = \hat{\lambda}^{\mu}_{(\nu)}\big[\exp(\underline{A}\tau/c)\big]^{\nu}_{\;(\kappa)} \,. \tag{2.111}$$

Since we expect $\underline{A}$ to correspond to the matrix $\tilde{A}$ in our previous discussion, we now know how we must define the matrix A by looking at (2.92) and (2.93) on p. 42:

$$A := \hat{\lambda}\underline{A}\hat{\lambda}^{-1} , \tag{2.112}$$

or in component form

$$A^{\nu}{}_{\sigma} := \hat{\lambda}^{\nu}_{(\gamma)}\underline{A}^{(\gamma)}{}_{(\kappa)}\big(\hat{\lambda}^{-1}\big)^{(\kappa)}{}_{\sigma} . \tag{2.113}$$

Note that A is constant, i.e., independent of τ, because the matrix $\hat{\lambda}$ is independent of τ. Now what we hope is that

$$c\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}u^{\nu} , \tag{2.114}$$

and that $\{\lambda_{(i)}\}_{i=1,2,3}$ is obtained by isometric transport, i.e., by solving

$$c\frac{\mathrm{d}\lambda^{\mu}_{(i)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(i)} , \quad i = 1,2,3 . \tag{2.115}$$

Since $u = \lambda_{(0)}$, satisfying the last two equations amounts to satisfying

$$c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(\kappa)} , \quad \kappa = 0,1,2,3 . \tag{2.116}$$

Let us drop indices and work from (2.111) and (2.112) in the form

$$\lambda = \hat{\lambda}\exp(\underline{A}\tau/c) , \qquad A = \hat{\lambda}\underline{A}\hat{\lambda}^{-1} . \tag{2.117}$$

These imply

$$\begin{aligned}\lambda\hat{\lambda}^{-1} &= \hat{\lambda}\exp(\underline{A}\tau/c)\hat{\lambda}^{-1}\\ &= \exp(\hat{\lambda}\underline{A}\hat{\lambda}^{-1}\tau/c)\\ &= \exp(A\tau/c) ,\end{aligned}$$

whence

$$\lambda = \exp(A\tau/c)\hat{\lambda} , \tag{2.118}$$

and this differentiates to give the required result

$$c\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = A\lambda .$$

Note that the above argument generalises in a certain sense to non-constant acceleration tensors [see Sect. 2.11.4, and in particular the discussion around (2.319) on p. 100 and following].

In the present case, the conclusion from this is that the rigid motion described earlier as superhelical is precisely the motion of fixed space points in the coordinate system constructed by Friedman and Scarr for an observer with generalised uniform acceleration (GUA) in the case where the rotational part of the acceleration matrix is not zero.

Note finally that we do expect the type (2,0) object $\tilde{A}_{(\kappa)(\nu)}$ to be antisymmetric, since

$$\begin{aligned}\lambda_{(\kappa)}\cdot\lambda_{(\nu)}=\eta_{\kappa\nu} \quad &\Longrightarrow \quad \dot{\lambda}_{(\kappa)}\cdot\lambda_{(\nu)}+\lambda_{(\kappa)}\cdot\dot{\lambda}_{(\nu)}=0\\ &\Longrightarrow \quad \Big[\lambda_{(\mu)}\tilde{A}^{(\mu)}{}_{(\kappa)}\Big]\cdot\lambda_{(\nu)}+\lambda_{(\kappa)}\cdot\Big[\lambda_{(\mu)}\tilde{A}^{(\mu)}{}_{(\nu)}\Big]=0\\ &\Longrightarrow \quad \eta_{\mu\nu}\tilde{A}^{(\mu)}{}_{(\kappa)}+\eta_{\kappa\mu}\tilde{A}^{(\mu)}{}_{(\nu)}=0\\ &\Longrightarrow \quad \tilde{A}_{(\nu)(\kappa)}+\tilde{A}_{(\kappa)(\nu)}=0\,. \end{aligned} \tag{2.119}$$

The type (2,0) object $\tilde{A}_{(\kappa)(\nu)}:=\eta_{\kappa\mu}\tilde{A}^{(\mu)}{}_{(\nu)}$ we obtain from (2.103) is

$$\tilde{A}_{(\kappa)(\nu)}=\begin{pmatrix}0 & \mathbf{a}_0^{\mathrm{T}}\\ -\mathbf{a}_0 & c\Omega\end{pmatrix}\,, \tag{2.120}$$

where κ labels rows and ν labels columns, and

$$\Omega:=\begin{pmatrix}\Omega_{11} & \Omega_{12} & \Omega_{13}\\ \Omega_{21} & \Omega_{22} & \Omega_{23}\\ \Omega_{31} & \Omega_{32} & \Omega_{33}\end{pmatrix}\,. \tag{2.121}$$

The matrix $\tilde{A}_{(\kappa)(\nu)}$ is indeed antisymmetric.

2.4.6 Reduction to FW Transport

We also deduce immediately from (2.120) that the isometric transport specified by (2.82) reduces to FW transport in the initial ICIF

$$\hat{\lambda}=\{\hat{\lambda}_{(\kappa)}\}=\big(u(0),n_1(0),n_2(0),n_3(0)\big)$$

precisely in the case where $\Omega_{ij}=0$ for all $i,j\in\{1,2,3\}$.

Note that, if the initial ICIF happens to be the original inertial laboratory frame K, i.e., if $u(0)=(1,0,0,0)$ and $A=\tilde{A}$, then we have

$$\begin{pmatrix}0 & \mathbf{g}^{\mathrm{T}}\\ \mathbf{g} & c\pi(\omega)\end{pmatrix}=\begin{pmatrix}0 & \mathbf{a}_0^{\mathrm{T}}\\ \mathbf{a}_0 & -c\Omega\end{pmatrix}\,,$$

using (2.75) on p. 38, whence it follows that $\mathbf{a}_0=\mathbf{g}$ and $\Omega=-\pi(\omega)$.

It is important to note here that we may have ordinary translational uniform acceleration (TUA) in the laboratory frame K, as in (2.77) on p. 39, but when we look at that in the initial ICIF, having carried out the Lorentz transformation (2.87), viz., $\tilde{A} := \hat{\lambda}^{-1} A \hat{\lambda}$, we will generally obtain an acceleration matrix $\tilde{A}$ with some rotational terms. The situation in which the isometric transport (2.82) reduces to FW transport is one in which $\tilde{A}$ has the purely translational form, and not in general one in which A has that form.

Note in this context that it may happen that the observer is at rest at some event in the laboratory frame K. We can then construct our frame from that event, taking the orthonormal basis of K to be the initial ICIF. In this case, still assuming that the motion looks like TUA motion in K, the isometric transport (2.82) will reduce to FW transport of that frame along the worldline. But this is not the only case when TUA motion in the laboratory frame can still be TUA motion in some initial ICIF. We merely require there to be some event E on the worldline and some ICIF at E such that the transformation $A \to \tilde{A} := \hat{\lambda}^{-1} A \hat{\lambda}$ preserves the translational form of the matrix.

A whole normal subgroup of such homogeneous Lorentz transformations, namely the little group associated with the direction of the 3-vector $\mathbf{g}$, will actually leave the acceleration matrix completely unchanged. This subgroup contains all rotations around the direction of $\mathbf{g}$ and all boosts in that direction. But in fact any pure space rotation will transform a purely translational acceleration matrix into another. Using L to denote the Lorentz transformation $\hat{\lambda}$, we would have

$$L = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} , \qquad L^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & R^{\mathrm{T}} \end{pmatrix} ,$$

where R is a 3×3 rotation matrix with inverse $R^{-1} = R^{\mathrm{T}}$, whence

$$\tilde{A} = L^{-1} A L = \begin{pmatrix} 1 & 0 \\ 0 & R^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} = \begin{pmatrix} 0 & (R^{\mathrm{T}}\mathbf{g})^{\mathrm{T}} \\ R^{\mathrm{T}}\mathbf{g} & 0 \end{pmatrix} . \tag{2.122}$$

We can also see that any boost not aligned with $\mathbf{g}$ will spoil the purely translational aspect of the acceleration matrix. Here it is instructive to look at some simple cases of the effects of Lorentz boosts on purely translational acceleration matrices. A Lorentz boost in the x direction has the form

$$L = \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

where $c := \cosh\beta$ and $s = \sinh\beta$, and $\beta = v/c$, with v the speed of the boost. This satisfies $L^{\mathrm{T}} = L$. Let us see first what happens when we boost a purely translational acceleration matrix of the form

$$\bar{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} ,$$

i.e., representing pure TUA in the same direction as the boost. We find

$$\begin{aligned} \tilde{\bar{A}} = L^{\mathrm{T}} \bar{A} L &= \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s & c & 0 & 0 \\ -c & -s & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \end{aligned} \tag{2.123}$$

since $c^2 - s^2 = 1$. So $\tilde{\bar{A}} = \bar{A}$ in this case. In other words, TUA motion in a given direction always looks exactly the same from any inertial frame boosted in that same direction.

Now let us see what happens when we view a purely translational acceleration matrix of the form

$$\bar{A} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} ,$$

from an inertial frame boosted in a space direction orthogonal to the direction of translational acceleration, e.g., using the above boost

$$L = \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

in the x direction. We shall see that this introduces terms into the rotational sector of the acceleration matrix. This time we calculate

$$\tilde{\tilde{A}} = L^{\mathrm{T}} \bar{A} L = \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} c & s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -c & -s & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 & 0 & c \\ 0 & 0 & 0 & s \\ 0 & 0 & 0 & 0 \\ -c & -s & 0 & 0 \end{pmatrix}, \tag{2.124}$$

so this does indeed introduce terms into the rotational sector of the acceleration matrix.

We can now claim that TUA motion in the laboratory frame K will be TUA motion in an initial ICIF at some event E on the worldline if the associated transformation $\hat{\lambda}$ from K to that initial ICIF involves only a boost along the 3-vector $\mathbf{g}$ associated with the acceleration matrix as expressed relative to K followed by a space rotation.

But suppose now that we have found such an event and initial ICIF and use the isometric transport (2.82) to carry the basis of that ICIF along to some other potential starting point E' on the worldline. We will obtain a tetrad ICIF$'$ there which could certainly have been used as the initial ICIF because the only thing we require of that ICIF is that the timelike basis vector should be the 4-velocity there, and the isometric transport (2.82) guarantees that by construction for this kind of worldline.

Now since the isometric transport (2.82) coincides with FW transport in this case, this suggests that the transformation directly from K to ICIF$'$ must keep the acceleration matrix in the TUA form. If this were not the case, we might have started with event E' and ICIF$'$, transformed our acceleration matrix A, and obtained an $\tilde{A}'$ with rotational entries. Then we would have said that the isometric transport (2.82) did *not* coincide with FW transport and we would appear to have a contradiction with the idea that the basis of the ICIF at E is getting FW transported along the worldline by the process laid down by (2.82).

To reassure ourselves that everything is consistent here, we need to prove the following. Whenever there exists an event E on the worldline and an initial ICIF $\hat{\lambda}$ such that the transformation from K to this ICIF involves only a boost along the 3-vector $\mathbf{g}$ associated with the acceleration matrix as expressed relative to K followed by a space rotation, then for any other event E' and the isometric transport (or equivalently the FW transport) of $\hat{\lambda}$ to a new ICIF denoted by ICIF$'$ at E', the transformation from K to ICIF$'$ also involves only a boost along the 3-vector $\mathbf{g}$ associated with the acceleration matrix as expressed relative to K followed by a space rotation, whence the expression $\tilde{A}'$ of the acceleration matrix relative to ICIF$'$ will still be purely translational.

Here is the proof. By (2.83) on p. 41, we have a basis $\hat{\lambda}'$ for ICIF$'$ given by

$$\hat{\lambda}' := \lambda(E') = \exp(A\tau_0/c)\hat{\lambda} ,$$

for some fixed $\tau_0 \in \mathbb{R}$. But then

$$\begin{aligned}\tilde{A}' &= \hat{\lambda}'^{-1}A\hat{\lambda}' \\ &= \big[\exp(A\tau_0/c)\hat{\lambda}\big]^{-1}A\big[\exp(A\tau_0/c)\hat{\lambda}\big] \\ &= \hat{\lambda}^{-1}\big[\exp(-A\tau_0/c)A\exp(A\tau_0/c)\big]\hat{\lambda} \\ &= \hat{\lambda}^{-1}A\hat{\lambda} = \tilde{A} . \end{aligned} \tag{2.125}$$

So the two matrices $\tilde{A}'$ and $\tilde{A}$ are actually one and the same, which proves the claim.

The reader should check that the homogeneous Lorentz transformation from K to the basis $\hat{\lambda}'$ for ICIF$'$ is indeed found by multiplying the two matrices $\exp(A\tau_0/c)$ and $\hat{\lambda}$, something that looks obvious only because we are using the notation $\hat{\lambda}'$ to denote two different things, viz., the basis for ICIF$'$ and the homogeneous Lorentz transformation from K to the basis $\hat{\lambda}'$.

Here we have been considering the case where both the acceleration matrix A and its version $\tilde{A}$ relative to some initial ICIF had purely translational form. But what concerns us when we seek cases in which the isometric transport (2.82) along the worldline coincides with FW transport is just that the second matrix $\tilde{A}$ should have purely translational form, *so that the motion looks like standard TUA in one of its instantaneous rest frames*.

There is a more general argument here. Whatever ICIF $\hat{\lambda}'$ we choose at a given event E on the worldline, it must have the same timelike basis vector as any other ICIF $\hat{\lambda}''$ at the same event, so the two ICIFs are related by a pure space rotation. So if the acceleration matrix has purely translational form $\tilde{A}'$ relative to the first ICIF, it will have purely translational form $\tilde{A}''$ relative to the second ICIF.

Likewise the argument (2.125) shows that, whatever the form of the matrix A, if we select some event E on the worldline and ICIF $\hat{\lambda}$ at that event, and if we obtain a purely translational $\tilde{A}$ there, then we will obtain a purely translational $\tilde{A}'$ at any other preselected event E' on the worldline for at least one choice of ICIF there, namely the ICIF we denoted by ICIF$'$ which was obtained by the isometric transport (2.82) (or equivalently by FW transport in this case) of $\hat{\lambda}$ from E. And hence by the last paragraph, we will obtain a purely translational $\tilde{A}'$ at E' for any choice of ICIF there.

These arguments reassure us that the conclusion that our isometric transport (2.82) along the worldline coincides with FW transport does not depend on the event E at which we choose to obtain $\tilde{A}$, nor on the choice of ICIF there.

2.4.7 Extent of Generalisation

Whatever acceleration matrix A gets associated with the worldline when we express this matrix relative to a given choice of laboratory inertial frame K, there may be another choice of inertial frame such that the matrix takes on a purely translational form. In other words, Friedman and Scarr's generalisation may look like a generalisation for one observer but not for another. However, we shall see here that there are plenty of cases where it really is a generalisation in the sense that there is no other choice of inertial frame such that the matrix takes on a purely translational form.

The point is that, relative to any other choice of laboratory inertial frame K_1, the acceleration matrix will have the form

$$A_1 = L^{-1} A L \,, \tag{2.126}$$

where L is the homogeneous Lorentz transformation from K to K_1. Now for certain forms of the antisymmetric matrix A, there will be some L such that A_1 has purely translational form. However, we can provide a general argument that shows why A will not always be reducible to purely translational form by any transformation of type (2.126).

An equivalent statement is this. If we begin with a purely translational A_1, we could generate a whole range of possible acceleration matrices A by considering matrices LA_1L^{-1} as L ranges over the whole group of homogeneous Lorentz transformations. We may as well begin with

$$A = \begin{pmatrix} 0 & g & 0 & 0 \\ g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \,, \tag{2.127}$$

for some arbitrary g. We can generate all matrices

$$A = \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & 0 \end{pmatrix} \,, \tag{2.128}$$

by suitable space rotations [for example, see (2.122) on p. 49]. This already gives us a 3-parameter subset of the 6-parameter set of all matrices of the kind we are interested in. [The type (1,1) matrices $A^{\mu}{}_{\nu}$ are not antisymmetric, but the space of such matrices obviously has the same dimension as the closely related space of antisymmetric matrices.] But the three other parameters introduced by Lorentz boosting in the three space directions will never completely fill out the rotational sector of A.

In brief, the key observation here is that, although the set of all possible acceleration matrices is specified by 6 parameters, and this is precisely the number of parameters required to specify all elements in the homogeneous Lorentz group, the matrix (2.127), or indeed the matrix (2.128), is rank 2, i.e., both these matrices have only two linearly independent columns, and the same goes for any matrix obtained from (2.127) by sandwiching it between L and L^{-1}. But the general antisymmetric 4×4 matrix has rank 4.

What can we conclude from this? It is important to be very clear about this, so let us think about what is involved in the claim that generalised uniform acceleration really is a generalisation. The first point to consider is that, given some TUA motion, i.e., standard uniform acceleration in a straight line, as described by (2.77) on p. 39, viz.,

$$c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = A^\mu{}_\nu u^\nu = \begin{pmatrix} 0 & \mathbf{g}^\mathrm{T} \\ \mathbf{g} & 0 \end{pmatrix} \begin{pmatrix} u^0 \\ \mathbf{u}/c \end{pmatrix} , \tag{2.129}$$

so that

$$c\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = (c^{-1}\mathbf{g}\cdot\mathbf{u}, \mathbf{g}u^0) , \tag{2.130}$$

we could always try to view this motion relative to some other inertial frame. We know from Sect. 2.4.2 how we would transform this equation of motion in order to describe the worldline relative to the new inertial frame: the answer is that we will get *the same equation* expressed relative to the new frame if we transform the object so suggestively written as $A^\mu{}_\nu$ as a type (1,1) tensor. Indeed, if we are to obtain the same equation expressed relative to the new frame, *it has to transform like this* because the left-hand side of (2.129) transforms as a contravector, and so does u.

So there is a sense in which, in many cases, we have not generalised anything at all. In many cases, all we have done is to recognise how to express uniform acceleration when it does not look like TUA motion because we are viewing it from the wrong kind of inertial frame. But we do know from the discussion above that Friedman and Scarr's generalisation really does generate other cases.

Furthermore, their frame construction is not always the same as the one obtained by FW transport of an initial ICIF, even when the motion looks like TUA motion in some inertial frame, *the point being that it has to look like TUA motion in an ICIF*. We can see exactly when it differs by considering the solution (2.80) of (2.129) on p. 39, viz.,

$$\mathbf{u} = \mathbf{g}t + \mathbf{u}_0 , \tag{2.131}$$

where $\mathbf{u}_0$ is the value of $\mathbf{u}$ at time $t = 0$. So let us suppose that we are in a case where there is some choice of inertial frame such that the generalised uniform acceleration (GUA) looks like (2.131).

Now if $\mathbf{g}$ and $\mathbf{u}_0$ happen to be parallel, there will even be a better choice of inertial frame, such that the motion has the form

$$\mathbf{u} = \mathbf{g}t , \tag{2.132}$$

in which the observer is at rest at inertial time $t = 0$. We can just carry out a simple time translation, or we can do a Lorentz boost in the $\mathbf{g}$ direction. We have seen in Sect. 2.4.6 that such a boost does not alter the acceleration matrix in the TUA case, a remarkable property of TUA motion [see (2.123) on p. 50].

In fact it is precisely in these cases, where $\mathbf{g}$ and $\mathbf{u}_0$ happen to be parallel in some inertial frame for which the motion looks like TUA motion, that Friedman and Scarr's isometric transport (2.82) on p. 41 is the same as FW transport. If **g** and $\mathbf{u}_0$ are not parallel in some inertial frame for which the motion looks like TUA motion, they will not be parallel in any inertial frame for which the motion looks like TUA motion, and then we obtain a superhelical frame construction that does not coincide with the standard rigid FW transported frame construction.

It is the two claims in the last sentence that need to be proven. The first claim says that, if there is some inertial frame for which the motion has the TUA form (2.132) with the observer coming to rest at some point, then in any other inertial frame for which the motion has the TUA form, she will also come to rest at some point. But as we have seen in Sect. 2.4.6, it is straightforward to establish that the TUA form of an acceleration matrix is preserved only by boosts in the relevant **g** direction followed by arbitrary space rotations, so this claim is clearly true.

The second claim follows from the first because, if there is an inertial frame for which the motion takes the form (2.131) with **g** not parallel to $\mathbf{u}_0$, then by the first claim there is no inertial frame in which the motion takes the TUA form and the observer comes to rest, and we know that there is always such a frame when the isometric transport (2.82) reduces to FW transport, because the latter happens precisely and only when the acceleration matrix has purely TUA form in some ICIF.

When we have the situation in (2.131) and **g** and $\mathbf{u}_0$ are not parallel, we can choose an inertial frame in which they are actually orthogonal in the spatial hypersurface, simply by carrying out a time translation. In order to get the initial velocity to zero, we then need to do a boost perpendicular to **g**, and we have seen in Sect. 2.4.6 that this would introduce terms into the rotational sector of the acceleration matrix [see in particular (2.124) on p. 51].

2.4.8 Metric for Friedman–Scarr Coordinates

These coordinates are obtained by transporting a tetrad from some initial point on the observer worldline to all other points along it and then carrying out the general construction for an SE coordinate frame. We can thus use the general theory developed in Sect. 2.3.8. We begin with the matrix $\tilde{A}$ given in (2.103) on p. 45, viz.,

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} , \tag{2.133}$$

where ν specifies the row and κ the column. Then relative to the coordinates $\{y^{(\kappa)}\}_{\kappa=0,1,2,3}$, the metric established in Sect. 2.3.8 takes the form

$$g_{(\mu)(\nu)} = \begin{pmatrix} \left[1+y^{(i)}a_{0i}\right]^2 - y^{(i)}y^{(j)}\Omega_{ik}\Omega_{jk} & y^{(i)}\Omega_{i1} & y^{(i)}\Omega_{i2} & y^{(i)}\Omega_{i3} \\ y^{(i)}\Omega_{i1} & -1 & 0 & 0 \\ y^{(i)}\Omega_{i2} & 0 & -1 & 0 \\ y^{(i)}\Omega_{i3} & 0 & 0 & -1 \end{pmatrix} . \tag{2.134}$$

Note how this matrix is always independent of the temporal coordinate $y^{(0)}$, and the a_{0i} and Ω_{ij} are just temporal constants for generalised uniform acceleration (GUA).

This is enough to conclude something that will be rather important for later discussions of the physical interpretation of such coordinate frames, namely that $\partial_{y^{(0)}}$ is a Killing vector field for every such coordinate construction for GUA motion. A Killing vector field X is one such that the Lie derivative $L_X g$ of the metric along the flow curves of X is zero.

To prove this claim, we may use the general coordinate formula for the Lie derivative as given in [27]. For any contravariant vector field X, we have

$$(L_X g)_{(\eta)(\phi)} = \frac{\partial g_{(\eta)(\phi)}}{\partial y^{(\iota)}} X^{(\iota)} + g_{(\iota)(\phi)} \frac{\partial X^{(\iota)}}{\partial y^{(\eta)}} + g_{(\eta)(\iota)} \frac{\partial X^{(\iota)}}{\partial y^{(\phi)}} . \tag{2.135}$$

We then take $X = \partial_{y^{(0)}}$ which has components $X^{(0)} = 1$, $X^{(i)} = 0$, $i = 1,2,3$, in these coordinates. Hence,

$$(L_X g)_{(\eta)(\phi)} = \frac{\partial g_{(\eta)(\phi)}}{\partial y^{(0)}} X^{(0)} = 0 ,$$

as claimed. We can thus say that all observers sitting at fixed space coordinate positions in these frames are Killing observers.

Explicit examples of this kind of metric are given for the case of translational uniform acceleration in Sect. 2.9 and for the case of uniform circular motion in Sect. 2.10.

Any spacetime with a metric of the form (2.134) has a globally defined timelike Killing vector field and is said to be stationary (see also Sect. 4.3.17). If in addition only the diagonal elements are nonzero, as happens when all the Ω_{ij} are zero and we have translational uniform acceleration, the spacetime is said to be static. Of course, this is the flat Minkowski spacetime so we already know that it is static. What we discover here is the plethora of Killing vector fields that can be used to get the Minkowski metric into the stationary or static forms.

Regarding singularities of the matrix of metric components (2.134), the comments at the end of Sect. 2.3.8 pertain exactly. In particular, we have (2.67) on p. 34, viz.,

$$\boxed{\det g_{\mathrm{SE}}^{\mathrm{Mink}} = -\left(1+\xi^i a_{0i}\right)^2} , \tag{2.136}$$

which is always independent of the rotation chosen for the space triad $\{n_i\}_{i=1,2,3}$, as specified by Ω_i, $i = 1,2,3$, but does depend on the acceleration of the worldline as specified by its absolute components a_{0i}, $i = 1,2,3$. This determinant is zero for all ξ^i satisfying

$$\xi^i a_{0i}(\sigma) = -1 ,$$

for some value of the proper time σ of the observer. The proposed coordinates could not be extended to such points.

2.4.9 More about Observers at Fixed Space Coordinates

A more general question is whether these Killing observers sitting at fixed space coordinates in the $\{y^{(\kappa)}\}_{\kappa=0,1,2,3}$ system actually have GUA motion. In order to tackle this, we need to know the proper time of these observers.

Here we can also use the general theory of semi-Euclidean coordinate systems in Sect. 2.3. Recall that this analysis considers a space triad $\{n_i\}_{i=1,2,3}$ that is smoothly transported along the observer worldline, without assuming anything other than smoothness about the transport. Furthermore, we have made the link with the quantities a_{0i} and Ω_{ij} in the relations (2.98) on p. 44, viz.,

$$c\dot{n}_i^{\ \mu} = a_{0i}u^\mu + c\Omega_{ij}n_j^{\ \mu} \ . \tag{2.137}$$

So as on p. 45 ff we have the correspondence $n_i \leftrightarrow \lambda_{(i)}$, $i = 1,2,3$, while $u \leftrightarrow \lambda_{(0)}$ and

$$a_{0i} = -cn_i \cdot \dot{u} \ , \tag{2.138}$$

which means that

$$c\dot{u} = a_{0i}n_i \ , \tag{2.139}$$

since $\dot{u}$ is orthogonal to u. Then, in this notation, which was a completely general construction using any smoothly chosen tetrad along the worldline, and for an arbitrary smooth timelike worldline, the relation

$$c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\nu)}\tilde{A}^{(\nu)}{}_{(\kappa)}$$

is replaced by

$$\begin{cases} c\dot{\lambda}_{(i)} = a_{0i}\lambda_{(0)} + c\Omega_{ij}\lambda_{(j)} \ , \\ c\dot{\lambda}_{(0)} = a_{0i}\lambda_{(i)} \ , \end{cases} \tag{2.140}$$

and we read off the matrix $\tilde{A}$ as

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \ , \tag{2.141}$$

with ν specifying the row and κ the column. The specific feature of GUA motion is that the a_{0i} and Ω_{ij} are actually independent of the proper time along the observer worldline. It is also important to note that the point of contact between this analysis and Friedman and Scarr's is through $\tilde{A}$, the expression for the acceleration matrix relative to any ICIF for the main observer, rather than through A, the expression for the acceleration matrix relative to some arbitrary laboratory inertial frame.

Now it is established in Sect. 2.3 that the 4-velocity of an observer sitting at fixed $\xi^i \leftrightarrow y^{(i)}$ in the accelerating frame is [see (2.15) on p. 22]

$$u^\mu(\xi,\tau) = \Big[\big(1+\xi^i a_{0i}\big)u_0{}^\mu + \xi^i \Omega_{ij} n_j{}^\mu\Big]\dot{\sigma}\ , \tag{2.142}$$

where τ is the proper time for the observer at ξ and $\sigma(\xi,\tau)$ is the corresponding proper time of the main observer, corresponding in the sense that, at that proper time, the main observer considers the observer at ξ to be simultaneous. The dot on σ denotes the derivative with respect to τ, keeping ξ fixed, so it is the time dilation effect between the two observers, something we encounter again in later sections.

In fact, it was shown in (2.16) on p. 22 that

$$\dot{\sigma} = \Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{-1/2}\ . \tag{2.143}$$

Note that $\dot{\sigma}$ is constant for GUA motion, because then a_{0i} and Ω_{ij} are constant and we have fixed the ξ^i. So the full formula for the 4-velocity of the observer sitting at fixed ξ is

$$u^\mu(\xi,\tau) = \frac{\big(1+\xi^i a_{0i}\big)u_0{}^\mu + \xi^i \Omega_{ij} n_j{}^\mu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}}\ . \tag{2.144}$$

We shall encounter this again as a special case of a result derived in later sections (see in particular Sect. 2.6). For the moment, the index μ on each side refers to components relative to some arbitrary laboratory inertial frame.

We must now obtain the 4-acceleration $a^\mu(\xi,\tau)$ by differentiating $u^\mu(\xi,\tau)$ with respect to τ for fixed ξ. The aim will be to see whether the 4-acceleration can be obtained by multiplying the 4-velocity by some constant matrix. We have

$$\begin{aligned}
a^\mu(\xi,\tau) &= \left.\frac{\partial u^\mu(\xi,\tau)}{\partial\tau}\right|_\xi \\
&= \frac{\big(1+\xi^i a_{0i}\big)\dot{u}_0{}^\mu + \xi^i \Omega_{ij} \dot{n}_j{}^\mu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}}\dot{\sigma} \\
&= \frac{\big(1+\xi^i a_{0i}\big)A^\mu{}_\nu u_0{}^\nu + \xi^i \Omega_{ij} A^\mu{}_\nu n_j{}^\nu}{\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}} \\
&= \frac{A^\mu{}_\nu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}} u^\nu(\xi,\tau)\ ,
\end{aligned} \tag{2.145}$$

using the fact that $\dot{u}_0{}^\mu = A^\mu{}_\nu u_0{}^\nu$ and $\dot{n}_j{}^\mu = A^\mu{}_\nu n_j{}^\nu$, where $A^\mu{}_\nu$ is the version of the constant acceleration matrix expressed relative to the laboratory inertial frame.

We conclude that an observer sitting at fixed ξ^i in the Friedman–Scarr accelerating frame would indeed have generalised uniform acceleration, with acceleration matrix

$$\boxed{A^\mu{}_\nu(\xi) = \frac{A^\mu{}_\nu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk}\Big]^{1/2}}}\,, \tag{2.146}$$

as expressed relative to the laboratory inertial frame. When the latter is the initial instantaneously comoving inertial frame ICIF(0) of the main observer (or as we know, any ICIF of the main observer and hence also of the observer at ξ), we have

$$\boxed{\tilde{A}^\mu{}_\nu(\xi) = \frac{1}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk}\Big]^{1/2}}\begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix}}\,, \tag{2.147}$$

although it is not necessary to see this form in order to prove the above claim.

In the above calculation, we could have complicated things by writing

$$\begin{aligned} a^\mu(\xi,\tau) &= \left.\frac{\partial u^\mu(\xi,\tau)}{\partial\tau}\right|_\xi \\ &= \frac{\big(1+\xi^i a_{0i}\big)\dot{u}_0{}^\mu + \xi^i\Omega_{ij}\dot{n}_j{}^\mu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk}\Big]^{1/2}}\dot{\sigma} \\ &= \frac{\big(1+\xi^i a_{0i}\big)a_0{}^\mu + \xi^i\Omega_{ij}\big(a_{0j}u_0{}^\mu + \Omega_{jk}n_k{}^\mu\big)}{\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk}}\,. \end{aligned} \tag{2.148}$$

It is interesting to consider what this looks like in an ICIF for the main observer, the obvious one being ICIF(σ), i.e., $\{u_0(\sigma), n_i(\sigma)\}_{i=1,2,3}$ for $\sigma(\xi,\tau)$, in which the observer at ξ appears to be simultaneous at her proper time τ. Of course, this ICIF will change with τ. If we can show that

$$a^\mu(\xi,\tau) = \tilde{A}^\mu{}_\nu(\xi,\tau)u^\nu(\xi,\tau) \quad \text{in ICIF}\big(\sigma(\xi,\tau)\big)\,,$$

where $\tilde{A}^\mu{}_\nu(\xi,\tau)$ is actually independent of τ, then we are still left with the problem that we are expressing $a^\mu(\xi,\tau)$ and $u^\mu(\xi,\tau)$ relative to *different* inertial frames for each value of τ. There are nevertheless ways around this, so let us get our hands dirty with the prospect of a slightly deeper insight.

First of all, relative to ICIF(σ), or any ICIF, u_0 has the form

$$u_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\,,$$

and hence

$$a_0 = \tilde{A}u_0 = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ a_{01} \\ a_{02} \\ a_{03} \end{pmatrix} ,$$

where we use the fact that the acceleration matrix always has the constant form given here for any ICIF of the main observer [see (2.125) on p. 52].

We now have, from (2.144),

$$u(\xi,\tau) = \frac{1}{\left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]^{1/2}} \left[\left(1+\xi^i a_{0i}\right)\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \xi^i \Omega_{i1} \\ \xi^i \Omega_{i2} \\ \xi^i \Omega_{i3} \end{pmatrix}\right] ,$$

whence

$$\boxed{u(\xi,\tau) = \frac{1}{\left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]^{1/2}} \begin{pmatrix} 1+\xi^i a_{0i} \\ \xi^i \Omega_{i1} \\ \xi^i \Omega_{i2} \\ \xi^i \Omega_{i3} \end{pmatrix}} . \tag{2.149}$$

This shows that, unless all the Ω_{ij} are not just constant but actually zero, the hyperplane of simultaneity for the main observer at $\sigma(\xi,\tau)$ is never a hyperplane of simultaneity for the observer at ξ at her proper time τ, since $u(\xi,\tau)$ is never orthogonal to it. When the Ω_{ij} are all zero and we have TUA motion, we recover the well known result (see Sect. 2.9) that observers at fixed ξ share hyperplanes of simultaneity with the main observer in this precise sense.

This result should be compared with the HOS sharing effect already discussed in conjunction with (2.33) on p. 27 at the beginning of Sect. 2.3.4. There we showed that it occurs for any FW transported space triad, i.e., purely translational acceleration matrix relative to a suitable tetrad field along the worldline, even if the translational acceleration is not uniform. We can thus say that HOS sharing is a consequence of the rigidity assumption for the general SE frame construction, while all the Friedman–Scarr frame constructions for GUA motion are rigid but we only have HOS sharing when there is no rotation in the GUA motion.

Returning to the problem at hand and using (2.148) for $a^\mu(\xi,\tau)$, we have

$$a^\mu(\xi,\tau) = \frac{1}{\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}} \left[\left(1+\xi^i a_{0i}\right)\begin{pmatrix} 0 \\ a_{01} \\ a_{02} \\ a_{03} \end{pmatrix} + \xi^i \Omega_{ij}\begin{pmatrix} a_{0j} \\ \Omega_{j1} \\ \Omega_{j2} \\ \Omega_{j3} \end{pmatrix}\right] ,$$

whence

$$a^\mu(\xi,\tau) = \frac{1}{\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}} \begin{pmatrix} \xi^i \Omega_{ij} a_{0j} \\ \big(1+\xi^i a_{0i}\big)a_{01} + \xi^i \Omega_{ij}\Omega_{j1} \\ \big(1+\xi^i a_{0i}\big)a_{02} + \xi^i \Omega_{ij}\Omega_{j2} \\ \big(1+\xi^i a_{0i}\big)a_{03} + \xi^i \Omega_{ij}\Omega_{j3} \end{pmatrix} .$$

It is now a simple matter to check that

$$a(\xi,\tau) = \frac{1}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}} \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix} u(\xi,\tau) , \tag{2.150}$$

when $a^\mu(\xi,\tau)$ and $u^\mu(\xi,\tau)$ are expressed in component form relative to the inertial frame ICIF$\big(\sigma(\xi,\tau)\big)$.

How can we deduce that the observer at ξ is uniformly accelerating? As mentioned before, even though the matrix in the last relation is independent of τ, the problem is that the two 4-vectors are expressed relative to different inertial frames at each value of τ. But let L be the Lorentz transformation from ICIF$\big(\sigma(\xi,\tau)\big)$ to ICIF(0). We know from previous investigations that

$$L = \exp\big[-\tilde{A}(\xi=0)\sigma(\xi,\tau)\big] ,$$

where

$$\tilde{A}(\xi=0) = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix} .$$

We have

$$\bar{a}^\kappa(\xi,\tau) = L^\kappa{}_\mu a^\mu(\xi,\tau) , \qquad u^\nu(\xi,\tau) = (L^{-1})^\nu{}_\lambda \bar{u}^\lambda(\xi,\tau) ,$$

where bars over a and u denote versions of these 4-vectors expressed relative to ICIF(0). The relation (2.150) can now be written

$$\bar{a}^\kappa(\xi,\tau) = L^\kappa{}_\mu \frac{\tilde{A}(\xi=0)^\mu{}_\nu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i\xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}} (L^{-1})^\nu{}_\lambda \bar{u}^\lambda(\xi,\tau) .$$

However, it is clear that

$$L\tilde{A}(\xi=0)L^{-1} = \tilde{A}(\xi=0) ,$$

since L is obtained as an exponential of $\tilde{A}(\xi=0)$. Hence (2.150) becomes

$$\bar{a}(\xi,\tau) = \frac{1}{\left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]^{1/2}} \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix} \bar{u}(\xi,\tau)\ , \tag{2.151}$$

when $\bar{a}^\mu(\xi,\tau)$ and $\bar{u}^\mu(\xi,\tau)$ are expressed in component form relative to the constant inertial frame ICIF(0), i.e., a frame independent of τ. This finalises the second proof of the claim that an observer sitting at fixed ξ has GUA motion.

2.5 Velocity Transformations

The aim here is to consider an object or observer with arbitrary motion and relate the description of its 4-velocity in the laboratory inertial frame to the description of its 4-velocity in a semi-Euclidean coordinate system $\{y^{(\kappa)}\}_{\kappa=0,1,2,3}$. Here we use the notation and analysis in [23] for the case of generalised uniform acceleration, although the considerations will apply equally to the general construction of SE coordinates.

We thus begin with the relation (2.97) on p. 44, viz.,

$$x^\mu = \hat{x}^\mu(\tau) + y^{(i)}\lambda_{(i)}(\tau)\ . \tag{2.152}$$

Basically, we have found τ such that the event X with laboratory coordinates x^μ can be written in this way, and the coordinates of event X in the accelerating frame K' are defined as $(c\tau, y^{(1)}, y^{(2)}, y^{(3)})$. The relation (2.152) then tells us how to convert from these coordinates to the original laboratory coordinate system K.

We use (2.152) to relate small changes in the $y^{(\kappa)}$ to small changes in the x^μ, which leads to

$$\mathrm{d}x^\mu = \frac{1}{c}\frac{\mathrm{d}\hat{x}^\mu(\tau)}{\mathrm{d}\tau}\mathrm{d}y^{(0)} + \lambda^\mu_{(i)}(\tau)\mathrm{d}y^{(i)} + y^{(i)}\frac{1}{c}\frac{\mathrm{d}\lambda^\mu_{(i)}(\tau)}{\mathrm{d}\tau}\mathrm{d}y^{(0)}\ . \tag{2.153}$$

Now $\mathrm{d}\hat{x}^\mu(\tau)/c\mathrm{d}\tau$ is the 4-velocity of the main observer at her proper time τ, so

$$\frac{1}{c}\frac{\mathrm{d}\hat{x}^\mu(\tau)}{\mathrm{d}\tau} = \lambda^\mu_{(0)}(\tau)\ . \tag{2.154}$$

We thus have

$$\mathrm{d}x^\mu = \lambda^\mu_{(\kappa)}(\tau)\mathrm{d}y^{(\kappa)} + y^{(i)}\frac{1}{c}\frac{\mathrm{d}\lambda^\mu_{(i)}(\tau)}{\mathrm{d}\tau}\mathrm{d}y^{(0)}\ . \tag{2.155}$$

We also know that

$$c\frac{\mathrm{d}\lambda^\mu_{(\kappa)}(\tau)}{\mathrm{d}\tau} = \lambda^\mu_{(\nu)}\tilde{A}^{(\nu)}{}_{(\kappa)}\ , \tag{2.156}$$

according to (2.94) on p. 43. Now we can rewrite the last term in (2.155) using

$$y^{(i)}\frac{1}{c}\frac{\mathrm{d}\lambda^{\mu}_{(i)}(\tau)}{\mathrm{d}\tau}=c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^{\mu}_{(\nu)}(\tau)\ , \tag{2.157}$$

where

$$\bar{y}:=x-\hat{x}(\tau)=y^{(i)}\lambda_{(i)}(\tau)\ . \tag{2.158}$$

The meaning of $(\tilde{A}\bar{y})^{(\nu)}$ is just

$$(\tilde{A}\bar{y})^{(\nu)}=\tilde{A}^{(\nu)}{}_{(\kappa)}\bar{y}^{(\kappa)}=\tilde{A}^{(\nu)}{}_{(i)}y^{(i)}\ .$$

Hence, finally,

$$\mathrm{d}x^{\mu}=\lambda^{\mu}_{(\kappa)}(\tau)\mathrm{d}y^{(\kappa)}+c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^{\mu}_{(\nu)}(\tau)\mathrm{d}y^{(0)}\ . \tag{2.159}$$

We now imagine a point object with motion $x^{\mu}(\tau_{\mathrm{p}})$, as described relative to the laboratory inertial frame K, where the parameter τ_{p} is the proper time of that object. The dimensionless 4-velocity u^{μ} as described in K is thus

$$\boxed{u^{\mu}:=\frac{\mathrm{d}x^{\mu}}{c\mathrm{d}\tau_{\mathrm{p}}}=\lambda^{\mu}_{(\kappa)}(\tau)\frac{\mathrm{d}y^{(\kappa)}}{c\mathrm{d}\tau_{\mathrm{p}}}+c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^{\mu}_{(\nu)}(\tau)\frac{\mathrm{d}y^{(0)}}{c\mathrm{d}\tau_{\mathrm{p}}}}\ . \tag{2.160}$$

The whole problem here is to establish a formula for τ_{p}. This is very similar to the problem of determining $\dot{\sigma}$ in (2.143) on p. 58.

The trick used to sort this out is basically the same too. We note that, if $\tilde{\tau}$ is any time parameter for the particle worldline $x^{\mu}(\tilde{\tau})$, then the quantity

$$\tilde{u}^{\mu}:=\frac{\mathrm{d}x^{\mu}}{c\mathrm{d}\tilde{\tau}} \tag{2.161}$$

is related to the dimensionless 4-velocity u by

$$u=\frac{\tilde{u}}{(\tilde{u}\cdot\tilde{u})^{1/2}}\ , \tag{2.162}$$

simply because u is a unit vector in the same direction as $\tilde{u}$. Furthermore,

$$u=\frac{1}{c}\frac{\mathrm{d}x}{\mathrm{d}\tau_{\mathrm{p}}}=\frac{1}{c}\frac{\mathrm{d}x}{\mathrm{d}\tilde{\tau}}\frac{\mathrm{d}\tilde{\tau}}{\mathrm{d}\tau_{\mathrm{p}}}=\tilde{u}\frac{\mathrm{d}\tilde{\tau}}{\mathrm{d}\tau_{\mathrm{p}}}\ , \tag{2.163}$$

whence

$$\frac{\mathrm{d}\tilde{\tau}}{\mathrm{d}\tau_{\mathrm{p}}}=\frac{1}{(\tilde{u}\cdot\tilde{u})^{1/2}}=:\tilde{\gamma}\,. \tag{2.164}$$

Now one option for the parameter $c\tilde{\tau}$ is just the time coordinate $y^{(0)}=c\tau$ in the accelerating frame. We define

$$\tilde{w}^{(\mu)}:=\frac{\mathrm{d}y^{(\mu)}}{\mathrm{d}y^{(0)}}\ . \tag{2.165}$$

This is the naive 4-velocity of the particle as measured relative to the coordinates of the accelerating frame. Then (2.159) becomes

$$\tilde{u}^\mu = \frac{1}{c}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \frac{\mathrm{d}x^\mu}{\mathrm{d}y^{(0)}} = \lambda^\mu_{(\kappa)}(\tau)\frac{\mathrm{d}y^{(\kappa)}}{\mathrm{d}y^{(0)}} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^\mu_{(\nu)}(\tau)$$
$$= \left[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\right]\lambda^\mu_{(\nu)}(\tau) \,. \tag{2.166}$$

This is the sum of the particle's naive 4-velocity $\tilde{w}$ as measured within the accelerating frame K' and a term

$$\tilde{u}_\mathrm{a} := c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^\mu_{(\nu)}(\tau) \,, \tag{2.167}$$

due to the acceleration of K' relative to the laboratory inertial frame K. Finally, the 4-velocity of the particle expressed relative to K is

$$\boxed{u = \frac{\tilde{u}}{(\tilde{u}\cdot\tilde{u})^{1/2}} = \frac{\tilde{w}+\tilde{u}_\mathrm{a}}{|\tilde{w}+\tilde{u}_\mathrm{a}|} = \frac{\left[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\right]\lambda^\mu_{(\nu)}(\tau)}{|\tilde{w}+c^{-2}\tilde{A}\bar{y}|}} \,. \tag{2.168}$$

To derive this, we have not assumed anything specific about the acceleration matrix $\tilde{A}$, so this is a completely general result for all the semi-Euclidean frame constructions of the kind discussed in Sect. 2.3.

We can express $\tilde{u}$ and the time dilation relation $\mathrm{d}\tau = \tilde{\gamma}\mathrm{d}\tau_\mathrm{p}$ in more detail by recalling the explicit form

$$\tilde{A} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \,, \tag{2.169}$$

whence

$$\tilde{A}\bar{y} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_0\cdot\mathbf{y} \\ c\left[\Omega_{21}y^{(2)} + \Omega_{31}y^{(3)}\right] \\ c\left[\Omega_{12}y^{(1)} + \Omega_{32}y^{(3)}\right] \\ c\left[\Omega_{13}y^{(1)} + \Omega_{23}y^{(2)}\right] \end{pmatrix} \,. \tag{2.170}$$

We can connect with the notation of [23] in (2.75) and (2.76) on p. 38, viz.,

$$\tilde{A}^{(\mu)}{}_{(\nu)}(\tilde{\mathbf{g}},\tilde{\omega}) = \begin{pmatrix} 0 & \tilde{\mathbf{g}}^\mathrm{T} \\ \tilde{\mathbf{g}} & c\pi(\tilde{\omega}) \end{pmatrix} \,, \qquad \pi(\tilde{\omega}) := \varepsilon_{ijk}\tilde{\omega}^k \,. \tag{2.171}$$

We are saying that

$$c\pi(\tilde{\omega}) := c\varepsilon_{ijk}\tilde{\omega}^k = c\begin{pmatrix} \Omega_{11} & \Omega_{21} & \Omega_{31} \\ \Omega_{12} & \Omega_{22} & \Omega_{32} \\ \Omega_{13} & \Omega_{23} & \Omega_{33} \end{pmatrix} , \tag{2.172}$$

and we know that

$$\begin{aligned}(c\mathbf{y}\times\tilde{\omega})_i = c\varepsilon_{ijk}y_j\tilde{\omega}_k &= c\begin{pmatrix} \Omega_{11} & \Omega_{21} & \Omega_{31} \\ \Omega_{12} & \Omega_{22} & \Omega_{32} \\ \Omega_{13} & \Omega_{23} & \Omega_{33} \end{pmatrix}\begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{pmatrix} \\ &= \begin{pmatrix} c\big[\Omega_{21}y^{(2)} + \Omega_{31}y^{(3)}\big] \\ c\big[\Omega_{12}y^{(1)} + \Omega_{32}y^{(3)}\big] \\ c\big[\Omega_{13}y^{(1)} + \Omega_{23}y^{(2)}\big] \end{pmatrix} ,\end{aligned}$$

so we have

$$\tilde{A}\bar{y} = \big(\tilde{\mathbf{g}}\cdot\mathbf{y}, c\mathbf{y}\times\tilde{\omega}\big) .$$

We now define $\mathbf{v}$ by

$$\tilde{w}^{(\nu)} = \frac{\mathrm{d}y^{(\nu)}}{\mathrm{d}y^{(0)}} =: (1, \mathbf{v}/c) . \tag{2.173}$$

Then by (2.166), $\tilde{u}^\mu$ is given by

$$\tilde{u}^\mu = \big(\tilde{w} + c^{-2}\tilde{A}\bar{y}\big)^{(\nu)}\lambda^{\mu}_{(\nu)}(\tau) ,$$

where

$$\tilde{w} + c^{-2}\tilde{A}\bar{y} = \Big(1 + \tilde{\mathbf{g}}\cdot\mathbf{y}/c^2, c^{-1}\big(\mathbf{v} + \mathbf{y}\times\tilde{\omega}\big)\Big) .$$

Referring to (2.164), it follows that

$$\tilde{\gamma} = \frac{1}{(\tilde{u}\cdot\tilde{u})^{1/2}} = \frac{1}{\sqrt{\left(1 + \dfrac{\tilde{\mathbf{g}}\cdot\mathbf{y}}{c^2}\right)^2 - \left(\dfrac{\mathbf{v} + \mathbf{y}\times\tilde{\omega}}{c}\right)^2}} . \tag{2.174}$$

Also by (2.164), we can now relate the proper time τ_{p} of the object to the time coordinate τ that the observer spreads over the region of spacetime around her worldline, with the result

$$\mathrm{d}\tau_{\mathrm{p}} = \sqrt{\left(1 + \frac{\tilde{\mathbf{g}}\cdot\mathbf{y}}{c^2}\right)^2 - \left(\frac{\mathbf{v} + \mathbf{y}\times\tilde{\omega}}{c}\right)^2}\,\mathrm{d}\tau . \tag{2.175}$$

It is interesting to see how readily we claim that this relates the time dilation between the particle and the observer in K'. As a matter of fact, the particle does not usually coincide with the observer worldline, so this is a relation between the proper time of the particle and a time coordinate the observer happens to have chosen in her spacetime neighbourhood, which is not at all the same thing. This relationship is coordinate dependent. There are other ways to spread time over this neighbourhood. Which one should the observer use? Which one would she most naturally use?

The answer may well be that she would use the spreading advocated so far by every semi-Euclidean frame construction of the kind we have been considering, viz., attributing her own proper time to all events in her instantaneous hyperplane of simultaneity, which is the uniquely defined 3-space orthogonal to her instantaneous 4-velocity. But this does need to be pointed out.

Note once again that the results of this section are completely general for all the semi-Euclidean frame constructions of the kind discussed in Sect. 2.3.

2.6 Four-Velocity of an Object at Fixed Space Coordinates in the Accelerating Frame

We can use the result in the last section to rederive (2.144) on p. 58, viz.,

$$u^\mu(\xi,\tau) = \frac{\big(1+\xi^i a_{0i}\big)u_0{}^\mu + \xi^i \Omega_{ij} n_j{}^\mu}{\Big[\big(1+\xi^i a_{0i}\big)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\Big]^{1/2}} . \tag{2.176}$$

We start from (2.153) on p. 62, viz.,

$$\mathrm{d}x^\mu = \frac{1}{c}\frac{\mathrm{d}\hat{x}^\mu(\tau)}{\mathrm{d}\tau}\mathrm{d}y^{(0)} + \lambda^\mu_{(i)}(\tau)\mathrm{d}y^{(i)} + y^{(i)}\frac{1}{c}\frac{\mathrm{d}\lambda^\mu_{(i)}(\tau)}{\mathrm{d}\tau}\mathrm{d}y^{(0)} , \tag{2.177}$$

considered to give an element of spacetime displacement along the object worldline. In this case we are assuming that it sits at fixed space coordinates in the frame K', i.e., that $\mathrm{d}y^{(i)} = 0$. Hence, the relation we require here is

$$\mathrm{d}x^\mu = \left[\frac{1}{c}\frac{\mathrm{d}\hat{x}^\mu(\tau)}{\mathrm{d}\tau} + y^{(i)}\frac{1}{c}\frac{\mathrm{d}\lambda^\mu_{(i)}(\tau)}{\mathrm{d}\tau}\right]\mathrm{d}y^{(0)} . \tag{2.178}$$

By the same arguments as those leading to (2.159) on p. 63, or directly from that relation, we have in this case

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}y^{(0)}} = \lambda^\mu_{(0)}(\tau) + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^\mu_{(\nu)}(\tau) . \tag{2.179}$$

If we now normalise this by dividing it by its own pseudolength, we obtain the 4-velocity of the object as expressed relative to the laboratory inertial frame K, which is what should correspond to $u^\mu(\xi,\tau)$ in (2.176).

Now $y^{(i)} \leftrightarrow \xi^i$ and

$$\tilde{A} \longleftrightarrow \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} ,$$

so, as in (2.170) on p. 64, we have

$$\tilde{A}\bar{y} \longleftrightarrow \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \xi^1 \\ \xi^2 \\ \xi^3 \end{pmatrix} = \begin{pmatrix} \mathbf{a}_0 \cdot \mathbf{y} \\ c\big(\Omega_{21}\xi^2 + \Omega_{31}\xi^3\big) \\ c\big(\Omega_{12}\xi^1 + \Omega_{32}\xi^3\big) \\ c\big(\Omega_{13}\xi^1 + \Omega_{23}\xi^2\big) \end{pmatrix} .$$

Then bearing in mind the correspondence $\lambda_{(i)} \leftrightarrow n_i$, we can thus rewrite

$$\begin{aligned} \lambda^{\mu}_{(0)}(\tau) + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\lambda^{\mu}_{(\nu)}(\tau) \longleftrightarrow\ & u^{\mu}_0(1 + c^{-2}\mathbf{a}_0 \cdot \mathbf{y}) + \frac{1}{c}\Big[\big(\Omega_{21}\xi^2 + \Omega_{31}\xi^3\big)n^{\mu}_1 \\ & + \big(\Omega_{12}\xi^1 + \Omega_{32}\xi^3\big)n^{\mu}_2 + \big(\Omega_{13}\xi^1 + \Omega_{23}\xi^2\big)n^{\mu}_3\Big] \\ =\ & (1 + a_{0i}\xi^i)u^{\mu}_0 + \xi^i\Omega_{ij}n^{\mu}_j . \end{aligned}$$

The denominator in (2.176) is clearly just the pseudolength of this, so we obtain that earlier result.

2.7 Acceleration Transformations for GUA Observers

The aim here is to consider an object or observer with arbitrary motion and relate the description of its 4-acceleration in the laboratory inertial frame to the description of its 4-acceleration in the $\{y^{(\kappa)}\}_{\kappa=0,1,2,3}$ coordinate system. In this section, we consider only the case where the accelerating frame is adapted to an observer with generalised uniform acceleration, whence the acceleration matrix $\tilde{A}$ in the analysis below is taken as independent of the proper time τ of that observer. Once again, the analysis is adapted from [23].

We begin by defining

$$\tilde{a} := c\frac{\mathrm{d}\tilde{u}}{\mathrm{d}\tau} = c^2\frac{\mathrm{d}\tilde{u}}{\mathrm{d}y^{(0)}} , \tag{2.180}$$

where we recall from (2.166) on p. 64 that

$$\tilde{u}^{\mu} := \frac{\mathrm{d}x^{\mu}}{\mathrm{d}y^{(0)}} = \big[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\big]\lambda^{\mu}_{(\nu)}(\tau) , \tag{2.181}$$

and from (2.165) that

$$\tilde{w}^{(\mu)} := \frac{\mathrm{d}y^{(\mu)}}{\mathrm{d}y^{(0)}} , \tag{2.182}$$

the naive 4-velocity of the particle as measured relative to the coordinates of the accelerating frame. We now write

$$\begin{aligned}\tilde{a} &= c\frac{\mathrm{d}\tilde{w}^{(\nu)}}{\mathrm{d}\tau}\lambda_{(\nu)}(\tau) + \left[\tilde{A}\frac{\mathrm{d}\bar{y}}{\mathrm{d}y^{(0)}}\right]^{(\nu)}\lambda_{(\nu)}(\tau) + c\big[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\big]\frac{\mathrm{d}\lambda^{\mu}_{(\nu)}(\tau)}{\mathrm{d}\tau} \\ &= \tilde{b} + \tilde{A}\bar{w} + \tilde{A}\tilde{u}\,,\end{aligned} \tag{2.183}$$

where we have made the definitions

$$\tilde{b} := c\frac{\mathrm{d}\tilde{w}^{(\nu)}}{\mathrm{d}\tau}\lambda_{(\nu)}(\tau) = c^2\frac{\mathrm{d}^2 y^{(\nu)}}{\mathrm{d}y^{(0)2}}\lambda_{(\nu)}(\tau)\,, \tag{2.184}$$

and

$$\bar{w} := \frac{\mathrm{d}\bar{y}^{(\nu)}}{\mathrm{d}y^{(0)}}\lambda_{(\nu)}(\tau)\,, \tag{2.185}$$

and we have made the deduction

$$\begin{aligned}c\big[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\big]\frac{\mathrm{d}\lambda^{\mu}_{(\nu)}}{\mathrm{d}\tau} &= \big[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\big]\lambda_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\nu)} \\ &= \lambda_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\nu)}\tilde{u}^{(\nu)} \\ &= \tilde{A}\tilde{u}\,,\end{aligned} \tag{2.186}$$

as stated above. In the last deduction, we may take the penultimate line as the obvious definition for the last line.

Care must be taken over the definition (2.184) of $\tilde{b}$. It is *not* true that

$$\tilde{b} = c\frac{\mathrm{d}\tilde{w}}{\mathrm{d}\tau} \quad \text{(not true)}\,,$$

since $\tilde{w} = \tilde{w}^{(\nu)}\lambda_{(\nu)}(\tau)$, so

$$c\frac{\mathrm{d}\tilde{w}}{\mathrm{d}\tau} = c\frac{\mathrm{d}\tilde{w}^{(\nu)}}{\mathrm{d}\tau}\lambda_{(\nu)} + c\tilde{w}^{(\nu)}\frac{\mathrm{d}\lambda_{(\nu)}}{\mathrm{d}\tau}\,.$$

Since $y^{(0)}$ is a length, $\tilde{b}^{(\mu)}$ is the naive acceleration as construed relative to the coordinates $\{y^{(\kappa)}\}_{\kappa=0,1,2,3}$ in the accelerating frame K'.

By the term $\tilde{A}\bar{w}$ we must understand

$$\tilde{A}\bar{w} = \lambda_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\nu)}\frac{\mathrm{d}\bar{y}^{(\nu)}}{\mathrm{d}y^{(0)}} = \left[\tilde{A}\frac{\mathrm{d}\bar{y}}{\mathrm{d}y^{(0)}}\right]^{(\kappa)}\lambda_{(\kappa)}\,,$$

which connects with the first line of (2.183). Note also that

$$\bar{w}^{(\mu)} = \frac{\mathrm{d}\bar{y}^{(\mu)}}{\mathrm{d}y^{(0)}} = \left(0, \frac{\mathrm{d}y^{(1)}}{\mathrm{d}y^{(0)}}, \frac{\mathrm{d}y^{(2)}}{\mathrm{d}y^{(0)}}, \frac{\mathrm{d}y^{(3)}}{\mathrm{d}y^{(0)}}\right) = (0, \mathbf{v}/c)\,, \tag{2.187}$$

with reference to the definition (2.158) on p. 63, viz.,

$$\bar{y} := x - \hat{x}(\tau) = y^{(i)}\lambda_{(i)}(\tau) \ , \tag{2.188}$$

and the definition (2.173) on p. 65, viz.,

$$\tilde{w}^{(\nu)} = \frac{\mathrm{d}y^{(\nu)}}{\mathrm{d}y^{(0)}} =: (1, \mathbf{v}/c) \ . \tag{2.189}$$

This shows the relation between $\bar{w}^{(\mu)}$ and $\tilde{w}^{(\mu)}$. In fact, $\bar{w}$ is the projection of $\tilde{w}$ onto the relevant hyperplane of simultaneity for the relevant value of τ.

We also define

$$\tilde{d} := \tilde{b} + \tilde{A}\bar{w} \ , \tag{2.190}$$

the first two terms in the first line of (2.183), since these are the terms that do not involve a differentation of $\lambda(\tau)$ in the step from $\tilde{u}$ to $\tilde{a}$. Explicitly,

$$\tilde{d} = \lambda_{(\nu)}(\tau)\left[c^2\frac{\mathrm{d}^2y^{(\nu)}}{\mathrm{d}y^{(0)2}} + \tilde{A}^{(\nu)}{}_{(\mu)}\frac{\mathrm{d}\bar{y}^{(\mu)}}{\mathrm{d}y^{(0)}}\right] \ . \tag{2.191}$$

We also define a naive 3-acceleration $\mathbf{a}_\mathrm{p}$ in K' by

$$\tilde{b} =: (0, \mathbf{a}_\mathrm{p}) \ , \tag{2.192}$$

using (2.184).

Now if $\tilde{A}$ has the form (2.171) on p. 64, viz.,

$$\tilde{A}^{(\mu)}{}_{(\nu)}(\tilde{\mathbf{g}}, \tilde{\boldsymbol{\omega}}) = \begin{pmatrix} 0 & \tilde{\mathbf{g}}^\mathrm{T} \\ \tilde{\mathbf{g}} & c\pi(\tilde{\boldsymbol{\omega}}) \end{pmatrix} \ , \qquad \pi(\tilde{\boldsymbol{\omega}}) := \varepsilon_{ijk}\tilde{\omega}^k \ , \tag{2.193}$$

this means that, for any 4-component object $r = (r^0, \mathbf{r})$, we have

$$(\tilde{A}r)^{(\nu)} = \begin{pmatrix} 0 & \tilde{\mathbf{g}}^\mathrm{T} \\ \tilde{\mathbf{g}} & c\pi(\tilde{\boldsymbol{\omega}}) \end{pmatrix}\begin{pmatrix} r^0 \\ \mathbf{r} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{r} \\ r^0\tilde{\mathbf{g}} + c\pi(\tilde{\boldsymbol{\omega}})\mathbf{r} \end{pmatrix} \ ,$$

and

$$\left[\pi(\tilde{\boldsymbol{\omega}})\mathbf{r}\right]_i = \varepsilon_{ijk}\tilde{\omega}^k r^j = (\mathbf{r}\times\tilde{\boldsymbol{\omega}})_i \ ,$$

so finally

$$(\tilde{A}r)^{(\nu)} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{r} \\ r^0\tilde{\mathbf{g}} + c\mathbf{r}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} \ . \tag{2.194}$$

We now return to (2.183), viz.,

$$\tilde{a} = \tilde{b} + \tilde{A}\bar{w} + \tilde{A}\tilde{u} \ ,$$

and insert

$$\tilde{b}^{(\nu)} = \begin{pmatrix} 0 \\ \mathbf{a}_\mathrm{p} \end{pmatrix} , \quad \tilde{w}^{(\nu)} = \begin{pmatrix} 1 \\ \mathbf{w} \end{pmatrix} , \quad \bar{w}^{(\nu)} = \begin{pmatrix} 0 \\ \mathbf{w} \end{pmatrix} , \quad \bar{y}^{(\nu)} = \begin{pmatrix} 0 \\ \mathbf{y} \end{pmatrix} , \tag{2.195}$$

$$\tilde{u}^{(\nu)} = \tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)} = \begin{pmatrix} 1 \\ \mathbf{w} \end{pmatrix} + c^{-2} \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{y} \\ c\mathbf{y}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} , \tag{2.196}$$

where the index on the object on the left of each relation indicates that these are components relative to $\lambda_{(\nu)}(\tau)$. To get (2.196), we used (2.194) with $r = \bar{y}$. Then (2.194) and (2.196) together imply that

$$\tilde{A}\tilde{u} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{u} \\ u^{(0)}\tilde{\mathbf{g}} + c\mathbf{u}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\big(\mathbf{w} + c^{-1}\mathbf{y}\times\tilde{\boldsymbol{\omega}}\big) \\ \big(1 + \tilde{\mathbf{g}}\cdot\mathbf{y}/c^2\big)\tilde{\mathbf{g}} + c\big(\mathbf{w} + c^{-1}\mathbf{y}\times\tilde{\boldsymbol{\omega}}\big)\times\tilde{\boldsymbol{\omega}} \end{pmatrix} .$$

Finally, by (2.194) once more,

$$\tilde{A}\bar{w} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{w} \\ c\mathbf{w}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} ,$$

whereupon

$$\tilde{a}^{(\nu)} = \begin{pmatrix} 0 \\ \mathbf{a}_\mathrm{p} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{w} \\ c\mathbf{w}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{g}}\cdot\big(\mathbf{w} + c^{-1}\mathbf{y}\times\tilde{\boldsymbol{\omega}}\big) \\ \big(1 + \tilde{\mathbf{g}}\cdot\mathbf{y}/c^2\big)\tilde{\mathbf{g}} + c\big(\mathbf{w} + c^{-1}\mathbf{y}\times\tilde{\boldsymbol{\omega}}\big)\times\tilde{\boldsymbol{\omega}} \end{pmatrix} ,$$

leading to

$$\boxed{\tilde{a}^{(\nu)} = \begin{pmatrix} \tilde{\mathbf{g}}\cdot\big(2\mathbf{w} + c^{-1}\mathbf{y}\times\tilde{\boldsymbol{\omega}}\big) \\ \mathbf{a}_\mathrm{p} + \big(1 + \tilde{\mathbf{g}}\cdot\mathbf{y}/c^2\big)\tilde{\mathbf{g}} + 2c\mathbf{w}\times\tilde{\boldsymbol{\omega}} + (\mathbf{y}\times\tilde{\boldsymbol{\omega}})\times\tilde{\boldsymbol{\omega}} \end{pmatrix}} . \tag{2.197}$$

But what can we do with this object? It gives us the quantity $\tilde{a} := c\mathrm{d}\tilde{u}/\mathrm{d}\tau$, where $\tilde{u} = \mathrm{d}x/c\mathrm{d}\tau$, and x is here the function describing the worldline of the chosen particle, while $c\tau$ is the zeroth coordinate $y^{(0)}$ for the accelerating coordinate frame K'. It is expressed in terms of the quantities $\tilde{\mathbf{g}}$ and $\tilde{\boldsymbol{\omega}}$ entering into the acceleration matrix $\tilde{A}$ together with the quantities $\mathbf{y}$, $\mathbf{w}$, and $\mathbf{a}_\mathrm{p}$ which describe the particle's 3-position, 3-velocity, and 3-acceleration relative to the coordinates of K'. Furthermore, it is in a component form relative to the tetrad $\{\lambda_{(\mu)}\}_{\mu=0,1,2,3}$, so that $\tilde{a} = \tilde{a}^{(\mu)}\lambda_{(\mu)}(\tau)$.

But we would like to express the 4-acceleration a of the particle, viz.,

$$a := c\frac{\mathrm{d}u}{\mathrm{d}\tau} = c\frac{\mathrm{d}u}{\mathrm{d}\tau_\mathrm{p}}\frac{\mathrm{d}\tau_\mathrm{p}}{\mathrm{d}\tau} , \tag{2.198}$$

where u is given by (2.168) on p. 64, viz.,

$$u = \frac{\tilde{u}}{(\tilde{u}\cdot\tilde{u})^{1/2}} = \frac{\big[\tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\bar{y})^{(\nu)}\big]\lambda^{\mu}_{(\nu)}(\tau)}{|\tilde{w} + c^{-2}\tilde{A}\bar{y}|} . \tag{2.199}$$

We obtained an expression for $\mathrm{d}\tau/\mathrm{d}\tau_\mathrm{p}$ in (2.164) on p. 63 and (2.174) on p. 65, viz.,

$$\frac{\mathrm{d}\tau_{\mathrm{p}}}{\mathrm{d}\tau} = \tilde{\gamma} = \frac{1}{(\tilde{u}\cdot\tilde{u})^{1/2}} = \frac{1}{\sqrt{\left(1+\dfrac{\tilde{\mathbf{g}}\cdot\mathbf{y}}{c^2}\right)^2 - \left(\dfrac{c\mathbf{w}+\mathbf{y}\times\tilde{\boldsymbol{\omega}}}{c}\right)^2}} \ . \tag{2.200}$$

Note that $\mathbf{v}$ in (2.174) was defined by (2.189) on p. 69, and this is exactly the same (up to a factor of c) as $\mathbf{w}$ defined by (2.195). Furthermore, by (2.163) and (2.164) on p. 63,

$$u = \tilde{\gamma}\tilde{u} \ . \tag{2.201}$$

We now have

$$a := c\frac{\mathrm{d}u}{\mathrm{d}\tau} = c\frac{\mathrm{d}u}{\mathrm{d}\tau_{\mathrm{p}}}\frac{\mathrm{d}\tau_{\mathrm{p}}}{\mathrm{d}\tau} = c\tilde{\gamma}\frac{\mathrm{d}}{\mathrm{d}\tau}(\tilde{\gamma}\tilde{u}) = \tilde{\gamma}^2\tilde{a} + c\tilde{\gamma}\tilde{u}\frac{\mathrm{d}\tilde{\gamma}}{\mathrm{d}\tau} \ ,$$

and

$$c\frac{\mathrm{d}\tilde{\gamma}}{\mathrm{d}\tau} = c\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{1}{(\tilde{u}\cdot\tilde{u})^{1/2}} = -\frac{1}{2}\frac{2c\tilde{u}\cdot\mathrm{d}\tilde{u}/\mathrm{d}\tau}{(\tilde{u}\cdot\tilde{u})^{3/2}} = -\tilde{a}\cdot\tilde{u}\tilde{\gamma}^3 = -\tilde{a}\cdot u\tilde{\gamma}^2 \ ,$$

whence

$$\boxed{a = \tilde{\gamma}^2\big[\tilde{a} - (\tilde{a}\cdot u)u\big]} \ . \tag{2.202}$$

Now we said that

$$\tilde{a} = \tilde{b} + \tilde{A}\bar{w} + \tilde{A}\tilde{u} = \tilde{d} + \tilde{A}\tilde{u} \ , \qquad \tilde{d} := \tilde{b} + \tilde{A}\bar{w} \ , \tag{2.203}$$

so

$$\begin{aligned} a &= \tilde{\gamma}^2(\tilde{d}+\tilde{A}\tilde{u}) - \tilde{\gamma}^2\big[\tilde{d}\cdot u + (\tilde{A}\tilde{u})\cdot u\big]u \\ &= \tilde{\gamma}^2(\tilde{d}+\tilde{A}\tilde{u}) - \tilde{\gamma}^2(\tilde{d}\cdot u)u - \big[\tilde{\gamma}(\tilde{A}u)\cdot u\big]u \ . \end{aligned}$$

But

$$(\tilde{A}u)\cdot u = u^{\mathrm{T}}\tilde{\tilde{A}}u = 0 \ ,$$

where $\tilde{\tilde{A}}$ is the type (2,0) tensor, because $\tilde{\tilde{A}}$ is an antisymmetric matrix. Hence,

$$\boxed{a = \tilde{\gamma}^2\Big[\tilde{A}\tilde{u} + \tilde{d} - (\tilde{d}\cdot u)u\Big]} \ . \tag{2.204}$$

If we define

$$\tilde{d}_\perp := \tilde{d} - (\tilde{d}\cdot u)u \ , \tag{2.205}$$

the component of $\tilde{d}$ orthogonal to the 4-velocity of the particle, this takes on the simpler form

$$\boxed{a = \tilde{\gamma}\tilde{A}u + \tilde{\gamma}^2\tilde{d}_\perp} \ . \tag{2.206}$$

Note that this is orthogonal to the 4-velocity u of the particle, as it should be. We just noted that $\tilde{A}u$ is orthogonal to u, and we see why the other term can only lie along the component of $\tilde{d}$ that is orthogonal to u. Remember that $\tilde{d}$ is supposed to describe the acceleration of the particle relative to the instantaneously comoving inertial frame $\{\lambda_{(\mu)}\}_{\mu=0,1,2,3}$.

The quantity $\tilde{A}u$ is basically the acceleration of a rest point in the accelerating frame K', i.e., the acceleration of a particle sitting at fixed space coordinates $y^{(i)}$, $i=1,2,3$, in that frame. This is because we then have $\tilde{d}=0$, since $\mathbf{a}_\mathrm{p}=0$, $\tilde{b}=0$, $\mathbf{w}=0$, and $\bar{w}=0$ in that case. But, of course, $\tilde{\gamma}\neq 1$ in general, even in that case. So we obtain

$$\boxed{a=\tilde{\gamma}\tilde{A}u\,,\quad \text{for particle at fixed } y^{(i)}}\,. \tag{2.207}$$

This is quite a nice result, since it says that the acceleration matrix $\tilde{A}$ can even be used to find the 4-acceleration of an arbitrary fixed space point in the coordinate frame K' by multiplying the 4-velocity of that point, provided we include the time dilation factor $\tilde{\gamma}$, given in this case by (2.200) above with $\mathbf{w}=0$, viz.,

$$\frac{\mathrm{d}\tau_\mathrm{p}}{\mathrm{d}\tau}=\tilde{\gamma}=\frac{1}{(\tilde{u}\cdot\tilde{u})^{1/2}}=\frac{1}{\sqrt{\left(1+\dfrac{\tilde{\mathbf{g}}\cdot\mathbf{y}}{c^2}\right)^2-\left(\dfrac{\mathbf{y}\times\tilde{\boldsymbol{\omega}}}{c}\right)^2}}\,,\quad \text{for particle at fixed } y^{(i)}\,. \tag{2.208}$$

Equation (2.207) is exactly the result (2.145) on p. 58. Equation (2.208) obviously corresponds to (2.143) on p. 58.

We can also rederive the result (2.10) on p. 19, viz.,

$$\frac{\mathrm{d}^2y^i}{\mathrm{d}y^{02}}+a^i+2\Omega^i{}_j\frac{\mathrm{d}y^j}{\mathrm{d}y^0}-2a^j\frac{\mathrm{d}y^j}{\mathrm{d}y^0}\frac{\mathrm{d}y^i}{\mathrm{d}y^0}=0\,, \tag{2.209}$$

which relates the coordinate acceleration $a^i_\mathrm{p}:=\mathrm{d}^2y^i/\mathrm{d}y^{02}$ of a free particle in a semi-Euclidean frame to the components a^i of the acceleration of the observer, the components $\Omega^i{}_j$ of the rotation tensor for the chosen space triad, and the components $v^i:=c\mathrm{d}y^i/\mathrm{d}y^0$ of the coordinate velocity, all expressed relative to the SE coordinate frame, at some event where the particle worldline just happens to intersect the worldline of the observer, i.e., at some event where $y^i=0$, $i=1,2,3$. [Note, however, that (2.209) is valid even for non-constant a^i and $\Omega^i{}_j$, whereas the present section is limited to the case where they are constant along the worldline.]

We begin with (2.202), setting $a=0$, since this is the equation of motion of a free particle:

$$\tilde{a}=(\tilde{a}\cdot u)u\,, \tag{2.210}$$

where $u=\tilde{\gamma}\tilde{u}$, according to (2.201). Setting $\mathbf{y}=0$ in (2.197), we have

$$\tilde{a}^{(\nu)}\Big|_{\mathbf{y}=0}=\begin{pmatrix}2\tilde{\mathbf{g}}\cdot\mathbf{w}\\ \mathbf{a}_\mathrm{p}+\tilde{\mathbf{g}}+2c\mathbf{w}\times\tilde{\boldsymbol{\omega}}\end{pmatrix} \tag{2.211}$$

Setting $\mathbf{y}=0$ in (2.196) on p. 70, we have

$$\tilde{u}^{(\nu)}\Big|_{\mathbf{y}=0}=\begin{pmatrix}1\\ \mathbf{w}\end{pmatrix}\,,\quad \mathbf{w}=\mathbf{v}/c\,, \tag{2.212}$$

and setting $\mathbf{y} = 0$ in (2.200) on p. 71, we have

$$\tilde{\gamma}|_{\mathbf{y}=0} = \frac{1}{\sqrt{1 - v^2/c^2}} .$$

The correspondence of notation is

$$\frac{v^i}{c} \longleftrightarrow \frac{\mathrm{d}y^i}{\mathrm{d}y^0} , \quad \mathbf{a}_\mathrm{p} \longleftrightarrow \frac{\mathrm{d}^2 y^i}{\mathrm{d}y^{02}} , \quad \tilde{g}^i \longleftrightarrow a^i , \quad c\varepsilon_{ijk}\tilde{\omega}^k \longleftrightarrow \Omega^i{}_j ,$$

where a^i and $\Omega^i{}_j$ are obtained by comparing (2.169) and (2.171) on pp. 64 ff [see, for example, (2.172) on p. 65]. This implies

$$(\mathbf{v}\times\tilde{\boldsymbol{\omega}})_i = \varepsilon_{ijk} v^j \tilde{\omega}^k = \frac{1}{c}\Omega^i{}_j v^j . \tag{2.213}$$

We thus obtain

$$\begin{aligned} \tilde{a}\cdot u &= \frac{1}{\sqrt{1 - v^2/c^2}} \left[2\tilde{\mathbf{g}}\cdot\mathbf{w} - \big(\mathbf{a}_\mathrm{p} + \tilde{\mathbf{g}} + 2c\mathbf{w}\times\tilde{\boldsymbol{\omega}}\big)\cdot\frac{\mathbf{v}}{c} \right] \\ &= \frac{1}{c\sqrt{1 - v^2/c^2}} (\mathbf{a} - \mathbf{a}_\mathrm{p})\cdot\mathbf{v} . \end{aligned}$$

The zero component of (2.210) reads

$$2\mathbf{a}\cdot\mathbf{v} = \frac{1}{1 - v^2/c^2} (\mathbf{a} - \mathbf{a}_\mathrm{p})\cdot\mathbf{v} , \tag{2.214}$$

while the other three components of (2.210) imply

$$\begin{aligned} \mathbf{a}_\mathrm{p} + \tilde{\mathbf{g}} + 2c\mathbf{w}\times\tilde{\boldsymbol{\omega}} &= \frac{1}{c^2(1 - v^2/c^2)} \Big[(\mathbf{a} - \mathbf{a}_\mathrm{p})\cdot\mathbf{v} \Big] \mathbf{v} \\ &= \frac{2\mathbf{a}\cdot\mathbf{v}}{c}\frac{\mathbf{v}}{c} , \end{aligned}$$

using (2.214) in the second step. Using $\tilde{\mathbf{g}} = \mathbf{a}$ and (2.213), we thus have

$$a_\mathrm{p}^i + a^i + \frac{2}{c}\Omega^i{}_j v^j - \frac{2}{c^2} a_j v^j v^i = 0 ,$$

and, by the above notational correspondence, this is the same as (2.209).

2.8 Summary

If we consider an arbitrary timelike worldline in special relativity, we can always find coordinate systems $\{\xi^i, \tau\}$ adapted to that worldline in the following sense:

- The worldline is given by $\xi^i = 0$, $i = 1,2,3$.
- The coordinate τ is equal to the proper time σ along the worldline.
- The metric is given at any event (ξ^i, τ) by

$$\boxed{g_{00} = \left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\ , \quad g_{0i} = \xi^j \Omega_{ji} = g_{i0}\ , \quad g_{ij} = -\delta_{ij}}\ , \tag{2.215}$$

  where $\Omega_{ij}(\sigma)$ is an antisymmetric 3×3 matrix describing the rotation of the spatial coordinate axes as one moves along the worldline [see (2.58) on p. 32] and $a_{0i}(\sigma)$ are the three nonzero components of the acceleration 4-vector in the instantaneously comoving inertial frame.
- The metric reduces to the standard form $\eta_{\mu\nu}$ of the Minkowski metric on the worldline and induces the Euclidean metric on the spacelike hypersurfaces of simultaneity $\tau = $ constant for these coordinates.
- The connection is given on the worldline itself by

$$\boxed{\Gamma^i_{00} = \Gamma^0_{0i} = \Gamma^0_{i0} = a_{0i}\ , \quad i = 1,2,3}\ , \tag{2.216}$$

$$\boxed{\Gamma^\mu_{ij} = 0\ , \quad \mu = 0,1,2,3,\ \ i = 1,2,3}\ , \tag{2.217}$$

$$\boxed{\Gamma^i_{0j} = \Gamma^i_{j0} = \Omega_{ij}\ , \quad i,\ j = 1,2,3}\ , \tag{2.218}$$

  and hence encodes the acceleration of the worldline and the rotation of the spatial coordinate axes.

Choosing spatial coordinate axes such that $\Omega_{ij}(\sigma) = 0$ for all proper times σ along the worldline amounts to selecting a spacelike triad orthogonal to the tangent to the worldline, which is just its 4-velocity, at some point, and then Fermi–Walker transporting that triad along the worldline to specify the spatial coordinate axes at other points along the worldline.

This also achieves rigidity of the coordinate system in the following sense. Any two observers sitting at two fixed neighbouring space coordinates ξ and $\xi + \delta\xi$ are always the same proper distance apart as measured by either in an instantaneously comoving inertial frame. This in turn means that a ruler satisfying the ruler hypothesis would always correctly indicate spatial coordinate separations when made to sit at fixed spatial coordinates in this system.

In the case of non-rotating spatial coordinate axes, the metric has the form

$$(g_{\mu\tau}) = \begin{pmatrix} \left(1+\xi^j a_{0j}\right)^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\ ,$$

which is static if and only if the absolute acceleration components a_{0i} are constants (independent of proper time along the worldline), a motion known as (translational) uniform acceleration.

The notion of uniform acceleration can be generalised and a rigid coordinate system obtained without the need to FW transport the initial instantaneously comoving frame (ICIF) along the worldline. Such motion satisfies

$$\boxed{c\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}u^{\nu}}\,, \tag{2.219}$$

with some specified initial value $u(0) = u_0$, where $A^{\mu}{}_{\nu}$ is a tensor under Lorentz transformations with the property that $A_{\mu\nu} := \eta_{\mu\sigma}A^{\sigma}{}_{\nu}$ is antisymmetric and with the further property of being independent of τ . The initial ICIF $\hat{\lambda} = \{\hat{\lambda}_{(\kappa)}\}$ is transported along the worldline by the isometry specified by

$$\boxed{c\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(\kappa)}\,, \qquad \lambda_{(\kappa)}(0) = \hat{\lambda}_{(\kappa)}}\,. \tag{2.220}$$

This generalised uniform acceleration (GUA) and the associated semi-Euclidean frame are Poincaré covariant constructions.

When the acceleration matrix A has the translational form

$$A = \begin{pmatrix} 0 & \mathbf{g}^{\mathrm{T}} \\ \mathbf{g} & 0 \end{pmatrix}\,, \tag{2.221}$$

where $\mathbf{g}$ is constant (independent of τ along the worldline), in some inertial frame relative to which the worldline comes to rest at some event, then the motion is pure translational uniform acceleration (TUA) according to the standard definition of uniform acceleration and the isometric transport (2.220) coincides with Fermi–Walker transport.

So we can construct semi-Euclidean (SE) coordinate systems for any observer motion and any smooth propagation of the space triad. In general, a fluid whose particles sit at fixed space coordinates in such a system will have rigid motion if and only if the space triad is Fermi–Walker (FW) transported along the observer worldline. However, for GUA motion with Friedman–Scarr (FS) isometric transport of the space triad, a fluid whose particles sit at fixed space coordinates will also have rigid (superhelical) motion. Indeed such superhelical motion can only be achieved for GUA motion of the main observer and FS transport of the space triad (see Sect. 2.4.5).

In the Friedman–Scarr coordinate construction for GUA motion, the Minkowski metric has the form (2.134) on p. 56, viz.,

$$g_{(\mu)(\nu)} = \begin{pmatrix} \left[1+y^{(i)}a_{0i}\right] - y^{(i)}y^{(j)}\Omega_{ik}\Omega_{jk} & y^{(i)}\Omega_{i1} & y^{(i)}\Omega_{i2} & y^{(i)}\Omega_{i3} \\ y^{(i)}\Omega_{i1} & -1 & 0 & 0 \\ y^{(i)}\Omega_{i2} & 0 & -1 & 0 \\ y^{(i)}\Omega_{i3} & 0 & 0 & -1 \end{pmatrix} . \quad (2.222)$$

An observer sitting at fixed space coordinates in the Friedman–Scarr construction for GUA motion is a Killing observer, i.e., the Lie derivative of the Minkowski metric is zero along her worldline (see Sect. 2.4.8). Furthermore, she herself will have GUA motion (see Sect. 2.4.9).

Observers sitting at fixed space coordinates in the FS frame of a main observer with GUA motion share hyperplanes of simultaneity with the latter, in the precise sense described just after (2.149) on p. 60, if and only if the motion of the main observer is actually TUA. But HOS sharing also occurs for a main observer with arbitrary motion provided she uses an FW transported tetrad to establish coordinates, regardless of whether her purely translational acceleration as viewed in this frame is uniform or not [see (2.33) on p. 27 and the ensuing discussion in Sect. 2.3.4].

These are very general considerations, but it is important to look at examples. Section 2.9 discusses standard translational uniform acceleration, while Sect. 2.10 shows that circular motion at constant angular velocity is an example of GUA motion and a rigid frame can therefore be constructed using isometric transport (2.220), and Sect. 2.11 gives a broad discussion of circular motion with varying angular velocity, including the frame construction with Fermi–Walker transport of the initial tetrad.

2.9 Translational Uniform Acceleration

We can deduce everything about TUA motion very quickly from the general case. We also make a host of simplifying assumptions, without loss of generality. So we consider the case $\mathbf{g} = (g,0,0)$, $\omega = 0$ and we assume that the observer is initially at rest at the origin of the laboratory frame K. Then we take the initial ICIF $\lambda(0) := \hat{\lambda}$ to be the orthonormal basis for K, whence $\lambda(0) = I$, the identity matrix. As can be seen from (2.87) on p. 42, this means that $\tilde{A} = A$. In this case then we know that the FS accelerating frame K' is found by FW transport of the ICIF along the worldline.

So we have

$$A^{\mu}{}_{\nu} = \begin{pmatrix} 0 & g & 0 & 0 \\ g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \quad (2.223)$$

and the ICIF $\lambda(\tau)$ is given by

$$\lambda(\tau) = \exp(A\tau/c)\hat{\lambda} = \exp(A\tau/c) = \begin{pmatrix} \cosh(g\tau/c) & \sinh(g\tau/c) & 0 & 0 \\ \sinh(g\tau/c) & \cosh(g\tau/c) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{2.224}$$

a well known Lorentz transformation for each value of τ, namely, a boost with rapidity $\tanh(g\tau/c)$ in the x direction.

We can read off the observer's 4-velocity $u(\tau) = \lambda_{(0)}(\tau)$ as a function of proper time τ. It is just the first column. Hence,

$$U(\tau) = \lambda_{(0)}(\tau) = \begin{pmatrix} \cosh(g\tau/c) \\ \sinh(g\tau/c) \\ 0 \\ 0 \end{pmatrix} . \tag{2.225}$$

Likewise, the $\lambda_{(i)}(\tau)$, $i = 1,2,3$, are just the other three columns:

$$\lambda_{(1)}(\tau) = \begin{pmatrix} \sinh(g\tau/c) \\ \cosh(g\tau/c) \\ 0 \\ 0 \end{pmatrix} , \qquad \lambda_{(2)}(\tau) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} , \qquad \lambda_{(3)}(\tau) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} . \tag{2.226}$$

Using the assumption that $\hat{x}(0) = 0$, we can integrate the expression for $u(\tau)$ to obtain the observer worldline $\hat{x}(\tau)$ in the form

$$\hat{x}(\tau) = \begin{pmatrix} \dfrac{c^2}{g}\sinh\dfrac{g\tau}{c} \\ \dfrac{c^2}{g}\left(\cosh\dfrac{g\tau}{c} - 1\right) \\ 0 \\ 0 \end{pmatrix} . \tag{2.227}$$

The extra factor of c comes in because u was dimensionless.

This worldline is a hyperbola in spacetime. Define

$$t(\tau) := \hat{x}^0(\tau)/c = \frac{c}{g}\sinh\frac{g\tau}{c} , \quad x(\tau) := \hat{x}^1(\tau) + c^2/g = \frac{c^2}{g}\cosh\frac{g\tau}{c} .$$

Then it is straightforward to check that the worldline has equation

$$x^2 - c^2t^2 = c^2/g , \tag{2.228}$$

and is thus a hyperbola. It is illustrated in Fig. 2.4, where we have put $c = 1$.

We now turn to (2.97) on p. 44, viz.,

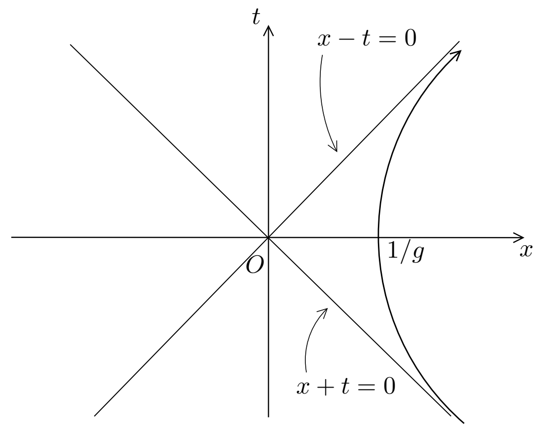

Fig. 2.4 The observer arrives from large positive x (*bottom right*), slows down to a halt at $x = 1/g$ (in this frame), actually $x = c^2/g$ if we reinstate c, then accelerates back up the x axis. The worldline is asymptotic to the null cones at the origin, i.e., it is asymptotic to $x+t=0$ for large negative times, and $x-t=0$ for large positive times. Naturally, the observer never actually reaches the speed of light

$$\boxed{x^\mu = \hat{x}^\mu(\tau) + y^{(i)}\lambda_{(i)}(\tau)}\,. \tag{2.229}$$

This is the transformation from the coordinates in the accelerating frame K' to the inertial coordinates in the laboratory frame K. In the present case, the above results allow us to read off immediately

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \dfrac{c^2}{g}\sinh\dfrac{g\tau}{c} \\ \dfrac{c^2}{g}\left(\cosh\dfrac{g\tau}{c} - 1\right) \\ 0 \\ 0 \end{pmatrix} + y^{(1)}\begin{pmatrix} \sinh\dfrac{g\tau}{c} \\ \cosh\dfrac{g\tau}{c} \\ 0 \\ 0 \end{pmatrix} + y^{(2)}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + y^{(3)}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\,,$$

which leads to

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \left[\dfrac{c^2}{g} + y^{(1)}\right]\sinh\dfrac{gy^{(0)}}{c^2} \\ \left[\dfrac{c^2}{g} + y^{(1)}\right]\cosh\dfrac{gy^{(0)}}{c^2} - \dfrac{c^2}{g} \\ y^{(2)} \\ y^{(3)} \end{pmatrix}\,, \tag{2.230}$$

since $\tau = y^{(0)}/c$.

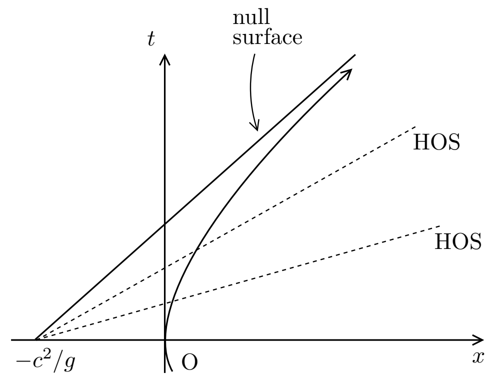

Fig. 2.5 The uniformly accelerating observer O passes through the origin of the inertial frame with coordinates x^μ. All hyperplanes of simultaneity of O intersect at $x = -c^2/g$ on the space axis

As we well know, this transformation is unlikely to be valid everywhere in spacetime. In fact, the coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ are only valid in a wedge-shaped region of spacetime, often called the Rindler wedge, with $x^1 > -c^2/g$ and bounded by the null lines $x^0 = x^1 + c^2/g$ and $x^0 = -x^1 - c^2/g$, which contains the hyperbolic path of the observer (see Fig. 2.5).

Let us examine this more closely and prove the above claim. First note from (2.134) on p. 56 that the Minkowski metric takes the component form

$$g_{(\mu)(\nu)} = \begin{pmatrix} \left[1+gy^{(1)}/c^2\right]^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \tag{2.231}$$

relative to the coordinates $y^{(\mu)}$.

As we have seen, the transformation from the semi-Euclidean coordinates y^0 and y^1 of a uniformly accelerating observer located at $y^1 = 0$ is

$$t = \frac{c}{g}\sinh\frac{gy^0}{c} + \frac{y^1}{c}\sinh\frac{gy^0}{c} , \tag{2.232}$$

$$x = \frac{c^2}{g}\left(\cosh\frac{gy^0}{c} - 1\right) + y^1\cosh\frac{gy^0}{c} , \tag{2.233}$$

ignoring the irrelevant coordinates, using t to denote x^0/c and x to denote x^1, and simplifying the notation for the SE coordinates somewhat. The coordinate y^0 here is just the proper time of the observer fixed at $y^1 = 0$, as can be seen directly from the metric, and the planes $y^0 =$ constant are the hyperplanes of simultaneity of the observer at $y^1 = 0$. So if we fix $y^0 = \kappa$ for some constant κ in the above transformation equations, allowing y^1 to vary, we pick out the relevant HOS in Minkowski spacetime. As we have suppressed two spatial dimensions, we shall consider the straight line in the (t,x) plane given parametrically by

$$x = \frac{c^2}{g}\left(\cosh\frac{g\kappa}{c} - 1\right) + y^1 \cosh\frac{g\kappa}{c}\,, \tag{2.234}$$

$$t = \frac{c}{g}\sinh\frac{g\kappa}{c} + \frac{y^1}{c}\sinh\frac{g\kappa}{c}\,, \tag{2.235}$$

with y^1 as variable. Eliminating y^1, we obtain the HOS in Minkowski coordinates as

$$x - \frac{c^2}{g}\left(\cosh\frac{g\kappa}{c} - 1\right) = c\coth\frac{g\kappa}{c}\left(t - \frac{c}{g}\sinh\frac{g\kappa}{c}\right)\,. \tag{2.236}$$

We do not have to do complicated algebra to find out that all these hyperplanes of simultaneity intersect on the $t = 0$ axis of Minkowski spacetime, since $t = 0$ in (2.236) implies that $x = -c^2/g$, regardless of the value of κ (see Fig. 2.5). In other words, if we looked at the HOS for $y^0 = \kappa_1$, we would get the Minkowski formula

$$x - \frac{c^2}{g}\left(\cosh\frac{g\kappa_1}{c} - 1\right) = c\coth\frac{g\kappa_1}{c}\left(t - \frac{c}{g}\sinh\frac{g\kappa_1}{c}\right)\,,$$

and this too crosses the $t = 0$ axis at $x = -c^2/g$. So the semi-Euclidean coordinate system breaks down at this point, or rather, on this plane, if we reinstate y^2 and y^3.

This region also corresponds to a singularity in the semi-Euclidean form of the Minkowski metric, since g_{00}^{SE} is zero there. Its determinant is thus zero in this region. Naturally, this is not due to any intrinsic singularity of the Minkowski metric, whose determinant is not zero anywhere when it is expressed as a matrix relative to any inertial coordinate system. The point is that the transformation to SE coordinates is singular in precisely this region.

Something else goes wrong outside the wedge-shaped region with $x > 0$ between the null cones at the origin in Fig. 2.4, and it is very easy to see by looking at the hyperplanes of simultaneity in the translated wedge of Fig. 2.5 as the observer's proper time tends to $\pm\infty$. In fact, for any event outside this wedge, there will be no proper time on the observer's hyperbolic worldline such that the observer considers it to be simultaneous! So no event outside the wedge will be attributed these SE coordinates.

We thus have three potential problems for any semi-Euclidean coordinate system adapted to an accelerating worldline:

- There is always a possibility that hyperplanes of simultaneity for different events on the worldline will intersect somewhere in spacetime, thereby making it impossible to extend this kind of adapted coordinate system beyond the intersection.
- The proposed transformation may be singular at some events, leading to a singular matrix of components of the Minkowski metric at those events.

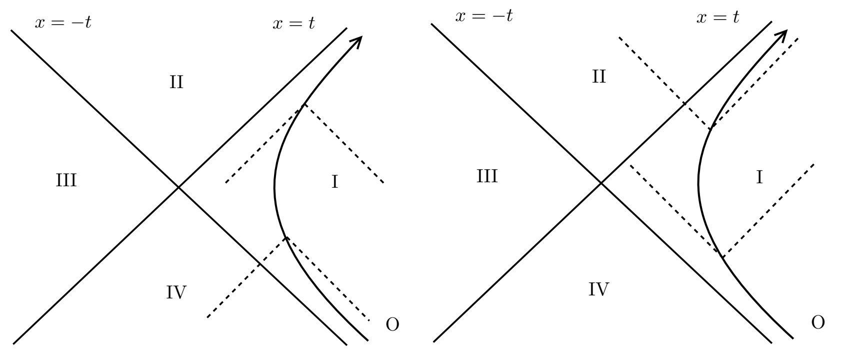

Fig. 2.6 Minkowski spacetime view of the TUA observer O. *Left*: The set of points from which O can receive signals is the union of all backward light cones of points on the worldline of O, i.e., the union of regions I and IV. *Right*: The set of points that O can signal to is the union of all forward light cones of points on the worldline of O, i.e., the union of regions I and II

- There may be some events that the given observer will *never* find to be simultaneous with her by borrowing the hyperplanes of simultaneity of instantaneously comoving inertial observers.

All three problems are illustrated by TUA motion.

As a relativist, the semi-Euclidean observer could of course use other coordinates and remove the singularity. However, as a real observer uniformly accelerating in a flat spacetime (with no gravitational effects), it should be noted that she could never receive information about events in the region $x-t<0$ and she could never send information to events in the region $x+t<0$, in the portrayal of Fig. 2.4. So there is a real physical significance to the boundaries of the wedge as far as this observer is concerned (see Fig. 2.6). These boundaries are referred to as a horizon.

Within the wedge (region I in Fig. 2.6), the coordinate transformation of (2.230), or (2.232) and (2.233), can of course be inverted, since it has nonzero determinant there. Indeed, it is easy to check that

$$y^0 = \frac{c^2}{g}\tanh^{-1}\frac{x^0}{x^1+c^2/g}\,, \quad y^1 = \left[\left(x^1+\frac{c^2}{g}\right)^2-(x^0)^2\right]^{1/2}-\frac{c^2}{g}\,, \tag{2.237}$$

$$y^2 = x^2\,, \quad y^3 = x^3\,, \tag{2.238}$$

in the appropriate region.

Let us now consider a particle at rest relative to the space coordinates of the frame K', therefore sitting at some fixed $(y^{(1)},y^{(2)},y^{(3)})$. Hence, $\tilde{w}=(1,0,0,0)$ and $\tilde{b}=0$ in (2.189) and (2.192) on p. 69, viz.,

$$\tilde{w}^{(\nu)} = \frac{\mathrm{d}y^{(\nu)}}{\mathrm{d}y^{(0)}} =: (1,\mathbf{v}/c)\,, \qquad \tilde{b} := c\frac{\mathrm{d}\tilde{w}^{(\nu)}}{\mathrm{d}\tau}\lambda_{(\nu)}(\tau) = c^2\frac{\mathrm{d}^2y^{(\nu)}}{\mathrm{d}y^{(0)2}}\lambda_{(\nu)} =: (0,\mathbf{a}_{\mathrm{p}})\,.$$

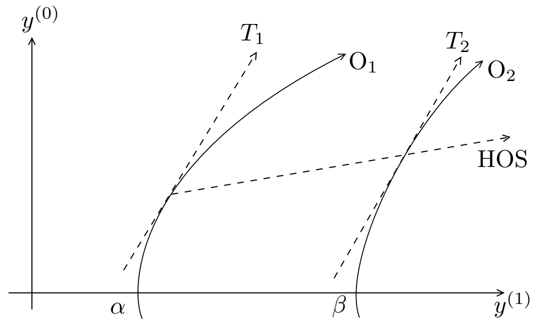

Fig. 2.7 Two observers O_1 and O_2 at fixed values α and β in the semi-Euclidean coordinate system, as viewed from a Minkowski frame. It turns out that both O_1 and O_2 are uniformly accelerating, but with different accelerations [see (2.245)]. The *dashed line* marked HOS is the hyperplane of simultaneity of O_1 at some event on its worldline, defined as the HOS of the ICIF at that event. It turns out that it is also a HOS for O_2 at the event where it intersects the worldline of O_2. The *dashed lines* marked T_1 and T_2 are time axes of the ICIOs for O_1 and O_2 at the two events. What we are saying is that they turn out to be parallel in Minkowski spacetime

Now (2.196) tells us that, in general,

$$\tilde{u}^{(\nu)} = \tilde{w}^{(\nu)} + c^{-2}(\tilde{A}\tilde{y})^{(\nu)} = \begin{pmatrix} 1 \\ \mathbf{w} \end{pmatrix} + c^{-2}\begin{pmatrix} \tilde{\mathbf{g}}\cdot\mathbf{y} \\ c\mathbf{y}\times\tilde{\boldsymbol{\omega}} \end{pmatrix} , \tag{2.239}$$

where $\mathbf{w}=0$ and $\omega=0$ in this case, with $\tilde{\mathbf{g}}\cdot\mathbf{y} = gy^{(1)}$. Hence,

$$\tilde{u} = \tilde{u}^{(\nu)}\lambda_{(\nu)} = \left[1+\frac{gy^{(1)}}{c^2}\right]\lambda_{(0)}(\tau) . \tag{2.240}$$

Then the time dilation factor is

$$\tilde{\gamma} = \frac{\mathrm{d}\tau}{\mathrm{d}\tau_{\mathrm{p}}} = \frac{1}{|\tilde{u}|} = \frac{1}{1+gy^{(1)}/c^2} . \tag{2.241}$$

We thus find the 4-velocity of this worldline as

$$u(y^{(1)},y^{(2)},y^{(3)}) = \tilde{\gamma}\tilde{u} = \lambda_{(0)}(\tau) = \begin{pmatrix} \cosh(g\tau/c) \\ \sinh(g\tau/c) \\ 0 \\ 0 \end{pmatrix} . \tag{2.242}$$

For a given τ, i.e., in the associated hyperplane of simultaneity of the accelerating observer, borrowed from the instantaneously comoving inertial frame (ICIF), each fixed space point relative to the frame K' has the same 4-velocity, viz., the 4-velocity of the observer herself at that proper time τ. This clearly illustrates the HOS sharing effect discussed on pp. 28 and 60 (see Fig. 2.7).

We can also get the 4-acceleration of the fixed space point at $(y^{(1)}, y^{(2)}, y^{(3)})$. For this, we use (2.207) on p. 72, viz.,

$$a = \tilde{\gamma}\tilde{A}u \ , \quad \text{for particle at fixed } y^{(i)} \ . \tag{2.243}$$

This gives

$$a(y^{(1)}, y^{(2)}, y^{(3)}) = \frac{1}{1 + gy^{(1)}/c^2} \begin{pmatrix} 0 & g & 0 & 0 \\ g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cosh(g\tau/c) \\ \sinh(g\tau/c) \\ 0 \\ 0 \end{pmatrix} \ ,$$

whence

$$a(y^{(1)}, y^{(2)}, y^{(3)}) = \frac{g}{1 + gy^{(1)}/c^2} \begin{pmatrix} \sinh(g\tau/c) \\ \cosh(g\tau/c) \\ 0 \\ 0 \end{pmatrix} = \frac{g}{1 + gy^{(1)}/c^2} \lambda_{(1)}(\tau) \ , \tag{2.244}$$

bearing in mind that the parameter τ here is not the proper time of an observer at the chosen space point, but the time coordinate $y^{(0)}$ in the SE frame. The Lorentzian pseudolength of this four-vector is

$$\sqrt{-a \cdot a} = \frac{g}{1 + gy^{(1)}/c^2} \ , \tag{2.245}$$

which is constant for fixed $y^{(1)}$, showing that this point once again has uniform acceleration.

However, each fixed space point has to have a different uniform acceleration in order to remain at the fixed point [see Fig. 2.7, where observers O_1 and O_2 at $y^{(1)} = \alpha$ and $y^{(1)} = \beta$ are shown to have curves of different steepness]. This shows that the term 'accelerating frame' is somewhat misleading. In the Newtonian world, it meant something to talk about an 'accelerating frame' because all space points in such a frame had the same acceleration relative to some other frame, and time never became part of the issue. But in relativity theories, a frame is just a frame, and what we have here are just convenient coordinates.

2.10 Uniform Circular Motion as Generalised Uniform Acceleration

Consider an observer following the worldline

$$t(\tau) = \gamma\tau \ , \quad x(\tau) = R\cos(\nu\gamma\tau) \ , \quad y(\tau) = R\sin(\nu\gamma\tau) \ , \quad z = 0 \ , \tag{2.246}$$

which is a regular spiral in spacetime with space radius R in the (x,y) plane, proper time τ, and constant speed $v := \nu R$. Of course,

$$\gamma := (1-v^2)^{-1/2} = (1-\nu^2 R^2)^{-1/2} .$$

The four-velocity u of the observer is

$$u(\tau) := \frac{\mathrm{d}x(\tau)}{\mathrm{d}\tau} = \gamma\big(1, -v\sin(\nu\gamma\tau), v\cos(\nu\gamma\tau), 0\big) , \tag{2.247}$$

and her four-acceleration is

$$a(\tau) := \frac{\mathrm{d}u(\tau)}{\mathrm{d}\tau} = -v\nu\gamma^2\big(0, \cos(\nu\gamma\tau), \sin(\nu\gamma\tau), 0\big) . \tag{2.248}$$

Note that $c = 1$ in this analysis. It can now be checked that we have

$$\frac{\mathrm{d}u^\mu}{\mathrm{d}\tau} = A^\mu{}_\nu u^\nu , \tag{2.249}$$

where the acceleration matrix $A = (A^\mu{}_\nu)$ is

$$A = \gamma\nu \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{2.250}$$

Since this is a constant matrix, we immediately deduce that we have a generalised uniform acceleration here.

It is easy to show that

$$\left(\frac{A}{\gamma\nu}\right)^{2n} = (-1)^n \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \quad \left(\frac{A}{\gamma\nu}\right)^{2n+1} = (-1)^n \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{2.251}$$

whence

$$\exp(A\tau) = I + \sum_{n=1}^{\infty} \frac{(-1)^n(\gamma v\tau)^{2n}}{(2n)!} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$+ \sum_{n=0}^{\infty} \frac{(-1)^n(\gamma v\tau)^{2n+1}}{(2n+1)!} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \cos(\gamma v\tau) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$+ \sin(\gamma v\tau) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

Consequently,

$$\boxed{\exp(A\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma v\tau) & -\sin(\gamma v\tau) & 0 \\ 0 & \sin(\gamma v\tau) & \cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}} . \tag{2.252}$$

We can now construct the isometrically transported frame, starting with the initial ICIF

$$\hat{\lambda}_{(0)} = u(0) = \gamma \begin{pmatrix} 1 \\ 0 \\ v \\ 0 \end{pmatrix} , \quad \hat{\lambda}_{(1)} = n_1(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} , \tag{2.253}$$

$$\hat{\lambda}_{(2)} = n_2(0) = \gamma \begin{pmatrix} v \\ 0 \\ 1 \\ 0 \end{pmatrix} , \quad \hat{\lambda}_{(3)} = n_3(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} , \tag{2.254}$$

where $\{n_i\}_{i=1,2,3}$ is the notation for the space triad in Sect. 2.3 [see (2.11) on p. 21]. The isometric transport equation is

$$\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(\kappa)} , \qquad \kappa = 0,1,2,3 , \tag{2.255}$$

which has solution

$$\lambda_{(\kappa)}(\tau) = \exp(A\tau)\hat{\lambda}_{(\kappa)} , \tag{2.256}$$

where $\exp(A\tau)$ is given by (2.252).

We can check that this gives back $u(\tau) = \lambda_{(0)}(\tau)$, since for $\kappa = 0$, (2.256) implies

$$\lambda_{(0)}(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma v\tau) & -\sin(\gamma v\tau) & 0 \\ 0 & \sin(\gamma v\tau) & \cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \gamma \begin{pmatrix} 1 \\ 0 \\ v \\ 0 \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ -v\sin(\gamma v\tau) \\ v\cos(\gamma v\tau) \\ 0 \end{pmatrix} ,$$

which is indeed $u(\tau)$ as given by (2.247).

More interestingly, we can now transport the space triad along the spiralling worldline in spacetime. We have

$$\lambda_{(1)}(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma v\tau) & -\sin(\gamma v\tau) & 0 \\ 0 & \sin(\gamma v\tau) & \cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \cos(\gamma v\tau) \\ \sin(\gamma v\tau) \\ 0 \end{pmatrix} , \tag{2.257}$$

$$\lambda_{(2)}(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma v\tau) & -\sin(\gamma v\tau) & 0 \\ 0 & \sin(\gamma v\tau) & \cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \gamma \begin{pmatrix} v \\ 0 \\ 1 \\ 0 \end{pmatrix} = \gamma \begin{pmatrix} v \\ -\sin(\gamma v\tau) \\ \cos(\gamma v\tau) \\ 0 \end{pmatrix} , \tag{2.258}$$

and

$$\lambda_{(3)}(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma v\tau) & -\sin(\gamma v\tau) & 0 \\ 0 & \sin(\gamma v\tau) & \cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} . \tag{2.259}$$

It is easy to check that $\{\lambda_{(\kappa)}(\tau)\}_{\kappa=0,1,2,3}$ form an orthonormal basis for all τ.

It is interesting to visualise this tetrad moving along the spiralling worldline in spacetime, suppressing the third space direction in a diagram. As viewed in the laboratory inertial frame, $\lambda_{(1)}(\tau)$ always points radially out from the center of the circular orbit [and lies in the hyperplane of simultaneity of $x(\tau)$ for the laboratory observer], while $\lambda_{(0)}(\tau)$ and $\lambda_{(2)}(\tau)$ are symmetrically arranged inside and outside the null cone at the point $x(\tau)$, respectively, and their projections on the instantaneous hyperplane of simultaneity are tangent to the projection of the spiral there.

There is no precession of the tetrad here, in contrast with the FW transported case to be considered in Sect. 2.11. Whenever τ is such that $x(\tau)$ and $y(\tau)$ get back to the same values, the $\{\lambda_{(\kappa)}(\tau)\}_{\kappa=0,1,2,3}$ will also have got back to the same values. This is because $x(\tau)$ and $y(\tau)$ are trigonometric functions with argument $\gamma v\tau$, and the same arguments arise in the trigonometric functions occurring in the expressions for the tetrad.

We can now find the picture $\tilde{A} := \hat{\lambda}^{-1}A\hat{\lambda}$ of the acceleration matrix in the initial ICIF, where

$$\hat{\lambda} := \left(\hat{\lambda}_{(0)}, \hat{\lambda}_{(1)}, \hat{\lambda}_{(2)}, \hat{\lambda}_{(3)}\right) .$$

But let us go the whole hog and find the picture $\tilde{A}(\tau) := \lambda^{-1}(\tau)A\lambda(\tau)$ of the acceleration matrix in the ICIF at time τ, since according to the calculation (2.125) on p. 52 this should be independent of τ. We have

$$\begin{aligned}\lambda(\tau) &= \big(\lambda_{(0)}(\tau), \lambda_{(1)}(\tau), \lambda_{(2)}(\tau), \lambda_{(3)}(\tau)\big) \\ &= \begin{pmatrix} \gamma & 0 & \gamma v & 0 \\ -\gamma v\sin(\gamma v\tau) & \cos(\gamma v\tau) & -\gamma\sin(\gamma v\tau) & 0 \\ \gamma v\cos(\gamma v\tau) & \sin(\gamma v\tau) & \gamma\cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .\end{aligned} \tag{2.260}$$

It is easy to check that this has unit determinant, as it should, since it is a Lorentz transformation from one inertial frame to another. It is also straightforward to show that

$$\lambda^{-1}(\tau) = \begin{pmatrix} \gamma & \gamma v\sin(\gamma v\tau) & -\gamma v\cos(\gamma v\tau) & 0 \\ 0 & \cos(\gamma v\tau) & \sin(\gamma v\tau) & 0 \\ -\gamma v & -\gamma\sin(\gamma v\tau) & \gamma\cos(\gamma v\tau) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{2.261}$$

A short matrix calculation leads to

$$\boxed{\tilde{A}(\tau) := \lambda^{-1}(\tau)A\lambda(\tau) = \gamma^2 v \begin{pmatrix} 0 & -v & 0 & 0 \\ -v & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}} . \tag{2.262}$$

This is indeed independent of τ. We note that, as viewed from any of these ICIFs, the acceleration matrix contains a translational component as well as the entries in the rotational sector, the latter being accompanied by an extra factor of γ as compared with the entries in the rotational sector of the original version of the acceleration matrix A. The reason for the translational component is that $\lambda_{(1)}(\tau)$ points radially outward from the center of the orbit in the hyperplane of simultaneity of the laboratory inertial observer, and the acceleration is actually in this direction for that observer.

We could have calculated the entries of the matrix $\tilde{A}$ directly using (2.103) on p. 45, viz.,

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} , \tag{2.263}$$

where the a_{0i} are given by (2.99) as

$$a_{0i} = -n_i \cdot \dot{u} = -\lambda_{(i)}(\tau) \cdot \dot{u} , \tag{2.264}$$

and the Ω_{ij} by the first relation of (2.102), which implies

$$\Omega_{ij} = -\dot{n}_i \cdot n_j = -\dot{\lambda}_{(i)}(\tau) \cdot \lambda_{(j)}(\tau) \ , \tag{2.265}$$

in the two notations. Since

$$\dot{u} = -v\nu\gamma^2 \begin{pmatrix} 0 \\ \cos(\gamma\nu\tau) \\ \sin(\gamma\nu\tau) \\ 0 \end{pmatrix} ,$$

by (2.248) on p. 84, the relations (2.264) give

$$a_{01} = v\nu\gamma^2 \begin{pmatrix} 0 \\ \cos(\gamma\nu\tau) \\ \sin(\gamma\nu\tau) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \cos(\gamma\nu\tau) \\ \sin(\gamma\nu\tau) \\ 0 \end{pmatrix} = -v\nu\gamma^2 \ ,$$

which is indeed constant and equal to the entry in row 0 and column 1 of (2.262), together with

$$a_{02} = v\nu\gamma^3 \begin{pmatrix} 0 \\ \cos(\gamma\nu\tau) \\ \sin(\gamma\nu\tau) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} v \\ -\sin(\gamma\nu\tau) \\ \cos(\gamma\nu\tau) \\ 0 \end{pmatrix} = 0$$

and

$$a_{03} = v\nu\gamma^2 \begin{pmatrix} 0 \\ \cos(\gamma\nu\tau) \\ \sin(\gamma\nu\tau) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = 0 \ ,$$

yielding the other two entries of (2.262) in row 0. Then using (2.265), we would obtain

$$\Omega_{12} = -\dot{\lambda}_{(1)}(\tau) \cdot \lambda_{(2)}(\tau) = -\gamma^2\nu \begin{pmatrix} 0 \\ -\sin(\gamma\nu\tau) \\ \cos(\gamma\nu\tau) \\ 0 \end{pmatrix} \cdot \begin{pmatrix} v \\ -\sin(\gamma\nu\tau) \\ \cos(\gamma\nu\tau) \\ 0 \end{pmatrix} = \gamma^2\nu \ ,$$

which is indeed the entry in row 2 and column 1 of the matrix (2.262).

In conclusion, circular motion at constant angular velocity provides a perfect example of what can be incorporated into the generalised notion of uniform acceleration, and at the same time it provides an example of DeWitt's superhelical motion [14]. A continuous medium whose material elements sit at constant space positions in the coordinates associated by the semi-Euclidean construction with the tetrad frame $\{\lambda_{(\kappa)}(\tau)\}_{\kappa=0,1,2,3}$ is moving rigidly.

We can also write down the components of the Minkowski metric in the above coordinate system, using the result in Sect. 2.4.8 [see (2.134) on p. 56] and inserting the values

$$a_{01} = -v\nu\gamma^2 \ , \quad a_{02} = 0 = a_{03} \ , \quad \Omega_{12} = \gamma^2 \nu = -\Omega_{21} \ ,$$

to obtain

$$g_{(\mu)(\nu)} = \begin{pmatrix} 1 - v\nu\gamma^2 y^{(1)} - \left[y^{(1)} - y^{(2)}\right]^2 \gamma^4 \nu^2 & -y^{(2)}\gamma^2\nu & y^{(1)}\gamma^2\nu & 0 \\ -y^{(2)}\gamma^2\nu & -1 & 0 & 0 \\ y^{(1)}\gamma^2\nu & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{2.266}$$

Here we used

$$\begin{aligned} y^{(i)}y^{(j)}\Omega_{ik}\Omega_{jk} &= y^{(1)}y^{(1)}\Omega_{1k}\Omega_{1k} + y^{(1)}y^{(2)}\Omega_{1k}\Omega_{2k} \\ &\qquad + y^{(2)}y^{(1)}\Omega_{2k}\Omega_{1k} + y^{(2)}y^{(2)}\Omega_{2k}\Omega_{2k} \\ &= \left[y^{(1)} - y^{(2)}\right]^2 \gamma^4 \nu^2 \ . \end{aligned}$$

2.11 General Circular Motion

An observer A moves round a circle of radius r centered on the origin in the (x, y) plane of an inertial frame L in flat spacetime. At $t = 0$, A is at azimuthal angle $\varphi_A = 0$. For $t > 0$, her motion is specified by the angular frequency $\hat{\Omega}_0(t) > 0$, so that her position is given by

$$\varphi_A(t) = \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' \ . \tag{2.267}$$

Although we still have motion in an exact circle, the angular speed is allowed to change here, whereas it was constant in Sect. 2.10. Let $\hat{\gamma}(t)$ be the Lorentz factor corresponding to the speed

$$\hat{v}(t) = r\hat{\Omega}_0(t) \ . \tag{2.268}$$

The notation here is due to Mashhoon [36]. The results of this section will be used later for further discussion.

The coordinates of A in L are

$$x_A^\mu = (t, r\cos\varphi_A, r\sin\varphi_A, 0) \ , \tag{2.269}$$

and the proper time along the worldline is

$$\tau = \int_0^t \sqrt{1 - \hat{v}^2(t')}\mathrm{d}t' \ , \tag{2.270}$$

where we take $\tau = 0$ at $t = 0$. We also assume that $\tau = \tau(t)$ has an inverse denoted by $t = F(\tau)$, whence

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{\mathrm{d}F}{\mathrm{d}\tau} = \gamma(\tau) := \big[1 - v(\tau)\big]^{-1/2} \tag{2.271}$$

is the Lorentz factor along the worldline of A. We are inventing a notation that distinguishes functions of t and τ, with $v(\tau) := \hat{v}(t)$ and $\gamma(\tau) := \hat{\gamma}(t)$.

We also define $\phi(\tau) := \varphi_A(t)$ and note that

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \gamma\Omega_0(\tau) \,, \tag{2.272}$$

where $\Omega_0(\tau) := \hat{\Omega}_0(t)$.

2.11.1 A First Tetrad for an Adapted Coordinate System

According to Mashhoon [36], the *natural* orthonormal tetrad frame along the worldline of A for $\tau > 0$ is given by

$$\begin{aligned} \lambda^{\mu}_{(0)} &= \gamma(1, -v\sin\phi, v\cos\phi, 0) \,, \\ \lambda^{\mu}_{(1)} &= (0, \cos\phi, \sin\phi, 0) \,, \\ \lambda^{\mu}_{(2)} &= \gamma(v, -\sin\phi, \cos\phi, 0) \,, \\ \lambda^{\mu}_{(3)} &= (0, 0, 0, 1) \,, \end{aligned} \tag{2.273}$$

where

$$\lambda^{\mu}_{(0)} = \frac{\mathrm{d}x^{\mu}_A}{\mathrm{d}\tau} \tag{2.274}$$

is the four-velocity of A. He claims that the spatial triad here is *natural*, but does not specify the criteria he applies for assessing naturalness.

As we know, for constant angular speed, we construct a rigid coordinate system with the above choice of tetrad (see Sect. 2.10), not just one with Euclidean spacelike hypersurfaces. But for varying angular speed, we know that the only rigid coordinate systems are ones in which the space triad is FW transported along the worldline, which is not the case here. There is no precession for the above space triad, and this may be what makes it appear natural.

Mashhoon effectively implies this in his justification. We note that, if $r = 0$ so that $v = 0$ and $\gamma = 1$, we obtain the *natural* tetrad of the fixed observer at the space origin who refers her observations to the axes of the (x', y', z') coordinate system obtained from (x, y, z) by a rotation about the z axis with frequency $\hat{\Omega}_0(t)$. It is just a space triad that rotates with the points A and B. We can then boost this tetrad with speed v along the second space axis, tangent to the circle of radius r, to give the tetrad in (2.273).

2.11.2 Acceleration Matrix for the Non-Precessing Tetrad

Another justification is just to choose some convenient space triad orthogonal to the four-velocity $\lambda_{(0)}$ at time $\tau = 0$, and then use Friedman and Scarr's construction with the acceleration matrix, which we now introduce. That is mathematically natural! Orthonormality of the tetrad system implies that the acceleration tensor $\mathscr{A}_{\alpha\beta}$, obtained by lowering one index of the object $\mathscr{A}_\alpha{}^\beta$ such that

$$\frac{\mathrm{d}\lambda^\mu_{(\alpha)}}{\mathrm{d}\tau} = \mathscr{A}_\alpha{}^\beta \lambda^\mu_{(\beta)} \,, \tag{2.275}$$

must be antisymmetric. Note, however, that this falls outside the FS construction unless the angular speed of rotation is constant! But the space triad in (2.273) does satisfy a constraint relating to the acceleration matrix, because (2.275) holds with the same acceleration matrix for all four values of μ.

Note also that we have changed notational conventions compared with the earlier discussion based on the paper [23] by Friedman and Scarr. In (2.275), the matrix $\mathscr{A}_\alpha{}^\beta$ corresponds to the transpose of $\tilde{A}^{(\alpha)}{}_{(\beta)}$ in (2.262) on p. 87, for example (see Sect. 2.10), as can be seen by comparing (2.281) below for the case $\mathrm{d}v/\mathrm{d}\tau$ with (2.262).

As we have seen, the translational acceleration of $\mathscr{A}$ is given by $a_i = \mathscr{A}_{0i}$, for $i \in \{1,2,3\}$, and the rotational acceleration by $\mathscr{A}_{jk}$ for $j,k \in \{1,2,3\}$, the latter giving us an angular velocity three-vector Ω defined by

$$\Omega_i := \frac{1}{2}\varepsilon_{ijk}\mathscr{A}^{jk} \,. \tag{2.276}$$

So let us determine $\mathscr{A}$. The first thing is to write down a matrix with entry $\lambda^\mu_{(\alpha)}$, where α labels rows and μ labels columns. So its rows are just the component forms of the four vectors in (2.273):

$$\lambda^\mu_{(\alpha)} = \begin{pmatrix} \gamma & -v\gamma\sin\phi & v\gamma\cos\phi & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ v\gamma & -\gamma\sin\phi & \gamma\cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{2.277}$$

This array is designed to be multiplied on the left by the matrix $\mathscr{A}_\alpha{}^\beta$ as in (2.275). The result should be the matrix with entries $\mathrm{d}\lambda^\mu_{(\alpha)}/\mathrm{d}\tau$, arranged in the same way, which we now calculate. We use the fact that

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\tau} = \gamma^3 v \frac{\mathrm{d}v}{\mathrm{d}\tau} \,, \qquad \frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \gamma\Omega_0 \,, \tag{2.278}$$

the first being standard and the second being (2.272). Note also that

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(\gamma v) = \gamma^3 v^2 \frac{\mathrm{d}v}{\mathrm{d}\tau} + \gamma \frac{\mathrm{d}v}{\mathrm{d}\tau} = (1+\gamma^2 v^2)\gamma \frac{\mathrm{d}v}{\mathrm{d}\tau} = \gamma^3 \frac{\mathrm{d}v}{\mathrm{d}\tau} \,. \tag{2.279}$$

Finally, we obtain

$$\frac{\mathrm{d}\lambda^{\mu}_{(\alpha)}}{\mathrm{d}\tau} = \begin{pmatrix} \gamma^3 v \dfrac{\mathrm{d}v}{\mathrm{d}\tau} & -\gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau}\sin\phi - v\gamma^2 \Omega_0 \cos\phi & \gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau}\cos\phi - v\gamma^2\Omega_0 \sin\phi & 0 \\ 0 & -\gamma\Omega_0 \sin\phi & \gamma\Omega_0\cos\phi & 0 \\ \gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} & -v\gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau}\sin\phi - \gamma^2\Omega_0\cos\phi & v\gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau}\cos\phi - \gamma^2\Omega_0\sin\phi & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{2.280}$$

once again with α labelling rows and μ labelling columns.

It is straightforward to check that the following matrix multiplies the one in (2.277) on the left to give the matrix in (2.280):

$$\boxed{\mathscr{A}_{\alpha}{}^{\beta} = \begin{pmatrix} 0 & -\gamma^2 v\Omega_0 & \gamma^2 \mathrm{d}v/\mathrm{d}\tau & 0 \\ -\gamma^2 v \Omega_0 & 0 & \gamma^2\Omega_0 & 0 \\ \gamma^2 \mathrm{d}v/\mathrm{d}\tau & -\gamma^2\Omega_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}} \,. \tag{2.281}$$

Multiplying on the left by the Minkowski metric $\eta := \mathrm{diag}(-1,1,1,1)$ to raise the index (using Mashhoon's convention for the metric signature), we have

$$\mathscr{A}^{\alpha\beta} = \begin{pmatrix} 0 & \gamma^2 v\Omega_0 & -\gamma^2 \mathrm{d}v/\mathrm{d}\tau & 0 \\ -\gamma^2 v \Omega_0 & 0 & \gamma^2\Omega_0 & 0 \\ \gamma^2 \mathrm{d}v/\mathrm{d}\tau & -\gamma^2\Omega_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \,. \tag{2.282}$$

Now we have

$$\Omega_1 = \frac{1}{2}(\varepsilon_{123}\mathscr{A}^{23} + \varepsilon_{132}\mathscr{A}^{32}) = \varepsilon_{123}\mathscr{A}^{23} = 0 \,,$$

$$\Omega_2 = \frac{1}{2}(\varepsilon_{231}\mathscr{A}^{31} + \varepsilon_{213}\mathscr{A}^{13}) = \varepsilon_{231}\mathscr{A}^{31} = 0 \,,$$

$$\Omega_3 = \frac{1}{2}(\varepsilon_{312}\mathscr{A}^{12} + \varepsilon_{321}\mathscr{A}^{21}) = \varepsilon_{312}\mathscr{A}^{12} = \gamma^2\Omega_0 \,.$$

Hence,

$$\Omega = \begin{pmatrix} 0 \\ 0 \\ \gamma^2\Omega_0 \end{pmatrix} , \qquad \mathbf{a} = \begin{pmatrix} -\gamma^2 v\Omega_0 \\ \gamma^2 \mathrm{d}v/\mathrm{d}\tau \\ 0 \end{pmatrix} \,. \tag{2.283}$$

Note in passing that, for any antisymmetric tensor $A^{\alpha\beta}$, i.e., of the form

$$A^{\alpha\beta}=\begin{pmatrix}0 & a_1 & a_2 & a_3\\ -a_1 & 0 & b_3 & -b_2\\ -a_2 & -b_3 & 0 & b_1\\ -a_3 & b_2 & -b_1 & 0\end{pmatrix}\,, \tag{2.284}$$

the quantities $\mathbf{a}\cdot\mathbf{b}$ and a^2-b^2 are Lorentz invariants. This is seen by calculating

$$I:=\frac{1}{2}A_{\alpha\beta}A^{\alpha\beta}\,,\quad I^*:=\frac{1}{2}A^*_{\alpha\beta}A^{\alpha\beta}\,, \tag{2.285}$$

where

$$A^*_{\alpha\beta}:=\frac{1}{2}\varepsilon_{\alpha\beta\gamma\delta}A^{\gamma\delta}\,, \tag{2.286}$$

and $\varepsilon_{\alpha\beta\gamma\delta}$ is the completely antisymmetric tensor density with $\varepsilon_{0123}=1$. These are obviously Lorentz invariants, by construction. The Faraday tensor provides a well known example, leading to invariants E^2-B^2 and $\mathbf{E}\cdot\mathbf{B}$.

In the present case the invariants are obtained from $\mathbf{a}$ and Ω in (2.283), which give $\mathbf{a}\cdot\Omega=0$ and

$$I:=-a^2+\Omega^2=-\gamma^4v^2\Omega_0^2-\gamma^4\left(\frac{\mathrm{d}v}{\mathrm{d}\tau}\right)^2+\gamma^4\Omega_0^2=-\gamma^4\left(\frac{\mathrm{d}v}{\mathrm{d}\tau}\right)^2+\gamma^2\Omega_0^2\,. \tag{2.287}$$

In the case of an EM field, if both invariants are zero, this is called a null EM field. It may be that both invariants are zero for the acceleration tensor [36].

We would like a physical interpretation of $\mathbf{a}$ and Ω. Since $\Omega_0=v/r$, the first component of $\mathbf{a}$ can be written $-\gamma^2v^2/r$, a centripetal acceleration, since this component is associated with the space tetrad vector $\lambda_{(1)}$, which points radially outward. The second component of $\mathbf{a}$, associated with the space tetrad vector $\lambda_{(2)}$, is $\gamma^2\mathrm{d}v/\mathrm{d}\tau$, and this is clearly some kind of tangential acceleration. Note that $\lambda_{(2)}$ has a time component in the inertial frame L, but its space components are tangential to the spiralling worldline.

It is important to understand why the components of $\mathbf{a}$, which is just part of a row of the acceleration matrix, should be associated with the directions of the space tetrad vectors $\lambda_{(i)}$, $i\in\{1,2,3\}$. This has already been discussed in Sect. 2.4.5, but will given specific attention again in Sect. 2.11.4.

Interpreting Ω is more intricate since we would like to say that it gives the frequency of rotation of the space frame relative to a non-rotating frame, and the closest we can come to a non-rotating frame is one that is FW transported. Here we shall also see precisely why the FW transported frame is sometimes said to be non-rotating: basically it is always got at one instant by a pure Lorentz boost from the frame at the previous instant, as viewed in that previous frame.

2.11.3 Fermi–Walker Transported Tetrad

Here we shall show that the FW transported space frame is $\{\tilde{\lambda}_{(i)}\}_{i=1,2,3}$, where

$$\begin{aligned}
\tilde{\lambda}^{\mu}_{(0)} &= \lambda^{\mu}_{(0)} ,\\
\tilde{\lambda}^{\mu}_{(1)} &= \cos\Phi\lambda^{\mu}_{(1)} - \sin\Phi\lambda^{\mu}_{(2)} ,\\
\tilde{\lambda}^{\mu}_{(2)} &= \sin\Phi\lambda^{\mu}_{(1)} + \cos\Phi\lambda^{\mu}_{(2)} ,\\
\tilde{\lambda}^{\mu}_{(3)} &= \lambda^{\mu}_{(3)} ,
\end{aligned} \tag{2.288}$$

with the tetrad $\{\lambda^{\mu}_{(\kappa)}\}_{\kappa=0,1,2,3}$ as defined in (2.273) and the angle Φ defined by

$$\Phi := \int_0^{\tau} \Omega(\tau')\mathrm{d}\tau' = \int_0^{\tau} \gamma^2(\tau')\Omega_0(\tau')\mathrm{d}\tau' , \qquad \Omega := |\Omega| , \tag{2.289}$$

so that

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\tau} = \Omega = \gamma^2\Omega_0 , \tag{2.290}$$

by (2.283). This gives a rather elegant picture, but remains to be justified (see below).

The point about the angular frequency Ω is that it will always be higher than Ω_0, but usually only a little bit higher. For most ordinary kinds of speed v, the quantity γ will be only slightly greater than unity. So this is a relativistic effect. Note that Ω is a proper time frequency. The corresponding inertial time frequency in L (the original inertial frame) would be

$$\Omega_{\text{inertial}} = \frac{\mathrm{d}\Phi}{\mathrm{d}t} = \frac{\mathrm{d}\Phi}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t} = \gamma\Omega_0 . \tag{2.291}$$

But what is rotating at this frequency, and relative to what? The relations (2.288) show that this is the frequency of rotation of the pair of space vectors $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$ relative to the other pair of space vectors $\lambda_{(1)}$ and $\lambda_{(2)}$. Note also that it is a rotation in the opposite direction to the rotation of the latter relative to L, and at almost the same frequency, as just mentioned, unless γ is very different from unity.

So the pair $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$ is of course busy undoing the rotation of $\lambda_{(1)}$ and $\lambda_{(2)}$, in order that the former may remain more or less constantly oriented in the inertial frame L. They would be exactly constantly oriented if we had $\gamma = 1$, but then we would have no motion at all! However, for non-relativistic v, we have $\gamma \approx 1$, and the FW transported pair $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$ will barely rotate relative to L.

Recalling that Ω_0 is an angular frequency relative to inertial time in L, whence $\mathrm{d}\phi/\mathrm{d}\tau = \gamma\Omega_0$ as in (2.272) on p. 90, and since Ω_0 is the rate of rotation of the pair of space vectors $\lambda_{(1)}$ and $\lambda_{(2)}$ relative to L, we obtain the estimate $(1-\gamma)\Omega_0$ for the (Thomas) precession of the pair of space vectors $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$ relative to the inertial frame L.

Let us now prove that the tetrad in (2.288) also satisfies the FW transport equation. With the convention $\eta = \mathrm{diag}(-1, 1, 1, 1)$ for the Minkowski metric, the FW transport equation for a four-vector A is

$$\dot{A} = (A \cdot \dot{u})u - (A \cdot u)\dot{u} , \tag{2.292}$$

where u is the four-velocity. In a case where we know that A is always orthogonal to u, as happens for $\tilde{\lambda}_{(1)}$ and $\tilde{\lambda}_{(2)}$, this equation reduces to

$$\dot{A} = (A \cdot \dot{u})u , \qquad \text{for } A \text{ orthogonal to } u . \tag{2.293}$$

We thus aim to show directly that

$$\dot{\tilde{\lambda}}_{(1)} = \left[\tilde{\lambda}_{(1)} \cdot \dot{u}\right] u , \qquad \dot{\tilde{\lambda}}_{(2)} = \left[\tilde{\lambda}_{(2)} \cdot \dot{u}\right] u , \tag{2.294}$$

where

$$u = \lambda_{(0)} = \tilde{\lambda}_{(0)} = \gamma \begin{pmatrix} 1 \\ -v\sin\phi \\ v\cos\phi \\ 0 \end{pmatrix} , \tag{2.295}$$

and

$$\dot{u} = \gamma^3 \frac{\mathrm{d}v}{\mathrm{d}\tau} \begin{pmatrix} v \\ -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} - v\gamma^2 \Omega_0 \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} = \begin{pmatrix} v\gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} \\ -\gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} \sin\phi - v\gamma^2 \Omega_0 \cos\phi \\ \gamma^3 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} \cos\phi - v\gamma^2 \Omega_0 \sin\phi \\ 0 \end{pmatrix} . \tag{2.296}$$

Check that $\dot{\tilde{\lambda}}_{(1)} = \left[\tilde{\lambda}_{(1)} \cdot \dot{u}\right] u$

First note that

$$\begin{aligned} \tilde{\lambda}_{(1)} \cdot \dot{u} &= \cos\Phi \left[\lambda_{(1)} \cdot \dot{u}\right] - \sin\Phi \left[\lambda_{(2)} \cdot \dot{u}\right] \\ &= -\cos\Phi \left(v\gamma^2 \Omega_0\right) + \sin\Phi \left[\gamma^4 \frac{\mathrm{d}v}{\mathrm{d}\tau}(v^2 - 1)\right] \\ &= -v\gamma^2 \Omega_0 \cos\Phi - \gamma^2 \frac{\mathrm{d}v}{\mathrm{d}\tau} \sin\Phi , \end{aligned} \tag{2.297}$$

while

$$
\begin{aligned}
\dot{\tilde{\lambda}}_{(1)} &= -\dot{\Phi}\sin\Phi\lambda_{(1)} + \cos\Phi\dot{\lambda}_{(1)} - \dot{\Phi}\cos\Phi\lambda_{(2)} - \sin\Phi\dot{\lambda}_{(2)} \\
&= -\gamma^2\Omega_0\sin\Phi \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} - \gamma^2\Omega_0\cos\Phi \begin{pmatrix} \gamma v \\ -\gamma\sin\phi \\ \gamma\cos\phi \\ 0 \end{pmatrix} \\
&\quad + \cos\Phi \begin{pmatrix} 0 \\ -\gamma\Omega_0\sin\phi \\ \gamma\Omega_0\cos\phi \\ 0 \end{pmatrix} - \sin\Phi \begin{pmatrix} \gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau} \\ -v\gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau}\sin\phi - \Omega_0\gamma^2\cos\phi \\ v\gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau}\cos\phi - \Omega_0\gamma^2\sin\phi \\ 0 \end{pmatrix} .
\end{aligned}
\tag{2.298}
$$

The latter has to be equal to

$$
\left[\tilde{\lambda}_{(1)}\cdot\dot{u}\right]u = -\left(v\gamma^2\Omega_0\cos\Phi + \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\sin\Phi\right)\gamma \begin{pmatrix} 1 \\ -v\sin\phi \\ v\cos\phi \\ 0 \end{pmatrix} . \tag{2.299}
$$

It is straightforward to check this row by row.

Check that $\dot{\tilde{\lambda}}_{(2)} = \left[\tilde{\lambda}_{(2)}\cdot\dot{u}\right]u$

First note that

$$
\begin{aligned}
\tilde{\lambda}_{(2)}\cdot\dot{u} &= \sin\Phi\left[\lambda_{(2)}\cdot\dot{u}\right] + \cos\Phi\left[\lambda_{(2)}\cdot\dot{u}\right] \\
&= -\sin\Phi\left(v\gamma^2\Omega_0\right) - \cos\Phi\left[\gamma^4\frac{\mathrm{d}v}{\mathrm{d}\tau}(v^2-1)\right] \\
&= -v\gamma^2\Omega_0\sin\Phi + \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\cos\Phi ,
\end{aligned}
\tag{2.300}
$$

while

$$\begin{aligned}
\dot{\tilde{\lambda}}_{(2)} &= \dot{\Phi}\cos\Phi\lambda_{(1)} + \sin\Phi\dot{\lambda}_{(1)} - \dot{\Phi}\sin\Phi\lambda_{(2)} + \cos\Phi\dot{\lambda}_{(2)} \\
&= \gamma^2\Omega_0\cos\Phi\begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} - \gamma^2\Omega_0\sin\Phi\begin{pmatrix} \gamma v \\ -\gamma\sin\phi \\ \gamma\cos\phi \\ 0 \end{pmatrix} \qquad (2.301) \\
&\quad + \sin\Phi\begin{pmatrix} 0 \\ -\gamma\Omega_0\sin\phi \\ \gamma\Omega_0\cos\phi \\ 0 \end{pmatrix} + \cos\Phi\begin{pmatrix} \gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau} \\ -v\gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau}\sin\phi - \Omega_0\gamma^2\cos\phi \\ v\gamma^3\dfrac{\mathrm{d}v}{\mathrm{d}\tau}\cos\phi - \Omega_0\gamma^2\sin\phi \\ 0 \end{pmatrix}.
\end{aligned}$$

The latter has to be equal to

$$\left[\tilde{\lambda}_{(2)}\cdot\dot{u}\right]u = -\left(v\gamma^2\Omega_0\sin\Phi - \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\cos\Phi\right)\gamma\begin{pmatrix} 1 \\ -v\sin\phi \\ v\cos\phi \\ 0 \end{pmatrix}. \qquad (2.302)$$

It is straightforward to check this row by row. ■

In the case of constant angular frequency ν, as discussed in Sect. 2.10, we would obtain

$$\boxed{n_1(\tau) = \begin{pmatrix} -v\gamma\sin(\nu\gamma^2\tau) \\ \cos(\nu\gamma\tau)\cos(\nu\gamma^2\tau) + \gamma\sin(\nu\gamma\tau)\sin(\nu\gamma^2\tau) \\ \sin(\nu\gamma\tau)\cos(\nu\gamma^2\tau) - \gamma\cos(\nu\gamma\tau)\sin(\nu\gamma^2\tau) \\ 0 \end{pmatrix}} \qquad (2.303)$$

and

$$\boxed{n_2(\tau) = \begin{pmatrix} v\gamma\cos(\nu\gamma^2\tau) \\ \cos(\nu\gamma\tau)\sin(\nu\gamma^2\tau) - \gamma\sin(\nu\gamma\tau)\cos(\nu\gamma^2\tau) \\ \sin(\nu\gamma\tau)\sin(\nu\gamma^2\tau) + \gamma\cos(\nu\gamma\tau)\cos(\nu\gamma^2\tau) \\ 0 \end{pmatrix}}. \qquad (2.304)$$

We can compare (2.303) with

$$
\begin{aligned}
\tilde{\lambda}_{(1)} &= \cos\Phi\lambda_{(1)} - \sin\Phi\lambda_{(2)} \\
&= \cos\Phi \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} - \sin\Phi \begin{pmatrix} \gamma v \\ -\gamma\sin\phi \\ \gamma\cos\phi \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -\gamma v\sin\Phi \\ \cos\phi\cos\Phi + \gamma\sin\phi\sin\Phi \\ \sin\phi\cos\Phi - \gamma\cos\phi\sin\Phi \\ 0 \end{pmatrix} ,
\end{aligned}
\tag{2.305}
$$

whence we find that $\phi \leftrightarrow \nu\gamma\tau$ and $\Phi \leftrightarrow \nu\gamma^2\tau$, with the extra factor of γ that we would expect to find with Φ as compared with ϕ. Likewise, we can compare (2.304) with

$$
\begin{aligned}
\tilde{\lambda}_{(2)} &= \sin\Phi\lambda_{(1)} + \cos\Phi\lambda_{(2)} \\
&= \sin\Phi \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} + \cos\Phi \begin{pmatrix} \gamma v \\ -\gamma\sin\phi \\ \gamma\cos\phi \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \gamma v\cos\Phi \\ \cos\phi\sin\Phi - \gamma\sin\phi\cos\Phi \\ \sin\phi\sin\Phi + \gamma\cos\phi\cos\Phi \\ 0 \end{pmatrix} ,
\end{aligned}
\tag{2.306}
$$

with the same correspondence.

2.11.4 Understanding Acceleration Matrices

The fact that the tetrad $\{\tilde{\lambda}_{(\mu)}\}_{\mu=0,1,2,3}$ satisfies the FW transport equation tells us something about the acceleration matrix for this tetrad, namely that it is pure translational. We can get that from the general analysis in Sect. 2.4.5, and in particular on p. 44 ff, although we were considering constant acceleration matrices in that case. The key relation in this general account was

$$
\dot{n}_i{}^{\mu} = a_{0i}u^{\mu} + \Omega_{ij}n_j{}^{\mu} . \tag{2.307}
$$

Now we have the correspondence $n_i \leftrightarrow \lambda_{(i)}$, $i = 1,2,3$, while $u \leftrightarrow \lambda_{(0)}$, where $\{\lambda_{(\mu)}\}_{\mu=0,1,2,3}$ can be any tetrad, either the rotating tetrad (2.273) on p. 90 or the FW transported tetrad in (2.288) on p. 94 which carries a tilde. We also have

$$
a_{0i} = n_i \cdot \dot{u} , \tag{2.308}
$$

for the convention $\eta = \mathrm{diag}(-1,1,1,1)$, which means that

$$\dot{u} = a_{0i} n_i \ , \tag{2.309}$$

since $\dot{u}$ is orthogonal to u. So, in the notation of Sect. 2.3, which was a completely general construction using any smoothly chosen tetrad along the worldline, and for an arbitrary smooth timelike worldline, the relation

$$\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\nu)} \tilde{A}^{(\nu)}{}_{(\kappa)}$$

is replaced by

$$\begin{cases} \dot{\lambda}_{(i)} = a_{0i}\lambda_{(0)} + \Omega_{ij}\lambda_{(j)} \ , \\ \dot{\lambda}_{(0)} = a_{0i}\lambda_{(i)} \ . \end{cases} \tag{2.310}$$

We can now read off the matrix $\tilde{A}$, obtaining

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix} \ , \tag{2.311}$$

with ν specifying the row and κ the column.

This allows us to answer the question on p. 93, viz., why should the components of $\mathbf{a}$, which is just part of a row of the acceleration matrix, be associated with the directions of the space tetrad vectors $\lambda_{(i)}$, $i \in \{1,2,3\}$? It is because of (2.308).

When the tetrad in question is FW transported, this immediately tells us that the Ω_{ij} in (2.307) are all zero, because this equation, with the help of the definition $a_{0i} := n_i \cdot \dot{u}$ from (2.308), then becomes the FW transport equation for any space vectors $n_i{}^{\mu}$ that are known to remain permanently orthogonal to the four-velocity u of the worldline. So the acceleration matrix for an FW transported tetrad always has the form

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & 0 & 0 \\ a_{02} & 0 & 0 & 0 \\ a_{03} & 0 & 0 & 0 \end{pmatrix} \qquad \text{(FW transported tetrad)} \ . \tag{2.312}$$

This is the case for the tetrad $\{\tilde{\lambda}_{(\mu)}\}_{\mu=0,1,2,3}$, as we carefully demonstrated in the last section. We know immediately then that there is a three-vector $\tilde{\mathbf{a}}$ such that

$$\frac{\mathrm{d}\tilde{\lambda}^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \tilde{\lambda}^{\mu}_{(\nu)} \tilde{A}^{(\nu)}{}_{(\kappa)} \ , \tag{2.313}$$

with $\tilde{A}^{(\nu)}{}_{(\kappa)}$ of the form (2.312). We thus have the relation

$$\boxed{\frac{\mathrm{d}\tilde{\lambda}^{\mu}_{(i)}}{\mathrm{d}\tau} = \tilde{a}_i \tilde{\lambda}^{\mu}_{(0)} \,, \qquad i = 1,2,3} \,, \tag{2.314}$$

and this brings us back to (2.307) in the case where the Ω_{ij} are all zero. The three quantities $\tilde{a}_i$, $i = 1,2,3$, are given by (2.308), i.e.,

$$\tilde{a}_i = \tilde{\lambda}_{(i)} \cdot \dot{u} \,, \tag{2.315}$$

and we have calculated the interesting cases in (2.297) and (2.300), viz.,

$$\tilde{a}_1 = \tilde{\lambda}_{(1)} \cdot \dot{u} = -v\gamma^2 \Omega_0 \cos\Phi - \gamma^2 \frac{\mathrm{d}v}{\mathrm{d}\tau} \sin\Phi \,, \tag{2.316}$$

$$\tilde{a}_2 = \tilde{\lambda}_{(2)} \cdot \dot{u} = -v\gamma^2 \Omega_0 \sin\Phi + \gamma^2 \frac{\mathrm{d}v}{\mathrm{d}\tau} \cos\Phi \,, \tag{2.317}$$

and, of course, $\tilde{a}_3 = 0$. So finally,

$$\boxed{\tilde{\mathbf{a}} = \begin{pmatrix} -v\gamma^2 \Omega_0 \cos\Phi - \gamma^2 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} \sin\Phi \\ -v\gamma^2 \Omega_0 \sin\Phi + \gamma^2 \dfrac{\mathrm{d}v}{\mathrm{d}\tau} \cos\Phi \\ 0 \end{pmatrix}} \,, \tag{2.318}$$

noting again that this is generally a function of the proper time.

We can make a general point about the notion of acceleration matrix here. As we have just seen, for arbitrary motion of the observer and arbitrary smooth choice of tetrad $\{\lambda_{(\kappa)}\}_{\kappa=0,1,2,3}$ along the observer worldline with $\lambda_{(0)}$ the observer four-velocity, there is a matrix $\tilde{A}^{(\nu)}{}_{(\kappa)}$ such that

$$\frac{\mathrm{d}\lambda^{\mu}_{(\kappa)}}{\mathrm{d}\tau} = \lambda^{\mu}_{(\nu)} \tilde{A}^{(\nu)}{}_{(\kappa)} \,, \tag{2.319}$$

where $\tilde{A}$ transforms as a type (1,1) tensor and its covariant associated tensor with components $\tilde{A}_{(\mu)(\nu)}$ is antisymmetric. Equation (2.319) just expresses the fact that the four-vector field $\mathrm{d}\lambda^{\mu}_{(\kappa)}/\mathrm{d}\tau$ along the worldline can be expressed as a linear combination of the tetrad for each τ, and antisymmetry follows from the argument (2.119) on p. 48, viz.,

$$\begin{aligned} \lambda_{(\kappa)} \cdot \lambda_{(\nu)} = \eta_{\kappa\nu} \quad &\Longrightarrow \quad \dot{\lambda}_{(\kappa)} \cdot \lambda_{(\nu)} + \lambda_{(\kappa)} \cdot \dot{\lambda}_{(\nu)} = 0 \\ &\Longrightarrow \quad \left[\lambda_{(\mu)} \tilde{A}^{(\mu)}{}_{(\kappa)}\right] \cdot \lambda_{(\nu)} + \lambda_{(\kappa)} \cdot \left[\lambda_{(\mu)} \tilde{A}^{(\mu)}{}_{(\nu)}\right] = 0 \\ &\Longrightarrow \quad \eta_{\mu\nu} \tilde{A}^{(\mu)}{}_{(\kappa)} + \eta_{\kappa\mu} \tilde{A}^{(\mu)}{}_{(\nu)} = 0 \\ &\Longrightarrow \quad \tilde{A}_{(\nu)(\kappa)} + \tilde{A}_{(\kappa)(\nu)} = 0 \,. \end{aligned} \tag{2.320}$$

But we may also ask whether there is always a matrix $A^{\mu}{}_{\nu}$ such that

$$\frac{\mathrm{d}\lambda^{\mu}_{(0)}}{\mathrm{d}\tau} = A^{\mu}{}_{\nu}\lambda^{\nu}_{(0)} \,. \tag{2.321}$$

The answer is affirmative. Of course, we simply define

$$A := \lambda\tilde{A}\lambda^{-1} \,, \tag{2.322}$$

where $\lambda(\tau)$ is the 4×4 matrix whose columns are the component forms of the four four-vectors $\lambda_{(\kappa)}$, $\kappa = 0,1,2,3$, in whatever inertial frame K is being used to express quantities. As explained in Sect. 2.4.4 for the case of constant acceleration matrices (see, in particular, p. 42), we are just reexpressing the tensor $\tilde{A}$ relative to the inertial frame K for each value of τ. In component form, the definition (2.322) reads

$$A^{\mu}{}_{\sigma} := \lambda^{\mu}{}_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\varepsilon)}(\lambda^{-1})^{(\varepsilon)}{}_{\sigma} \,. \tag{2.323}$$

It is in fact easy to show that we have, not only (2.321), but the general result

$$\frac{\mathrm{d}\lambda^{\mu}_{(\alpha)}}{\mathrm{d}\tau} = A^{\mu}{}_{\sigma}\lambda^{\sigma}_{(\alpha)} \,. \tag{2.324}$$

Indeed,

$$\begin{aligned}
A^{\mu}{}_{\sigma}\lambda^{\sigma}_{(\alpha)} &= \lambda^{\mu}{}_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\varepsilon)}(\lambda^{-1})^{(\varepsilon)}{}_{\sigma}\lambda^{\sigma}_{(\alpha)} \\
&= \lambda^{\mu}{}_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\varepsilon)}\delta^{(\varepsilon)}{}_{(\alpha)} \\
&= \lambda^{\mu}{}_{(\kappa)}\tilde{A}^{(\kappa)}{}_{(\alpha)} \\
&= \frac{\mathrm{d}\lambda^{\mu}_{(\alpha)}}{\mathrm{d}\tau} \,,
\end{aligned}$$

by (2.319). Then by the general argument (2.84) on p. 41, we have an isometric transport of four-vectors along the observer worldline, in the sense that, for

$$\frac{\mathrm{d}}{\mathrm{d}\tau}v(\tau) = Av(\tau) \,, \quad \frac{\mathrm{d}}{\mathrm{d}\tau}w(\tau) = Aw(\tau) \,, \tag{2.325}$$

we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\big[v(\tau)\cdot w(\tau)\big] &= \big[Av(\tau)\big]\cdot\big[w(\tau)\big] + \big[v(\tau)\big]\cdot\big[Aw(\tau)\big] \\
&= \eta_{\mu\sigma}A^{\mu}{}_{\nu}v^{\nu}w^{\sigma} + \eta_{\mu\sigma}v^{\mu}A^{\sigma}{}_{\nu}w^{\nu} \\
&= A_{\sigma\nu}v^{\nu}w^{\sigma} + A_{\mu\nu}v^{\mu}w^{\nu} \\
&= A_{\mu\nu}(v^{\nu}w^{\mu} + v^{\mu}w^{\nu}) = 0 \,.
\end{aligned} \tag{2.326}$$

The kind of transport satisfying (2.325) preserves scalar products of four-vectors because the covariant acceleration matrix appearing in (2.325) is antisymmetric.

As mentioned on p. 43, this shows that, whatever tetrad $\{\lambda_{(\kappa)}\}_{\kappa=0,1,2,3}$ we choose, it can be considered to be isometrically transported along the observer worldline in the sense of (2.82) on p. 41, or indeed (2.95) and (2.96) on p. 43, for the matrix $A^{\mu}{}_{\sigma}$. Naturally, this matrix depends on the choice of space triad, as can be seen from (2.322). The difference with that earlier discussion is that we are now considering quite general acceleration matrices which may not be constant along the worldline.

In the parallel discussion for constant acceleration matrices on p. 46 and following pages, we made the definition [see (2.112) on p. 47]

$$A^{\nu}{}_{\sigma} := \hat{\lambda}^{\nu}_{(\gamma)}\tilde{A}^{(\gamma)}{}_{(\kappa)}\big(\hat{\lambda}^{-1}\big)^{(\kappa)}{}_{\sigma}\,, \tag{2.327}$$

where $\hat{\lambda}$ is the initial ICIF $\lambda(0)$, rather than $\lambda(\tau)$ as in the above argument. But if $\tilde{A}$ is constant, this amounts to exactly the same definition since we have

$$\lambda(0)\tilde{A}\lambda(0)^{-1} = \lambda(\tau)\tilde{A}\lambda(\tau)^{-1}\,,\quad \forall\tau\,,\quad \text{when } \tilde{A} \text{ is constant}\,.$$

This follows because, by (2.117) on p. 47,

$$\lambda(\tau) = \lambda(0)\exp(\tilde{A}\tau/c) = \hat{\lambda}\exp(\tilde{A}\tau/c)\,,$$

whence

$$\begin{aligned}\lambda(\tau)\tilde{A}\lambda(\tau)^{-1} &= \hat{\lambda}\exp(\tilde{A}\tau/c)\tilde{A}\exp(-\tilde{A}\tau/c)\hat{\lambda}^{-1}\\ &= \hat{\lambda}\tilde{A}\hat{\lambda}^{-1}\,,\end{aligned}$$

as claimed.

2.11.5 Geodesic Coordinates for Rotating Tetrad

We now use the rotating tetrad $\{\lambda_{(\mu)}\}_{\mu=0,1,2,3}$ of (2.273) on p. 90 to construct coordinates according to the discussion in Sect. 2.3. Such coordinates are often called geodesic coordinates. This is an idea that generalises to curved spacetimes, as we shall see later (Sect. 3.2.1). We choose any smooth tetrad frame along the worldline, with the first member of the tetrad being the four-velocity and the other three members being orthogonal to it, hence spacelike. For any finite stretch of the worldline, there is then some neighbourhood of that part of the worldline such that, for any event chosen in that neighbourhood, there is a unique proper time τ on the worldline such that there exists a spacelike geodesic from the event corresponding to τ on the worldline to the chosen event with the property that this geodesic is orthogonal to the worldline at the event corresponding to τ.

The spacelike geodesic found in this way has a tangent vector at the event on the worldline corresponding to τ which can be uniquely expressed as a linear combination $X^i\lambda^\mu_{(i)}(\tau)$ of the three spacelike members $\lambda^\mu_{(i)}(\tau)$, $i=1,2,3$, of the tetrad. We then attribute geodesic coordinates (τ, X^1, X^2, X^3) to the originally chosen event.

In a flat spacetime, of the kind considered here, the inertial coordinates x^μ of the chosen event can be related algebraically to the geodesic coordinates $(\tau, \mathbf{X})$ by

$$x^\mu = x^\mu_A(\tau) + X^i\lambda^\mu_{(i)}(\tau)\;, \qquad T := \tau\;, \tag{2.328}$$

where $x_A(\tau)$ specifies the worldline of the revolving observer (or object) A in the inertial coordinate system. This works because the geodesics are straight lines in the inertial coordinate system. Equation (2.328) is just an example of (2.18) on p. 23.

Note, however, that there is no particular reason to choose the rotating tetrad $\{\lambda_{(\mu)}\}_{\mu=0,1,2,3}$ which Mashhoon finds so natural, rather than, say, the FW transported tetrad $\{\tilde{\lambda}_{(\mu)}\}_{\mu=0,1,2,3}$, for this construction. This is the real problem with attempts to make an observer picture: there is too much choice. Note also that, in the present case, where the angular speed of the observer A may not be constant, only an FW transported tetrad will deliver a rigid coordinate system, and these are likely to provide a better agreement with the approximation hoped for by the locality hypothesis to be discussed later.

We now write $\mathbf{X} = (X, Y, Z)$ and recall that we assumed on p. 90 that $\tau = \tau(t)$ has an inverse $t = F(\tau)$. We have the notation $\varphi_A(t) = \phi(\tau)$ [see the definition just prior to (2.272) on p. 90]. Then we also need the coordinates x^μ_A of A in L as given by (2.269) on p. 89, viz.,

$$x^\mu_A = (t, r\cos\varphi_A, r\sin\varphi_A, 0) = \big(F(T), r\cos\phi(T), r\sin\phi(T), 0\big)\;, \tag{2.329}$$

and the components of the tetrad $\{\lambda_{(\kappa)}\}_{\kappa=0,1,2,3}$ in L, as given by (2.273) on p. 90, viz.,

$$\begin{aligned} \lambda^\mu_{(0)} &= \gamma(1, -v\sin\phi, v\cos\phi, 0)\;,\\ \lambda^\mu_{(1)} &= (0, \cos\phi, \sin\phi, 0)\;,\\ \lambda^\mu_{(2)} &= \gamma(v, -\sin\phi, \cos\phi, 0)\;,\\ \lambda^\mu_{(3)} &= (0,0,0,1)\;. \end{aligned} \tag{2.330}$$

With $\mu = 0$ in (2.328),

$$t = x^0_A(\tau) + X^i\lambda^0_{(i)}(\tau) = F(T) + \gamma(T)v(T)Y\;.$$

With $\mu = 1$, we obtain

$$x = x^1_A(\tau) + X^i\lambda^1_{(i)}(\tau) = (X+r)\cos\phi(T) - \gamma(T)Y\sin\phi(T)\;.$$

With $\mu = 2$, we obtain

$$x = x^2_A(\tau) + X^i\lambda^2_{(i)}(\tau) = (X+r)\sin\phi(T) + \gamma(T)Y\cos\phi(T)\;.$$

With $\mu = 3$, we obtain

$$x = x_A^3(\tau) + X^i \lambda_{(i)}^3(\tau) = Z \; .$$

Summing all that up,

$$\boxed{\begin{aligned} &t = F(T) + \gamma(T)v(T)Y \; , \\ &x = (X+r)\cos\phi(T) - \gamma(T)Y\sin\phi(T) \; , \\ &y = (X+r)\sin\phi(T) + \gamma(T)Y\cos\phi(T) \; , \\ &z = Z \; . \end{aligned}} \tag{2.331}$$

When $r = 0$, so that $v = 0$ and $\gamma = 1$, this becomes

$$\begin{aligned} &t = F(T) + \gamma(T)v(T)Y = T \; , \\ &x = X\cos\phi(T) - \gamma(T)Y\sin\phi(T) = X\cos\phi(T) - Y\sin\phi(T) \; , \\ &y = X\sin\phi(T) + \gamma(T)Y\cos\phi(T) = X\sin\phi(T) + Y\cos\phi(T) \; , \\ &z = Z \; , \end{aligned} \tag{2.332}$$

which is just the transformation from the coordinates $(t' = t, x', y', z')$ obtained from (x, y, z) by rotation about the z axis with frequency $\hat{\Omega}_0(t)$. Interestingly, Mashhoon describes this as the geodesic coordinate system constructed along the worldline of the *non-inertial* observer at rest at the origin of the space coordinates in L. Such an observer is of course *inertial*, if she merely chooses a rotating space tetrad! She would be non-inertial in some sense if she spun round with the frequency $\hat{\Omega}_0(t)$, assuming that she had spatial extent.

2.11.6 Metric for Rotating Tetrad

We can now obtain the components of the Minkowski metric tensor relative to the coordinates (T, X, Y, Z). We shall do this in two ways. First of all, we can make the connection with the general results in Sect. 2.3.8:

$$g_{\mu\nu} = \begin{pmatrix} \left(1+\xi^i a_{0i}\right)^2 - \xi^i\xi^j\Omega_{ik}\Omega_{jk} & \xi^j\Omega_{j1} & \xi^j\Omega_{j2} & \xi^j\Omega_{j3} \\ \xi^j\Omega_{j1} & -1 & 0 & 0 \\ \xi^j\Omega_{j2} & 0 & -1 & 0 \\ \xi^j\Omega_{j3} & 0 & 0 & -1 \end{pmatrix} \; . \tag{2.333}$$

It remains only to identify the notation. In the earlier construction, the relation

$$x^\mu(\xi, \tau) = x^\mu(0, \sigma) + \xi^i n_i{}^\mu(\sigma) \tag{2.334}$$

corresponds to the above relation (2.328), viz.,

$$x^\mu = x_A^\mu(\tau) + X^i \lambda_{(i)}^\mu(\tau)\,, \qquad T := \tau\,. \tag{2.335}$$

In (2.334), τ is the proper time of the point with fixed spatial SE coordinates ξ^i, $i = 1,2,3$, while $\sigma(\xi,\tau)$ is the proper time of the main observer at the space origin $\xi = 0$ at which the latter considers herself simultaneous with (ξ,τ). So σ in (2.334) corresponds to τ, and hence T, in (2.335). Naturally, ξ^i corresponds to X^i and n_i to $\lambda_{(i)}$.

To rewrite the above metric (2.333), we just have to identify a_{0i} and Ω_{ij} using the discussion in Sect. 2.11.4. Comparing the acceleration matrix in (2.311) on p. 99, viz.,

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & \Omega_{21} & \Omega_{31} \\ a_{02} & \Omega_{12} & 0 & \Omega_{32} \\ a_{03} & \Omega_{13} & \Omega_{23} & 0 \end{pmatrix}\,, \tag{2.336}$$

with (2.281) on p. 92, viz.,

$$\mathscr{A}_\alpha{}^\beta = \begin{pmatrix} 0 & -\gamma^2 v\Omega_0 & \gamma^2 \mathrm{d}v/\mathrm{d}\tau & 0 \\ -\gamma^2 v\Omega_0 & 0 & \gamma^2\Omega_0 & 0 \\ \gamma^2 \mathrm{d}v/\mathrm{d}\tau & -\gamma^2\Omega_0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\,, \tag{2.337}$$

we have the correspondence

$$a_{01} = -\gamma^2 v\Omega_0\,, \quad a_{02} = \gamma^2 \dot{v}\,, \quad a_{03} = 0\,, \tag{2.338}$$

$$\Omega_{12} = -\gamma^2\Omega_0 = -\Omega_{21}\,, \quad \Omega_{23} = 0 = \Omega_{31}\,, \tag{2.339}$$

where $\dot{v} := \mathrm{d}v/\mathrm{d}\tau$. Noting that $\xi^i\xi^j\Omega_{ik}\Omega_{jk}$ is the sum of the squares of the terms $\xi^j\Omega_{j1}$, $\xi^j\Omega_{j2}$, and $\xi^j\Omega_{j3}$, which gives

$$\xi^i\xi^j\Omega_{ik}\Omega_{jk} = (X^2+Y^2)\gamma^4\Omega_0^2\,,$$

we thus obtain our first version of the metric for this coordinate construction:

$$g_{\mu\nu} = \begin{pmatrix} \left(1 - X\gamma^2 v\Omega_0 + Y\gamma^2\dot{v}\right)^2 - (X^2+Y^2)\gamma^4\Omega_0^2 & Y\gamma^2\Omega_0 & -X\gamma^2\Omega_0 & 0 \\ Y\gamma^2\Omega_0 & -1 & 0 & 0 \\ -X\gamma^2\Omega_0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\,. \tag{2.340}$$

Note that the components are functions of T as well as X and Y. This dependence occurs through the functions $v(T)$, $\dot{v}(T)$, $\gamma(T)$ and $\Omega_0(T)$.

To check that everything is working, and in particular that we are justified in quoting the general result from Sect. 2.3.8, let us go back to the coordinate transformation relations (2.331) on p. 104, viz.,

$$\begin{aligned} t &= F(T) + \gamma(T)v(T)Y \,, \\ x &= (X+r)\cos\phi(T) - \gamma(T)Y\sin\phi(T) \,, \\ y &= (X+r)\sin\phi(T) + \gamma(T)Y\cos\phi(T) \,, \\ z &= Z \,, \end{aligned} \tag{2.341}$$

and use the standard transformation formula

$$g_{\mu\nu} = \frac{\partial x^\sigma}{\partial X^\mu}\frac{\partial x^\tau}{\partial X^\nu}\eta_{\sigma\tau} \,, \tag{2.342}$$

where the capital letter X^μ denotes (T,X,Y,Z).

Recall from (2.278) and (2.279) on p. 91 that

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\tau} = \gamma^3 v\frac{\mathrm{d}v}{\mathrm{d}\tau} \,, \quad \frac{\mathrm{d}}{\mathrm{d}\tau}(\gamma v) = \gamma^3\frac{\mathrm{d}v}{\mathrm{d}\tau} \,, \quad \frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \gamma\Omega_0 \,,$$

and by (2.271) on p. 90 that

$$\frac{\mathrm{d}F}{\mathrm{d}\tau} = \gamma(\tau) = \gamma(T) \,.$$

We obtain

$$\begin{pmatrix} \partial x^0/\partial X^0 \\ \partial x^1/\partial X^0 \\ \partial x^2/\partial X^0 \\ \partial x^3/\partial X^0 \end{pmatrix} = \begin{pmatrix} \gamma + \gamma^3\dot{v}Y \\ -(X+r)\gamma\Omega_0\sin\phi - Y\gamma^3 v\dot{v}\sin\phi - Y\gamma^2\Omega_0\cos\phi \\ (X+r)\gamma\Omega_0\cos\phi + Y\gamma^3 v\dot{v}\cos\phi - Y\gamma^2\Omega_0\sin\phi \\ 0 \end{pmatrix} \,,$$

$$\begin{pmatrix} \partial x^0/\partial X^1 \\ \partial x^1/\partial X^1 \\ \partial x^2/\partial X^1 \\ \partial x^3/\partial X^1 \end{pmatrix} = \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} \,, \quad \begin{pmatrix} \partial x^0/\partial X^2 \\ \partial x^1/\partial X^2 \\ \partial x^2/\partial X^2 \\ \partial x^3/\partial X^2 \end{pmatrix} = \begin{pmatrix} \gamma v \\ -\gamma\sin\phi \\ \gamma\cos\phi \\ 0 \end{pmatrix} \,,$$

and

$$\begin{pmatrix} \partial x^0/\partial X^3 \\ \partial x^1/\partial X^3 \\ \partial x^2/\partial X^3 \\ \partial x^3/\partial X^3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \,.$$

According to (2.342), the components of $g_{\mu\nu}$ are found by taking scalar products of these four-component objects for the metric $\eta_{\mu\nu}$. For example, we find that

$$\begin{aligned}
g_{00} = &\ \left(\gamma + \gamma^2 \dot{v} Y\right)^2 - \Big[(X+r)\gamma\Omega_0 \sin\phi + Y\gamma^3 v\dot{v}\sin\phi + Y\gamma^2\Omega_0\cos\phi\Big]^2 \\
&- \Big[(X+r)\gamma\Omega_0\cos\phi + Y\gamma^3 v\dot{v}\cos\phi - Y\gamma^2\Omega_0\sin\phi\Big]^2 \\
= &\ \gamma^2 + 2\gamma^4\dot{v}Y + \gamma^6\dot{v}^2Y^2 - (X+r)^2\gamma^2\Omega_0^2\sin^2\phi - 2(X+r)\gamma^4\Omega_0 v\dot{v}Y\sin^2\phi \\
&-2(X+r)\gamma^3\Omega_0^2 Y\sin\phi\cos\phi - 2v\dot{v}\gamma^5\Omega_0 Y^2\sin\phi\cos\phi \\
&-\gamma^6 v^2\dot{v}^2 Y^2\sin^2\phi - \gamma^4\Omega_0^2 Y^2\cos^2\phi \\
&-(X+r)^2\gamma^2\Omega_0^2\cos^2\phi - 2(X+r)\gamma^4\Omega_0 v\dot{v}Y\cos^2\phi \\
&+2(X+r)\gamma^3\Omega_0^2 Y\sin\phi\cos\phi + 2v\dot{v}\gamma^5\Omega_0 Y^2\sin\phi\cos\phi \\
&-\gamma^6 v^2\dot{v}^2 Y^2\cos^2\phi - \gamma^4\Omega_0^2 Y^2\sin^2\phi \\
= &\ \gamma^2 + 2\gamma^4\dot{v}Y + \gamma^6\dot{v}^2Y^2 - (X+r)^2\gamma^2\Omega_0^2 - 2(X+r)\gamma^4\Omega_0 v\dot{v}Y \\
&-\gamma^4 v^2\dot{v}^2Y^2 - \gamma^4\Omega_0^2Y^2 \\
= &\ \gamma^2 + 2\gamma^4\dot{v}Y + \gamma^6\dot{v}^2Y^2 - (X+r)^2\gamma^2\Omega_0^2 - 2(X+r)\gamma^4\Omega_0 v\dot{v}Y - \gamma^4\Omega_0^2Y^2 \ .
\end{aligned}$$

The reason why this does not look like g_{00} in (2.340) is simply that we need to eliminate $r = v/\Omega_0$ [see (2.268) on p. 89]. Eliminating r from the last version of g_{00}, we do indeed recover the first version.

Taking the Lorentzian scalar product of the first and second four-component objects, we also have

$$\begin{aligned}
g_{01} = &\begin{pmatrix} \partial x^0/\partial X^0 \\ \partial x^1/\partial X^0 \\ \partial x^2/\partial X^0 \\ \partial x^3/\partial X^1 \end{pmatrix} \cdot \begin{pmatrix} \partial x^0/\partial X^1 \\ \partial x^1/\partial X^1 \\ \partial x^2/\partial X^1 \\ \partial x^3/\partial X^1 \end{pmatrix} \\
= &\ \cos\phi\Big[(X+r)\gamma\Omega_0\sin\phi + Y\gamma^3 v\dot{v}\sin\phi + Y\gamma^2\Omega_0\cos\phi\Big] \\
&+ \sin\phi\Big[-(X+r)\gamma\Omega_0\cos\phi - Y\gamma^3 v\dot{v}\cos\phi + Y\gamma^2\Omega_0\sin\phi\Big] \\
= &\ Y\gamma^2\Omega_0 \ ,
\end{aligned}$$

as we found previously. Taking the Lorentzian scalar product of the first and third four-component objects, we have

$$\begin{aligned}
g_{02} &= \begin{pmatrix} \partial x^0/\partial X^0 \\ \partial x^1/\partial X^0 \\ \partial x^2/\partial X^0 \\ \partial x^3/\partial X^0 \end{pmatrix} \cdot \begin{pmatrix} \partial x^0/\partial X^2 \\ \partial x^1/\partial X^2 \\ \partial x^2/\partial X^2 \\ \partial x^3/\partial X^2 \end{pmatrix} \\
&= \gamma^2 v(1+\gamma^2 \dot{v} Y) - \gamma \sin\phi \Big[(X+r)\gamma\Omega_0 \sin\phi + Y\gamma^3 v\dot{v}\sin\phi + Y\gamma^2\Omega_0\cos\phi\Big] \\
&\quad -\gamma\cos\phi\Big[(X+r)\gamma\Omega_0\cos\phi + Y\gamma^3 v\dot{v}\cos\phi - Y\gamma^2\Omega_0\sin\phi\Big] \\
&= \gamma^2 v + \gamma^4 v\dot{v}Y - \gamma^2\Omega_0(X+r) - \gamma^4 v\dot{v}Y \\
&= -\gamma^2\Omega_0(X+r) + \gamma^2 v \\
&= -X\gamma^2\Omega_0 \,,
\end{aligned}$$

as before, having substituted in $r = v/\Omega_0$. Other components are trivial.

2.11.7 Metric for FW Transported Tetrad

We can also apply the general result (2.333) to the SE coordinates that would be constructed using the FW transported tetrad (2.288) of Sect. 2.11.3. We simply read off the values of a_{0i}, $i = 1,2,3$, from (2.318) on p. 100, viz.,

$$a_{01} = -v\gamma^2\Omega_0\cos\Phi - \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\sin\Phi\,, \quad a_{02} = -v\gamma^2\Omega_0\sin\Phi + \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\cos\Phi\,,$$

and $a_{03} = 0$, and insert $\Omega_{ij} = 0$, $\forall i,j \in \{1,2,3\}$. For the record, the result is

$$g_{\mu\nu}^{\mathrm{FW}} = \begin{pmatrix} g_{00}^{\mathrm{FW}} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\,, \tag{2.343}$$

where

$$g_{00}^{\mathrm{FW}} = \Bigg[1 + \left(-v\gamma^2\Omega_0\cos\Phi - \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\sin\Phi\right)X^{\mathrm{FW}} + \left(-v\gamma^2\Omega_0\sin\Phi + \gamma^2\frac{\mathrm{d}v}{\mathrm{d}\tau}\cos\Phi\right)Y^{\mathrm{FW}}\Bigg]^2\,, \tag{2.344}$$

with X^{FW} and Y^{FW} the first and second space coordinates. It should be noted that a_{01} and a_{02} are not constant. Indeed, the component g_{00}^{FW} is a function of the time coordinate T^{FW}. It depends on this coordinate through each of the functions $v(\tau)$, $\dot{v}(\tau)$, $\gamma(\tau)$, $\Phi(\tau)$ as given by (2.289) on p. 94, and $\Omega_0(\tau)$.

Regarding this time coordinate T^{FW}, it is worth remembering that it is just the proper time of the observer in circular motion, spread over her hyperplanes of simultaneity in just the same way as when setting up the rotating tetrad coordinates in Sect. 2.11.5. It is merely by choosing a suitable space triad along the worldline that we get the components g_{0i}^{FW}, $i = 1,2,3$, equal to zero.

2.12 Range of Validity of SE Coordinates

As usual, the component g_{00} of the metric (2.340) in Sect. 2.11.6 goes to zero on a surface beyond which it is negative. There is no fundamental problem with its being negative, although the interpretation of the coordinates has to change then (see below), but the coordinates are not useful in places where they make the matrix of metric components singular. Note, however, that this is the Minkowski metric, which is defined throughout spacetime, so the problem has to come from the matrix $(\partial x/\partial X)$ which we used to transform the matrix of metric components, as already explained in Sect. 2.3.8.

To recapitulate that discussion, note that

$$\det \frac{\partial x}{\partial X} = 0$$

precisely where

$$\det g = 0 \,.$$

For the most general semi-Euclidean form of the metric, viz., (2.333) on p. 104, which is

$$g_{\mu\nu} = \begin{pmatrix} \left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk} & \xi^j \Omega_{j1} & \xi^j \Omega_{j2} & \xi^j \Omega_{j3} \\ \xi^j \Omega_{j1} & -1 & 0 & 0 \\ \xi^j \Omega_{j2} & 0 & -1 & 0 \\ \xi^j \Omega_{j3} & 0 & 0 & -1 \end{pmatrix} , \tag{2.345}$$

the matrix of components of the Minkowski metric has determinant

$$\boxed{\det g_{\mathrm{SE}}^{\mathrm{Mink}} = -\left(1+\xi^i a_{0i}\right)^2} \,. \tag{2.346}$$

This is always independent of the rotation chosen for the space triad $\{n_i\}_{i=1,2,3}$, as specified by Ω_i, $i = 1,2,3$, but it does depend on the acceleration of the worldline as specified by its absolute components a_{0i}, $i = 1,2,3$. It is zero for all ξ^i satisfying

$$\xi^i a_{0i}(\sigma) = -1 \,,$$

for some value of the proper time σ of the observer. This specifies a 2-plane of the 3D space of ξ^i for each proper time σ. Note that we also have

$$g_{00} \leq 0 \,, \quad \text{when} \;\; \det g_{\text{SE}}^{\text{Mink}} = 0 \,. \tag{2.347}$$

Other things can go wrong with this kind of coordinate construction, especially in a context where one would like each point of constant $\{\xi^i\}_{i=1,2,3}$ to correspond to a particle in a medium, as was the case for the initial discussion in Sect. 2.3. Recall that we had

$$x^\mu(\xi,\tau) = x^\mu(0,\sigma) + \xi^i n_i^\mu(\sigma) \,, \quad u^\mu := \dot{x}^\mu(\xi,\tau) = (u_0^\mu + \xi^i \dot{n}_i^\mu)\dot{\sigma} \,, \tag{2.348}$$

where $x^\mu(\xi,\tau)$ specifies the worldline of the point of constant $\{\xi^i\}_{i=1,2,3}$, with τ being its proper time, a dot on this symbol refers to its partial derivative with respect to τ for fixed $\{\xi^i\}_{i=1,2,3}$ (thereby delivering the four-velocity u^μ of this point), and σ is a proper time of the main observer at $\xi = 0$, depending on both ξ^i and τ, such that the event $x^\mu(\xi,\tau)$ is simultaneous with $x^\mu(0,\sigma)$ in the reckoning of the main observer at $\xi = 0$.

This construction assumes first and foremost that the latter always exists, i.e., for the event $x^\mu(\xi,\tau)$, there exists a point on the main worldline that the main observer considers to be simultaneous with it. A case where this assumption breaks down is translational uniform acceleration, if we choose events outside region I in Fig. 2.6 on p. 81. It also assumes that there is a *unique* event on the main worldline that the main observer considers to be simultaneous with it. This assumption also breaks down for translational uniform acceleration, and it breaks down in general for accelerating observers (see the discussion in Sect. 2.2.2).

Yet another problem is that the object u^μ in (2.348) may not have the credentials of a four-velocity, i.e., it may be impossible for it to satisfy $u^2 = 1$. This happens when the four-vector

$$W^\mu := u_0^\mu + \xi^i \dot{n}_i^\mu$$

is null or spacelike. When we write

$$\dot{n}_i^\mu = a_{0i} u_0^\mu + \Omega_{ij} n_j^\mu \,,$$

this becomes

$$W^\mu = (1 + \xi^i a_{0i}) u_0^\mu + \xi^i \Omega_{ij} n_j^\mu \,,$$

and there should be no great surprise to note that the condition for u^μ to be a four-velocity is precisely

$$g_{00} = \left(1 + \xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik} \Omega_{jk} > 0 \,.$$

So this is never satisfied, for example, when $\det g_{\text{SE}}^{\text{Mink}} = 0$ [see (2.347) above]. Note, however, that there is no fundamental problem with g_{00} being zero or negative. It just means that we cannot interpret worldlines of fixed ξ as being worldlines of particles or observers when this is the case. Basically, it disallows the first property on the list back on p. 12.

2.12.1 Range of Validity for General Circular Motion and Rotating Tetrad

In the case of an observer with general circular motion and the FS transported tetrad leading to the metric (2.340) on p. 105,

$$a_{01} = -\gamma^2 v\Omega_0 \,, \quad a_{02} = \gamma^2 \dot{v} \,, \quad a_{03} = 0 \,, \tag{2.349}$$

$$\Omega_{12} = -\gamma^2 \Omega_0 = -\Omega_{21} \,, \quad \Omega_{23} = 0 = \Omega_{31} \,. \tag{2.350}$$

Note that each of these quantities is a function of T, but neither X, Y, nor Z. The coordinate transformation to (T,X,Y,Z) is singular when

$$\det g_{\mathrm{SE}} = 0 \iff 1 + \xi^i a_{0i} = 0 \iff 1 - \gamma^2 v\Omega_0 X + \gamma^2 \dot{v} Y = 0 \,. \tag{2.351}$$

This can be written

$$\det g_{\mathrm{SE}} = 0 \iff 1 - v^2 + \dot{v} Y = v\Omega_0 X \,. \tag{2.352}$$

For each value of T, this is a plane in the (T,X,Y,Z) system.

It is more difficult to express relative to the original inertial system (t,x,y,z) in the completely general case. However, since $v = r\Omega_0$, we have

$$1 + \dot{v} Y = v(X + r)\Omega_0 \,, \tag{2.353}$$

and we have

$$\begin{cases} X + r = x\cos\phi + y\sin\phi \,, \\ \gamma Y = -x\sin\phi + y\cos\phi \,, \end{cases} \tag{2.354}$$

whence the condition for coordinate singularity becomes

$$1 + \dot{v}\gamma^{-1}(-x\sin\phi + y\cos\phi) = v(x\cos\phi + y\sin\phi)\Omega_0 \,. \tag{2.355}$$

Unfortunately, this still refers to the coordinate T through the functions $\phi(T)$, $v(T)$ and $\dot{v}(T)$. We can in principle obtain $T = T(t,Y)$ by inverting the relation $t = F(T) + \gamma(T)v(T)Y$, then eliminate it from (2.355), but this brings back Y and we still have T in the expression if we eliminate Y using the second relation of (2.354). Of course, in specific cases, one might obtain an explicit expression for the surface of singularity of the coordinate transformation in the (t,x,y,z) coordinate system.

The condition for g_{00} to be positive discussed above, viz.,

$$\left(1+\xi^i a_{0i}\right)^2 > \xi^i \xi^j \Omega_{ik}\Omega_{jk} ,$$

becomes

$$(1-\gamma^2 v\Omega_0 X + \gamma^2 \dot{v} Y)^2 > (X^2+Y^2)\gamma^4\Omega_0^2 . \tag{2.356}$$

This expands out to

$$1 + X^2\gamma^4 v^2\Omega_0^2 + Y^2\gamma^4\dot{v}^2 - 2XY\gamma^4 v\dot{v}\Omega_0 + 2Y\gamma^2\dot{v} - 2X\gamma^2 v\Omega_0 \\ -X^2\gamma^4\Omega_0^2 - Y^2\gamma^4\Omega_0^2 > 0 .$$

The second and seventh terms on the left-hand side join to give

$$1 - X^2\gamma^2\Omega_0^2 + Y^2\gamma^4\dot{v}^2 - 2XY\gamma^4 v\dot{v}\Omega_0 + 2Y\gamma^2\dot{v} - 2X\gamma^2 v\Omega_0 - Y^2\gamma^4\Omega_0^2 > 0 .$$

This rearranges to

$$\boxed{X^2\Omega_0^2 + \gamma^2(\Omega_0^2 - \dot{v}^2)Y^2 + 2XY\gamma^2 v\dot{v}\Omega_0 < 1 - v^2 + 2Y\dot{v} - 2Xv\Omega_0} . \tag{2.357}$$

For each value of T, entering this through the functions $\Omega_0(T)$, $v(T)$, $\gamma(T)$, and $\dot{v}(T)$, we obtain a cylindrical surface in the (X,Y,Z) space, since Z is free to vary. Its section in the (X,Y) plane will be a conic section, whose nature depends on the relative signs of the terms.

Note that it is not sufficient to examine the coefficient of Y^2 in the above. The sign of this term would depend on whether Ω_0^2 was bigger than, equal to, or smaller than $\dot{v}^2$, and one might naively think that this would lead to an ellipse, a parabola, or a hyperbola, respectively. However, a more careful analysis is needed (see Sect. 2.13 and in particular Sect. 2.13.5). The actual condition is

$$\begin{aligned} I &< 0 , && \text{hyperbola} , \\ I &= 0 , && \text{parabola} , \\ I &> 0 , && \text{ellipse} , \end{aligned} \tag{2.358}$$

where $I := \gamma^2(\Omega_0^2 - \gamma^2\dot{v}^2)$ is the Lorentz invariant quantity we encountered in (2.287) on p. 93. Note the extra factor of γ^2 compared with the naive suggestion above. But of course, one would expect the condition for the shape of these surfaces to be Lorentz invariant, and hence involve a Lorentz invariant quantity like I.

2.12.2 Light Cylinder for Uniform Circular Motion and Rotating Tetrad

We can analyse one very simple case in which Ω_0^2 is definitely bigger than $\gamma^2\dot{v}^2$, namely the case of uniform angular speed, so that $\dot{v} = 0$. The condition $g_{00}^{\mathrm{SE}} = 0$, viz.,

$$X^2\Omega_0^2+\gamma^2(\Omega_0^2-\dot{v}^2)Y^2+2XY\gamma^2 v\dot{v}\Omega_0=1-v^2+2Y\dot{v}-2Xv\Omega_0\ , \tag{2.359}$$

then just specifies the light cylinder

$$(x^2+y^2)^{1/2}=r/v=:r_{\mathrm{LC}}\ . \tag{2.360}$$

Anyone rotating around the same centre with the same angular speed but at this radius would be moving at the speed of light, because if we multiply this by the frequency $\Omega_0=v/r$ to get the actual speed, we obtain unity, the speed of light in these units.

Let us just prove that claim. First, when $\dot{v}=0$, (2.359) becomes

$$X^2\Omega_0^2+\gamma^2\Omega_0^2Y^2=1-v^2-2Xv\Omega_0\ . \tag{2.361}$$

Let us transform this to the original inertial coordinates (t,x,y). We go to the conversion (2.331) on p. 104 from (T,X,Y) to (t,x,y), viz.,

$$\begin{aligned}
t&=F(T)+\gamma(T)v(T)Y\ ,\\
x&=(X+r)\cos\phi(T)-\gamma(T)Y\sin\phi(T)\ ,\\
y&=(X+r)\sin\phi(T)+\gamma(T)Y\cos\phi(T)\ ,
\end{aligned} \tag{2.362}$$

bearing in mind that v and γ are constant, and that $F(T)=\gamma T$ because $\mathrm{d}F/\mathrm{d}T=\gamma$. Recall also that $\phi(T)=\Omega_0\gamma T$. In this calculation, we shall just use ϕ for $\phi(T)$, since it will be the same everywhere. Hence

$$t=\gamma T+\gamma vY\ , \tag{2.363}$$

and we can invert the second and third relations of (2.362) to obtain

$$\begin{aligned}
t&=\gamma T+\gamma vY\ ,\\
X+r&=x\cos\phi+y\sin\phi\ ,\\
\gamma Y&=-x\sin\phi+y\cos\phi\ .
\end{aligned} \tag{2.364}$$

The last two are inserted in (2.361) to give

$$\begin{aligned}
(x\cos\phi+y\sin\phi-r)^2\Omega_0^2&+(-x\sin\phi+y\cos\phi)^2\Omega_0^2\\
&=1-v^2-2v\Omega_0(x\cos\phi+y\sin\phi-r)\\
&=1-v^2-2v\Omega_0x\cos\phi-2v\Omega_0y\sin\phi+2vr\Omega_0\ .
\end{aligned}$$

Multiplying out the left-hand side and using $\Omega_0=v/r$, we soon obtain

$$(x^2+y^2+r^2-2xr\cos\phi-2yr\sin\phi)\frac{v^2}{r^2}=1-v^2-\frac{2v^2x}{r}\cos\phi-\frac{2v^2y}{r}\sin\phi+2v^2\ ,$$

and this in turn quickly boils down to

$$(x^2+y^2)\frac{v^2}{r^2} = 1 ,$$

which is just the equation (2.360) for the light cylinder.

In [35, 36], Mashhoon asserts that the SE coordinates (T, X, Y, Z) we have been discussing for general circular motion are valid within the cylindrical region whose nature is specified by (2.358), but he does not specify what goes wrong outside this region. He merely talks about a region outside of which the SE coordinates are not permissible. However, it should be stressed that there is nothing *mathematically* problematic until we reach the surface where the coordinate transformation from inertial coordinates to SE coordinates is singular. In fact, Mashhoon's problem here is with the *physical interpretation* of the SE coordinates in regions where g_{00} is zero or negative. We shall return to that issue later.

2.12.3 Intersecting Hyperplanes of Simultaneity for Circular Motion

A general discussion of the condition $g_{00} = 0$ for these coordinate constructions is given in Sect. 2.13. However, we have not yet given any consideration to regions where we are sure that the SE coordinates become problematic, because the construction process breaks down, viz., regions where there is no unique point on the worldline of the main observer such that this observer considers a given point in those regions to be simultaneous (the existence of a point and its uniqueness are both crucial).

That does not seem to be a tractable problem in the completely general case, although we have already considered how this non-uniqueness problem arises from a qualitative point of view in Sect. 2.2.2. One case we considered was an observer who changes direction. When this observer has the same speed but in the opposite direction at two events on her worldline, her HOSs will meet somewhere to the right if the change of speed is to the left, and vice versa. In the case of circular motion, this happens all the time, although here we need a second space dimension, because the observer will have moved (a distance $2r$) in that other space direction between two velocity reversals.

To consider general circular motion would obscure the essential issue here, so let us deal only with constant angular velocity. If we picture the spiralling worldline with constant gradient for uniform rotation, the HOSs of two velocity-reversed events always meet outside the spiral in the direction to which the observer was originally moving (at the lower of the two events in the time dimension). It looks as though the intersections of all HOSs may generate a circular cylinder containing the spiral. The steeper its gradient, i.e., the smaller the angular speed, the further away this cylinder will be and the bigger the region in which we can build the SE coordinate system.

But if the gradient of the spiral were shallow, lying close to, but just above the null gradient, so that the HOSs were steep, lying just below the null gradient, could the HOSs for two consecutive velocity-reversed events on the worldline actually meet within the spiral? The two events would be closer in time coordinate. Note, however, that there is a limit to how big we can make the spiral radius if we are to keep that time difference between the two events equal, since the observer cannot move faster than light.

Let us analyse this quantitatively for the case of uniform rotation. Mashhoon does not consider this in [35, 36]. We shall show that things go wrong outside a cylinder of radius r/v^2. Since $v < 1$, this is always *outside* the light cylinder, which lies at $r_{\mathrm{LC}} = r/v$.

First consider the event on the spiralling worldline at time $t = 0$. This event is

$$(t,x,y) = (0,r,0) \ ,$$

dropping the z coordinate throughout the calculation. The HOS here is spanned by the vectors $\lambda_{(1)}(\phi = 0)$ and $\lambda_{(2)}(\phi = 0)$ as given by (2.330) on p. 103, viz.,

$$\lambda_{(1)}(\phi = 0) = (0,1,0) \ , \quad \lambda_{(2)}(\phi = 0) = (\gamma v,0,\gamma) \ .$$

The plane through the relevant event and spanned by these vectors is given by

$$\left\{ \lambda \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} v \\ 0 \\ 1 \end{pmatrix} : \lambda,\, \mu \in \mathbb{R} \right\} \ ,$$

or

$$\mathrm{HOS}_{\phi=0} = \left\{ \begin{pmatrix} \mu v \\ \lambda \\ \mu \end{pmatrix} : \lambda,\, \mu \in \mathbb{R} \right\} \ . \tag{2.365}$$

Now consider the event on the spiralling worldline at $t = \phi/\Omega_0$. This event is

$$(t,x,y) = \big(\phi/\Omega_0, r\cos\phi, r\sin\phi\big) \ .$$

The HOS here is spanned by the vectors $\lambda_{(1)}(\phi)$ and $\lambda_{(2)}(\phi)$ as given by (2.330) on p. 103, viz.,

$$\lambda_{(1)}(\phi = 0) = (0,\cos\phi,\sin\phi) \ , \quad \lambda_{(2)}(\phi = 0) = (\gamma v, -\gamma\sin\phi, \gamma\cos\phi) \ .$$

The plane through the relevant event and spanned by these vectors is given by

$$\left\{ \begin{pmatrix} \phi/\Omega_0 \\ r\cos\phi \\ r\sin\phi \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \end{pmatrix} + \beta \begin{pmatrix} v \\ -\sin\phi \\ \cos\phi \end{pmatrix} : \alpha,\, \beta \in \mathbb{R} \right\} \ ,$$

or

$$\mathrm{HOS}_{\phi} = \left\{ \begin{pmatrix} \phi/\Omega_0 + \beta v \\ -\beta \sin\phi + (r+\alpha)\cos\phi \\ \beta\cos\phi + (r+\alpha)\sin\phi \end{pmatrix} : \alpha, \beta \in \mathbb{R} \right\} . \tag{2.366}$$

We now find the line of intersection of $\mathrm{HOS}_{\phi=0}$ and HOS_{ϕ} as given by (2.365) and (2.366), by solving the three simultaneous equations

$$\begin{cases} \dfrac{\phi}{\Omega_0} + \beta v = \mu v \,, \\ -\beta \sin\phi + (r+\alpha)\cos\phi = \lambda \,, \\ \beta\cos\phi + (r+\alpha)\sin\phi = \mu \,, \end{cases} \tag{2.367}$$

in the four unknowns λ, μ, α, and β. The first and third can give us β in terms of α, viz.,

$$\beta = \frac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{v(1-\cos\phi)} \,. \tag{2.368}$$

We can then get λ and μ in terms of α alone, viz.,

$$\lambda = \frac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)}\sin\phi + (r+\alpha)\cos\phi \,, \tag{2.369}$$

$$\mu = \frac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{v(1-\cos\phi)}\cos\phi + (r+\alpha)\sin\phi \,, \tag{2.370}$$

and the line of intersection of the two hyperplanes of simultaneity is given by inserting these in the vector

$$\begin{pmatrix} \mu v \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \dfrac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{1-\cos\phi}\cos\phi + (r+\alpha)v\sin\phi \\ \dfrac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)}\sin\phi + (r+\alpha)\cos\phi \\ \dfrac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{v(1-\cos\phi)}\cos\phi + (r+\alpha)\sin\phi \end{pmatrix} ,$$

to give finally

$$\mathrm{L}_{0,\phi} = \left\{ \begin{pmatrix} \dfrac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{1-\cos\phi}\cos\phi + (r+\alpha)v\sin\phi \\ \dfrac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)}\sin\phi + (r+\alpha)\cos\phi \\ \dfrac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{v(1-\cos\phi)}\cos\phi + (r+\alpha)\sin\phi \end{pmatrix} : \alpha \in \mathbb{R} \right\} , \tag{2.371}$$

where $\mathrm{L}_{0,\phi}$ denotes the line of intersection of the hyperplanes of simultaneity at these two events.

We now consider the distance D of this line from the space origin, still ignoring the z dimension. We are going to minimise D to obtain the closest point to the space origin at which the coordinate construction breaks down, and this will give the closest point for any z. Indeed, it is the freedom in z that makes the final surface of breakdown a cylinder in the spacelike hypersurface covered by (x,y,z) for each value of t. But there is some work to do to get this picture clear (see below).

First note that

$$D^2 = \left[\frac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)}\sin\phi + (r+\alpha)\cos\phi\right]^2 + \left[\frac{(r+\alpha)v\sin\phi - \phi/\Omega_0}{v(1-\cos\phi)}\cos\phi + (r+\alpha)\sin\phi\right]^2 ,$$

whence

$$\boxed{D^2 = \left[\frac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)}\right]^2 + (r+\alpha)^2} . \tag{2.372}$$

We now minimise that with respect to α, which will also deliver the corresponding value of t when the line comes closest to the space origin, by inserting the relevant value of α into the first component of the vector in (2.371).

To minimise $D^2(\alpha)$, we consider

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}D^2 = 2\frac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v(1-\cos\phi)} \times -\frac{v\sin\phi}{v(1-\cos\phi)} + 2(r+\alpha) .$$

Then, recalling that $\Omega_0 = v/r$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}D^2 = 0 &\iff \frac{\phi/\Omega_0 - (r+\alpha)v\sin\phi}{v^2(1-\cos\phi)^2} \times v\sin\phi = r+\alpha \\
&\iff \left[1 + \frac{v^2\sin^2\phi}{v^2(1-\cos\phi)^2}\right](r+\alpha) = \frac{\phi r v\sin\phi}{v^3(1-\cos\phi)} \\
&\iff r+\alpha = \frac{\phi r\sin\phi}{v^2(1-\cos\phi)^2\left[1+\dfrac{\sin^2\phi}{(1-\cos\phi)^2}\right]} \\
&\qquad\quad = \frac{\phi r\sin\phi}{v^2\left[(1-\cos\phi)^2 + \sin^2\phi\right]} \\
&\qquad\quad = \frac{\phi r\sin\phi}{2v^2(1-\cos\phi)} .
\end{aligned}$$

For the record, the value $\alpha_{\min}$ of the parameter α that minimises the distance of the line $\mathrm{L}_{0,\phi}$ from the space origin in the (x,y) plane is given by

$$\boxed{r+\alpha_{\min}=\frac{\phi r\sin\phi}{2v^2(1-\cos\phi)}}\;. \tag{2.373}$$

Note that this works even for values of ϕ like π or $\pi/2$. For $\phi=\pi$, the observer has reversed its velocity with respect to $\phi=0$, while $\phi=\pi/2$ represents a rotation of the velocity through one quarter turn relative to $\phi=0$.

We feed the above value of α back into (2.372) to obtain the minimum distance $D^2_{\min}$ as

$$\begin{aligned}
D^2_{\min} &= \frac{\phi^2 r^2\sin^2\phi}{4v^4(1-\cos\phi)^2}+\left[\frac{\dfrac{\phi r}{v}-\dfrac{\phi r\sin^2\phi}{2v(1-\cos\phi)}}{v(1-\cos\phi)}\right]^2\\
&=\frac{\phi^2 r^2}{v^4}\left\{\frac{\sin^2\phi}{4(1-\cos\phi)^2}+\left[\frac{1-\dfrac{\sin^2\phi}{2(1-\cos\phi)}}{1-\cos\phi}\right]^2\right\}\,,
\end{aligned}$$

whence

$$D^2_{\min}=\frac{\phi^2 r^2}{4v^4(1-\cos\phi)^2}\left[\sin^2\phi+\left(2-\frac{\sin^2\phi}{1-\cos\phi}\right)^2\right]\,.$$

This simplifies enormously using

$$\sin^2\phi=1-\cos^2\phi=(1-\cos\phi)(1+\cos\phi)\,,$$

whence

$$\begin{aligned}
\sin^2\phi+\left(2-\frac{\sin^2\phi}{1-\cos\phi}\right)^2 &= \sin^2\phi+\big[2-(1+\cos\phi)\big]^2\\
&=\sin^2\phi+(1-\cos\phi)^2\\
&=\sin^2\phi+1-2\cos\phi+\cos^2\phi=2(1-\cos\phi)\,.
\end{aligned}$$

Hence finally,

$$\boxed{D^2_{\min}=\frac{\phi^2 r^2}{2v^4(1-\cos\phi)}}\;. \tag{2.374}$$

We now obtain the value $t_{\min}$ of t when the line comes closest to the space origin, by inserting the relevant value $\alpha_{\min}$ of α into the first component of the vector in (2.371) to obtain

$$
\begin{aligned}
t_{\min} &= \frac{\dfrac{\phi r \sin^2\phi}{2v(1-\cos\phi)} - \dfrac{\phi r}{v}}{1-\cos\phi}\cos\phi + \frac{\phi r \sin^2\phi}{2v(1-\cos\phi)} \\
&= \frac{\phi r \sin^2\phi}{2v(1-\cos\phi)}\left(\frac{\cos\phi}{1-\cos\phi}+1\right) - \frac{\phi r\cos\phi}{v(1-\cos\phi)} \\
&= \frac{\phi r}{2v(1-\cos\phi)}\left(\frac{\sin^2\phi}{1-\cos\phi} - 2\cos\phi\right) \\
&= \frac{\phi r}{2v(1-\cos\phi)}(1+\cos\phi - 2\cos\phi)\,,
\end{aligned}
$$

whence finally

$$
\boxed{t_{\min} = \frac{\phi r}{2v} = \frac{1}{2}\frac{\phi}{\Omega_0}}\,. \tag{2.375}
$$

Now the observer reaches the event parametrised by ϕ at inertial time ϕ/Ω_0, so the minimum distance $D_{\min}$ from the space origin of the (x,y) plane at which these two hyperplanes of simultaneity coincide occurs at an event with inertial time halfway between the two events. This could be expected from the symmetry of the situation.

Also by the symmetry of the situation, if we had started by considering hyperplanes of simultaneity at events for $-\phi/2$ and $\phi/2$, the relevant intersection of hyperplanes of simultaneity would have occurred at inertial time zero. Then varying ϕ from π down toward zero, we would get a whole set of minimum distances of HOS intersection occurring at inertial time $t=0$. The minimum of these, if there is one, sets the distance at which the coordinate construction breaks down for this value of the inertial time. But all values of the inertial time are equivalent by the symmetry of the situation, so we get this distance (at which the coordinate construction breaks down) for all values of t.

Now we need to find the x and y coordinates of the point of breakdown, returning to the calculation above for angles 0 and ϕ. This is important, because we have only calculated the distance from the origin of the (x,y) plane. We would like to show that the minimum of this distance occurs at angle $\phi/2$. The point is that the observer reaches this angle precisely at the inertial time of the coordinate breakdown due to this point of closest HOS intersection. That observer event on the observer's worldline then lies on the straight line from the space origin at that inertial time to the breakdown event at that inertial time.

Inserting the relevant value $\alpha_{\min}$ of α into the second component of the vector in (2.371), we obtain

$$
\begin{aligned}
x_{\min} &= \frac{\dfrac{\phi r}{v} - \dfrac{\phi r \sin^2\phi}{2v(1-\cos\phi)}}{v(1-\cos\phi)} \sin\phi + \frac{\phi r \sin\phi\cos\phi}{2v^2(1-\cos\phi)} \\
&= \frac{\phi r \sin\phi}{2v^2(1-\cos\phi)} \left(2 - \frac{\sin^2\phi}{1-\cos\phi} + \cos\phi \right) \\
&= \frac{\phi r \sin\phi}{2v^2(1-\cos\phi)} \big[2 - (1+\cos\phi) + \cos\phi \big] \\
&= \frac{\phi r \sin\phi}{2v^2(1-\cos\phi)} ,
\end{aligned}
$$

whence finally

$$
\boxed{x_{\min} = \frac{\phi r \sin\phi}{2v^2(1-\cos\phi)}} . \tag{2.376}
$$

Inspection of the first and third components of (2.371) shows that

$$
\boxed{y_{\min} = \frac{t_{\min}}{v} = \frac{\phi r}{2v^2}} . \tag{2.377}
$$

It is easy to check that

$$
\begin{aligned}
D_{\min}^2 = x_{\min}^2 + y_{\min}^2 &= \frac{\phi^2 r^2}{4v^4} \left[\frac{\sin^2\phi}{(1-\cos\phi)^2} + 1 \right] \\
&= \frac{\phi^2 r^2}{4v^4} \left[\frac{1+\cos\phi}{1-\cos\phi} + 1 \right] \\
&= \frac{\phi^2 r^2}{2v^4(1-\cos\phi)} ,
\end{aligned}
$$

agreeing with (2.374).

For $\phi = \pi$, we obtain $D_{\min} = \pi r/2v^2$. On the other hand, for $\phi = \pi/2$, we obtain $D_{\min} = \pi r/2\sqrt{2}v^2$. It turns out that $D_{\min}$ decreases to a finite limit as $\phi \to 0$. Since

$$
\cos\phi = 1 - \frac{\phi^2}{2} + O(\phi^4) ,
$$

we obtain

$$
\boxed{\lim_{\phi\to 0} D_{\min} = \frac{r}{v^2}} , \tag{2.378}
$$

as claimed earlier. This is well outside the light cylinder at r/v.

Concerning the angle $\phi_{\min}$ of the closest breakdown event to the space origin (and also to the observer), as measured from the x axis in the instantaneous (x,y) plane at the inertial time $t = \phi/2\Omega_0 = \phi r/2v = t_{\min}$, we have

$$\tan\phi_{\min} = \frac{y_{\min}}{x_{\min}} = \frac{1-\cos\phi}{\sin\phi} = \frac{1-\cos 2\dfrac{\phi}{2}}{\sin 2\dfrac{\phi}{2}} = \frac{1-\left(1-2\sin^2\dfrac{\phi}{2}\right)}{2\sin\dfrac{\phi}{2}\cos\dfrac{\phi}{2}} = \tan\frac{\phi}{2}\ ,$$

which is enough to show that

$$\boxed{\phi_{\min} = \frac{\phi}{2}}\ , \tag{2.379}$$

as claimed above.

2.12.4 A Brief Conclusion for Uniform Rotation

The problem of intersecting hyperplanes of simultaneity which complicates the construction of accelerating coordinate systems by this method occurs well outside the light cylinder. The problem considered by Mashhoon to invalidate the construction beyond a certain distance begins right on the light cylinder, where the (00) component of the semi-Euclidean metric goes to zero, even though there is no particular problem with the matrix of components of the metric relative to these coordinates either here or beyond the light cylinder, unless one wishes for observers to be able to sit at fixed SE space coordinates (a condition he does not mention explicitly).

Things do go seriously wrong with the matrix of components of the metric relative to these coordinates in regions where $\det g_{\mathrm{SE}} = 0$, and this happens because the transformation from inertial to SE coordinates has a singularity. For uniform rotation ($\dot{v} = 0$), the condition for singularity is (2.352) on p. 111, viz.,

$$X = \frac{1}{v\Omega_0\gamma^2} = \frac{r}{v^2\gamma^2} = \frac{r}{v^2} - r\ . \tag{2.380}$$

To visualise this in the inertial coordinate system, we require an inverse for the transformation given in (2.331) on p. 104, viz.,

$$\begin{aligned} t &= \gamma T + \gamma v Y\ ,\\ x &= (X+r)\cos\gamma\Omega_0 T - \gamma Y \sin\gamma\Omega_0 T\ ,\\ y &= (X+r)\sin\gamma\Omega_0 T + \gamma Y \cos\gamma\Omega_0 T\ ,\\ z &= Z\ . \end{aligned} \tag{2.381}$$

This is quite a complicated problem, even for this uniform rotation case. Note, however, that the SE spatial plane given by (2.380) lies at large X values for low speeds $v \ll 1$, but moves in toward $X = 0$ as $v \to 1$. One would expect it to occur a long way from the observer for low speeds, and become closer as v increases.

2.13 General Analysis of the Surface $g_{00} = 0$

Mashhoon asserts in [36] that the SE coordinates (T,X,Y,Z) are valid within a cylindrical region with boundary an elliptic cylinder for $I > 0$, a parabolic cylinder for $I = 0$, and a hyperbolic cylinder for $I < 0$, where I is the invariant given by (2.287) on p. 93, viz.,

$$I := -a^2 + \Omega^2 = \gamma^2 \Omega_0^2 - \gamma^4 \dot{v}^2 \ . \tag{2.382}$$

It should be borne in mind, however, that the only thing that becomes impossible is for observers to sit at fixed (X,Y,Z) in this region, i.e., these are no longer like spatial inertial coordinates in that respect. But they are still coordinates.

The aim in this section is to give a general discussion of the conditions on g_{00}, based on Mashhoon's account in [35]. He constructs the usual SE coordinate system along the arbitrary worldline D, specified by $x_D(\tau)$, where τ is the proper time. The orthornormal tetrad is denoted by $\{\lambda^{\mu}_{(\alpha)}\}_{\alpha=0,1,2,3}$, where $\lambda_{(0)} = u_D$ is the four-velocity. The tetrad is a function of the proper time and we define the SE coordinates $\{X^{\alpha}\}_{\alpha=0,1,2,3}$ as usual by

$$\tau = X^0 \ , \quad x^{\mu} = x^{\mu}_D(\tau) + X^i \lambda^{\mu}_{(i)}(\tau) \ , \tag{2.383}$$

at least as far as this construction is allowed by non-intersection of the spacelike geodesics orthogonal to the worldline at different proper times.

In Mashhoon's notation, the Minkowski metric now has components

$$g_{00} = -S \ , \qquad g_{0i} = U_i \ , \qquad g_{ij} = \delta_{ij} \ , \tag{2.384}$$

where

$$S := (1 + \mathbf{a} \cdot \mathbf{X})^2 - U^2 \ , \qquad \mathbf{U} = \Omega \times \mathbf{X} \ , \tag{2.385}$$

and $\mathbf{a}$ and Ω are the usual three-vector functions of the proper time defined as the components of the four-acceleration relative to the space triad $\{\lambda^{\mu}_{(i)}\}_{i=1,2,3}$, in the case of $\mathbf{a}$, and the rotational frequency of the same space triad defined by examining the expression of its proper time derivative relative to this same triad.

In the notation of Sect. 2.3,

$$S = (1 + \xi^i a_{0i})^2 - \xi^i \xi^j \Omega_{ik} \Omega_{jk} \ , \qquad U^i = \xi^j \Omega_{ji} \ , \tag{2.386}$$

with the correspondence

$$\xi^i \longleftrightarrow X^i \ , \qquad a_{0i} \longleftrightarrow a_i \ , \qquad \Omega_{ij} \longleftrightarrow \varepsilon_{ijk} \Omega_k \ .$$

We can then check that

$$\xi^j \Omega_{ji} \longleftrightarrow X^j \varepsilon_{jik} \Omega_k = \varepsilon_{ijk} \Omega_j X^k = (\Omega \times \mathbf{X})_i = U^i \ . \tag{2.387}$$

Note that the matrix of metric components has determinant

$$g := \det g_{\mathrm{SE}}^{\mathrm{Mink}} = -(1 + \mathbf{a} \cdot \mathbf{X})^2 \,, \tag{2.388}$$

as proven in (2.67) on p. 34. It is easy to show that the inverse of this matrix is given by

$$g^{00} = \frac{1}{g} \,, \qquad g^{0i} = -\frac{U^i}{g} \,, \qquad g^{ij} = \delta^{ij} + \frac{1}{g} U^i U^j \,, \tag{2.389}$$

in the case where $g \neq 0$.

We use the signature -2 here, following Mashhoon. In this picture, Mashhoon's view is that things go wrong when $g_{00} \geq 0$, although he does not explain why. In fact it is just because he would like observers to be able to sit at fixed $\mathbf{X}$. It will then be possible to pretend under some circumstances that these SE coordinates are inertial, to a degree of approximation that will remain concealed by experimental error. As we shall see, such a pretence is licensed by his locality 'hypothesis'.

So the boundary of the region that is admissible in this sense is a 3-surface specified by

$$S = 0 \,. \tag{2.390}$$

We write this in the form

$$S(X^0, \mathbf{X}) = 1 + 2a_i(X^0)X^i + M_{ij}(X^0)X^i X^j = 0 \,, \tag{2.391}$$

where M_{ij} is a symmetric matrix found by expanding out the expression for S in (2.385), and bearing in mind that X^0 is just proper time along the observer worldline. Clearly,

$$S = 1 + 2a_i X^i + a_i a_j X^i X^j - (\Omega \times \mathbf{X}) \cdot (\Omega \times \mathbf{X}) \,.$$

Using the result (2.387) established above, the last term is

$$\begin{aligned}
(\Omega \times \mathbf{X}) \cdot (\Omega \times \mathbf{X}) &= X^i X^j \Omega_{ik} \Omega_{jk} \\
&= X^i X^j \varepsilon_{ikl} \Omega_l \varepsilon_{jkm} \Omega_m \\
&= X^i X^j \Omega_l \Omega_m (\delta_{ij} \delta_{lm} - \delta_{im} \delta_{lj}) \\
&= \Omega^2 X^i X^j \delta_{ij} - \Omega_i \Omega_j X^i X^j \,,
\end{aligned}$$

whence finally,

$$\boxed{M_{ij} = a_i a_j + \Omega_i \Omega_j - \Omega^2 \delta_{ij}} \,, \tag{2.392}$$

with $\Omega^2 := \Omega_1^2 + \Omega_2^2 + \Omega_3^2$.

2.13.1 Eigenvalues and Determinant of M

The last relation gives a symmetric matrix

$$M = \begin{pmatrix} a_1^2+\Omega_1^2-\Omega^2 & a_1a_2+\Omega_1\Omega_2 & a_1a_3+\Omega_1\Omega_3 \\ a_1a_2+\Omega_1\Omega_2 & a_2^2+\Omega_2^2-\Omega^2 & a_2a_3+\Omega_2\Omega_3 \\ a_1a_3+\Omega_1\Omega_3 & a_2a_3+\Omega_2\Omega_3 & a_3^2+\Omega_3^2-\Omega^2 \end{pmatrix}\,, \tag{2.393}$$

which has characteristic equation $\det(M-\alpha I)=0$ given by

$$\det\begin{pmatrix} a_1^2+\Omega_1^2-\Omega^2-\alpha & a_1a_2+\Omega_1\Omega_2 & a_1a_3+\Omega_1\Omega_3 \\ a_1a_2+\Omega_1\Omega_2 & a_2^2+\Omega_2^2-\Omega^2-\alpha & a_2a_3+\Omega_2\Omega_3 \\ a_1a_3+\Omega_1\Omega_3 & a_2a_3+\Omega_2\Omega_3 & a_3^2+\Omega_3^2-\Omega^2-\alpha \end{pmatrix} = 0\,. \tag{2.394}$$

Let us define $\beta := \alpha+\Omega^2$ and solve this cubic equation in β, viz.,

$$\det\begin{pmatrix} a_1^2+\Omega_1^2-\beta & a_1a_2+\Omega_1\Omega_2 & a_1a_3+\Omega_1\Omega_3 \\ a_1a_2+\Omega_1\Omega_2 & a_2^2+\Omega_2^2-\beta & a_2a_3+\Omega_2\Omega_3 \\ a_1a_3+\Omega_1\Omega_3 & a_2a_3+\Omega_2\Omega_3 & a_3^2+\Omega_3^2-\beta \end{pmatrix} = 0\,. \tag{2.395}$$

This reads

$$\begin{aligned} 0 = {}& (a_1^2+\Omega_1^2-\beta)\Big[(a_2^2+\Omega_2^2-\beta)(a_3^2+\Omega_3^2-\beta)-(a_2a_3+\Omega_2\Omega_3)^2\Big] \\ &+(a_1a_2+\Omega_1\Omega_2)\Big[(a_2a_3+\Omega_2\Omega_3)(a_1a_3+\Omega_1\Omega_3) \\ &\qquad\qquad -(a_1a_2+\Omega_1\Omega_2)(a_3^2+\Omega_3^2-\beta)\Big] \\ &+(a_1a_3+\Omega_1\Omega_3)\Big[(a_1a_2+\Omega_1\Omega_2)(a_2a_3+\Omega_2\Omega_3) \\ &\qquad\qquad -(a_1a_3+\Omega_1\Omega_3)(a_2^2+\Omega_2^2-\beta)\Big]\,. \end{aligned}$$

This becomes

$$\begin{aligned} 0 = {}& (a_1^2+\Omega_1^2-\beta)\Big[(a_2^2+\Omega_2^2)(a_3^2+\Omega_3^2)-\beta(a_2^2+\Omega_2^2+a_3^2+\Omega_3^2) \\ &\qquad\qquad +\beta^2-(a_2a_3+\Omega_2\Omega_3)^2\Big] \\ &+2(a_1a_3+\Omega_1\Omega_3)(a_1a_2+\Omega_1\Omega_2)(a_2a_3+\Omega_2\Omega_3) \\ &-(a_1a_2+\Omega_1\Omega_2)^2(a_3^2+\Omega_3^2)+\beta(a_1a_2+\Omega_1\Omega_2)^2 \\ &-(a_1a_3+\Omega_1\Omega_3)^2(a_2^2+\Omega_2^2)+\beta(a_1a_3+\Omega_1\Omega_3)^2\,. \end{aligned}$$

The coefficient of β^2 in this is just

$$\text{coefficient of } \beta^2 = a^2+\Omega^2\,, \tag{2.396}$$

where $a^2 := a_1^2+a_2^2+a_3^2$ and $\Omega^2 := \Omega_1^2+\Omega_2^2+\Omega_3^2$. The coefficient of β is

$$\text{coefficient of } \beta = -(a_2^2+\Omega_2^2)(a_3^2+\Omega_3^2) - (a_3^2+\Omega_3^2)(a_1^2+\Omega_1^2) - (a_1^2+\Omega_1^2)(a_2^2+\Omega_2^2) + (a_2a_3+\Omega_2\Omega_3)^2 + (a_1a_2+\Omega_1\Omega_2)^2 + (a_1a_3+\Omega_1\Omega_3)^2$$

Note the necessary symmetry in permutations of (123). Multiplying out, we soon find that

$$\text{coefficient of } \beta = -(a_1\Omega_2 - a_2\Omega_1)^2 - (a_2\Omega_3 - a_3\Omega_2)^2 - (a_3\Omega_1 - a_1\Omega_3)^2 \ .$$

This can be written in the form

$$\text{coefficient of } \beta = -(\mathbf{a}\times\Omega)^2 = (\mathbf{a}\cdot\Omega)^2 - a^2\Omega^2 \ . \tag{2.397}$$

The constant term in the cubic is

$$\begin{aligned}\text{constant term} = {} & (a_1^2+\Omega_1^2)(a_2^2+\Omega_2^2)(a_3^2+\Omega_3^2) - (a_1^2+\Omega_1^2)(a_2a_3+\Omega_2\Omega_3)^2 \\ & +2(a_1a_3+\Omega_1\Omega_3)(a_1a_2+\Omega_1\Omega_2)(a_2a_3+\Omega_2\Omega_3) \\ & -(a_2^2+\Omega_2^2)(a_1a_3+\Omega_1\Omega_3)^2 - (a_3^2+\Omega_3^2)(a_1a_2+\Omega_1\Omega_2)^2 \ .\end{aligned}$$

After a long-winded expansion of the terms, we find that the constant term is zero.

The characteristic equation for M thus becomes

$$\boxed{0 = \beta^3 - (a^2+\Omega^2)\beta^2 + \left[a^2\Omega^2 - (\mathbf{a}\cdot\Omega)^2\right]\beta \ , \qquad \beta = \alpha + \Omega^2} \ . \tag{2.398}$$

One solution is $\beta = 0$, leading to the eigenvalue $\alpha_0 := -\Omega^2$. The other two eigenvalues are $\alpha_\pm := \beta_\pm - \Omega^2$, where $\beta_\pm$ satisfy the quadratic equation

$$0 = \beta^2 - (a^2+\Omega^2)\beta + a^2\Omega^2 - (\mathbf{a}\cdot\Omega)^2 \ . \tag{2.399}$$

We thus find

$$\beta_\pm = \frac{1}{2}(a^2+\Omega^2) \pm \frac{1}{2}\sqrt{(a^2+\Omega^2)^2 - 4\left[a^2\Omega^2 - (\mathbf{a}\cdot\Omega)^2\right]} \ ,$$

whence

$$\alpha_\pm = \frac{1}{2}(a^2-\Omega^2) \pm \frac{1}{2}\sqrt{(a^2-\Omega^2)^2 + 4(\mathbf{a}\cdot\Omega)^2} \ .$$

Now recall the Lorentz invariants associated with a tensor of the form of (2.284) on p. 93. Mashhoon defines

$$I = -a^2 + \Omega^2 \ , \qquad I^* = -\mathbf{a}\cdot\Omega \ , \tag{2.400}$$

whence the eigenvalues of M can be written

$$\boxed{\alpha_\pm = -\frac{1}{2}I \pm \sqrt{\left(\frac{1}{2}I\right)^2 + I^*}\,, \qquad \alpha_0 = -\Omega^2}\,, \tag{2.401}$$

and not as Mashhoon gives them, without the remaining factors of 1/2. It would have been neater here to consider the invariants $I/2$ and I^*.

So we have three eigenvalues $\alpha_+ \geq 0$, $\alpha_0 \leq 0$, and $\alpha_- \leq 0$, and the determinant of M is

$$\begin{aligned}\det M = \alpha_-\alpha_0\alpha_+ &= -\Omega^2\left[\frac{1}{4}I^2 - \left(\frac{1}{4}I^2 + I^{*2}\right)\right]\\ &= +\Omega^2 I^{*2} = \Omega^2(\mathbf{a}\cdot\Omega)^2\,,\end{aligned}$$

or for reference

$$\boxed{\det M = \Omega^2(\mathbf{a}\cdot\Omega)^2}\,. \tag{2.402}$$

Now we need to talk about changing the space coordinates (X^1, X^2, X^3) in order to make the matrix M look simpler, and hence simplify the condition $g_{00} = 0$.

2.13.2 Simplifying the Condition $g_{00} = 0$

The matrix M can be diagonalized at any time X^0 (recall that this coordinate is the proper time of the observer distributed equally over each hyperplane of simultaneity). In fact, this can be done by a simple rotation of the space coordinates because M is a real symmetric 3×3 matrix. So let $R(X^0)$ be the orthogonal matrix such that

$$R^{-1} = R^{\mathrm{T}}\,, \qquad R^{-1}MR = \begin{pmatrix}\alpha_+ & 0 & 0\\ 0 & \alpha_0 & 0\\ 0 & 0 & \alpha_-\end{pmatrix}\,. \tag{2.403}$$

Now change to the coordinates

$$\hat{X}^0 := X^0\,, \qquad \hat{\mathbf{X}} := R^{-1}\mathbf{X}\,, \tag{2.404}$$

assuming that R is a smooth function of X^0. Note that

$$\hat{g}_{00} = \frac{\partial X^\mu}{\partial \hat{X}^0}\frac{\partial X^\nu}{\partial \hat{X}^0}g_{\mu\nu} \neq g_{00}\,, \tag{2.405}$$

in general. This is because

$$X^0 = \hat{X}^0\,, \qquad \mathbf{X} = R(\hat{X}^0)\hat{\mathbf{X}}\,. \tag{2.406}$$

This means that

$$\frac{\partial X^\mu}{\partial \hat{X}^0} = \frac{\partial R(\hat{X}^0)^\mu{}_\nu}{\partial \hat{X}^0} \hat{X}^\nu \neq 0 \,. \tag{2.407}$$

Hence the condition $\hat{g}_{00} = 0$ is not the same as $g_{00} = 0$. Does this matter? Not as long as we bear in mind that we are just rewriting g_{00}, and not really considering the metric induced on the hyperplanes of simultaneity by a new coordinate system.

In the end, the aim is to describe the surface geometrically. Suppose for example that we show that it intersects each HOS in an elliptical cylinder as described relative to the space coordinates $(\hat{X}^1, \hat{X}^2, \hat{X}^3)$. Since we have just done a rotation, which preserves this kind of property, it means that the surface intersects each HOS in an elliptical cylinder as described relative to the space coordinates (X^1, X^2, X^3). Later we shall do a time dependent translation of coordinates in each HOS, and the same reasoning will apply.

The metric component g_{00} will now look simpler. Recall that we had (2.391), viz.,

$$0 = 1 + 2a_i(X^0)X^i + M_{ij}(X^0)X^i X^j = 1 + 2\mathbf{a}\cdot\mathbf{X} + \mathbf{X}^{\mathrm{T}} M \mathbf{X} \,. \tag{2.408}$$

But of course

$$\mathbf{X}^{\mathrm{T}} M \mathbf{X} = \hat{\mathbf{X}}^{\mathrm{T}} R^{\mathrm{T}} M R \hat{\mathbf{X}} = \alpha_+ \left(\hat{X}^1\right)^2 + \alpha_0 \left(\hat{X}^2\right)^2 + \alpha_- \left(\hat{X}^3\right)^2 \,. \tag{2.409}$$

Regarding the term linear in $\mathbf{X}$, we define

$$\hat{\mathbf{a}} := R^{-1}\mathbf{a} \,, \tag{2.410}$$

so that

$$\mathbf{a}\cdot\mathbf{X} = \mathbf{a}^{\mathrm{T}}\mathbf{X} = \hat{\mathbf{a}}^{\mathrm{T}} R^{\mathrm{T}} R \hat{\mathbf{X}} = \hat{\mathbf{a}}^{\mathrm{T}} \hat{\mathbf{X}} \,, \tag{2.411}$$

and our condition $g_{00} = 0$ becomes

$$\boxed{1 + 2\hat{\mathbf{a}}\cdot\hat{\mathbf{X}} + \alpha_+ \left(\hat{X}^1\right)^2 + \alpha_0 \left(\hat{X}^2\right)^2 + \alpha_- \left(\hat{X}^3\right)^2 = 0} \,. \tag{2.412}$$

2.13.3 *The Case* $\det M \neq 0$

Later we shall consider the two cases where $\det M = 0$, but for the moment, let us assume that each of α_0 and $\alpha_\pm$ is nonzero, so that $\det M \neq 0$.

We now observe that

$$2\hat{a}_i \hat{X}^i + \alpha \left(\hat{X}^i\right)^2 = \alpha \left[\left(\hat{X}^i\right)^2 + \frac{2\hat{a}_i}{\alpha}\hat{X}^i \right] = \alpha \left[\left(\hat{X}^i + \frac{\hat{a}_i}{\alpha}\right)^2 - \left(\frac{\hat{a}_i}{\alpha}\right)^2 \right] ,$$

so we can write

$$g_{00} = 1 - \frac{\hat{a}_1^2}{\alpha_+} - \frac{\hat{a}_2^2}{\alpha_0} - \frac{\hat{a}_3^2}{\alpha_-} + \alpha_+ \left(\hat{X}^1 + \frac{\hat{a}_1}{\alpha_+}\right)^2 + \alpha_0 \left(\hat{X}^2 + \frac{\hat{a}_2}{\alpha_0}\right)^2 + \alpha_- \left(\hat{X}^3 + \frac{\hat{a}_3}{\alpha_-}\right)^2 . \tag{2.413}$$

In a moment, we shall show that

$$1 = \frac{\hat{a}_1^2}{\alpha_+} + \frac{\hat{a}_2^2}{\alpha_0} + \frac{\hat{a}_3^2}{\alpha_-} , \tag{2.414}$$

whence it follows that

$$g_{00} = \alpha_+ \left(\hat{X}^1 + \frac{\hat{a}_1}{\alpha_+}\right)^2 + \alpha_0 \left(\hat{X}^2 + \frac{\hat{a}_2}{\alpha_0}\right)^2 + \alpha_- \left(\hat{X}^3 + \frac{\hat{a}_3}{\alpha_-}\right)^2 . \tag{2.415}$$

We now do another transformation of the space coordinates:

$$\xi := \hat{X}^1 + \frac{\hat{a}_1}{\alpha_+} , \qquad \eta := \hat{X}^2 + \frac{\hat{a}_2}{\alpha_0} , \qquad \zeta := \hat{X}^3 + \frac{\hat{a}_3}{\alpha_-} . \tag{2.416}$$

As above, we are not talking here about obtaining a new component version $g'_{\mu\nu}$ of the metric and considering the condition $g'_{00} = 0$, which would be a different condition to $g_{00} = 0$. We are merely describing the surface $g_{00} = 0$, or rather its intersection with each HOS, relative to new coordinates on the given HOS. We are interested in geometric shape, but both rotation and translation preserve shape within the HOS.

Anyway, the upshot of these space transformations within a given HOS is a surface given by

$$0 = \alpha_+ \xi^2 + \alpha_0 \eta^2 + \alpha_- \zeta^2 , \tag{2.417}$$

at least once we have shown (2.414) (see below). Recalling that $\alpha_+ \geq 0$, $\alpha_0 \leq 0$, and $\alpha_- \leq 0$, we have

$$\boxed{0 = \alpha_+ \xi^2 - |\alpha_0| \eta^2 - |\alpha_-| \zeta^2} . \tag{2.418}$$

Let us therefore prove (2.414). First we prove

$$\left(M^{-1}\right)_{ij} a^i a^j = 1 . \tag{2.419}$$

Note that it would be equivalent to show that

$$\big(\text{matrix of cofactors of } M\big)_{ij} a^i a^j = \det M = \Omega^2 (\mathbf{a} \cdot \boldsymbol{\Omega})^2 . \tag{2.420}$$

The matrix of cofactors of

$$M = \begin{pmatrix} a_1^2 + \Omega_1^2 - \Omega^2 & a_1 a_2 + \Omega_1 \Omega_2 & a_1 a_3 + \Omega_1 \Omega_3 \\ a_1 a_2 + \Omega_1 \Omega_2 & a_2^2 + \Omega_2^2 - \Omega^2 & a_2 a_3 + \Omega_2 \Omega_3 \\ a_1 a_3 + \Omega_1 \Omega_3 & a_2 a_3 + \Omega_2 \Omega_3 & a_3^2 + \Omega_3^2 - \Omega^2 \end{pmatrix} \tag{2.421}$$

is clearly a symmetric matrix of the form

$$\text{matrix of cofactors of } M = \begin{pmatrix} A & B & C \\ B & D & E \\ C & E & F \end{pmatrix} , \tag{2.422}$$

where

$$\begin{aligned} A &= (a_2^2 + \Omega_2^2 - \Omega^2)(a_3^2 + \Omega_3^2 - \Omega^2) - (a_2 a_3 + \Omega_2 \Omega_3)^2 \\ &= a_2^2 \Omega_3^2 - a_2^2 \Omega^2 + a_3^2 \Omega_2^2 - \Omega_2^2 \Omega^2 - a_3^2 \Omega^2 - \Omega_3^2 \Omega^2 + (\Omega^2)^2 - 2 a_2 a_3 \Omega_2 \Omega_3 , \end{aligned} \tag{2.423}$$

$$\begin{aligned} D &= (a_3^2 + \Omega_3^2 - \Omega^2)(a_1^2 + \Omega_1^2 - \Omega^2) - (a_3 a_1 + \Omega_3 \Omega_1)^2 \\ &= a_3^2 \Omega_1^2 - a_3^2 \Omega^2 + a_1^2 \Omega_3^2 - \Omega_3^2 \Omega^2 - a_1^2 \Omega^2 - \Omega_1^2 \Omega^2 + (\Omega^2)^2 - 2 a_3 a_1 \Omega_3 \Omega_1 , \end{aligned} \tag{2.424}$$

$$\begin{aligned} F &= (a_1^2 + \Omega_1^2 - \Omega^2)(a_2^2 + \Omega_2^2 - \Omega^2) - (a_1 a_2 + \Omega_1 \Omega_2)^2 \\ &= a_1^2 \Omega_2^2 - a_1^2 \Omega^2 + a_2^2 \Omega_1^2 - \Omega_1^2 \Omega^2 - a_2^2 \Omega^2 - \Omega_2^2 \Omega^2 + (\Omega^2)^2 - 2 a_1 a_2 \Omega_1 \Omega_2 , \end{aligned} \tag{2.425}$$

$$\begin{aligned} B &= (a_1 a_3 + \Omega_1 \Omega_3)(a_2 a_3 + \Omega_2 \Omega_3) - (a_1 a_2 + \Omega_1 \Omega_2)(a_3^2 + \Omega_3^2 - \Omega^2) \\ &= a_1 a_3 \Omega_2 \Omega_3 + a_2 a_3 \Omega_1 \Omega_3 - a_1 a_2 \Omega_3^2 + a_1 a_2 \Omega^2 - \Omega_1 \Omega_2 a_3^2 + \Omega_1 \Omega_2 \Omega^2 , \end{aligned} \tag{2.426}$$

$$\begin{aligned} C &= (a_1 a_2 + \Omega_1 \Omega_2)(a_2 a_3 + \Omega_2 \Omega_3) - (a_1 a_3 + \Omega_1 \Omega_3)(a_2^2 + \Omega_2^2 - \Omega^2) \\ &= a_3 a_2 \Omega_1 \Omega_2 + a_1 a_2 \Omega_3 \Omega_2 - a_3 a_1 \Omega_2^2 + a_3 a_1 \Omega^2 - \Omega_3 \Omega_1 a_2^2 + \Omega_3 \Omega_1 \Omega^2 , \end{aligned} \tag{2.427}$$

$$\begin{aligned} E &= (a_1 a_3 + \Omega_1 \Omega_3)(a_1 a_2 + \Omega_1 \Omega_2) - (a_2 a_3 + \Omega_2 \Omega_3)(a_1^2 + \Omega_1^2 - \Omega^2) \\ &= a_2 a_1 \Omega_3 \Omega_1 + a_3 a_1 \Omega_2 \Omega_1 - a_2 a_3 \Omega_1^2 + a_2 a_3 \Omega^2 - \Omega_2 \Omega_3 a_1^2 + \Omega_2 \Omega_3 \Omega^2 . \end{aligned} \tag{2.428}$$

Now observe that

$$
\begin{aligned}
\big(\text{matrix of cofactors of } M\big)_{ij} a^i a^j &= (a_1\ a_2\ a_3) \begin{pmatrix} A & B & C \\ B & D & E \\ C & E & F \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \\
&= (a_1\ a_2\ a_3) \begin{pmatrix} Aa_1 + Ba_2 + Ca_3 \\ Ba_1 + Da_2 + Ea_3 \\ Ca_1 + Ea_2 + Fa_3 \end{pmatrix} \\
&= Aa_1^2 + Da_2^2 + Fa_3^2 + 2Ba_1a_2 + 2Ea_2a_3 + 2Ca_1a_3 \ .
\end{aligned}
$$

Inserting the values for A, B, C, D, E, and F, we find

$$
\begin{aligned}
(M_{\text{cof}})_{ij} a^i a^j = \; & a_1^2 a_2^2 \Omega_3^2 - a_1^2 a_2^2 \Omega^2 + a_1^2 a_3^2 \Omega_2^2 - a_1^2 \Omega_2^2 \Omega^2 \\
& - a_1^2 a_3^2 \Omega^2 - a_1^2 \Omega_3^2 \Omega^2 + a_1^2 (\Omega^2)^2 - 2a_1^2 a_2 a_3 \Omega_2 \Omega_3 \\
& + a_2^2 a_3^2 \Omega_1^2 - a_2^2 a_3^2 \Omega^2 + a_2^2 a_1^2 \Omega_3^2 - a_2^2 \Omega_3^2 \Omega^2 \\
& - a_2^2 a_1^2 \Omega^2 - a_2^2 \Omega_1^2 \Omega^2 + a_2^2 (\Omega^2)^2 - 2a_2^2 a_3 a_1 \Omega_3 \Omega_1 \\
& + a_3^2 a_1^2 \Omega_2^2 - a_3^2 a_1^2 \Omega^2 + a_3^2 a_2^2 \Omega_1^2 - a_3^2 \Omega_1^2 \Omega^2 \\
& - a_3^2 a_2^2 \Omega^2 - a_3^2 \Omega_2^2 \Omega^2 + a_3^2 (\Omega^2)^2 - 2a_3^2 a_1 a_2 \Omega_1 \Omega_2 \\
& + 2a_1^2 a_2 a_3 \Omega_2 \Omega_3 + 2a_1 a_2^2 a_3 \Omega_1 \Omega_3 - 2a_1^2 a_2^2 \Omega_3^2 \\
& + 2a_1^2 a_2^2 \Omega^2 - 2a_1 a_2 \Omega_1 \Omega_2 a_3^2 + 2a_1 a_2 \Omega_1 \Omega_2 \Omega^2 \\
& + 2a_2^2 a^3 a_1 \Omega_3 \Omega_1 + 2a_2 a_3^2 a_1 \Omega_2 \Omega_1 - 2a_2^2 a_3^2 \Omega_1^2 \\
& + 2a_2^2 a_3^2 \Omega^2 - 2a_2 a_3 a_1^2 \Omega_2 \Omega_3 + 2a_2 a_3 \Omega_2 \Omega_3 \Omega^2 \\
& + 2a_1 a_3^2 a_2 \Omega_1 \Omega_2 + 2a_1^2 a_3 a_2 \Omega_3 \Omega_2 - 2a_3^2 a_1^2 \Omega_2^2 \\
& + 2a_3^2 a_1^2 \Omega^2 - 2a_1 a_3 a_2^2 \Omega_3 \Omega_1 + 2a_1 a_3 \Omega_3 \Omega_1 \Omega^2 \ .
\end{aligned}
$$

Now note that all terms in the following cancel:

$$
a_1^2 a_2^2 \Omega_3^2 \ , \quad a_1^2 a_2^2 \Omega^2 \ , \quad a_1^2 a_3^2 \Omega_2^2 \ , \quad a_1^2 a_3^2 \Omega^2 \ , \quad a_2^2 a_3^2 \Omega_1^2 \ , \quad a_2^2 a_3^2 \Omega^2 \ , \quad a_1^2 a_2^2 \Omega^2 \ ,
$$

$$
2a_1^2 a_2 a_3 \Omega_2 \Omega_3 \ , \quad 2a_1 a_2^2 a_3 \Omega_1 \Omega_3 \ , \quad 2a_1 a_2 a_3^2 \Omega_1 \Omega_2 \ .
$$

This leaves us with

$$
\begin{aligned}
(M_{\text{cof}})_{ij} a^i a^j = \Omega^2 \Big[& (a_1^2 + a_2^2 + a_3^2) \Omega^2 - a_1^2 \Omega_2^2 - a_1^2 \Omega_3^2 - a_2^2 \Omega_3^2 - a_2^2 \Omega_1^2 \\
& - a_3^2 \Omega_1^2 - a_3^2 \Omega_2^2 + 2a_1 a_2 \Omega_1 \Omega_2 + 2a_2 a_3 \Omega_2 \Omega_3 + 2a_1 a_3 \Omega_1 \Omega_3 \Big] \ .
\end{aligned}
$$

This soon boils down to

$$
(M_{\text{cof}})_{ij} a^i a^j = \Omega^2 (\mathbf{a} \cdot \Omega)^2 \ ,
$$

as claimed in (2.420).

We now use this to prove (2.414), viz.,

$$1 = \frac{\hat{a}_1^2}{\alpha_+} + \frac{\hat{a}_2^2}{\alpha_0} + \frac{\hat{a}_3^2}{\alpha_-} \ . \tag{2.429}$$

We have shown that

$$a^{\mathrm{T}} M^{-1} a = 1 \ ,$$

and we defined $\hat{a} := R^{\mathrm{T}} a$, where R is the orthogonal matrix such that

$$R^{\mathrm{T}} M R = \mathrm{diag}\,(\alpha_+, \alpha_0, \alpha_-) \ .$$

Taking the inverse of each side of the latter,

$$R^{\mathrm{T}} M^{-1} R = \mathrm{diag}\,(1/\alpha_+, 1/\alpha_0, 1/\alpha_-) \ .$$

Hence,

$$\begin{aligned} \hat{a}^{\mathrm{T}} R^{\mathrm{T}} M^{-1} R \hat{a} &= (\hat{a}_1\ \hat{a}_2\ \hat{a}_3) \begin{pmatrix} 1/\alpha_+ & 0 & 0 \\ 0 & 1/\alpha_0 & 0 \\ 0 & 0 & 1/\alpha_- \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} \\ &= \frac{\hat{a}_1^2}{\alpha_+} + \frac{\hat{a}_2^2}{\alpha_0} + \frac{\hat{a}_3^2}{\alpha_-} \ . \end{aligned}$$

But the left-hand side of the latter is

$$\hat{a}^{\mathrm{T}} R^{\mathrm{T}} M^{-1} R \hat{a} = (R\hat{a})^{\mathrm{T}} M^{-1} (R\hat{a}) = a^{\mathrm{T}} M^{-1} a = 1 \ ,$$

as required.

This ends discussion of the case where none of α_0 and $\alpha_\pm$ are zero, i.e., the case where M is non-singular. We obtain the quadric surface (in fact, an elliptic cone) given by (2.418), viz.,

$$\boxed{0 = \alpha_+ \xi^2 - |\alpha_0| \eta^2 - |\alpha_-| \zeta^2} \ . \tag{2.430}$$

Remember exactly what this means: in each hyperplane of simultaneity (spacelike surface of constant X^0) for this coordinate system (X^0, X^1, X^2, X^3), the hypersurface $g_{00} = 0$ intersects that 3D space in a quadric cone, but these cones generally change smoothly as we move from one HOS to the next.

2.13.4 The Case $\det M = 0$

Since $\det M = \Omega^2 (\mathbf{a} \cdot \Omega)^2$, the matrix M is singular iff $\Omega = 0$ or $\mathbf{a}$ and Ω are orthogonal.

The Case $\Omega = 0$

Consider first $\Omega = 0$, whence

$$M = \begin{pmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} (a_1\ a_2\ a_3)\ . \qquad (2.431)$$

It is interesting to note that we can always arrange for this by FW transporting the chosen tetrad, no matter what rotation there may be in the observer worldline.

In this case, $\alpha_0 = 0$, $\alpha_+ = 0$, and $\alpha_- = -I = a^2$, using (2.401) and the fact that $I^* = 0$. The diagonalised form of M is thus

$$R^{-1}MR = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & a^2 \end{pmatrix}\ .$$

Of course, the diagonalised form of M has to have the same trace as M, which is a^2. Defining $\hat{a} := R^{-1}a$, we have

$$\hat{a} = \begin{pmatrix} 0 \\ 0 \\ \hat{a}_3 \end{pmatrix}\ ,$$

where $\hat{a}_3 = |\mathbf{a}|$, which concords with the second relation in (2.431). What does the hypersurface $g_{00} = 0$ look like relative to the space coordinates $\hat{\mathbf{X}} = R^{-1}\mathbf{X}$ on some prechosen HOS of fixed X^0 (we are considering here the intersection of this hypersurface with the given HOS)? We have

$$1 + 2|\mathbf{a}|\hat{X}^3 + |\mathbf{a}|^2\big(\hat{X}^3\big)^2 = 0\ ,$$

equivalent to

$$\big(1 + |\mathbf{a}|\hat{X}^3\big)^2 = 0\ ,$$

which is in turn equivalent to

$$\hat{X}^3 = -1/|\mathbf{a}|\ , \qquad (2.432)$$

a plane in the 3D space.

Mashhoon says that the surface $g_{00} = 0$ degenerates into coincident planes. As we have seen in Sect. 2.9, this is indeed the case for eternal uniform linear acceleration of the kind much discussed in the context of the radiation problem (Chap. 11) or the Unruh effect (Chap. 14). This corresponds to the case where $\mathbf{a}$ is constant, i.e., independent of the proper time of the observer. It then picks out a space direction in the original inertial frame, and the $\hat{X}^3$ coordinate axis lies in this direction.

In Sect. 2.9, it is shown that all the hyperplanes of simultaneity of the observer intersect in the spacetime region $t = 0$, $x = -c^2/g$, using the notation there (see Fig. 2.5). This is actually a space plane (2D region) in the HOS $t = 0$ of the original inertial frame, since we have complete freedom in the values of y and z. As a matter of fact, it is precisely the same as the region $g_{00} = 0$ in this case. This is where the SE coordinate system even breaks down mathematically because there are many different SE time values that could be assigned to any given event in this region!

But what happens when $\Omega = 0$, i.e., we have an FW transported tetrad, as we are discussing here, but $\mathbf{a}$ is a changing function of the proper time of the observer? Are the planes in (2.432) coincident, as suggested by Mashhoon? It certainly does not look like it, since the specification of the plane in this relation now becomes

$$\hat{X}^3 = -1/|\mathbf{a}(\hat{X}^0)|\,, \tag{2.433}$$

where $\hat{X}^0$ is just X^0, the proper time of the observer distributed over the neighbourhood of her worldline by borrowing hyperplanes of simultaneity from instantaneously comoving inertial observers.

For one thing, the 2D spacelike region specified in (2.433) is contained in a HOS of preselected X^0 in the SE coordinate system (X^0, X^1, X^2, X^3), and although these HOSs inevitably intersect somewhere for nonzero acceleration of the observer, we do not know where they intersect. They will not generally intersect precisely where g_{00} happens to be zero. That is a special result for eternal uniform linear acceleration.

The Case $\Omega \neq 0$

Consider now $\Omega \neq 0$, but $\mathbf{a}$ and Ω orthogonal. This is the case that concerns us when we consider an observer revolving around a circle of constant radius with arbitrary (possibly changing) angular speed. This can be seen directly from (2.283) on p. 92. We still have $I^* = 0$, but now $\alpha_0 = -\Omega^2 \neq 0$, by hypothesis, and $\mu_+ = 0$, $\mu_- = -I = a^2 - \Omega^2$, giving $\mathrm{Tr}\,M = a^2 - 2\Omega^2$, as one would expect from the general form (2.393) for M on p. 124.

There exist coordinates $\hat{X}^1$, $\hat{X}^2$, $\hat{X}^3$, found by rotating the original spatial coordinate axes, such that

$$\mathbf{a} = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0 \\ \Omega \\ 0 \end{pmatrix},$$

and hence

$$M = \begin{pmatrix} a^2 - \Omega^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Omega^2 \end{pmatrix}. \tag{2.434}$$

The condition $g_{00} = 0$ becomes

$$\boxed{0 = 1 + 2a\hat{X}^1 + (a^2 - \Omega^2)\big(\hat{X}^1\big)^2 - \Omega^2\big(\hat{X}^3\big)^2}\,. \tag{2.435}$$

If $a^2 = \Omega^2$, this condition becomes

$$0 = 1 + 2a\hat{X}^1 - \Omega^2 \left(\hat{X}^3\right)^2 , \tag{2.436}$$

which rearranges to

$$\hat{X}^1 = \frac{\Omega^2}{2a}\left(\hat{X}^3\right)^2 - \frac{1}{2a} . \tag{2.437}$$

This is a typical parabola. Since it was obtained by a space rotation from the coordinates X^1, X^2, X^3, the surface $g_{00} = 0$ in spacetime intersects this particular (and hence every) HOS of fixed X^0 in a parabolic cylinder. Note, however, that this shape is expressed relative to the SE coordinates X^1, X^2, X^3, and these *are* the space coordinates of a global inertial frame, viz., one that is instantaneously comoving with the observer at her proper time X^0.

Now consider the situation where $a^2 - \Omega^2 \neq 0$. We have

$$\begin{aligned}
&(a^2 - \Omega^2)\left(\hat{X}^1\right)^2 + 2a\hat{X}^1 + 1 - \Omega^2\left(\hat{X}^3\right)^2 \\
&\qquad = (a^2 - \Omega^2)\left[\left(\hat{X}^1\right)^2 + \frac{2a}{a^2 - \Omega^2}\hat{X}^1\right] + 1 - \Omega^2\left(\hat{X}^3\right)^2 \\
&\qquad = (a^2 - \Omega^2)\left[\left(\hat{X}^1 + \frac{a}{a^2 - \Omega^2}\right)^2 - \frac{a^2}{(a^2 - \Omega^2)^2}\right] + 1 - \Omega^2\left(\hat{X}^3\right)^2 \\
&\qquad = (a^2 - \Omega^2)\left(\hat{X}^1 + \frac{a}{a^2 - \Omega^2}\right)^2 + 1 - \frac{a^2}{a^2 - \Omega^2} - \Omega^2\left(\hat{X}^3\right)^2 ,
\end{aligned}$$

and finally, the condition $g_{00} = 0$ becomes

$$(a^2 - \Omega^2)\left(\hat{X}^1 + \frac{a}{a^2 - \Omega^2}\right)^2 - \Omega^2\left(\hat{X}^3\right)^2 = \frac{\Omega^2}{a^2 - \Omega^2} . \tag{2.438}$$

If $a^2 - \Omega^2 > 0$, then the terms in

$$\left(\hat{X}^1 + \frac{a}{a^2 - \Omega^2}\right)^2 \quad \text{and} \quad \left(\hat{X}^3\right)^2$$

have opposite signs, so we have a hyperbolic cylinder. If $a^2 - \Omega^2 < 0$, then these terms have the same sign and we have an elliptical cylinder.

The Case $\mathbf{a = 0}$

The condition $g_{00} = 0$ expressed by (2.435) relative to the special choice of coordinates $\hat{X}^1$, $\hat{X}^2$, $\hat{X}^3$, becomes

$$0 = 1 + 2a\hat{X}^1 - \Omega^2\left(\hat{X}^1\right)^2 - \Omega^2\left(\hat{X}^3\right)^2 , \tag{2.439}$$

which clearly describes a circular cylinder.

Interestingly, this is not the case when an observer revolving around a circle of constant radius with arbitrary (possibly changing) angular speed uses rotating tetrad coordinates, as can be seen from (2.283) on p. 92, because **a** is clearly not zero there. This in turn is partly due to the possibility of acceleration $\dot{v}$ tangential to the circle, but perhaps also partly due to the choice of tetrad. In the case of constant Ω_0, is there a choice of tetrad such that $\mathbf{a} = 0$?

When Ω_0 and hence $v = \Omega_0 r$ are constant, **a** has a nonzero component in the radial direction λ_1 [see (2.277) and then (2.283) on p. 92]. So what situation is described by $\Omega \neq 0$ and $\mathbf{a} = 0$. We can answer this question by looking at (2.102) and (2.103) on p. 45, viz.,

$$\begin{cases} c\dot{\lambda}_{(i)} = a_{0i}\lambda_{(0)} + c\Omega_{ij}\lambda_{(j)} \;, \\ c\dot{\lambda}_{(0)} = a_{0i}\lambda_{(i)} \;, \end{cases} \tag{2.440}$$

and

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & a_{01} & a_{02} & a_{03} \\ a_{01} & 0 & c\Omega_{21} & c\Omega_{31} \\ a_{02} & c\Omega_{12} & 0 & c\Omega_{32} \\ a_{03} & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \;, \tag{2.441}$$

with ν specifying the row and κ the column, which become

$$\begin{cases} c\dot{\lambda}_{(i)} = c\Omega_{ij}\lambda_{(j)} \;, \\ c\dot{\lambda}_{(0)} = 0 \;, \end{cases} \tag{2.442}$$

and

$$\tilde{A}^{(\nu)}{}_{(\kappa)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & c\Omega_{21} & c\Omega_{31} \\ 0 & c\Omega_{12} & 0 & c\Omega_{32} \\ 0 & c\Omega_{13} & c\Omega_{23} & 0 \end{pmatrix} \;. \tag{2.443}$$

Clearly, the observer is stationary in some inertial frame, since $\mathrm{d}v/\mathrm{d}\tau = \dot{\lambda}_{(0)}$ is zero. But since $\dot{\lambda}_{(i)} = \Omega_{ij}\lambda_{(j)}$, she has chosen a rotating space triad.

2.13.5 The Case of Rotating Tetrad Coordinates

Let us now consider the specific case of an observer revolving around a circle of constant radius with arbitrary (possibly changing) angular speed and using rotating tetrad coordinates. The condition $g_{00} = 0$ was found in (2.357) on p. 112 to be

$$\boxed{X^2\Omega_0^2 + \gamma^2(\Omega_0^2 - \dot{v}^2)Y^2 + 2XY\gamma^2 v\dot{v}\Omega_0 = 1 - v^2 + 2Y\dot{v} - 2Xv\Omega_0} \;. \tag{2.444}$$

We now know how to do a proper analysis of this. First we write it in the form

$$(X\,Y)\begin{pmatrix}\Omega_0^2 & \gamma^2 v\dot{v}\Omega_0\\ \gamma^2 v\dot{v}\Omega_0 & \gamma^2(\Omega_0^2-\dot{v}^2)\end{pmatrix}\begin{pmatrix}X\\ Y\end{pmatrix}=1-v^2+2Y\dot{v}-2Xv\Omega_0\ . \tag{2.445}$$

We then diagonalise the 2×2 matrix

$$M:=\begin{pmatrix}\Omega_0^2 & \gamma^2 v\dot{v}\Omega_0\\ \gamma^2 v\dot{v}\Omega_0 & \gamma^2(\Omega_0^2-\dot{v}^2)\end{pmatrix}\ . \tag{2.446}$$

We begin by writing the characteristic equation

$$(\Omega_0^2-\lambda)\big[\gamma^2(\Omega_0^2-\dot{v}^2)-\lambda\big]=\gamma^4v^2\dot{v}^2\Omega_0^2\ . \tag{2.447}$$

This is

$$\lambda^2-\lambda\big[\gamma^2(\Omega_0^2-\dot{v}^2)+\Omega_0^2\big]+\gamma^2\Omega_0^2(\Omega_0^2-\dot{v}^2)=\gamma^4v^2\dot{v}^2\Omega_0^2\ ,$$

and completing the square,

$$\left\{\lambda-\frac{1}{2}\big[(\gamma^2+1)\Omega_0^2-\gamma^2\dot{v}^2\big]\right\}^2-\frac{1}{4}\big[(\gamma^2+1)\Omega_0^2-\gamma^2\dot{v}^2\big]^2+\gamma^2\Omega_0^2(\Omega_0^2-\dot{v}^2)$$
$$=\gamma^4v^2\dot{v}^2\Omega_0^2\ ,$$

whence we must solve

$$\left\{\lambda-\frac{1}{2}\big[(\gamma^2+1)\Omega_0^2-\gamma^2\dot{v}^2\big]\right\}^2$$
$$=\frac{1}{4}\big[(\gamma^2+1)\Omega_0^2-\gamma^2\dot{v}^2\big]^2-\gamma^2\Omega_0^2(\Omega_0^2-\dot{v}^2)+\gamma^4v^2\dot{v}^2\Omega_0^2\ .$$

The constant on the right-hand side is

$$\begin{aligned}
&\frac{1}{4}\Big[(\gamma^2+1)^2\Omega_0^4+\gamma^4\dot{v}^4-2(\gamma^2+1)\gamma^2\Omega_0^2\dot{v}^2\Big]-\gamma^2\Omega_0^2(\Omega_0^2-\dot{v}^2)+\gamma^4v^2\dot{v}^2\Omega_0^2\\
&\qquad=\dot{v}^2\Omega_0^2\left[-\frac{1}{2}(\gamma^2+1)\gamma^2+\gamma^2+\gamma^4v^2\right]+\left[\frac{1}{4}(\gamma^2+1)^2-\gamma^2\right]\Omega_0^4+\frac{1}{4}\gamma^4\dot{v}^4\\
&\qquad=\frac{1}{2}\gamma^4\Omega_0^2v^2\dot{v}^2+\frac{1}{4}(\gamma^2-1)^2\Omega_0^4+\frac{1}{4}\gamma^4\dot{v}^4\\
&\qquad=\frac{1}{4}\gamma^4\big(\dot{v}^4+2\Omega_0^2v^2\dot{v}^2+v^4\Omega_0^4\big)=\frac{1}{4}\gamma^4(\dot{v}^2+\Omega_0v^2)^2\ ,
\end{aligned}$$

using relations like

$$\gamma^2-1=v^2\gamma^2\ .$$

The characteristic equation is thus

$$\boxed{\left\{\lambda - \frac{1}{2}\left[(\gamma^2+1)\Omega_0^2 - \gamma^2\dot{v}^2\right]\right\}^2 = \frac{1}{4}\gamma^4(\dot{v}^2 + \Omega_0 v^2)^2}\,, \tag{2.448}$$

which has two solutions

$$\lambda_\pm = \frac{1}{2}\left[(\gamma^2+1)\Omega_0^2 - \gamma^2\dot{v}^2\right] \pm \frac{1}{2}\gamma^2(\dot{v}^2 + \Omega_0 v^2)\,. \tag{2.449}$$

These can be simplified. We soon find

$$\boxed{\lambda_+ = \Omega_0^2\gamma^2\,, \qquad \lambda_- = \Omega_0^2 - \gamma^2\dot{v}^2}\,. \tag{2.450}$$

As a quick test, note that

$$\lambda_+ + \lambda_- = \Omega_0^2 + \gamma^2(\Omega_0^2 - \dot{v}^2)\,,$$

which is indeed the trace of M in (2.446).

We must now find the eigenvectors of M, since the 2×2 matrix whose columns are these eigenvectors is the rotation matrix that diagonalises M, and we shall need that to get (2.444) into a suitable form. The eigenvectors are

$$\begin{pmatrix} a_\pm \\ b_\pm \end{pmatrix} \qquad \text{such that} \quad \begin{pmatrix} \Omega_0^2 & \gamma^2 v\dot{v}\Omega_0 \\ \gamma^2 v\dot{v}\Omega_0 & \gamma^2(\Omega_0^2 - \dot{v}^2) \end{pmatrix}\begin{pmatrix} a_\pm \\ b_\pm \end{pmatrix} = \lambda_\pm \begin{pmatrix} a_\pm \\ b_\pm \end{pmatrix}.$$

The components of the eigenvector for λ_+ thus satisfy

$$\begin{cases} \Omega_0^2(1-\gamma^2)a_+ + \gamma^2 v\dot{v}\Omega_0 b_+ = 0\,, \\ \gamma^2 v\dot{v}\Omega_0 a_+ - \gamma^2\dot{v}^2 b_+ = 0\,, \end{cases}$$

whence the unit eigenvector for λ_+ is

$$\mathbf{e}_+ = \frac{1}{(\dot{v}^2 + v^2\Omega_0^2)^{1/2}}\begin{pmatrix} \dot{v} \\ v\Omega_0 \end{pmatrix}. \tag{2.451}$$

The components of the eigenvector for λ_- satisfy

$$\begin{cases} \gamma^2\dot{v}^2 a_- + \gamma^2 v\dot{v}\Omega_0 b_- = 0\,, \\ \gamma^2 v\dot{v}\Omega_0 a_- + (\gamma^2-1)\Omega_0^2 b_- = 0\,, \end{cases}$$

whence the unit eigenvector for λ_- is

$$\mathbf{e}_- = \frac{1}{(\dot{v}^2 + v^2\Omega_0^2)^{1/2}}\begin{pmatrix} v\Omega_0 \\ -\dot{v} \end{pmatrix}. \tag{2.452}$$

Now M in (2.446) is a real symmetric matrix, and can be diagonalised by an orthogonal matrix R. The columns of R, which is a rotation matrix, are the unit eigenvectors $\mathbf{e}_+$ and $\mathbf{e}_-$, which are of course orthogonal, since they correspond to different eigenvalues. Let us check that this works. We have

$$\boxed{R := \frac{1}{(\dot{v}^2 + v^2\Omega_0^2)^{1/2}} \begin{pmatrix} \dot{v} & v\Omega_0 \\ v\Omega_0 & -\dot{v} \end{pmatrix}} , \tag{2.453}$$

and we can check that

$$\begin{aligned} MR &= \frac{1}{(\dot{v}^2 + v^2\Omega_0^2)^{1/2}} \begin{pmatrix} \Omega_0^2 & \gamma^2 v\dot{v}\Omega_0 \\ \gamma^2 v\dot{v}\Omega_0 & \gamma^2(\Omega_0^2 - \dot{v}^2) \end{pmatrix} \begin{pmatrix} \dot{v} & v\Omega_0 \\ v\Omega_0 & -\dot{v} \end{pmatrix} \\ &= \frac{1}{(\dot{v}^2 + v^2\Omega_0^2)^{1/2}} \begin{pmatrix} \lambda_+\dot{v} & \lambda_- v\Omega_0 \\ \lambda_+ v\Omega_0 & -\lambda_-\dot{v} \end{pmatrix} , \end{aligned}$$

and

$$\begin{aligned} R^{\mathrm{T}}MR &= \frac{1}{\dot{v}^2 + v^2\Omega_0^2} \begin{pmatrix} \dot{v} & v\Omega_0 \\ v\Omega_0 & -\dot{v} \end{pmatrix} \begin{pmatrix} \lambda_+\dot{v} & \lambda_- v\Omega_0 \\ \lambda_+ v\Omega_0 & -\lambda_-\dot{v} \end{pmatrix} \\ &= \frac{1}{\dot{v}^2 + v^2\Omega_0^2} \begin{pmatrix} \lambda_+(\dot{v}^2 + v^2\Omega_0^2) & 0 \\ 0 & \lambda_-(\dot{v}^2 + v^2\Omega_0^2) \end{pmatrix} = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} , \end{aligned}$$

in the usual way.

Now that we can diagonalise M, let us return to the condition $g_{00} = 0$ as expressed by (2.445) on p. 136, viz.,

$$(X\ Y) \begin{pmatrix} \Omega_0^2 & \gamma^2 v\dot{v}\Omega_0 \\ \gamma^2 v\dot{v}\Omega_0 & \gamma^2(\Omega_0^2 - \dot{v}^2) \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 1 - v^2 + 2Y\dot{v} - 2Xv\Omega_0 . \tag{2.454}$$

Put another way

$$(X\ Y)M \begin{pmatrix} X \\ Y \end{pmatrix} = 1 - v^2 + 2Y\dot{v} - 2Xv\Omega_0 . \tag{2.455}$$

This in turn means that

$$(X\ Y)RR^{\mathrm{T}}MRR^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix} = 1 - v^2 + 2Y\dot{v} - 2Xv\Omega_0 ,$$

and the left-hand side is

$$(X\ Y)RR^{\mathrm{T}}MRR^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix} = \left[R^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix}\right]^{\mathrm{T}} \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \left[R^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix}\right] .$$

So if we define new coordinates W and U by

$$\begin{pmatrix} W \\ U \end{pmatrix} := R^{\mathrm{T}} \begin{pmatrix} X \\ Y \end{pmatrix} , \qquad \begin{pmatrix} X \\ Y \end{pmatrix} := R \begin{pmatrix} W \\ U \end{pmatrix} , \tag{2.456}$$

the left-hand side of (2.455) becomes

$$(X\ Y)M \begin{pmatrix} X \\ Y \end{pmatrix} = (W\ U) \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \begin{pmatrix} W \\ U \end{pmatrix} = \lambda_+ W^2 + \lambda_- U^2 . \tag{2.457}$$

The second relation of (2.456) gives

$$\begin{pmatrix} X \\ Y \end{pmatrix} := R \begin{pmatrix} W \\ U \end{pmatrix} = \frac{1}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} \begin{pmatrix} \dot{v} & v\Omega_0 \\ v\Omega_0 & -\dot{v} \end{pmatrix} \begin{pmatrix} W \\ U \end{pmatrix} , \tag{2.458}$$

whence

$$\begin{cases} X = \dfrac{1}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} (\dot{v}W + v\Omega_0 U) , \\ Y = \dfrac{1}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} (v\Omega_0 W - \dot{v}U) . \end{cases} \tag{2.459}$$

We can now write the condition $g_{00} = 0$, or at least its intersection with the chosen hypersurface of simultaneity for these coordinates, in the form

$$\begin{aligned} \Omega_0^2 \gamma^2 W^2 + (\Omega_0^2 - \gamma^2 \dot{v}^2) U^2 &= 1 - v^2 + \frac{2\dot{v}(v\Omega_0 W - \dot{v}U)}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} - \frac{2v\Omega_0(\dot{v}W + v\Omega_0 U)}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} \\ &= 1 - v^2 - \frac{2\dot{v}^2 + 2v^2\Omega_0^2}{(\dot{v}^2 + v^2 \Omega_0^2)^{1/2}} U \\ &= 1 - v^2 - 2(\dot{v}^2 + v^2 \Omega_0^2)^{1/2} U , \end{aligned} \tag{2.460}$$

recalling that the other space dimension labelled by Z is left unconstrained by the condition and it is this that gives the surface its cylindrical aspect. So the final expression for this surface is

$$\Omega_0^2 \gamma^2 W^2 + (\Omega_0^2 - \gamma^2 \dot{v}^2) U^2 = 1 - v^2 - 2(\dot{v}^2 + v^2 \Omega_0^2)^{1/2} U . \tag{2.461}$$

As we expected, this is an ellipse when $\Omega_0^2 > \gamma^2 \dot{v}^2$, a hyperbola when $\Omega_0^2 < \gamma^2 \dot{v}^2$, and a parabola when $\Omega_0^2 = \gamma^2 \dot{v}^2$. So it is the sign of the invariant I in (2.287) on p. 93 that decides, as claimed in the discussion around (2.358) on p. 112.

Chapter 3
Adapted Frames in General Relativity

At any point p in the spacetime manifold $\mathscr{M}$ of general relativity, under the usual assumption that the covariant derivative of the metric is zero and the connection is torsion free, i.e., symmetric in its two lower indices when expressed relative to a coordinate frame (see Sect. 4.2), coordinates can be found on some neighbourhood of p such that the metric has the Minkowski form at p and the connection and first derivatives of the metric are zero at p.

This result is often called the *weak equivalence principle*, although it is actually built into the standard construction of general relativity in the torsion free case. To be of any practical use, it has to have a counterpart in observation, so the principle here is really the statement that such locally inertial frames correspond to freely falling, non-rotating laboratory frames. The reader is referred to Chap. 6.5 for a more critical appraisal of this idea. For the moment, we concentrate on the mathematical statement in the last paragraph.

There are two standard approaches for showing this. The first is to begin with arbitrary coordinates and then specify a new set in terms of them for which the conditions hold. The second is to determine such coordinates directly by means of a canonical map from the tangent space at p, and this is the approach which will be described here. The aim will be to show how such canonical coordinates can be used to specify a *normal frame*, extending the idea of semi-Euclidean frames in flat spacetime as far as possible.

In Sect. 3.1, we establish special coordinates called *normal coordinates* at an arbitrary point of an arbitrary differentiable manifold $\mathscr{M}$ of n dimensions with metric and metric connection. In this section, we use Latin indices throughout. In Sect. 3.2, we describe special coordinate systems adapted to arbitrary timelike worldlines in spacetime and return to the convention that Greek indices run over $\{0,1,2,3\}$ and Latin indices run over $\{1,2,3\}$.

3.1 Normal Coordinates

For a suitably differentiable connection, the standard existence theorems for ordinary differential equations applied to the coordinate form of the geodesic equation show that, for any point $p \in \mathscr{M}$ and any vector X_p at p, there exists a maximal geodesic $\lambda_X(v)$ in $\mathscr{M}$ with starting point p and initial direction X_p, i.e., such that [27]

$$\lambda_X(0) = p \qquad \text{and} \qquad (\partial/\partial v)_\lambda|_{v=0} = X_p \; .$$

This geodesic is unique and depends continuously on p and X_p. We use this fact to construct a map

$$\exp : T_p \longrightarrow \mathscr{M} \; ,$$

where for each $X \in T_p$, $\exp(X)$ is the point in $\mathscr{M}$ a unit parameter distance along the geodesic λ_X from p.

This map may not be defined for all $X \in T_p$, since the geodesic $\lambda_X(v)$ may not be defined for all v. If v does take all values, the geodesic $\lambda(v)$ will be said to be a *complete* geodesic. The manifold itself is said to be *geodesically complete* if all geodesics on it are complete, that is, if exp is defined on all of T_p for every point p, as will be supposed in the following.

It is intuitively obvious that the map $\exp_p$ will be of rank $n = \dim \mathscr{M}$ at p. Therefore, by the implicit function theorem, there exists an open neighbourhood $\mathscr{N}_0$ of the origin in T_p and an open neighbourhood $\mathscr{N}_p$ of $p \in \mathscr{M}$ such that the map exp is a diffeomorphism of $\mathscr{N}_0$ onto $\mathscr{N}_p$. Such a neighbourhood $\mathscr{N}_p$ is called a *normal neighbourhood* of p. In addition, $\mathscr{N}_p$ can be chosen to be *convex*, i.e., to be such that any point $q \in \mathscr{N}_p$ can be joined to any other point $r \in \mathscr{N}_p$ by a unique geodesic starting at q and totally contained in $\mathscr{N}_p$. Within a convex normal neighbourhood $\mathscr{N}$, coordinates $(x^1, \ldots, x^n)$ can be defined by choosing any point $q \in \mathscr{N}$, choosing a basis $\{E_a\}$ of T_q, and defining the coordinates of the point $r \in \mathscr{N}$ by the relation

$$r = \exp(x^a E_a) \; .$$

The procedure, then, is to expand $\exp^{-1}(r) \in T_q$ in terms of the basis $\{E_a\}$, taking the vector components as the coordinates of r.

We shall now demonstrate that

$$(\partial/\partial x^i)\big|_q = E_i \qquad \text{and} \qquad \Gamma^i_{(jk)}\Big|_q = 0 \, , \tag{3.1}$$

where the notation (jk) indicates symmetrisation. This will be done by showing first that the n curves λ_i given by

$$\lambda_i(t) = (0, \ldots, t, \ldots, 0) \; ,$$

with t in the ith place, are geodesics when the coordinates are constructed as above (see Fig. 3.1). For simplicity, we shall suppose that, for the normal neighbourhood $\mathscr{N}_q$ of q,

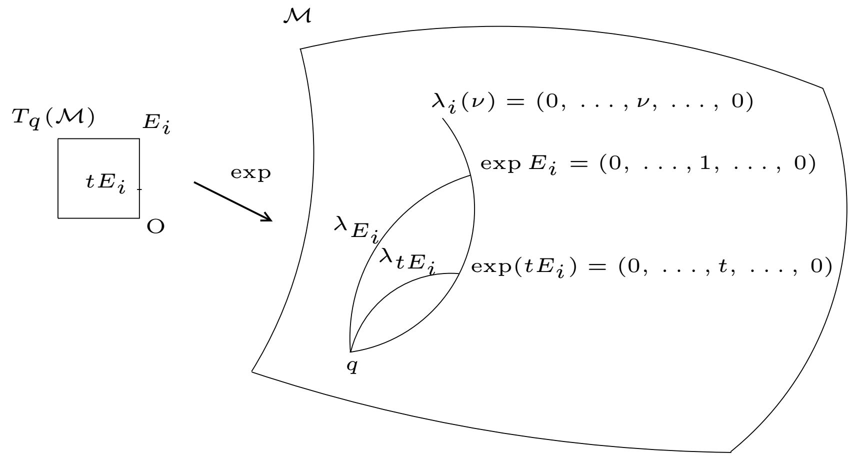

Fig. 3.1 Picture of the exponential map from the tangent space $T_q(\mathscr{M})$ to the manifold at some point $q \in \mathscr{M}$

$$E_i \in \exp^{-1} \mathscr{N}_q \ , \forall i \ ,$$

where $\{E_i\}$ is the basis of $T_q(\mathscr{M})$ chosen to determine the normal coordinates.

For any $X \in T_q(\mathscr{M})$, λ_X is the geodesic with

$$\lambda_X(0) = q \ , \qquad \left.\frac{\partial}{\partial t}\right|_{\lambda_X} = X \ ,$$

that is, the direction of λ_X at q is X. This is the geodesic used to define

$$\exp X = \lambda_X(1) \ .$$

For each t, $\lambda_{tE_i}(1)$ determines $\exp(tE_i)$. But since we are using normal coordinates, $\exp(tE_i)$ has coordinates

$$(0, \ldots, t, \ldots, 0) = \lambda_i(t) \ .$$

This therefore shows that

$$\lambda_{tE_i}(1) = \lambda_i(t) \ , \ \ \forall t \ .$$

Now for each t, consider the curve μ_{it} given by

$$\mu_{it}(w) = \lambda_{E_i}(tw) \ , \ \ \forall w \ ,$$

so that μ_{it} is the same curve as λ_{E_i}, but with a different parametrisation. From the homogeneity of the geodesic equation, which is satisfied by λ_{E_i}, we see that μ_{it} is also a geodesic, satisfying the same equation. Furthermore,

$$\left.\frac{\partial}{\partial w}\right|_{\mu_{it}} = t \left.\frac{\partial}{\partial v}\right|_{\lambda_{E_i}} = tE_i \ .$$

This means that μ_{it} is the unique geodesic through q with direction tE_i at q, and this in turn implies that

$$\exp(tE_i) = \mu_{it}(1) \ .$$

But from the definition of μ_{it},

$$\mu_{it}(1) = \lambda_{E_i}(t) \ ,$$

and consequently,

$$\lambda_{E_i}(t) = \exp(tE_i) = (0, \dots, t, \dots, 0) = \lambda_i(t) \ .$$

Finally, we deduce

$$\lambda_{E_i} = \lambda_i \ , \ \ \forall i \ .$$

Hence λ_i is a geodesic, satisfying

$$\frac{\mathrm{d}^2 x^a\big(\lambda_i(t)\big)}{\mathrm{d}t^2} + \Gamma^a_{bc} \frac{\mathrm{d}x^b\big(\lambda_i(t)\big)}{\mathrm{d}t} \frac{\mathrm{d}x^c\big(\lambda_i(t)\big)}{\mathrm{d}t} = 0 \ .$$

But we have an explicit formula for the curve when $\{x^a\}$ are normal coordinates, and the functions are linear in t, so

$$\frac{\mathrm{d}^2 x^a\big(\lambda_i(t)\big)}{\mathrm{d}t^2} = 0 \ , \ \ \forall a \ .$$

Furthermore, from this explicit formula,

$$\frac{\mathrm{d}x^b\big(\lambda_i(t)\big)}{\mathrm{d}t} = \delta^b_i \ , \qquad \frac{\mathrm{d}x^c\big(\lambda_i(t)\big)}{\mathrm{d}t} = \delta^c_i \ ,$$

which implies

$$\Gamma^a_{ii} = 0 \ , \ \ \forall i, a \ .$$

Notice that this result holds for a given i all along the curve λ_i. It therefore holds for all i at the point q. However, it is not quite the symmetry result we require to demonstrate (3.1). There is a little more work to do to show that.

We can show that $E_i = \partial/\partial x^i\big|_q$ in two different ways:

1. The direction of λ_i at q is $\partial/\partial x^i\big|_q$, so $\lambda_i(1)$ must be $\exp(\partial/\partial x^i\big|_q)$. But we know that $\lambda_i(1) = \exp E_i$, and exp is one-to-one.
2. λ_i has direction $\partial/\partial x^i\big|_q$ at q, and λ_{E_i} has direction E_i at q. But $\lambda_i = \lambda_{E_i}$.

So we have at least proven the first result stated in (3.1).

These constructions are not quite enough to show that the symmetric part of the connection is zero at q in these coordinates, but they point the way. For each pair

i, j such that $i \neq j$, define a curve

$$\lambda_{ij}(t) = (0, \dots, t, \dots, t, \dots, 0) .$$

By definition of normal coordinates,

$$\exp t(E_i + E_j) = \lambda_{ij}(t) , \ \forall t .$$

Consider the geodesics $\lambda_{E_iE_jt}$ with

$$\lambda_{E_iE_jt}(0) = q , \qquad \left.\frac{\partial}{\partial v}\right|_{\lambda_{E_iE_jt}} = t(E_i + E_j) ,$$

used to define $\exp t(E_i + E_j)$ by their value at $v = 1$, and hence giving

$$\lambda_{E_iE_jt}(1) = \exp t(E_i + E_j) .$$

Again define μ_{ijt} by

$$\mu_{ijt}(w) = \lambda_{E_iE_j1}(tw) ,$$

for each t. By homogeneity of the geodesic equation, μ_{ijt} is a geodesic through q, with direction there given by

$$\left.\frac{\partial}{\partial w}\right|_{\mu_{ijt}} = t(E_i + E_j) .$$

Consequently,

$$\mu_{ijt}(1) = \exp t(E_i + E_j) = \lambda_{ij}(t) .$$

This means that

$$\lambda_{E_iE_j1} = \lambda_{ij} , \ \forall i \neq j .$$

Therefore λ_{ij} satisfies

$$\frac{\mathrm{d}^2 x^a\big(\lambda_{ij}(t)\big)}{\mathrm{d}t^2} + \Gamma^a_{bc}\frac{\mathrm{d}x^b\big(\lambda_{ij}(t)\big)}{\mathrm{d}t}\frac{\mathrm{d}x^c\big(\lambda_{ij}(t)\big)}{\mathrm{d}t} = 0 .$$

Once again

$$\frac{\mathrm{d}^2 x^a\big(\lambda_{ij}(t)\big)}{\mathrm{d}t^2} = 0 , \ \forall a ,$$

and

$$\frac{\mathrm{d}x^b\big(\lambda_{ij}(t)\big)}{\mathrm{d}t} = \delta^b_i + \delta^b_j , \qquad \frac{\mathrm{d}x^c\big(\lambda_{ij}(t)\big)}{\mathrm{d}t} = \delta^c_i + \delta^c_j ,$$

which implies

$$0 = \Gamma^a_{bc}\frac{\mathrm{d}x^b\big(\lambda_{ij}(t)\big)}{\mathrm{d}t}\frac{\mathrm{d}x^c\big(\lambda_{ij}(t)\big)}{\mathrm{d}t} = \Gamma^a_{ij} + \Gamma^a_{ji} , \ \forall i \neq j ,$$

on the intersection of all these curves. The second result of (3.1) thus follows at q alone.

In the torsion free case (see Sect. 4.2), we thus have all connection coefficients equal to zero at the point q where we have chosen to set up these normal coordinates. Furthermore, since there is a metric, we can choose the basis $\{E_a\}$ to be orthonormal, whence the metric is in Minkowski form at q for these coordinates.

In the context of the 4D spacetime manifold of general relativity, this is therefore a *locally inertial coordinate frame*, and we have proven that the weak equivalence principle always holds in such a manifold, i.e., equipped with a Lorentzian metric and the unique symmetric (torsion free) connection that guarantees the so-called metric condition, i.e., the condition that the metric have zero covariant derivative.

3.2 Normal Frames

In the context of general relativity, there are no privileged inertial coordinate systems, unless there is no matter anywhere, since the curvature is not generally zero. And neither are there any rigid Euclidean frames, because the metric is semi-Riemannian, and hence entirely determined by the connection. However, we can associate a particularly useful kind of coordinate frame, called a *normal frame*, with any given timelike (not necessarily geodesic) worldline $\sigma(\tau)$. In this section, we return to the convention that Greek indices run over $\{0,1,2,3\}$, while Latin indices run only over $\{1,2,3\}$.

The normal coordinate system $\{y^\mu\}$ has the following properties:

1. Curves $y^i =$ constant ($i = 1, 2, 3$) are timelike, and any curve with $y^0 =$ constant is spacelike.
2. $y^0 = \tau$ along $\sigma(\tau)$.
3. The metric assumes the usual Minkowski form along $\sigma(\tau)$.
4. Any purely spacelike geodesic through σ on a $y^0 =$ constant hypersurface satisfies $\mathrm{d}^2y^i/\mathrm{d}s^2 = 0$.
5. $\{y^\mu\}$ is adapted to σ, in the sense that σ is given by $y^i = 0$, $i = 1, 2, 3$.

The difference between a normal frame and a semi-Euclidean frame in flat spacetime, as described in Sect. 2.2.1, lies in (4). In a semi-Euclidean frame, (4) is true for any purely spacelike geodesic, and not just for spacelike geodesics intersecting σ. Semi-Euclidean frames have Euclidean spatial geometry on every $y^0 =$ constant hypersurface. Normal frames do not necessarily have Euclidean spatial geometry.

3.2.1 Construction

Let us see how such a normal frame can be constructed, with the help of the normal coordinate systems in the last section. Referring to Fig. 3.2, for any value of τ,

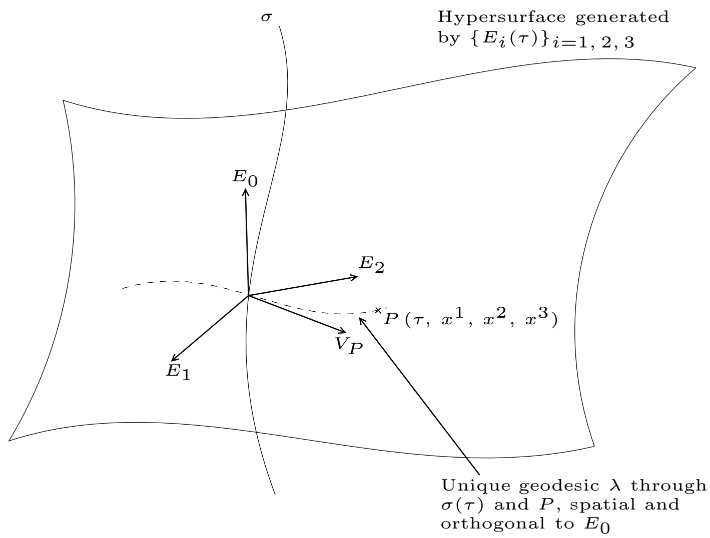

Fig. 3.2 Construction of a normal frame along an arbitrary timelike worldline in general relativity

choose a quadruple of orthonormal vectors E_0, E_1, E_2, E_3 at $\sigma(\tau)$, in such a way that each E_μ is a smooth function of τ, i.e., it is a vector field along the curve. By orthonormal, we mean that

$$\big(g_{\sigma(\tau)}(E_\mu, E_\nu)\big) = \mathrm{diag}(1, -1, -1, -1)\ ,$$

for each τ. The timelike vector E_0 is chosen to be the unit tangent to the curve, for each τ. Now we can construct our normal coordinates around each point of the curve, but how are we going to fit them together?

Firstly, for each τ, consider the hypersurface occupied by all spatial geodesics through $\sigma(\tau)$ and which are orthogonal to σ at their intersection with it. It is an assumption that we do get a hypersurface in this way, but one which will certainly be true in some neighbourhood of $\sigma(\tau)$, because spacetime manifolds always look locally like some neighbourhood of $\mathbb{R}^4$. Since we only require this construction on some neighbourhood of σ, let us assume that we do get a hypersurface orthogonal to the worldline for each value of τ.

These spacelike hypersurfaces are clearly the natural generalisation of the hyperplanes of simultaneity (HOS) we used in flat spacetime, and we shall use them to reinstate a vestige of the notion of simultaneity in our coordinate frame constructions below. We may thus continue to refer to them as hypersurfaces of simultaneity (HOS).

Now if we take any point P near enough to σ (see Fig. 3.2), there will be a unique spacelike geodesic λ through P which intersects σ and is also orthogonal to the curve there. The tangent to λ at $\sigma(\tau)$ is a spacelike vector V_P lying in the subspace of the tangent space spanned by E_1, E_2, E_3.

It is important to see what is involved in the existence and uniqueness claim regarding the curve λ. First note that, given a point P near σ but not actually on it, there may be several geodesics through it which intersect σ. There is at least one, if it is near enough, from the theorem that the map exp is onto. But this geodesic may be timelike. Intuitively, there will be some τ such that there is a geodesic linking P to $\sigma(\tau)$ which is spacelike, but then without further analysis, it remains open as to whether this geodesic will be orthogonal to σ.

What we are talking about here is whether or not the union of all the hypersurfaces of simultaneity we defined a moment ago actually contains a neighbourhood of the worldline σ. This is of course the case under suitable conditions, and can be proven by consideration of the exponential map. However, we shall not do so here since it is a mathematical issue and the aim is rather to understand physics.

For an example of what can go wrong, and what might be suitable conditions for such a theorem, the reader may refer to the case of translational uniform acceleration discussed in Sect. 2.9. It is noted there that, for any event outside region I in Fig. 2.6, there will be no proper time on the observer's hyperbolic worldline such that the observer considers it to be simultaneous, so the chosen point does not lie on any hyperplane of simultaneity for the observer.

Thus if P is near enough to σ, it will lie in one of the above HOSs, which are generated by geodesics starting out from σ in such a way that they are orthogonal to σ at their intersection with it. The question of uniqueness nevertheless remains open, for even if P lies near enough to σ to be within the range over which exp is injective, there may be some τ' such that $\sigma(\tau')$ is also linked to P by a spacelike geodesic, and this latter may even be orthogonal to σ where it intersects it. Even in flat spacetime, we remarked that hyperplanes of simultaneity would always intersect somewhere for accelerating observers (see in particular Sects. 2.2.2 and 2.12.3).

In a rigorous investigation, the need for more detailed analysis is clear. However, for the the present purposes we shall assume that, for P sufficiently close to σ, a unique spacelike geodesic orthogonal to σ can be found to link it with $\sigma(\tau)$ for some unique τ. What we are doing then is mapping our subspaces $\{E_i\}_{i=1,2,3}$ of $T_{\sigma(\tau)}$ onto 3D submanifolds containing $\sigma(\tau)$ for all τ and only coordinatising those P which lie on a single geodesic whose tangent at its intersection with σ lies in the subspace mentioned.

The rest of the construction is straightforward. Given the geodesic λ, with an affine parametrisation ν adjusted so that $P = \lambda(1)$, find

$$V_P = \left.\frac{\partial \lambda}{\partial \nu}\right|_{\sigma(\tau)} ,$$

and write it (uniquely) in the form

$$V_P = y^1 E_1 + y^2 E_2 + y^3 E_3 .$$

Then the point P will be attributed the coordinates $(0, y^1, y^2, y^3)$ in the scheme of normal coordinates devised above. Of course, V_P cannot have a nonzero component $y^0 E_0$, since λ was chosen orthogonal to σ.

The coordinates of P for the purposes of our normal frame will be (τ, y^1, y^2, y^3), where τ is assumed to be the unique value allowing the construction for the given P. We also know that

$$P = \exp V_P \ .$$

Let us restate this in terms of the metric. Let λ' be the same curve as λ but reparametrised by proper distance s. Note that λ has been parametrised so that $P = \lambda(1)$. Now both s and ν are affine parameters, with the same origin, and in this situation there is always some constant κ such that $s = \kappa\nu$. What we are proposing then is to define

$$\lambda'(s) := \lambda(\nu) = \lambda(s/\kappa) \ ,$$

and this implies

$$\left.\frac{\partial \lambda'}{\partial s}\right|_{\sigma(\tau)} = \frac{1}{\kappa} V_P = n \ .$$

The symbol n denotes the unit tangent vector to our geodesic λ. We can now write the space coordinates y^i in terms of this vector n, the proper distance separating P from σ, as measured along the unique geodesic λ, and the metric:

$$y^i = -g_{\sigma(\tau)}(n, E_i) s \ .$$

Note that the value of s here is just κ, because P is given when $\nu = 1$. In other words, the spatial coordinates of P are found by identifying the components of its direction from $\sigma(\tau)$, relative to the basis at this point, and multiplying by its proper distance from this point, along the unique geodesic. The temporal coordinate is obtained as the value of the unique τ such that the spacelike geodesic from P intersects σ orthogonally. This works because we diagonalised the metric by the choice of basis $\{E_i\}_{i=1,2,3}$ along σ.

We have also established the coordinate formula for the unique spacelike geodesic from $\sigma(\tau)$ to P:

$$\big(y^\mu(\lambda_P(\nu))\big) = (\tau, \nu y^1, \nu y^2, \nu y^3) \ , \tag{3.2}$$

where $(y^i, i = 1, 2, 3)$ are the coordinates of P. As mentioned at the beginning of Sect. 2.11.5, such coordinates are also called geodesic coordinates, and it is not difficult to see why!

3.2.2 Checking Properties

The main task now is to check that we have satisfied all the claims made about this coordinate system. It is clear that σ satisfies $y^\mu = 0$, $i = 1, 2, 3$ and that $y^0 = \tau$ along the curve. To see why the metric is diagonalised at any point of σ, we must evaluate

$$g(\partial/\partial y^\mu, \partial/\partial y^\nu)\ , \quad \forall\, \mu, \nu\ .$$

The point is, of course, that

$$\left.\frac{\partial}{\partial y^i}\right|_{\sigma(\tau)} = E_i\ , \ i = 1, 2, 3\ ,$$

as shown in Sect. 3.1 on normal coordinates, and

$$\left.\frac{\partial}{\partial y^0}\right|_{\sigma(\tau)} = E_0\ ,$$

is the tangent vector to σ. But $\{E_i\}$ was an orthonormal basis for the tangent space at this point.

Concerning the spacelike geodesics through σ on some $y^0 =$ constant hypersurface, we have specifically arranged for them to be linear functions of the spatial coordinates, as can be seen from (3.2), so that they must satisfy

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}s^2} = 0\ .$$

Note also that curves with $y^0 =$ constant which intersect σ but are not necessarily geodesics must be orthogonal to σ where they intersect it. Suppose the curve is $\lambda(v)$, so that

$$y^0 \circ \lambda(v) = \text{constant}\ ,$$

and only $y^i \circ \lambda(v)$ vary for $i = 1, 2, 3$. Then the tangent at $\sigma(\tau)$ has zero zeroth component, as claimed.

What about the claim that the curves $y^0 =$ constant are spacelike, and the curves $y^i =$ constant are timelike? In the case of the semi-Euclidean frame in flat spacetime, we eventually examined this question explicitly by constructing the matrix of components of the Minkowski metric relative to these coordinates in Sect. 2.3.8. Of course, it is clear in both the flat and curved spacetime cases that the $y^0 =$ constant curves are spacelike, because they lie in spacelike hypersurfaces. For the semi-Euclidean frame, the spacelike hypersurface in question is the plane of simultaneity of an inertial frame, and we can check explicitly that any curve contained completely within such a hypersurface must have tangent everywhere spacelike. In the present case, we would have to show that the hypersurface spanned by the spacelike geodesics from a given $\sigma(\tau)$ is indeed spacelike.

Intuitively speaking, i.e., without going into the mathematical details, the metric varies little from its diagonalised form, provided we remain close enough to the point at which it has been diagonalised, and this is sufficient to establish the result. A similar argument will work for the curves $y^i =$ constant. So here we can capitalise on the idea that we only need to set up our coordinate system in a neighbourhood of the worldline. In specific cases, of course, we can be more specific!

3.2.3 Connection Coefficients

Let us now examine the connection coefficients in such a frame. They will not generally be zero anywhere. This is no surprise, because we did not adopt the timelike normal coordinate, which can get all the connection coefficients equal to zero at *one* spacetime event, but replaced it by proper time along the curve. We shall analyse these coefficients along σ.

Firstly, note that, in the new coordinates, σ has 4-velocity and 4-acceleration

$$u = (1, 0, 0, 0) , \qquad a = (0, a^1, a^2, a^3) .$$

Here we use the fact that the 4-velocity has constant pseudolength, and hence is always orthogonal to the 4-acceleration, a consequence of the metric condition, which says that the covariant derivative of the metric is zero. A general proof follows from the fact that

$$a^0 = g(T_\sigma, \mathrm{D}_{T_\sigma} T_\sigma) = \frac{1}{2}\mathrm{D}_{T_\sigma} g(T_\sigma, T_\sigma) = 0 ,$$

where we have used $\mathrm{D}g = 0$ and $g(T_\sigma, T_\sigma) = 1$. Of course, the 4-velocity is just the unit tangent vector to σ.

Now if we consider any spacelike geodesic intersecting σ orthogonally, we have seen that

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}s^2} = 0 ,$$

for $i = 1, 2, 3$. We also know that the first derivatives are constants. Hence,

$$\boxed{\Gamma^\mu_{ij}\Big|_\sigma = 0 , \quad \forall \mu \quad \text{and} \quad \forall\, i,\, j = 1, 2, 3} . \tag{3.3}$$

This follows from the geodesic equations for these spatial geodesics. Note that here arises the main difference with the semi-Euclidean frame. In the hyperplane of simultaneity of an inertial frame, all the spacelike geodesics, even those not intersecting σ, are given by linear functions of their parameter, and consequently, the connection coefficients specified above are zero everywhere (see the discussion on p. 36). In the present case, those coefficients can only be guaranteed to go to zero on σ itself.

Concerning the equation for σ itself, which is specified by $y^0 = \tau$, $y^i = 0$, for $i = 1, 2, 3$, we can say that the 4-acceleration is

$$a^\mu = \frac{\mathrm{d}^2 y^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\nu\sigma} \frac{\mathrm{d}y^\nu}{\mathrm{d}\tau} \frac{\mathrm{d}y^\sigma}{\mathrm{d}\tau} = \Gamma^\mu_{00} ,$$

whence

$$\boxed{\Gamma^0_{00}\big|_\sigma = 0 , \qquad \Gamma^i_{00}\big|_\sigma = a^i , \quad i = 1,2,3} . \tag{3.4}$$

The remaining nonzero coefficients are $\Gamma^i_{j0}\Big|_\sigma = \Gamma^i_{0j}\Big|_\sigma$ for $i, j = 1, 2, 3$, and also $\Gamma^0_{i0}\big|_\sigma = \Gamma^0_{0i}\big|_\sigma$, for $i = 1, 2, 3$. We can deduce something about these from the fact that the connection is the unique metric connection, and therefore $\mathrm{D}_{T_\sigma} g = 0$. Recall that

$$0 = g_{\mu\nu;\sigma} = g_{\mu\nu,\sigma} - \Gamma^\lambda_{\mu\sigma} g_{\lambda\nu} - \Gamma^\lambda_{\nu\sigma} g_{\mu\lambda} \ ,$$

and, when this is contracted with the tangent vector to the curve σ, the first term on the right-hand side gives the derivative of the metric component along the curve, which is zero. Furthermore, we know the values of $g_{\lambda\nu}$ and $g_{\mu\lambda}$ along σ, and we can deduce that

$$\Gamma^\lambda_{\mu 0}\Big|_\sigma \eta_{\lambda\nu} + \Gamma^\lambda_{\nu 0}\Big|_\sigma \eta_{\mu\lambda} = 0 \ ,$$

whence

$$\boxed{\Gamma^i_{j0}\big|_\sigma = \Gamma^i_{0j}\big|_\sigma = \Omega^i{}_j \ , \quad i, j = 1, 2, 3} \ , \tag{3.5}$$

where Ω is an antisymmetric (rotation) matrix, and also

$$\boxed{\Gamma^0_{i0}\big|_\sigma = \Gamma^0_{0i}\big|_\sigma = \Gamma^i_{00}\big|_\sigma = a^i \ , \quad i = 1, 2, 3} \ . \tag{3.6}$$

We thus obtain the same system of connection coefficients as for SE coordinates in flat spacetimes (see Sect. 2.3.9).

3.2.4 Eliminating the Rotation Matrix

We can guess how it will be possible to remove the rotation matrix $\Omega^i{}_j$ and make the coefficients Γ^i_{0j} equal to zero on the observer worldline. Let us suppose that we select the tetrad $\{T_\sigma, E_i\}_{i=1,2,3}$ at some point $\sigma(\tau_0)$ on the worldline, then Fermi–Walker (FW) transport it along the worldline to obtain a tetrad everywhere along the worldline. If we use this tetrad, we find that the connection coefficients Γ^i_{0j} are zero. To show this, consider the equation for FW transport

$$\dot{A} = -(A \cdot \dot{T}_\sigma) T_\sigma + (A \cdot T_\sigma) \dot{T}_\sigma \ , \tag{3.7}$$

which works exactly like (2.28) on p. 26 in the flat spacetime context. We shall be inserting $A = E_i$, for $i = 1,2,3$. Now

$$\dot{T}_\sigma = \frac{\mathrm{D}T_\sigma}{\mathrm{d}\tau} = a \ ,$$

the 4-acceleration of the worldline, which has components $a = (0, a^1, a^2, a^3)$, while T_σ has the component form $T_\sigma = (1,0,0,0)$. Furthermore, $E_i^\mu = \delta_i^\mu$. Now

$$\dot{E}_i^j := \frac{\mathrm{D}E_i^j}{\mathrm{d}\tau} := \frac{\mathrm{d}E_i^j}{\mathrm{d}\tau} + \Gamma^j_{\mu\nu} E_i^\mu T_\sigma^\nu = \Gamma^j_{i0} \ ,$$

and the FW equation (3.7) sets this equal to

$$-(E_i \cdot \dot{T}_\sigma)T_\sigma^j + (E_i \cdot T_\sigma)\dot{T}_\sigma^j = 0 \, .$$

We conclude that $\Gamma^j_{i0}\big|_\sigma = 0$ for all $i, j = 1, 2, 3$. Note that the converse is true. If we assume the latter, the above argument works backwards and we conclude that the E_i are FW transported.

3.2.5 Viewing Free Particles

We can now write down the equation of motion of a free particle on σ, assuming that it intersects the observer's worldline, and only at the unique event at which it coincides with the observer. We obtain exactly the same result as in (2.10) on p. 19, viz.,

$$\boxed{\frac{\mathrm{d}^2 y^i}{\mathrm{d}y^{02}} + a^i + 2\Omega^i_j \frac{\mathrm{d}y^j}{\mathrm{d}y^0} - 2a^j \frac{\mathrm{d}y^j}{\mathrm{d}y^0}\frac{\mathrm{d}y^i}{\mathrm{d}y^0} = 0} \, , \tag{3.8}$$

where indices run over values 1, 2, 3. It contains a term a^i which could be referred to as an inertial force, a term proportional to the semi-Euclidean three-velocity which corresponds to a Coriolis force, and finally a relativistic correction. As mentioned at the end of the last section, the Coriolis term can be removed by Fermi–Walker transporting the coordinate axes along the worldline.

This formula is established in the same way as (2.10). We start with the geodesic equation for $i = 1, 2, 3$,

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}\tau'^2} + a^i \left(\frac{\mathrm{d}y^0}{\mathrm{d}\tau'}\right)^2 + 2\Omega^i_j \frac{\mathrm{d}y^j}{\mathrm{d}\tau'}\frac{\mathrm{d}y^0}{\mathrm{d}\tau'} = 0 \, ,$$

where τ' is proper time along this geodesic. If we change parameter from the affine parameter τ' to the non-affine parameter y^0, we simplify the left-hand side here, but the right-hand side is now

$$-\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2}\frac{\mathrm{d}y^i}{\mathrm{d}y^0} \Big/ \left(\frac{\mathrm{d}y^0}{\mathrm{d}\tau'}\right)^2 \, .$$

The zero component of the geodesic equation is

$$\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2} + \Gamma^0_{jk}\frac{\mathrm{d}y^j}{\mathrm{d}\tau'}\frac{\mathrm{d}y^k}{\mathrm{d}\tau'} = 0 \, ,$$

which gives, since $\Gamma^0_{00} = 0$,

$$\frac{\mathrm{d}^2 y^0}{\mathrm{d}\tau'^2} = -2a^j \frac{\mathrm{d}y^j}{\mathrm{d}\tau'}\frac{\mathrm{d}y^0}{\mathrm{d}\tau'} \, ,$$

where j is summed from 1 to 3. The result follows.

3.3 Locally Inertial Frames

Let us note finally what happens if the acceleration a^i and the rotation $\Omega^i{}_j$ are all zero. In this case, the normal frame becomes a local inertial frame, and the law of motion (3.8) for a free particle takes the familiar form

$$\frac{\mathrm{d}^2 y^i}{\mathrm{d}y^{02}} = 0 \, .$$

Note that $a^i = 0$ for $i = 1,2,3$, if and only if the curve σ followed by the observer setting up these coordinates is a geodesic, while $\Omega^i{}_j = 0$ for $i, j = 1,2,3$, if and only if the spacelike triad $\{E_i\}$ used to construct the normal frame has been FW transported along σ. As pointed out earlier, this in turn means that $\{E_i\}$ has in fact been parallel transported along the geodesic σ, since FW transport reduces to parallel transport when the worldline in question is a geodesic. We also know from the earlier discussion that a^i and the rotation $\Omega^i{}_j$ are all zero if and only if

$$\Gamma^{\mu}_{\nu\gamma}\big|_{\sigma} = 0 \, , \quad \forall \mu, \nu, \gamma \, .$$

In short, we can arrange for all the connection coefficients to be zero right along the worldline σ if and only if σ is a geodesic, i.e., if and only if the observer is freely falling. Then all freely falling particles whose worldlines intersect σ follow coordinate-straight worldlines in the normal frame coordinates, at least close to the intersection.

Chapter 4
Holonomic and Non-Holonomic Frames in General Relativity

A frame in a region U of a spacetime $\mathscr{M}$ is a basis $\{e_\alpha\}_{\alpha=0,1,2,3}$ for the tangent space $T_p(\mathscr{M})$ at each event p in U. A holonomic frame is one that arises from coordinates $\{x^\mu\}_{\mu=0,1,2,3}$ over U, observing that the four contravectors

$$\partial_\mu := \left.\frac{\partial}{\partial x^\mu}\right|_p$$

form a basis for $T_p(\mathscr{M})$, for each $p \in U$.

So far we have constructed coordinates, and hence coordinate frames, that are adapted to observers with arbitrary motion either in flat or curved spacetime. All our frames have therefore effectively been holonomic frames. However, non-holonomic frames can also be of interest for formulating physical problems.

A well known example are tetrad frames in which the four contravectors e_α actually constitute an orthonormal basis for $T_p(\mathscr{M})$ at each event p for the given spacetime metric g:

$$(e_\alpha|e_\beta) := g(e_\alpha, e_\beta) = \eta_{\alpha\beta} \ .$$

We shall stick to the convention $\eta := \mathrm{diag}(-1,1,1,1)$ until further notice. If a tetrad frame turns out to be holonomic, i.e., if there exist coordinates x^α, $\alpha = 0,1,2,3$, such that

$$e_\alpha = \partial_\alpha \ , \quad \alpha = 0,1,2,3 \ ,$$

then the spacetime is flat in the given region.

This is a broad subject that can be found in many standard textbooks (good references are [12, 14]). The aim here will be to review some of the essentials, for completeness, then look at how those essentials relate to the problem of physical observation that are the main concern here. As mentioned in the title, we discuss this in the context of general relativity, considering the manifold of special relativity (Minkowski spacetime) as a special case when there are coordinates such that the metric is just η.

We shall talk about arbitrary (possibly non-holonomic) contravector frames $\{e_\alpha\}_{\alpha=0,1,2,3}$ in a region U of spacetime, forming a basis for the tangent space

$T_p(\mathscr{M})$ at each $p \in U$, with dual covector (1-form) frame $\{\mathbf{e}^\alpha\}_{\alpha=0,1,2,3}$, defined by

$$\mathbf{e}^\alpha(e_\beta) = \delta^\alpha_\beta \ , \tag{4.1}$$

forming a basis for the cotangent space $T^*_p(\mathscr{M})$ at each $p \in U$. Under a change of such frame to $\{e'_\alpha\}_{\alpha=0,1,2,3}$, there always exists a field of invertible matrices A over U such that

$$e'_\alpha = A_\alpha{}^\beta e_\beta \ , \tag{4.2}$$

and it is easy to check that

$$\mathbf{e}^{\alpha\prime} = \mathbf{e}^\beta A^{-1}{}_\beta{}^\alpha \ , \tag{4.3}$$

since

$$\begin{aligned}
\mathbf{e}^\gamma A^{-1}{}_\gamma{}^\alpha (e'_\beta) &= \mathbf{e}^\gamma (A_\beta{}^\delta e_\delta) A^{-1}{}_\gamma{}^\alpha \\
&= A_\beta{}^\delta \mathbf{e}^\gamma(e_\delta) A^{-1}{}_\gamma{}^\alpha \\
&= A_\beta{}^\delta \delta^\gamma_\delta A^{-1}{}_\gamma{}^\alpha \\
&= \delta^\alpha_\beta \ .
\end{aligned}$$

4.1 Lie Bracket and Structure Coefficients

Associated with any frame $\{e_\alpha\}_{\alpha=0,1,2,3}$ is a set of numbers $c^\alpha{}_{\beta\gamma}$ called structure coefficients which are useful for understanding the frame in an abstract sense. To define them we need first to define the Lie bracket of two vector fields X and Y, variously denoted $\mathscr{L}_X Y$, $XY - YX$, and $[X,Y]$. The Lie bracket of two vector fields is another vector field. We shall define it here by its action on functions f specified on the relevant region U of the spacetime manifold. There are many other ways, some more elegant, but our purpose here is not an exhaustive presentation of differential geometry.

The Lie bracket $\mathscr{L}_X Y$ is defined at $p \in U$ by the following action on a smooth function f in the neighbourhood of p:

$$\begin{aligned}
(\mathscr{L}_X Y)(p) f &= \Big[X\big(Y(f)\big) - Y\big(X(f)\big) \Big](p) \\
&= X^\mu (Y^\nu f_{,\nu})_{,\mu} - Y^\mu (X^\nu f_{,\nu})_{,\mu} \ ,
\end{aligned} \tag{4.4}$$

where indices μ and ν refer to a local coordinate system and commas denote partial coordinate derivatives. It is easy to check that this has the properties of a vector at each p, viz., linearity as a map on the local function space $\mathscr{F}$,

$$(\mathscr{L}_X Y)(af + bg) = a(\mathscr{L}_X Y) f + b(\mathscr{L}_X Y) g \ , \quad \forall a,b \in \mathbb{R} \ , \quad \forall f,g \in \mathscr{F} \ ,$$

and Leibniz rule,

$$(\mathscr{L}_X Y)(fg) = \big[(\mathscr{L}_X Y)f\big]g + f(\mathscr{L}_X Y)g\ , \quad \forall f, g \in \mathscr{F}\ .$$

From (4.4), the component form of $\mathscr{L}_X Y = [X,Y]$ relative to a coordinate frame $\{\partial_\mu\}_{\mu=0,1,2,3}$ is clearly

$$[X,Y]^\mu = X^\nu \partial_\nu Y^\mu - Y^\nu \partial_\nu X^\mu\ . \tag{4.5}$$

It is also clear that

$$\mathscr{L}_X Y = -\mathscr{L}_Y X\ , \quad \text{or} \quad [X,Y] = -[Y,X]\ . \tag{4.6}$$

Given a vector field X, one can also define the Lie derivative $\mathscr{L}_X T$ of any type (r,s) tensor field T. This gives another type (r,s) tensor field. Note that the derivatives in the definition (4.4) are coordinate derivatives, not covariant derivatives. No metric structure or connection is needed to define the Lie derivative, and more sophisticated presentations show why this is the case [14, 27].

A very important property of the Lie bracket is the Jacobi identity:

$$\big[X,[Y,Z]\big] + \big[Z,[X,Y]\big] + \big[Y,[Z,X]\big] = 0\ . \tag{4.7}$$

This is straightforward but somewhat tedious to prove. The highly suggestive notation $[X,Y] = XY - YX$ makes it look misleadingly obvious! Another result that can be proven straight from the definition is

$$[fX, gY] = fg[X,Y] + fX(g)Y - gY(f)X\ , \quad \forall f, g \in \mathscr{F}\ , \quad \forall X, Y \in T(\mathscr{M})\ . \tag{4.8}$$

Then, given any frame $\{e_\alpha\}_{\alpha=0,1,2,3}$ and any contravector fields X and Y in some region U of spacetime, we can expand them as $X = X^\alpha e_\alpha$ and $Y = Y^\beta e_\beta$, where X^α and Y^β are functions in $\mathscr{F}$, and (4.8) implies

$$[X,Y] = X^\alpha Y^\beta [e_\alpha, e_\beta] + X^\alpha e_\alpha(Y^\beta) e_\beta - Y^\beta e_\beta(X^\alpha) e_\alpha\ . \tag{4.9}$$

This idea throws up the Lie bracket of pairs of vector fields selected from the frame $\{e_\alpha\}_{\alpha=0,1,2,3}$. But such a Lie bracket always delivers another contravector field, and since the frame provides a basis everywhere in the given region, we must be able to express these Lie brackets as linear combinations of the e_α. Hence there exist functions $C^\gamma{}_{\alpha\beta}$ on U such that

$$\boxed{[e_\alpha, e_\beta] = C^\gamma{}_{\alpha\beta} e_\gamma}\ . \tag{4.10}$$

These are the structure coefficients of the frame $\{e_\alpha\}_{\alpha=0,1,2,3}$. They are obviously antisymmetric in the lower indices:

$$C^\gamma{}_{\alpha\beta} = -C^\gamma{}_{\beta\alpha} = C^\gamma{}_{[\alpha\beta]}\ , \tag{4.11}$$

where the square bracket around indices denotes antisymmetrisation over those indices.

It should be noted that these sets of 4^3 numbers do not transform as tensor components under change of frame $e_\alpha \to e'_\alpha = A_\alpha{}^\beta e_\beta$. Indeed, the new coefficients $C'^\alpha{}_{\beta\gamma}$ are defined by

$$[e'_\alpha, e'_\beta] = C'^\gamma{}_{\alpha\beta} e'_\gamma \,,$$

whence

$$\begin{aligned}
C'^\gamma{}_{\alpha\beta} A_\gamma{}^\delta e_\delta &= [A_\alpha{}^\gamma e_\gamma, A_\beta{}^\delta e_\delta] \\
&= A_\alpha{}^\gamma A_\beta{}^\delta [e_\gamma, e_\delta] + A_\alpha{}^\gamma e_\gamma(A_\beta{}^\delta) e_\delta - A_\beta{}^\delta e_\delta(A_\alpha{}^\gamma) e_\gamma \\
&= A_\alpha{}^\gamma A_\beta{}^\delta C^\varepsilon{}_{\gamma\delta} e_\varepsilon + A_\alpha{}^\gamma e_\gamma(A_\beta{}^\delta) e_\delta - A_\beta{}^\delta e_\delta(A_\alpha{}^\gamma) e_\gamma \,,
\end{aligned}$$

where the second step uses (4.8). Now equating coefficients of e_δ on both sides, we deduce that

$$C'^\gamma{}_{\alpha\beta} A_\gamma{}^\varepsilon = A_\alpha{}^\gamma A_\beta{}^\delta C^\varepsilon{}_{\gamma\delta} + A_\alpha{}^\gamma e_\gamma(A_\beta{}^\varepsilon) - A_\beta{}^\delta e_\delta(A_\alpha{}^\varepsilon) \,,$$

and hence

$$C'^\lambda{}_{\alpha\beta} = A_\alpha{}^\gamma A_\beta{}^\delta C^\varepsilon{}_{\gamma\delta} A^{-1}{}_\varepsilon{}^\lambda + A_\alpha{}^\gamma e_\gamma(A_\beta{}^\varepsilon) A^{-1}{}_\varepsilon{}^\lambda - A_\beta{}^\delta e_\delta(A_\alpha{}^\varepsilon) A^{-1}{}_\varepsilon{}^\lambda \,,$$

so that, finally,

$$\boxed{C'^\lambda{}_{\alpha\beta} = A_\alpha{}^\gamma A_\beta{}^\delta C^\varepsilon{}_{\gamma\delta} A^{-1}{}_\varepsilon{}^\lambda + 2A_{[\alpha}{}^\gamma e_\gamma(A_{\beta]}{}^\delta) A^{-1}{}_\delta{}^\lambda} \,, \tag{4.12}$$

where the antisymmetrization in the second term (denoted by square brackets) is over α and β alone. If the structure coefficients transformed tensorially, we would expect only the first term on the right-hand side.

It is clear straight from (4.5) that the structure coefficients are zero for a holonomic frame:

$$[\partial_\mu, \partial_\nu] = 0 \,, \quad C^\rho{}_{\mu\nu} = 0 \quad \text{(holonomic frame)} \,. \tag{4.13}$$

It turns out that the converse is also true, i.e., if we have a frame $\{e_\alpha\}_{\alpha=0,1,2,3}$ such that $C^\gamma{}_{\alpha\beta} = 0$ for all $\alpha, \beta, \gamma \in \{0,1,2,3\}$ on the region U of spacetime, then there exist coordinates $\{x^\alpha\}_{\alpha=0,1,2,3}$ on the region U such that $e_\alpha = \partial_\alpha$. This is one consequence of the Frobenius theorems [12]. Looking back at (4.9) applied to a holonomic frame, we immediately rederive (4.5).

4.2 Connection and Torsion

Since

$$\begin{aligned}
e_\gamma(A_\beta{}^\delta) A^{-1}{}_\delta{}^\lambda &= e_\gamma(A_\beta{}^\delta A^{-1}{}_\delta{}^\lambda) - A_\beta{}^\delta e_\gamma(A^{-1}{}_\delta{}^\lambda) \\
&= e_\gamma(\delta_\beta^\lambda) - A_\beta{}^\delta e_\gamma(A^{-1}{}_\delta{}^\lambda) \\
&= -A_\beta{}^\delta e_\gamma(A^{-1}{}_\delta{}^\lambda) \,,
\end{aligned}$$

the transformation equation (4.12) for the structure coefficients can be written

$$\boxed{C'^{\lambda}{}_{\alpha\beta} = A_\alpha{}^\gamma A_\beta{}^\delta C^\varepsilon{}_{\gamma\delta} A^{-1}{}_\varepsilon{}^\lambda + 2A_\alpha{}^\gamma A_\beta{}^\delta e_{[\delta}(A^{-1}{}_{\gamma]}{}^\lambda)} \,. \tag{4.14}$$

It is interesting to compare this with the transformation rule for the connection, viz.,

$$\boxed{\Gamma'^{\lambda}{}_{\alpha\beta} = A_\alpha{}^\gamma A_\beta{}^\delta \Gamma^\varepsilon{}_{\gamma\delta} A^{-1}{}_\varepsilon{}^\lambda - A_\alpha{}^\gamma A_\beta{}^\delta e_{\delta}(A^{-1}{}_{\gamma}{}^\lambda)} \,. \tag{4.15}$$

This transformation law ensures that the covariant derivative of any type (r,s) tensor defined in the usual way is a type $(r,s+1)$ tensor:

- For a contravector field $Y = Y^\alpha e_\alpha$,

$$\nabla_\alpha Y^\beta := e_\alpha(Y^\beta) + \Gamma^\beta_{\gamma\alpha} Y^\gamma \,. \tag{4.16}$$

- For a covector field $\omega = \omega_\alpha \mathbf{e}^\alpha$,

$$\nabla_\alpha \omega_\beta := e_\alpha(\omega_\beta) - \Gamma^\gamma_{\beta\alpha}\omega_\gamma \,. \tag{4.17}$$

- For a type (r,s) tensor $T = T^{\alpha\beta\cdots}{}_{\mu\nu\ldots} e_\alpha \otimes e_\beta \otimes \cdots \otimes \mathbf{e}^\mu \otimes \mathbf{e}^\nu \otimes \cdots$,

$$\begin{aligned}\nabla_\alpha T^{\beta\gamma\cdots}{}_{\mu\nu\ldots} := e_\alpha(T^{\beta\gamma\cdots}{}_{\mu\nu\ldots}) + \Gamma^\beta_{\delta\alpha} T^{\delta\gamma\cdots}{}_{\mu\nu\ldots} + \Gamma^\gamma_{\delta\alpha} T^{\beta\delta\cdots}{}_{\mu\nu\ldots} + \cdots \\ -\Gamma^\rho_{\mu\alpha} T^{\alpha\beta\cdots}{}_{\rho\nu\ldots} - \Gamma^\rho_{\nu\alpha} T^{\alpha\beta\cdots}{}_{\mu\rho\ldots} - \cdots \,.\end{aligned} \tag{4.18}$$

Of course, we define

$$\nabla_X Y := (X^\alpha \nabla_\alpha Y^\beta) e_\beta = \left[X(Y^\beta) + X^\alpha \Gamma^\beta_{\gamma\alpha} Y^\gamma\right] e_\beta \,, \tag{4.19}$$

whence

$$\nabla_Y X - \nabla_X Y = \left[Y(X^\beta) - X(Y^\beta) + 2\Gamma^\beta_{[\alpha\gamma]} X^\alpha Y^\gamma\right] e_\beta \,. \tag{4.20}$$

But looking at (4.9),

$$\left[Y(X^\beta) - X(Y^\beta)\right] e_\beta = [Y,X] + C^\beta{}_{\alpha\gamma} X^\alpha Y^\gamma e_\beta \,. \tag{4.21}$$

So if we define the type (1,2) torsion tensor T by

$$\boxed{T(X,Y) := \nabla_Y X - \nabla_X Y - [Y,X]} \,, \tag{4.22}$$

it follows that

$$\boxed{T(Y,X) = \left(2\Gamma^\beta_{[\alpha\gamma]} + C^\beta{}_{\alpha\gamma}\right) X^\alpha Y^\gamma e_\beta} \,, \tag{4.23}$$

or in component form,

$$\boxed{T^\beta{}_{\alpha\gamma} = 2\Gamma^\beta_{[\alpha\gamma]} + C^\beta{}_{\alpha\gamma}} \,. \tag{4.24}$$

We know already that this is a tensor by construction, but we can immediately see why it transforms correctly from (4.14) and (4.15).

Relative to a holonomic basis $\{\partial_\mu\}_{\mu=0,1,2,3}$, the structure coefficients are zero and the torsion tensor has component form

$$T^\rho{}_{\mu\nu} = 2\Gamma^\rho_{[\mu\nu]} \, , \tag{4.25}$$

showing that this tensor is entirely determined by the connection. A torsion free connection is then one for which the torsion tensor is zero. This happens if and only if the connection coefficients are symmetric in the two lower indices whenever the connection is expressed relative to a coordinate frame. In Sect. 3.1, we assumed a torsion free connection in order to construct normal coordinates at any given point in spacetime.

In building relativistic theories of gravitation, it is commonplace to assume a certain compatibility between metric and connection. Basically, one assumes that parallel transport, determined by the connection, preserves inner products. This in turn is valid if and only if the covariant derivative of the metric is zero, often called the metric condition:

$$\nabla_\alpha g_{\beta\gamma} = 0 \, . \tag{4.26}$$

From the definition (4.18), this means that

$$e_\alpha(g_{\beta\gamma}) = \Gamma^\delta_{\beta\alpha} g_{\delta\gamma} + \Gamma^\delta_{\gamma\alpha} g_{\beta\delta} \, . \tag{4.27}$$

If in addition the connection is torsion free, then we have seen from (4.24) that

$$2\Gamma^\beta_{[\alpha\gamma]} = -C^\beta{}_{\alpha\gamma} \, . \tag{4.28}$$

It is now straightforward to check that the connection is fully determined by the metric, since (4.27) and (4.28) imply that

$$\begin{aligned} \Gamma^\gamma_{\alpha\beta} = \frac{1}{2} g^{\gamma\delta} \Big[e_\beta(g_{\alpha\delta}) + e_\alpha(g_{\delta\beta}) - e_\delta(g_{\alpha\beta}) \Big] \\ + \frac{1}{2} \Big[C^\gamma{}_{\beta\alpha} + g^{\gamma\delta} g_{\alpha\varepsilon} C^\varepsilon{}_{\delta\beta} + g^{\gamma\delta} g_{\varepsilon\beta} C^\varepsilon{}_{\delta\alpha} \Big] \, . \end{aligned} \tag{4.29}$$

Expressed relative to a coordinate frame, where the structure coefficients are zero, the connection thus has components

$$\boxed{\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma} \Big(\partial_\nu g_{\mu\sigma} + \partial_\mu g_{\sigma\nu} - \partial_\sigma g_{\mu\nu} \Big)} \, , \tag{4.30}$$

the well known formula for the Christoffel symbols.

From a physical point of view, torsion is sourced by spinning matter and energy. In fact, it is sourced by the spin currents of matter in such a way that, in contrast to curvature, it does not propagate in spacetime, so it could only be nonzero in regions where there is matter or energy with some rotation. A very clear, though somewhat sophisticated account can be found in [25, Chap. 5].

4.3 Timelike Congruences and the Tetrad Formalism

As mentioned above, a tetrad frame $\{e_\alpha\}_{\alpha=0,1,2,3}$ is one in which the four contravector fields e_α constitute an orthonormal basis for the tangent space $T_p(\mathscr{M})$ at each event p where they take values. The aim here is to review this formalism, closely following the presentation in [12, Chap. 4].

The starting point is a congruence of timelike curves γ with tangent vector field X such that $(X|X) = -1$. We use the proper time τ for the parameter on each curve, with

$$X = \dot{\gamma}\,,$$

and synchronise suitably over the curves. This congruence can model the worldlines of particles making up a material, e.g., a fluid, or a physical observer, or the observer's laboratory. The aim will be eventually to see how this kind of formalism can actually be used to relate the theory to physical measurement.

Synchronising the proper time τ suitably over the curves is an issue here. In Minkowski spacetime, when setting up coordinates for an observer identified here with one worldline in the congruence, with proper time τ_0, we favour hyperplanes of simultaneity $\mathrm{HOS}(\tau_0)$ borrowed from instantaneously comoving inertial observers at each value of τ_0, which seems a natural enough choice. We then attribute the same value τ_0 of the time coordinate to each worldline in the congruence at its intersection with $\mathrm{HOS}(\tau_0)$. We noted in Sect. 2.3.4 that this time coordinate would not generally be equal to the proper time along each worldline, even when we had synchronised proper times on some particular hyperplane of simultaneity [see (2.35) on p. 28].

Still in flat spacetime, another problem is that the hyperplanes of simultaneity for one worldline in the congruence will not generally intersect the other worldlines orthogonally, so they will not generally be a natural choice for the hyperplanes of simultaneity for observers following other worldlines in the congruence. A case where hyperplanes of simultaneity are shared between all members of the congruence, in the flat spacetime context, is when those members move rigidly together, an effect referred to as HOS sharing in Sect. 2.3.4.

The first difference in a curved spacetime is that there is not even a natural choice for the hypersurfaces of simultaneity, at least over regions where the curvature is significant. In flat spacetime, it was natural enough to borrow a hyperplane of simultaneity from an instantaneously comoving inertial observer. But in a curved spacetime, there are no hyperplanes, only hypersurfaces, and we may expect there to be several of these. So which should we choose? The only obvious constraint is that the hypersurfaces be orthogonal to one of the worldlines in the congruence, but that leaves plenty of freedom over what the hypersurface should do elsewhere in the congruence.

In a curved spacetime, if we find a series of spacelike hypersurfaces intersecting each member of the congruence at the same proper time for all members, there is no obvious reason why those hypersurfaces should be orthogonal to the members of the congruence. Conversely, for a given timelike congruence, if we can find shared hypersurfaces of simultaneity, i.e., spacelike hypersurfaces that are orthogonal to all members of the congruence, they are unlikely to intersect the members of the congruence at equal values of their proper times.

Note that, whatever the curvature, spacetime will always look flat to within measurement accuracy over some small enough region. This is basically the import of the weak equivalence principle (WEP) mentioned at the very beginning of Chap. 3 (see p. 141). At any point p in the spacetime manifold $\mathscr{M}$ of general relativity, under the usual assumption that the covariant derivative of the metric is zero and the connection is torsion free, i.e., symmetric in its two lower indices when expressed relative to a coordinate frame, coordinates can be found on some neighbourhood of p such that the metric has the Minkowski form at p and the connection and first derivatives of the metric are zero at p. This mathematical facet of the weak equivalence principle is thus built into the standard construction of general relativity in the torsion-free case.

So given any event in spacetime, one can always find a locally inertial frame containing it, in which the spacetime looks flat to within measurement accuracy, and where the timelike congruence intersects this small region, we can always treat the spacetime as flat, e.g., borrow hyperplanes of simultaneity of instantaneously comoving inertial observers, where the notion of hyperplane is an approximation specified by the locally inertial coordinate system. We shall have more to say about this kind of approximation later.

In what follows we may assume that, for some preselected event on one worldline of the congruence, we can find a spacelike hypersurface σ through that event intersecting all the worldlines of the congruence orthogonally, then for each worldline γ, assign zero proper time for γ at the event where it intersects this hypersurface σ. The details of the strategy adopted here will not be immediately important in what follows. For example, we shall not worry for the moment about whether other hypersurfaces of simultaneity intersect any of the worldlines orthogonally. However, in Sect. 4.3.14, we shall find a simple condition that is mathematically equivalent to the existence of a family of 3D hypersurfaces everywhere orthogonal to the congruence.

The aim here will just be to construct a tetrad frame over the whole region occupied by the congruence in such a way that the tangent to the congruence is always one of the vectors in the frame. Note the difference with the discussion in Sect. 2.3. There we also had a timelike congruence, with each label ξ denoting a different worldline. However, we only constructed a tetrad along one timelike worldline, viz., $\xi = 0$, which we took to be the worldline of a pointlike observer, then used that to construct convenient coordinates (adapted to the observer $\xi = 0$). In the present case, the tetrad frame will occupy all events covered by the congruence.

Of course, the three spacelike members of the tetrad at each event are orthogonal to the worldline through that event, so in a very local sense (so local that it is pointwise), we are implementing a notion of simultaneity, and presumably the most natural one possible.

For any $p \in \gamma$ with γ in the congruence, let $X\big|_p = u$ and define projection operators h and π in T_pM by [12]

$$\pi(w) := -(u|w)u\ , \qquad h(w) = w - \pi(w) = w + (u|w)u\ , \quad \forall w \in \mathrm{T}_pM\ .$$

In component form, relative to a coordinate frame,

$$\pi(w)^\nu = \pi^\nu_\mu w^\mu\ , \qquad \pi^\nu_\mu = -u_\mu u^\nu\ ,$$

$$h(w)^\nu = h^\nu_\mu w^\mu\ , \qquad h^\nu_\mu = \delta^\nu_\mu + u_\mu u^\nu\ .$$

In this notation, h^ν_μ corresponds to the projection operator $P^\nu{}_\mu$ introduced in Sect. 2.3.1 and used elsewhere. We then express the tangent space at p as a direct sum of a 1D subspace spanned by u and a 3D subspace orthogonal to it under the metric:

$$\mathrm{T}_pM = \mathrm{T}_{\|p}M \oplus \mathrm{T}_{\perp p}M\ ,$$

in the obvious way.

4.3.1 Tetrad

We now construct the tetrad. For each p on some γ in the congruence, we choose three mutually orthogonal unit spacelike vectors $n_{\hat{a}} \in \mathrm{T}_{\perp p}M$, with $\hat{a} = 1,2,3$, so that

$$(n_{\hat{a}}|n_{\hat{b}})_p = \delta_{\hat{a}\hat{b}}\ .$$

The reader will understand shortly why we put hats on the new indices. These triads are chosen smoothly along curves, and from one curve to the next, but for the moment, no other constraint is put on them, e.g., parallel transport or Fermi–Walker transport along γ, etc. We now set $X =: n_{\hat{0}}$, so that the set $\{n_{\hat{\alpha}}\}$, $\hat{\alpha} = 0,1,2,3$, forms an orthonormal basis. This is the tetrad, satisfying

$$(n_{\hat{\alpha}}|n_{\hat{\beta}}) = \eta_{\hat{\alpha}\hat{\beta}}\ .$$

We now introduce the matrices transforming between a coordinate basis $\{\partial_\mu\}$ and the tetrad, viz.,

$$n_{\hat{\alpha}} = n^\mu_{\hat{\alpha}}\partial_\mu\ , \qquad \partial_\mu = n^{\hat{\beta}}_\mu n_{\hat{\beta}}\ . \tag{4.31}$$

The position of different indices at the top or bottom constitutes a sufficient notational difference to distinguish these two matrices, which are obviously mutual inverses:

$$n^\mu_{\hat{\alpha}} n^{\hat{\beta}}_\mu = \delta^{\hat{\beta}}_{\hat{\alpha}}\ , \qquad n^\mu_{\hat{\alpha}} n^{\hat{\alpha}}_\nu = \delta^\mu_\nu\ . \tag{4.32}$$

Orthonormality of the tetrad gives more relations. The first is

$$\eta_{\hat{\alpha}\hat{\beta}} = (n_{\hat{\alpha}}|n_{\hat{\beta}}) = n^{\mu}_{\hat{\alpha}} n^{\nu}_{\hat{\beta}} (\partial_\mu|\partial_\nu) = n^{\mu}_{\hat{\alpha}} n^{\nu}_{\hat{\beta}} g_{\mu\nu} \ . \tag{4.33}$$

Another is

$$g_{\mu\nu} = (\partial_\mu|\partial_\nu) = \eta_{\hat{\alpha}\hat{\beta}} n^{\hat{\alpha}}_{\mu} n^{\hat{\beta}}_{\nu} \ . \tag{4.34}$$

Of course, we also have the interesting

$$\boxed{g_{\mu\nu} n^{\nu}_{\hat{\alpha}} = (\partial_\mu|n_{\hat{\alpha}}) = \eta_{\hat{\alpha}\hat{\beta}} n^{\hat{\beta}}_{\mu}} \ , \tag{4.35}$$

and it is tempting to define

$$n_{\hat{\alpha}\mu} := \eta_{\hat{\alpha}\hat{\beta}} n^{\hat{\beta}}_{\mu} \ , \tag{4.36}$$

or

$$n_{\hat{\alpha}\mu} := g_{\mu\nu} n^{\nu}_{\hat{\alpha}} \ . \tag{4.37}$$

Fortunately, the right-hand sides of the last two definitions are the same, since we have (4.35).

This is an important point in the following manipulations. We raise and lower indices on the transformation matrices in (4.31) paying no regard to whether they appear left or right in this index formulation. (A proper matrix formulation would pay attention to this.) We can do this with impunity because any two objects which end up looking the same are in fact the same. For example,

$$n^{\hat{\alpha}}_{\mu} = \eta^{\hat{\alpha}\hat{\beta}} g_{\mu\nu} n^{\nu}_{\hat{\beta}} \ . \tag{4.38}$$

We really do get the inverse of $n^{\nu}_{\hat{\beta}}$ by raising and lowering its indices, as can be seen from (4.35).

4.3.2 Form Tetrad

We can also define a form tetrad $\{\mathbf{n}^{\hat{\alpha}}\}$ by

$$\mathbf{n}^{\hat{\alpha}} := n^{\hat{\alpha}}_{\mu} \mathrm{d}x^{\mu} \ . \tag{4.39}$$

We have arranged this to be the dual basis to $\{n_{\hat{\alpha}}\}$, since

$$\mathbf{n}^{\hat{\alpha}}(n_{\hat{\beta}}) = n^{\hat{\alpha}}_{\mu} \mathrm{d}x^{\mu}(n_{\hat{\beta}}) = n^{\hat{\alpha}}_{\mu} n^{\mu}_{\hat{\beta}} = \delta^{\hat{\alpha}}_{\hat{\beta}} \ .$$

Now we really do have to be careful. The problem is not that the new objects, i.e., the 1-forms in (4.39), have their own components, viz., $\mathbf{n}^{\hat{\alpha}}_{\mu}$, because these numbers are the same as $n^{\hat{\alpha}}_{\mu}$, from (4.39). The problem is that we can define forms by

$$\mathbf{n}_{\hat{\alpha}} = \eta_{\hat{\alpha}\hat{\beta}} \mathbf{n}^{\hat{\beta}} \ , \tag{4.40}$$

and these are obviously not the same objects as $n_{\hat{\alpha}}$, because they are forms, whereas the latter are contravectors. Note, however, that

$$\mathbf{n}_{\hat{\alpha}\mu} = \eta_{\hat{\alpha}\hat{\beta}}\mathbf{n}^{\hat{\beta}}_{\mu} = \eta_{\hat{\alpha}\hat{\beta}}n^{\hat{\beta}}_{\mu} = n_{\hat{\alpha}\mu} \ .$$

We have to be particularly careful when taking covariant derivatives of these objects, as we shall see later.

4.3.3 Tetrad and Projection Operators

There are some relations between the tetrad and the projection operators h_μ^ν and π_μ^ν. In fact,

$$\boxed{n^{\hat{a}}_{\mu}n^{\nu}_{\hat{a}} = h^{\nu}_{\mu} \ , \qquad n^{\hat{0}}_{\mu}n^{\nu}_{\hat{0}} = \pi^{\nu}_{\mu}} \ , \tag{4.41}$$

where it is understood that $\hat{a}$ only runs over space indices $\{1,2,3\}$. We now see the utility of the hat notation. Although we already distinguish arbitrary tetrad indices from arbitrary coordinate indices by using letters in the first part of the alphabet for the former and letters in the middle of the alphabet for the latter, there are times when a specific index appears, in this case $\hat{0}$, and we could soon become confused. There are other examples below.

4.3.4 Tetrad and Determinant of the Metric

Taking the determinant of both sides of (4.34), we have

$$\sqrt{-g} = \det n^{\hat{\alpha}}_{\mu} \ . \tag{4.42}$$

4.3.5 Tetrad Components of a Tensor

Any tensor $T^{\mu\nu\dots}_{\rho\sigma\dots}$ has components relative to the tetrad frame given by

$$T^{\hat{\alpha}\hat{\beta}\dots}_{\hat{\delta}\hat{\gamma}\dots} = T^{\mu\nu\dots}_{\rho\sigma\dots}n^{\hat{\alpha}}_{\mu}n^{\hat{\beta}}_{\nu}\dots n^{\rho}_{\hat{\delta}}n^{\sigma}_{\hat{\gamma}}\dots \ .$$

This is indeed how tensor components change under change of frame.

4.3.6 Some Other Relations

For any contravector v, we have

$$v_{\perp}^{\mu} = h_{\nu}^{\mu} v^{\nu} = n_{\hat{a}}^{\mu} n_{\nu}^{\hat{a}} v^{\nu} = n_{\hat{a}}^{\mu} v^{\hat{a}} \,, \tag{4.43}$$

by the first relation of (4.41). By the second relation of (4.41),

$$v_{\parallel}^{\mu} = \pi_{\nu}^{\mu} v^{\nu} = n_{\hat{0}}^{\mu} n_{\nu}^{\hat{0}} v^{\nu} = n_{\hat{0}}^{\mu} v^{\hat{0}} \,. \tag{4.44}$$

Here is another situation where the hat on the index 0 is crucial, since v has a zero coordinate component, too, denoted by v^0.

By (4.43), and since h_{μ}^{ν} is a projection operator, we have

$$h_{\mu\nu} v^{\mu} v^{\nu} = (v_{\perp}|v_{\perp}) = g_{\mu\nu} n_{\hat{a}}^{\mu} n_{\hat{b}}^{\nu} v^{\hat{a}} v^{\hat{b}} = \eta_{\hat{a}\hat{b}} v^{\hat{a}} v^{\hat{b}} = \delta_{\hat{a}\hat{b}} v^{\hat{a}} v^{\hat{b}} \,, \tag{4.45}$$

recalling that $\eta_{\hat{a}\hat{b}} = \delta_{\hat{a}\hat{b}}$ because $\hat{a}$, $\hat{b} \in \{1,2,3\}$. The last result in this sequence is, from (4.44),

$$\pi_{\mu\nu} v^{\mu} v^{\nu} = (v_{\parallel}|v_{\parallel}) = g_{\mu\nu} n_{\hat{0}}^{\mu} v^{\hat{0}} n_{\hat{0}}^{\nu} v^{\hat{0}} = -(v^{\hat{0}})^2 \,. \tag{4.46}$$

4.3.7 Tetrad as Coordinate Frame

The tetrad frame could not generally be a coordinate frame, i.e., one cannot generally find coordinates $\hat{x}^{\hat{\alpha}}$ such that

$$n_{\hat{\alpha}} = \frac{\partial}{\partial \hat{x}^{\hat{\alpha}}} \,, \qquad \hat{\alpha} = 0,1,2,3 \,. \tag{4.47}$$

The degree of non-integrability of these equations is described by the structure constants $C^{\hat{\alpha}}_{\hat{\beta}\hat{\gamma}}$ given by [12]

$$\left[n_{\hat{\beta}}, n_{\hat{\gamma}}\right] = C^{\hat{\alpha}}_{\hat{\beta}\hat{\gamma}} n_{\hat{\alpha}} \,. \tag{4.48}$$

The Frobenius theorem tells us that coordinates satisfying (4.47) exist if and only if $C^{\hat{\alpha}}_{\hat{\beta}\hat{\gamma}} = 0$ for all $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$. In terms of 1-forms, this means that there are coordinates $\hat{x}^{\hat{\alpha}}$ such that $\mathbf{n}^{\hat{\alpha}} = \mathrm{d}\hat{x}^{\hat{\alpha}}$ if and only if $\mathrm{d}\mathbf{n}^{\hat{\alpha}} = 0$ for all $\hat{\alpha}$, i.e., if and only if $\partial_{[\nu} n^{\hat{\alpha}}_{\mu]} = 0$ for all $\hat{\alpha}$, μ, and ν.

4.3.8 Propagation of Tetrads

We denote the absolute derivative of $n_{\hat{\alpha}}$ along a γ in the congruence by a dot. Recall that we only required the tetrad to vary smoothly along each γ (and from one γ to the next), so there is a priori no condition on the $\dot{n}_{\hat{\alpha}}$. However, since it is itself a vector field, it can be expressed in terms of the tetrad, thus determining sixteen coefficients $\Lambda_{\hat{\alpha}}{}^{\hat{\beta}}$ by

$$\dot{n}_{\hat{\alpha}} = \dot{\gamma}^{\mu}\nabla_{\mu} n_{\hat{\alpha}} = \Lambda_{\hat{\alpha}}{}^{\hat{\beta}} n_{\hat{\beta}} \ . \tag{4.49}$$

Of course, we can raise and lower the indices using the Minkowski metric η, so we also have

$$\dot{n}_{\hat{\alpha}} = \Lambda_{\hat{\alpha}\hat{\beta}} n^{\hat{\beta}} \ . \tag{4.50}$$

The left or right location of an index on Λ now matters. Indeed, $(\Lambda_{\hat{\alpha}\hat{\beta}})$ is an antisymmetric matrix. This is because the tetrad is always orthonormal by definition, whence

$$\begin{aligned}
0 &= \frac{\mathrm{D}}{\mathrm{D}\tau} g(n_{\hat{\alpha}}, n_{\hat{\beta}}) \\
&= g\left(\frac{\mathrm{D}n_{\hat{\alpha}}}{\mathrm{D}\tau}, n_{\hat{\beta}}\right) + g\left(n_{\hat{\alpha}}, \frac{\mathrm{D}n_{\hat{\beta}}}{\mathrm{D}\tau}\right) \\
&= g\big(\Lambda_{\hat{\alpha}\hat{\gamma}}\eta^{\hat{\gamma}\hat{\delta}} n_{\hat{\delta}}, n_{\hat{\beta}}\big) + g\big(n_{\hat{\alpha}}, \Lambda_{\hat{\beta}\hat{\varepsilon}}\eta^{\hat{\varepsilon}\hat{\phi}} n_{\hat{\phi}}\big) \\
&= \Lambda_{\hat{\alpha}\hat{\gamma}}\eta^{\hat{\gamma}\hat{\delta}} g(n_{\hat{\delta}}, n_{\hat{\beta}}) + \Lambda_{\hat{\beta}\hat{\varepsilon}}\eta^{\hat{\varepsilon}\hat{\phi}} g(n_{\hat{\alpha}}, n_{\hat{\phi}}) \\
&= \Lambda_{\hat{\alpha}\hat{\gamma}}\eta^{\hat{\gamma}\hat{\delta}} \eta_{\hat{\delta}\hat{\beta}} + \Lambda_{\hat{\beta}\hat{\varepsilon}}\eta^{\hat{\varepsilon}\hat{\phi}} \eta_{\hat{\alpha}\hat{\phi}} \\
&= \Lambda_{\hat{\alpha}\hat{\beta}} + \Lambda_{\hat{\beta}\hat{\alpha}} \ .
\end{aligned}$$

We now have

$$\dot{X} = \dot{n}_0 = \Lambda_{\hat{0}\hat{b}} n^{\hat{b}} \ , \qquad \dot{n}_{\hat{a}} = \Lambda_{\hat{a}\hat{b}} n^{\hat{b}} - \Lambda_{\hat{a}\hat{0}} X \ . \tag{4.51}$$

The first of these tells us that

$$\Lambda_{\hat{0}\hat{a}} = \dot{X}_{\hat{a}} := \eta_{\hat{a}\hat{b}} \dot{X}^{\hat{b}} \ , \tag{4.52}$$

where $\dot{X}^{\hat{b}}$ is the $\hat{b}$ component of the four-acceleration along the chosen curve γ of the congruence. Note that this four-acceleration has no component along the curve itself, i.e., $\dot{X}^{\hat{0}} = 0$, because we know that $(\dot{X}|X) = 0$. It is also clear from the second equation of (4.52) that

$$\Lambda_{\hat{a}\hat{b}} = (\dot{n}_{\hat{a}}|n_{\hat{b}}) \ . \tag{4.53}$$

Comparing with Sect. 2.3, we see that $n_{\hat{a}}$ corresponds to n_i, with $i \leftrightarrow \hat{a}$, and the antisymmetric 3×3 matrix $\Lambda_{\hat{a}\hat{b}}$ corresponds to Ω_{ij}. Here we have merely generalised to arbitrary spacetimes, although it is clear that the presentation, and in particular the notation, in Sect. 2.3 generalise in a straightforward way.

4.3.9 Fermi Rotation Coefficients

The antisymmetric 3×3 matrix $\Lambda_{\hat{a}\hat{b}}$ generates an instantaneous rotation. At the given instant, all the triad vectors $\{n_{\hat{a}}\}$ are rotating together about some axis which can be determined from the matrix. $\Lambda_{\hat{a}\hat{b}}$ are called the Fermi rotation coefficients. There is a coordinate component version:

$$\begin{aligned}\Lambda_{[\mu\nu]} &= \Lambda_{\hat{a}\hat{b}} n^{\hat{a}}_{\mu} n^{\hat{b}}_{\nu} \\ &= (\dot{n}_{\hat{a}}|n_{\hat{b}}) n^{\hat{a}}_{\mu} n^{\hat{b}}_{\nu} \\ &= g\big(\dot{n}_{\hat{a}}, n_{\hat{b}}\big) n^{\hat{a}}_{\mu} n^{\hat{b}}_{\nu} \\ &= g_{\rho\tau} X^{\sigma} n^{\rho}_{\hat{a};\sigma} n^{\tau}_{\hat{b}} n^{\hat{b}}_{\nu} n^{\hat{a}}_{\mu} \\ &= X^{\sigma} n_{\hat{a}\rho;\sigma} h^{\rho}_{\nu} n^{\hat{a}}_{\mu} \ .\end{aligned} \tag{4.54}$$

When all the coefficients $\Lambda_{\hat{a}\hat{b}}$ are zero along a given γ, the tetrad is said to be Fermi–Walker transported. In a sense there is no rotation of the tetrad frame as it moves along γ. In fact this lack of rotation is only a lack of rotation relative to the frame itself! A rotation may be happening relative to the manifold as a whole, as can be seen by the Thomas precession, even in a Minkowski spacetime (see Sects. 2.3.3 and 2.11.3).

When the tetrad is indeed FW transported along each worldline in the congruence, the second equation of (4.51) together with (4.52) gives

$$\dot{n}_{\hat{a}} = \dot{X}_{\hat{a}} X \ . \tag{4.55}$$

If a vector w has constant components in an FW transported frame, then by (4.55),

$$\dot{w} = w^{\hat{a}} \dot{n}_{\hat{a}} + w^{\hat{0}} \dot{X} = (w|\dot{X})X - (w|X)\dot{X} \ , \tag{4.56}$$

whence

$$\boxed{\dot{w} = w^{\mu}\big(\dot{X}_{\mu} X - X_{\mu}\dot{X}\big)} \ . \tag{4.57}$$

Such a vector is itself said to be FW transported. This is the general relativistic generalisation of (2.28) on p. 26, also discussed in Sect. 3.2.4. As we have seen, one vector that is always FW transported is the tangent to the curve itself, viz., X. It obviously satisfies the last equation. Furthermore, this kind of transport coincides with parallel transport when the curve is a geodesic, i.e., when $\dot{X} = 0$.

The other important observation made earlier is that FW transport preserves scalar products: if u and v are two vector fields that are FW transported along the curve γ, then $g(u,v)$ is constant along γ. Hence, orthogonality to the curve is preserved: if u is FW transported along γ and orthogonal to it at some point, then since $\dot{\gamma}$ is always FW transported along γ, we have $g(u,\dot{\gamma}) = 0$ all along the curve.

4.3.10 Ricci Rotation Coefficients

We consider the covariant derivatives of the dual basis $\mathbf{n}^{\hat{\delta}}$ along the tetrad vectors $n_{\hat{\alpha}}$, viz.,

$$(\nabla_{\hat{\alpha}}\mathbf{n}^{\hat{\gamma}})_{\hat{\beta}} = -\Gamma^{\hat{\gamma}}_{\hat{\beta}\hat{\alpha}} , \tag{4.58}$$

where $\nabla_{\hat{\alpha}}$ means $\nabla_{n_{\hat{\alpha}}}$. This follows from the general rule

$$(\nabla_{\hat{\alpha}}\phi)_{\hat{\beta}} = n_{\hat{\alpha}}(\phi_{\hat{\beta}}) - \Gamma^{\hat{\gamma}}_{\hat{\beta}\hat{\alpha}}\phi_{\hat{\gamma}} , \tag{4.59}$$

for any 1-form ϕ [see (4.17) on p. 159], and the fact that, when $\phi = \mathbf{n}^{\hat{\delta}}$,

$$\phi_{\hat{\beta}} = \mathbf{n}^{\hat{\delta}}(n_{\hat{\beta}}) = \delta^{\hat{\delta}}_{\hat{\beta}} .$$

We can lower the index to obtain

$$\Gamma_{\hat{\gamma}\hat{\beta}\hat{\alpha}} = \eta_{\hat{\gamma}\hat{\delta}}\Gamma^{\hat{\delta}}_{\hat{\beta}\hat{\alpha}} = -(\nabla_{\hat{\alpha}}\mathbf{n}_{\hat{\gamma}})_{\hat{\beta}} . \tag{4.60}$$

These are the Ricci rotation coefficients. They are just the connection coefficients relative to the tetrad frame.

At first sight it seems important to keep $\mathbf{n}_{\hat{\gamma}}$ in the last expression, as opposed to $n_{\hat{\gamma}}$, because that would change the covariant derivative. In fact, it appears to make no sense to replace $\mathbf{n}_{\hat{\gamma}}$ by $n_{\hat{\gamma}}$ in (4.60) since the index $\hat{\beta}$ is lowered. However, since

$$(\nabla_{\hat{\alpha}}u)^{\hat{\beta}} = n_{\hat{\alpha}}(u^{\hat{\beta}}) + n^{\hat{\varepsilon}}_{\hat{\alpha}}\Gamma^{\hat{\beta}}_{\hat{\delta}\hat{\varepsilon}}u^{\hat{\delta}} ,$$

for any contravector u, according to (4.16) on p. 159, we also have the result

$$(\nabla_{\hat{\alpha}}n_{\hat{\gamma}})^{\hat{\beta}} = n_{\hat{\alpha}}(\delta^{\hat{\beta}}_{\hat{\gamma}}) + \delta^{\hat{\varepsilon}}_{\hat{\alpha}}\Gamma^{\hat{\beta}}_{\hat{\delta}\hat{\varepsilon}}\delta^{\hat{\delta}}_{\hat{\gamma}} = \Gamma^{\hat{\beta}}_{\hat{\gamma}\hat{\alpha}} . \tag{4.61}$$

We can now lower the index $\hat{\beta}$ to obtain

$$\Gamma_{\hat{\delta}\hat{\gamma}\hat{\alpha}} = \eta_{\hat{\delta}\hat{\beta}}\Gamma^{\hat{\beta}}_{\hat{\gamma}\hat{\alpha}} = (\nabla_{\hat{\alpha}}n_{\hat{\gamma}})_{\hat{\delta}} , \tag{4.62}$$

or for direct comparison with (4.60),

$$\Gamma_{\hat{\gamma}\hat{\beta}\hat{\alpha}} = (\nabla_{\hat{\alpha}}n_{\hat{\beta}})_{\hat{\gamma}} . \tag{4.63}$$

There are two differences: the opposite sign, and two indices transposed. It may seem confusing to have a vector with a lowered index, but we do this all the time with the coordinate metric components, and no problems arise, precisely because, for the metric connection, covariant differentiation commutes with raising and lowering indices.

In fact, (4.60) and (4.63) together effectively constitute a proof that the Ricci rotation coefficients are antisymmetric in the first two indices, viz.,

$$\Gamma_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = \Gamma_{[\hat{\alpha}\hat{\beta}]\hat{\gamma}}\,, \tag{4.64}$$

which is in fact the case. It can also be proved as follows. Since the tetrad is orthonormal everywhere, the covariant derivative of the constant $\eta_{\hat{\alpha}\hat{\beta}} = (n_{\hat{\alpha}}|n_{\hat{\beta}})$ along any $n_{\hat{\gamma}}$ is zero, i.e.,

$$\begin{aligned} 0 = \nabla_{n_{\hat{\gamma}}}(n_{\hat{\alpha}}|n_{\hat{\beta}}) &= (\nabla_{n_{\hat{\gamma}}} n_{\hat{\alpha}}|n_{\hat{\beta}}) + (n_{\hat{\alpha}}|\nabla_{n_{\hat{\gamma}}} n_{\hat{\beta}}) \\ &= \Gamma_{\hat{\beta}\hat{\alpha}\hat{\gamma}} + \Gamma_{\hat{\alpha}\hat{\beta}\hat{\gamma}}\,. \end{aligned}$$

We have taken the covariant derivative past the metric brackets with impunity. In fact, what the last calculation amounts to is this:

$$\begin{aligned} 0 &= (\nabla_{\hat{\gamma}}\eta)_{\hat{\alpha}\hat{\beta}} \\ &= n_{\hat{\gamma}}(\eta_{\hat{\alpha}\hat{\beta}}) - \Gamma^{\hat{\delta}}_{\hat{\alpha}\hat{\gamma}}\eta_{\hat{\delta}\hat{\beta}} - \Gamma^{\hat{\delta}}_{\hat{\beta}\hat{\gamma}}\eta_{\hat{\alpha}\hat{\delta}} \\ &= -\Gamma_{\hat{\beta}\hat{\alpha}\hat{\gamma}} - \Gamma_{\hat{\alpha}\hat{\beta}\hat{\gamma}}\,. \end{aligned}$$

Finally, note that the coordinate form of the relations we are dealing with here is

$$\Gamma_{\hat{\gamma}\hat{\beta}\hat{\alpha}} = n_{\hat{\beta}\nu;\mu} n^{\nu}_{\hat{\gamma}} n^{\mu}_{\hat{\alpha}} = -n_{\hat{\gamma}\nu;\mu} n^{\nu}_{\hat{\beta}} n^{\mu}_{\hat{\alpha}}\,, \tag{4.65}$$

which are immediate from (4.60) and (4.63), respectively.

4.3.11 Torsion-Free Connection

In (4.24) on p. 159, we obtained the components of the torsion tensor with respect to any frame as

$$T^{\beta}{}_{\alpha\gamma} = 2\Gamma^{\beta}_{[\alpha\gamma]} + C^{\beta}{}_{\alpha\gamma}\,. \tag{4.66}$$

When the connection is torsion-free, this means that

$$2\Gamma^{\hat{\beta}}_{[\hat{\alpha}\hat{\gamma}]} = C^{\hat{\beta}}{}_{\hat{\gamma}\hat{\alpha}}\,, \tag{4.67}$$

in the present orthonormal frame.

In (4.29) on p. 160, it was shown that, in the torsion-free case, we have for any frame $\{e_{\alpha}\}$,

$$\Gamma^{\gamma}_{\alpha\beta} = \frac{1}{2}g^{\gamma\delta}\Big[e_{\beta}(g_{\alpha\delta}) + e_{\alpha}(g_{\delta\beta}) - e_{\delta}(g_{\alpha\beta})\Big] + \frac{1}{2}\Big[C^{\gamma}{}_{\beta\alpha} + g^{\gamma\delta}g_{\alpha\varepsilon}C^{\varepsilon}{}_{\delta\beta} + g^{\gamma\delta}g_{\varepsilon\beta}C^{\varepsilon}{}_{\delta\alpha}\Big] . \tag{4.68}$$

Applying to the present orthonormal frame, where the metric coefficients are constants, and lowering indices, we have

$$\Gamma_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = \frac{1}{2}(C_{\hat{\alpha}\hat{\gamma}\hat{\beta}} + C_{\hat{\beta}\hat{\alpha}\hat{\gamma}} + C_{\hat{\gamma}\hat{\alpha}\hat{\beta}}) . \tag{4.69}$$

Note that we can deduce the antisymmetry of $\Gamma_{\hat{\alpha}\hat{\beta}\hat{\gamma}}$ in $\hat{\alpha}$ and $\hat{\beta}$ from this, using the antisymmetry of $C_{\hat{\alpha}\hat{\beta}\hat{\gamma}}$ in its last two indices.

4.3.12 Fermi Rotation Coefficients as Ricci Rotation Coefficients

In fact, we have the result

$$\boxed{\Lambda_{\hat{\alpha}\hat{\beta}} = -\Gamma_{\hat{\alpha}\hat{\beta}\hat{0}}} . \tag{4.70}$$

This is proved as follows. We have

$$\begin{aligned}
\Gamma_{\hat{\alpha}\hat{\beta}\hat{0}} &= (\nabla_X n_{\hat{\beta}})_{\hat{\alpha}} \qquad \text{[by (4.63)]} \\
&= (\nabla_X n_{\hat{\beta}}|n_{\hat{\alpha}}) \\
&= (\dot{n}_{\hat{\beta}}|n_{\hat{\alpha}}) \\
&= (\Lambda_{\hat{\beta}\hat{\gamma}}n^{\hat{\gamma}}|n_{\hat{\alpha}}) \qquad \text{[by (4.50) on p. 167]} \\
&= \Lambda_{\hat{\beta}\hat{\gamma}}\delta^{\hat{\gamma}}_{\hat{\alpha}} = \Lambda_{\hat{\beta}\hat{\alpha}} ,
\end{aligned} \tag{4.71}$$

as claimed.

4.3.13 Expansion and Vorticity

We now consider the congruence of timelike curves as the flow of a material. Following the account in [12], we may define the expansion by

$$\boxed{\Theta_{\hat{a}\hat{b}} := -\Gamma_{\hat{0}(\hat{a}\hat{b})} = -\frac{1}{2}\big(\Gamma_{\hat{0}\hat{a}\hat{b}} + \Gamma_{\hat{0}\hat{b}\hat{a}}\big) , \quad a,b \in \{1,2,3\}} , \tag{4.72}$$

where round brackets among indices indicate symmetrization. This is not a very elegant way of defining it. Things looks nicer if we refer to arbitrary coordinates $\{x^{\mu}\}$ and make the definition

$$\boxed{\Theta_{\hat{a}\hat{b}} := \Theta_{\mu\nu} n_{\hat{a}}^{\mu} n_{\hat{b}}^{\nu} \,, \quad a,b \in \{1,2,3\}} \,, \tag{4.73}$$

where

$$\boxed{\Theta_{\mu\nu} := X_{(\rho;\sigma)} h_{\mu}^{\rho} h_{\nu}^{\sigma} = \frac{1}{2}\big(X_{\rho;\sigma} + X_{\sigma;\rho}\big) h_{\mu}^{\rho} h_{\nu}^{\sigma}} \,. \tag{4.74}$$

This is in fact the rate of strain tensor for a continuous medium whose constituent particles or fluid elements follow the worldlines in the congruence, as we shall see in Sect. 4.3.16.

Let us show the equivalence of the last definition with (4.72). First a technical detail. If we take the μ component of $\mathbf{n}_{\hat{0}}$, we obtain $n_{\hat{0}\mu} = \eta_{\hat{0}\hat{\alpha}} n_{\mu}^{\hat{\alpha}}$. We do have to check this from the definition (4.39) on p. 164, viz.,

$$\mathbf{n}^{\hat{\alpha}} := n_{\mu}^{\hat{\alpha}} \mathrm{d}x^{\mu} \,,$$

and the definition

$$\mathbf{n}_{\hat{\beta}} := \eta_{\hat{\beta}\hat{\alpha}} \mathbf{n}^{\hat{\alpha}} = \eta_{\hat{\beta}\hat{\alpha}} n_{\mu}^{\hat{\alpha}} \mathrm{d}x^{\mu} \,.$$

We also have $n_{\hat{0}\mu} = X_{\mu} := g_{\mu\nu} X^{\nu}$, because $X := g_{\mu\nu} X^{\nu} \mathrm{d}x^{\mu}$ has coordinate components $g_{\mu\nu} n_{\hat{0}}^{\nu}$, which are equal to $\eta_{\hat{0}\hat{\alpha}} n_{\mu}^{\hat{\alpha}}$. Indeed, in general, by (4.35) on p. 164,

$$\eta_{\hat{\alpha}\hat{\beta}} n_{\mu}^{\hat{\beta}} = g_{\mu\nu} n_{\hat{\alpha}}^{\nu} \,.$$

To sum up,

$$(\mathbf{n}_{\hat{0}})_{\mu} = n_{\hat{0}\mu} = X_{\mu} \,. \tag{4.75}$$

Now by the first relation of (4.41) on p. 165,

$$h_{\mu}^{\rho} n_{\hat{a}}^{\mu} = n_{\mu}^{\hat{b}} n_{\hat{b}}^{\rho} n_{\hat{a}}^{\mu} = n_{\hat{a}}^{\rho} \,, \tag{4.76}$$

so the definitions (4.73) and (4.74) imply that

$$\Theta_{\hat{a}\hat{b}} = \frac{1}{2}\big(X_{\rho;\sigma} + X_{\sigma;\rho}\big) h_{\mu}^{\rho} h_{\nu}^{\sigma} n_{\hat{a}}^{\mu} n_{\hat{b}}^{\nu} = \frac{1}{2}\big(X_{\rho;\sigma} + X_{\sigma;\rho}\big) n_{\hat{a}}^{\rho} n_{\hat{b}}^{\sigma} \,. \tag{4.77}$$

But

$$\begin{aligned}
-\Gamma_{\hat{0}(\hat{a}\hat{b})} &= -\frac{1}{2}\big(\Gamma_{\hat{0}\hat{a}\hat{b}} + \Gamma_{\hat{0}\hat{b}\hat{a}}\big) \\
&= \frac{1}{2}\Big[(\nabla_{\hat{b}} \mathbf{n}_{\hat{0}})_{\hat{a}} + (\nabla_{\hat{a}} \mathbf{n}_{\hat{0}})_{\hat{b}}\Big] \qquad \text{[by (4.60) on p. 169]} \\
&= \frac{1}{2}\Big[n_{\hat{a}}^{\mu} n_{\hat{b}}^{\nu} (n_{\hat{0}\mu,\nu} - \Gamma_{\nu\mu}^{\rho} n_{\hat{0}\rho}) + n_{\hat{b}}^{\mu} n_{\hat{a}}^{\nu} (n_{\hat{0}\mu,\nu} - \Gamma_{\nu\mu}^{\rho} n_{\hat{0}\rho})\Big] \\
&= \frac{1}{2} n_{\hat{a}}^{\mu} n_{\hat{b}}^{\nu} \big(n_{\hat{0}\mu,\nu} + n_{\hat{0}\nu,\mu} - \Gamma_{\nu\mu}^{\rho} n_{\hat{0}\rho} - \Gamma_{\mu\nu}^{\rho} n_{\hat{0}\rho}\big) \\
&= \frac{1}{2} n_{\hat{a}}^{\mu} n_{\hat{b}}^{\mu} \big(X_{\mu;\nu} + X_{\nu;\mu}\big) \,,
\end{aligned}$$

as required. We do not need to assume a torsion-free connection here.

It can also be shown in the torsion-free case that

$$\boxed{\Theta_{\hat{a}\hat{b}} = -n_{\mu(\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b})}}\ . \tag{4.78}$$

To prove this, we start with the observation that

$$\mathscr{L}_X n_{\hat{b}} = \nabla_X n_{\hat{b}} - \nabla_{\hat{b}} X\ , \tag{4.79}$$

which follows from (4.22) on p. 159 in the torsion-free case. Note that we are not assuming that the $n_{\hat{\beta}}$ are parallel transported along the curves of the congruence, so in general we have $\nabla_X n_{\hat{b}} \neq 0$. Of course, this can be written

$$\mathscr{L}_X n_{\hat{b}} = \dot{n}_{\hat{b}} - \nabla_{\hat{b}} X\ , \tag{4.80}$$

or again

$$\mathscr{L}_X n_{\hat{b}} = \Lambda_{\hat{b}\hat{\gamma}} n^{\hat{\gamma}} - \nabla_{\hat{b}} X\ , \tag{4.81}$$

by the definition (4.50) of $\Lambda_{\hat{b}\hat{\gamma}}$ on p. 167. We now have

$$\begin{aligned}
-n_{\mu(\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b})} &= -\frac{1}{2}\Big[n_{\mu\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b}} + n_{\mu\hat{b}}\mathscr{L}_X n^{\mu}_{\hat{a}}\Big] \\
&= +\frac{1}{2}\Big[n^{\mu}_{\hat{a}}(\nabla_{\hat{b}}X)_{\mu} + n^{\mu}_{\hat{b}}(\nabla_{\hat{a}}X)_{\mu}\Big] \\
&= \frac{1}{2}\Big[n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}(X_{\mu,\nu} - \Gamma^{\rho}_{\nu\mu}X_{\rho}) + n^{\mu}_{\hat{b}} n^{\nu}_{\hat{a}}(X_{\mu,\nu} - \Gamma^{\rho}_{\nu\mu}X_{\rho})\Big]\ ,
\end{aligned}$$

from which we retrieve a previous expression for $\Theta_{\hat{a}\hat{b}}$. It should be noted here that $\mathscr{L}_X n^{\mu}_{\hat{b}}$ means $(\mathscr{L}_X n_{\hat{b}})^{\mu}$. In the second step, the terms in $\Lambda_{\hat{b}\hat{\gamma}}$ in (4.81) drop out due to their antisymmetry.

We now turn to the vorticity, defined by

$$\boxed{\omega_{\hat{a}\hat{b}} := -\Gamma_{\hat{0}[\hat{a}\hat{b}]} = -\frac{1}{2}\big(\Gamma_{\hat{0}\hat{a}\hat{b}} - \Gamma_{\hat{0}\hat{b}\hat{a}}\big)\ , \quad a,b \in \{1,2,3\}}\ . \tag{4.82}$$

As before, this is not a very elegant way of defining it. It looks nicer if we make the definition

$$\boxed{\omega_{\hat{a}\hat{b}} := \omega_{\mu\nu} n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}\ , \quad a,b \in \{1,2,3\}}\ , \tag{4.83}$$

where

$$\boxed{\omega_{\mu\nu} := X_{[\rho;\sigma]} h^{\rho}_{\mu} h^{\sigma}_{\nu} = \frac{1}{2}\big(X_{\rho;\sigma} - X_{\sigma;\rho}\big) h^{\rho}_{\mu} h^{\sigma}_{\nu}}\ . \tag{4.84}$$

These definitions give

$$\omega_{\hat{a}\hat{b}} = \frac{1}{2}\big(X_{\rho;\sigma} - X_{\sigma;\rho}\big) h^{\rho}_{\mu} h^{\sigma}_{\nu} n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}} = \frac{1}{2}\big(X_{\rho;\sigma} - X_{\sigma;\rho}\big) n^{\rho}_{\hat{a}} n^{\sigma}_{\hat{b}}\ . \tag{4.85}$$

But

$$\begin{aligned}
-\Gamma_{\hat{0}[\hat{a}\hat{b}]} &= -\frac{1}{2}\left(\Gamma_{\hat{0}\hat{a}\hat{b}} - \Gamma_{\hat{0}\hat{b}\hat{a}}\right) \\
&= \frac{1}{2}\left[(\nabla_{\hat{b}}\mathbf{n}_{\hat{0}})_{\hat{a}} - (\nabla_{\hat{a}}\mathbf{n}_{\hat{0}})_{\hat{b}}\right] \\
&= \frac{1}{2}\left[n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}(n_{\hat{0}\mu,\nu} - \Gamma^{\rho}_{\nu\mu} n_{\hat{0}\rho}) - n^{\mu}_{\hat{b}} n^{\nu}_{\hat{a}}(n_{\hat{0}\mu,\nu} - \Gamma^{\rho}_{\nu\mu} n_{\hat{0}\rho})\right] \\
&= \frac{1}{2} n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}\left(n_{\hat{0}\mu;\nu} - n_{\hat{0}\nu;\mu}\right) \\
&= \frac{1}{2} n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}\left(X_{\mu;\nu} - X_{\nu;\mu}\right) ,
\end{aligned}$$

as required to show equivalence with (4.82). Once again, we do not need to assume a torsion-free connection to get this.

Note that, immediately from the definitions (4.72) and (4.82), we have

$$\boxed{\Theta_{\hat{a}\hat{b}} + \omega_{\hat{a}\hat{b}} = -\Gamma_{\hat{0}\hat{a}\hat{b}} , \qquad \Theta_{\hat{a}\hat{b}} - \omega_{\hat{a}\hat{b}} = -\Gamma_{\hat{0}\hat{b}\hat{a}}} . \tag{4.86}$$

There is another expression for $\omega_{\hat{a}\hat{b}}$ in the torsion-free case, viz.,

$$\boxed{\omega_{\hat{a}\hat{b}} = -\Lambda_{\hat{a}\hat{b}} - n_{\mu[\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b}]}} . \tag{4.87}$$

Once again, we use (4.81), viz.,

$$\mathscr{L}_X n_{\hat{b}} = \Lambda_{\hat{b}\hat{\gamma}} n^{\hat{\gamma}} - \nabla_{\hat{b}} X , \tag{4.88}$$

which is true in the torsion-free case. This time the terms in $\Lambda_{\hat{b}\hat{\gamma}}$ do not cancel out. We have

$$\begin{aligned}
-n_{\mu[\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b}]} &= -\frac{1}{2}\left[n_{\mu\hat{a}}\mathscr{L}_X n^{\mu}_{\hat{b}} - n_{\mu\hat{b}}\mathscr{L}_X n^{\mu}_{\hat{a}}\right] \\
&= +\frac{1}{2}\left[n^{\mu}_{\hat{a}}(\nabla_{\hat{b}}X)_{\mu} - n^{\mu}_{\hat{b}}(\nabla_{\hat{a}}X)_{\mu}\right] - \frac{1}{2}\left(n^{\mu}_{\hat{a}}\Lambda_{\hat{b}\hat{\gamma}} n^{\hat{\gamma}}_{\mu} - n^{\mu}_{\hat{b}}\Lambda_{\hat{a}\hat{\gamma}} n^{\hat{\gamma}}_{\mu}\right) \\
&= \frac{1}{2}\left[n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}(X_{\mu,\nu} - \Gamma^{\rho}_{\nu\mu}X_{\rho}) - n^{\mu}_{\hat{b}} n^{\nu}_{\hat{a}}(X_{\mu,\nu} - \Gamma^{\rho}_{\nu\mu}X_{\rho})\right] - \frac{1}{2}\left(\Lambda_{\hat{b}\hat{a}} - \Lambda_{\hat{a}\hat{b}}\right) \\
&= \frac{1}{2} n^{\mu}_{\hat{a}} n^{\nu}_{\hat{b}}(X_{\mu;\nu} - X_{\mu;\nu}) + \Lambda_{\hat{a}\hat{b}} \\
&= \omega_{\hat{a}\hat{b}} + \Lambda_{\hat{a}\hat{b}} ,
\end{aligned}$$

as required.

There is yet another expression for the vorticity in the torsion-free case, namely

$$\boxed{\omega_{\hat{a}\hat{b}} = -\frac{1}{2} X_{\mu}(\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}})^{\mu}} . \tag{4.89}$$

Once again, for a torsion-free connection,

$$\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}} = \nabla_{n_{\hat{b}}} n_{\hat{a}} - \nabla_{n_{\hat{a}}} n_{\hat{b}} , \tag{4.90}$$

whence

$$\begin{aligned} -\frac{1}{2} X_\mu (\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}})^\mu &= -\frac{1}{2} n^\mu_{\hat{0}} (\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}})_\mu \\ &= -\frac{1}{2}\Big[(\nabla_{\hat{b}} n_{\hat{a}})_\mu - (\nabla_{\hat{a}} n_{\hat{b}})_\mu \Big] n^\mu_{\hat{0}} \\ &= -\frac{1}{2}\Big[(\nabla_{\hat{b}} n_{\hat{a}})_{\hat{0}} - (\nabla_{\hat{a}} n_{\hat{b}})_{\hat{0}} \Big] \\ &= -\frac{1}{2}\big(\Gamma_{\hat{0}\hat{a}\hat{b}} - \Gamma_{\hat{0}\hat{b}\hat{a}} \big) \quad \text{[by (4.63) on p. 169]} \\ &= \Gamma_{\hat{0}[\hat{b}\hat{a}]} = \omega_{\hat{a}\hat{b}} , \end{aligned}$$

as claimed.

4.3.14 Vorticity-Free Congruence

Equation (4.89) can also be written

$$\omega_{\hat{a}\hat{b}} = -\frac{1}{2} (\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}})^{\hat{0}} . \tag{4.91}$$

Furthermore, we have structure constants $C^{\hat{\gamma}}{}_{\hat{b}\hat{a}}$ such that

$$\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}} = \big[n_{\hat{b}}, n_{\hat{a}} \big] = C^{\hat{\gamma}}{}_{\hat{b}\hat{a}} n_{\hat{\gamma}} , \tag{4.92}$$

whence

$$\boxed{\omega_{\hat{a}\hat{b}} = -\frac{1}{2} C^{\hat{0}}{}_{\hat{b}\hat{a}} = \frac{1}{2} C^{\hat{0}}{}_{\hat{a}\hat{b}}} . \tag{4.93}$$

Let us consider what happens if the vorticity is zero. It is clear from the last relation that

$$\omega_{\hat{a}\hat{b}} = 0 \quad \Longleftrightarrow \quad C^{\hat{0}}_{\hat{a}\hat{b}} = 0 . \tag{4.94}$$

If all these structure constants are zero, we can apply the Frobenius theorem mentioned on p. 158 and presented in [12]. According to this theorem, there are coordinates $\{x^\mu\}$ such that $n^0_{\hat{a}} = 0$ for $\hat{a} = 1,2,3$. This means that the 3D hypersurfaces $x^0 =$ constant are orthogonal to all the curves in the congruence.

How do we prove this? We have to show that, at any point, the tangent $X = n_{\hat{0}}$ to the congruence there is orthogonal to any vector tangent to the hypersurface. But the vectors $\{\partial_1, \partial_2, \partial_3\}$ span the tangent space to the hypersurface $x^0 =$ constant, and each of the linearly independent vectors $n_{\hat{1}}$, $n_{\hat{2}}$, and $n_{\hat{3}}$ can be expressed as a linear combination of ∂_1, ∂_2, and ∂_3 (without the need for ∂_0), so the three vectors $n_{\hat{1}}$, $n_{\hat{2}}$, and $n_{\hat{3}}$ span the space of vectors tangent to the hypersurface at the point in question. But these are orthogonal to $X = n_{\hat{0}}$ and so we are through.

Conversely, one can say that, if there is a family of 3D hypersurfaces orthogonal to the congruence, then the congruence has zero vorticity. Let us see how to prove that. First note that each hypersurface intersects each curve of the congruence at a single point. For if a short section of some curve lies in the given hypersurface, then the tangent to that curve cannot be orthogonal to all vectors tangent to the hypersurface, and indeed, the hypersurface cannot be entirely spacelike. This means that we can label the hypersurfaces smoothly by the parameter τ (e.g., proper time) of a chosen curve in the congruence.

Now choose coordinates on each hypersurface in such a way that they vary smoothly as we move along any curve in the congruence. Attribute to any point the coordinate $x^0 = \tau$ labelling the hypersurface it is in, and the values x^1, x^2, x^3 of the coordinates of the point in the given hypersurface. At any point, because of the orthogonality of the hypersurfaces to the congruence at that point, ∂_0 will be proportional to $n_{\hat{0}}$ and ∂_1, ∂_2, and ∂_3 will be linear combinations of $n_{\hat{1}}$, $n_{\hat{2}}$, and $n_{\hat{3}}$. This in turn implies that $n^0_{\hat{a}} = 0$ everywhere.

As always, we have

$$\left(\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}}\right)^{\mu} = \left[n_{\hat{b}}, n_{\hat{a}}\right]^{\mu} = n^{\nu}_{\hat{b}} n^{\mu}_{\hat{a},\nu} - n^{\nu}_{\hat{a}} n^{\mu}_{\hat{b},\nu} \, . \tag{4.95}$$

If we calculate $X_\mu \left(\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}}\right)^{\mu}$, then since X only has a zero component in the coordinate frame, being tangent to the congruence everywhere, what we find will be proportional to

$$\left[n_{\hat{b}}, n_{\hat{a}}\right]^{0} = n^{\nu}_{\hat{b}} n^{0}_{\hat{a},\nu} - n^{\nu}_{\hat{a}} n^{0}_{\hat{b},\nu} \, , \tag{4.96}$$

and the two terms on the right-hand side of (4.96) contain a derivative of $n^0_{\hat{a}}$ and a derivative of $n^0_{\hat{b}}$, respectively. But these two objects are zero everywhere and we conclude that $\omega_{\hat{a}\hat{b}} = -X_\mu (\mathscr{L}_{n_{\hat{b}}} n_{\hat{a}})^\mu / 2$ is also zero everywhere, as claimed.

Recall the labelling coordinates (ξ, τ) introduced in the case of flat spacetime in Sect. 2.3. That approach is interesting even in a general curved spacetime, when we have a congruence made up of timelike worldlines of real particles in a medium. Each particle carries a label ξ^i, $i = 1,2,3$, and its worldline is expressed as $x^\mu(\xi, \tau)$, $\mu = 0,1,2,3$, where τ is its proper time. The x^μ are arbitrary coordinates in spacetime. We can in principle invert this to obtain ξ^i and τ as functions of the x^μ in the domain of spacetime occupied by the medium. These coordinates $\{\tau, \xi^1, \xi^2, \xi^3\}$ look rather like the kind of coordinates envisaged above.

But what are the differences? To set this system up, we need to synchronise the proper times somewhere. Then 'later' can be taken as an attribution of some common, larger value of the proper time to all material particles. But do we get a spacelike hypersurface by looking at all the particles at some value of τ, and more importantly, is this hypersurface orthogonal to the congruence (which would ensure its being spacelike)? In fact there is no guarantee here. If we did, we would have the kind of coordinates mentioned above, and the vorticity would be zero. So it would be enough to show that the vorticity is not necessarily zero in order to prove that these hypersurfaces are not necessarily orthogonal to the particle worldlines. But then, the motion of the medium is presumably quite arbitrary (unless we assume rigidity, and we shall generalise that idea to curved spacetimes in the next section), so the vorticity is not necessarily going to be zero.

On the other hand, we ought to be able to see explicitly what might go wrong with our intuitive idea that the medium occupies a spacelike hypersurface orthogonal to the particle worldlines for any given value of τ. Referring back to Sect. 2.3.1, and in particular (2.19) on p. 23, we examined the 4-vector representing the spacetime displacement between a particle and one of its neighbours for the same value of the synchronised proper time along their worldlines. The problem we tackled there was the fact that this infinitesimal 4-vector is not generally orthogonal to either worldline. In (2.20), we established the infinitesimal proper time difference that must be taken in order for either particle to view the other as simultaneous.

4.3.15 Rigid Motions in Curved Spacetimes

The aim here is to generalise the notion of proper metric introduced for the flat spacetime case in Sect. 2.3.1. In fact, everything carries over with very minor changes. We return to the continuous medium in which particles are labelled by ξ^i, $i = 1, 2, 3$. Then arbitrary coordinates x^μ can be considered as functions of the ξ^i and τ, and conversely, at least in the region of spacetime occupied by the medium.

Let ξ and $\xi^i + \delta\xi^i$ label neighbouring particles in the medium. The worldline of the particle with label $\xi^i + \delta\xi^i$ is given by the functions

$$x^\mu(\xi + \delta\xi, \tau) = x^\mu(\xi, \tau) + x^\mu{}_{,i}(\xi, \tau)\delta\xi^i \, , \tag{4.97}$$

where the comma followed by a Latin index denotes partial differentiation with respect to the corresponding ξ as before. Once again, the quantity $x^\mu{}_{,i}(\xi, \tau)\delta\xi^i$, representing the difference between the two sets of worldline functions, is formally a 4-vector, being basically an infinitesimal coordinate difference. However, it is not generally orthogonal to the worldline of ξ. In other words, it does not lie in the approximate 'hyperplane' of simultaneity of either particle.

To get such a vector one applies the projection tensor

$$P^{\mu\nu} = g^{\mu\nu} - \dot{x}^\mu \dot{x}^\nu \, ,$$

where the dot denotes partial differentiation with respect to τ keeping the ξ^i fixed. The result is

$$\delta x^\mu := P^\mu{}_\nu x^\nu{}_{,i}(\xi, \tau)\delta\xi^i = x^\mu{}_{,i}\delta\xi^i - \dot{x}^\mu \dot{x}_\nu x^\nu{}_{,i}\delta\xi^i \, . \tag{4.98}$$

We find that application of the projection tensor corresponds to a simple proper time shift of amount

$$\delta\tau = -g_{\mu\nu}\dot{x}^{\mu}\dot{x}^{\nu}{}_{,i}\delta\xi^{i} , \tag{4.99}$$

so that

$$\delta x^{\mu} = x^{\mu}(\xi + \delta\xi, \tau + \delta\tau) - x^{\mu}(\xi, \tau) .$$

Indeed,

$$x^{\mu}(\xi + \delta\xi, \tau + \delta\tau) = x^{\mu}(\xi, \tau) + x^{\mu}{}_{,i}\delta\xi^{i} + \dot{x}^{\mu}\delta\tau ,$$

and feeding in the proposed expression for $\delta\tau$, we obtain

$$\begin{aligned}\delta x^{\mu} &= x^{\mu}(\xi, \tau) + x^{\mu}{}_{,i}\delta\xi^{i} + \dot{x}^{\mu}\delta\tau - x^{\mu}(\xi, \tau) \\ &= x^{\mu}{}_{,i}\delta\xi^{i} - g_{\nu\sigma}\dot{x}^{\nu}\dot{x}^{\sigma}{}_{,i}\delta\xi^{i}\dot{x}^{\mu} ,\end{aligned}$$

which is precisely δx^{μ} as defined in (4.98).

As in Sect. 2.3.1, we conclude that the two particles ξ and $\xi + \delta\xi$ appear, in the instantaneous rest frame of either, to be separated by a distance δs given by

$$(\delta s)^2 = (\delta x)^2 = \gamma_{ij}\delta\xi^{i}\delta\xi^{j} , \tag{4.100}$$

where

$$\boxed{\gamma_{ij} := P_{\mu\nu}x^{\mu}{}_{,i}x^{\nu}{}_{,j}} . \tag{4.101}$$

This follows because

$$(\delta x)^2 = P_{\mu\sigma}x^{\sigma}{}_{,i}\delta\xi^{i}P^{\mu}{}_{\nu}x^{\nu}{}_{,j}\delta\xi^{j} ,$$

and

$$\begin{aligned}P_{\mu\sigma}P^{\mu}{}_{\nu} &= (g_{\mu\sigma} - \dot{x}_{\mu}\dot{x}_{\sigma})(\delta^{\mu}{}_{\nu} - \dot{x}^{\mu}\dot{x}_{\nu}) \\ &= g_{\nu\sigma} - \dot{x}_{\nu}\dot{x}_{\sigma} - \dot{x}_{\sigma}\dot{x}_{\nu} + (\dot{x}_{\mu}\dot{x}^{\mu})\dot{x}_{\sigma}\dot{x}_{\nu} \\ &= g_{\nu\sigma} - \dot{x}_{\nu}\dot{x}_{\sigma} = P_{\nu\sigma} ,\end{aligned}$$

whence

$$(\delta x)^2 = P_{\nu\sigma}x^{\sigma}{}_{,i}x^{\nu}{}_{,j}\delta\xi^{i}\delta\xi^{j} = \gamma_{ij}\delta\xi^{i}\delta\xi^{j} ,$$

for the given γ_{ij}, as claimed. As before, we call the quantity γ_{ij} the proper metric of the medium.

So once again the point about γ_{ij} is that the two particles or observers labelled by ξ and $\xi + \delta\xi$ appear in the instantaneous rest frame of either to be separated by a proper distance δs as they would measure it given by

$$\boxed{(\delta s)^2 = \gamma_{ij}\delta\xi^{i}\delta\xi^{j}} . \tag{4.102}$$

We shall say that the set of particles or observers undergoes rigid motion if and only if the proper metric is everywhere independent of τ. This is expressed by

$$\boxed{\dot{\gamma}_{ij} = 0} . \tag{4.103}$$

Under rigid motion, the instantaneous separation distance between any pair of neighbouring observers is constant in time as they would see it in an instantaneously comoving inertial frame.

4.3.16 Rate of Strain Tensor

The aim here will be twofold:

- To understand the relationship between ordinary non-relativistic concepts of fluid motion and the expansion and vorticity of a timelike congruence as described in Sects. 4.3.13 and 4.3.14.
- To express the rigid motion condition $\dot{\gamma}_{ij} = 0$ of Sect. 4.3.15 [see (4.103)] in terms of derivatives with respect to the arbitrary coordinates x^μ by introducing the relativistic analog of the rate of strain tensor in ordinary continuum mechanics.

The non-relativistic strain tensor can be defined by

$$e_{ij} := \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right) ,$$

where $u_i(x)$ are the components of the displacement vector of the medium, describing the motion of the point originally at x when the material is deformed. One also defines the antisymmetric tensor

$$\omega_{ij} := \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} - \frac{\partial u_i}{\partial x_j}\right) ,$$

which describes the rotation occurring when the material is deformed. Clearly,

$$e_{ij} - \omega_{ij} = \frac{\partial u_i}{\partial x_j} ,$$

and hence, if all distortions are small,

$$\Delta u_i = (e_{ij} - \omega_{ij})\Delta x_j .$$

We can consider that e_{ij} describes non-rotational distortions, i.e., stretching, compression, and shear.

In the present discussion, u_i is replaced by a velocity field v_i and we have a rate of strain tensor. The nonrelativistic rate of strain tensor is

$$r_{ij} = v_{i,j} + v_{j,i} , \tag{4.104}$$

where v_i is a 3-velocity field and the differentiation is with respect to ordinary Cartesian coordinates. Let us look for a moment at this tensor. The nonrelativistic condition for rigid motion is

$$r_{ij} = 0 \quad \text{everywhere} .$$

This equation implies

$$0 = r_{ij,k} = v_{i,jk} + v_{j,ik} \,, \tag{4.105}$$

$$0 = r_{jk,i} = v_{j,ki} + v_{k,ji} \,. \tag{4.106}$$

Subtracting (4.106) from (4.105) and commuting the partial derivatives, we find

$$v_{i,jk} - v_{k,ji} = 0 \,, \tag{4.107}$$

which, upon permutation of the indices j and k, yields also

$$v_{i,kj} - v_{j,ki} = 0 \,. \tag{4.108}$$

Adding (4.105) and (4.108), we obtain

$$v_{i,jk} = 0 \,,$$

which has the general solution

$$v_i = -\omega_{ij} x_j + \beta_i \,, \tag{4.109}$$

where ω_{ij} and β_i are functions of time only. The condition $r_{ij} = 0$ constrains ω_{ij} to be antisymmetric, i.e.,

$$\omega_{ij} = -\omega_{ji} \,,$$

and nonrelativistic rigid motion is seen to be, at each instant, a uniform rotation with angular velocity

$$\omega_i = \frac{1}{2} \varepsilon_{ijk} \omega_{jk}$$

about the coordinate origin, superimposed upon a uniform translation with velocity β_i. Because the coordinate origin may be located arbitrarily at each instant, rigid motion may alternatively be described as one in which an arbitrary particle in the medium moves in an arbitrary way while at the same time the medium as a whole rotates about this point in an arbitrary (but uniform) way. Such a motion has six degrees of freedom.

Note that when r_{ij} is zero, we can also deduce that $v_{i,i} = 0$, i.e., $\operatorname{div} v = 0$, which is the condition for an incompressible fluid. This is evidently a weaker condition than rigidity.

Let us see how this generalises to special and general relativity. We return to the continuous medium in which particles are labelled by ξ^i, $i = 1,2,3$. Just as the coordinates x^μ are functions of the ξ^i and τ, so the ξ^i and τ can be regarded as functions of the x^μ, at least in the region of spacetime occupied by the medium. Following [14], we write

$$u^\mu := \dot{x}^\mu \;, \qquad u^2 = -1 \;, \qquad P_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu \;,$$

where, for the present purposes, the dot over a symbol denotes partial differentiation with respect to τ, keeping the ξ^i fixed. If f is an arbitrary function in the region occupied by the medium then

$$f_{,\mu} = f_{,i}\xi^i_{,\mu} + \dot{f}\tau_{,\mu} \;,$$

where the comma followed by a Greek index μ denotes partial differentiation with respect to the coordinate x^μ. This is just the good old chain rule.

Let us begin by considering the flat spacetime case of special relativity with inertial coordinates $\{x^\mu\}$, whence $P_{\mu\nu} = \eta_{\mu\nu} + u_\mu u_\nu$. It is very important to note that the dot over a symbol corresponds to the covariant derivative along the relevant worldline in the congruence, because the connection is zero for these coordinates. Hence, $\dot{u}$ is the four-acceleration, but it will *not* be for the curved spacetime case considered below. We thus have the following relations:

$$\dot{x}\cdot\ddot{x} = 0 \quad \text{or} \quad u\cdot\dot{u} = 0 \;,$$

since $u^2 = -1$, and

$$\begin{gathered}
u_\mu u^\mu{}_{,\nu} = 0 \;, \qquad \dot{u}_\mu = u_{\mu,\nu}u^\nu \;, \qquad u_\mu u^\mu{}_{,i} = 0 \;, \\
x^\mu{}_{,i}\xi^i{}_{,\nu} + \dot{x}^\mu\tau_{,\nu} = \delta^\mu{}_\nu \;, \\
\xi^i{}_{,\mu}x^\mu{}_{,j} = \delta^i_j \;, \qquad \xi^i{}_{,\mu}\dot{x}^\mu = 0 \;, \\
\tau_{,\mu}x^\mu{}_{,i} = 0 \;, \qquad \tau_{,\mu}\dot{x}^\mu = 1 \;, \\
P_{\mu\nu}\dot{x}^\nu{}_{,i} = P_{\mu\nu}u^\nu{}_{,i} = u_{\mu,i} \;.
\end{gathered} \tag{4.110}$$

Note that some of these relations will change when we use arbitrary coordinates $\{x^\mu\}$ in a curved spacetime (see below).

We now define the rate of strain tensor for the medium by

$$\boxed{r_{\mu\nu} := \dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu}} \;, \tag{4.111}$$

where γ_{ij} is defined as in (4.101) on p. 178, viz.,

$$\boxed{\gamma_{ij} := P_{\mu\nu}x^\mu{}_{,i}x^\nu{}_{,j}} \;. \tag{4.112}$$

We then have the following analysis in the flat spacetime case with $\{x^\mu\}$ being inertial coordinates:

$$
\begin{aligned}
r_{\mu\nu} &:= \dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu} \\
&= \left(\dot{P}_{\sigma\tau}x^\sigma{}_{,i}x^\tau{}_{,j} + P_{\sigma\tau}\dot{x}^\sigma{}_{,i}x^\tau{}_{,j} + P_{\sigma\tau}x^\sigma{}_{,i}\dot{x}^\tau{}_{,j}\right)\xi^i{}_{,\mu}\xi^j{}_{,\nu} \\
&= (\dot{u}_\sigma u_\tau + u_\sigma \dot{u}_\tau)(\delta^\sigma{}_\mu - u^\sigma \tau_{,\mu})(\delta^\tau{}_\nu - u^\tau \tau_{,\nu}) \\
&\quad + u_{\tau,i}\xi^i{}_{,\mu}(\delta^\tau{}_\nu - u^\tau\tau_{,\nu}) + (\delta^\sigma{}_\mu - u^\sigma\tau_{,\mu})u_{\sigma,j}\xi^j{}_{,\nu} \\
&= \dot{u}_\mu u_\nu + u_\mu \dot{u}_\nu + \dot{u}_\mu \tau_{,\nu} + \tau_{,\mu}\dot{u}_\nu + u_{\nu,\mu} - \dot{u}_\nu \tau_{,\mu} + u_{\mu,\nu} - \dot{u}_\mu \tau_{,\nu} \\
&= u_{\mu,\sigma}u^\sigma u_\nu + u_\mu u^\sigma u_{\nu,\sigma} + u_{\nu,\mu} + u_{\mu,\nu} \\
&= P_\mu{}^\sigma P_\nu{}^\tau (u_{\sigma,\tau} + u_{\tau,\sigma}) \, .
\end{aligned}
\tag{4.113}
$$

This is to be compared with (4.104) to justify calling it the rate of strain tensor. At any event x^μ, it lies entirely in the instantaneous hyperplane of simultaneity of the particle ξ^i that happens to coincide with that event.

This generalises to curved spacetimes. We define

$$
r_{\mu\nu} := \dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu} \, , \tag{4.114}
$$

as before, where the dot over γ_{ij} still denotes partial differentiation with respect to τ keeping the ξ^i fixed. We note that $r_{\mu\nu}$ is a tensor, since γ_{ij}, $\dot{\gamma}_{ij}$, ξ^i, and ξ^j are scalars under change of coordinates. At any x, there are coordinates such that $g_{\mu\nu,\sigma}\big|_x = 0$, whence covariant derivatives with respect to the Levi-Civita connection are just coordinate derivatives at x, and it follows immediately by the above flat spacetime argument that

$$
\boxed{r_{\mu\nu} = P_\mu{}^\sigma P_\nu{}^\tau (u_{\sigma;\tau} + u_{\tau;\sigma})} \, , \tag{4.115}
$$

where semi-colons denote covariant derivatives and $P^{\mu\nu}$ is given by

$$
P^{\mu\nu} = g^{\mu\nu} + \dot{x}^\mu \dot{x}^\nu \, ,
$$

for metric $g^{\mu\nu}$.

This is the neat way to establish (4.115) but it is instructive to see how the above argument changes in the details when we use arbitrary coordinates $\{x^\mu\}$ and take into account the nonzero connection. We must now be altogether more careful about some of the relations in (4.110).

To begin with, recall that the dot over a symbol denotes the *ordinary partial derivative* with respect to τ, keeping the ξ^i fixed, so $\dot{u}$ is *not* the four-acceleration. For the purposes only of this section, let us define a new symbol by

$$
\boxed{\overset{\odot}{u}{}^\mu := u^\nu u^\mu{}_{;\nu} = u^\nu (u^\mu{}_{,\nu} + \Gamma^\mu_{\rho\nu}u^\rho) = \dot{u}^\mu + \Gamma^\mu_{\rho\nu}u^\rho u^\nu} \, . \tag{4.116}
$$

Likewise,

$$\boxed{\overset{\odot}{u}_\mu := u^\nu u_{\mu;\nu} = u^\nu(u_{\mu,\nu} - \Gamma^\rho_{\mu\nu}u_\rho) = \dot{u}_\mu - \Gamma^\rho_{\mu\nu}u_\rho u^\nu} \ . \tag{4.117}$$

So we still have $u \cdot \overset{\odot}{u} = 0$, but we do *not* have $u \cdot \dot{u} = 0$. The point is that

$$\boxed{\dot{u}^\mu := u^\nu u^\mu{}_{,\nu} \ , \quad \dot{u}_\mu := u^\nu u_{\mu,\nu}} \ , \tag{4.118}$$

are just directional derivatives along the congruence. We thus have the argument

$$\begin{aligned}(u_\mu u^\mu)_{,\nu} = 0 &\Longleftrightarrow u_\mu u^\mu{}_{,\nu} + u_{\mu,\nu}u^\mu = 0 \\ &\Longleftrightarrow u_\mu u^\mu{}_{,\nu} + (g_{\mu\rho}u^\rho)_{,\nu}u^\mu = 0 \\ &\Longleftrightarrow u_\mu u^\mu{}_{,\nu} + g_{\mu\rho,\nu}u^\rho u^\mu + g_{\mu\rho}u^\rho{}_{,\nu}u^\mu = 0 \ ,\end{aligned}$$

whence finally,

$$\boxed{u_\mu u^\mu{}_{,\nu} = -\frac{1}{2}g_{\mu\rho,\nu}u^\rho u^\mu} \ . \tag{4.119}$$

An exactly analogous argument shows that

$$\boxed{u_\mu u^\mu{}_{,i} = -\frac{1}{2}g_{\mu\rho,i}u^\rho u^\mu} \ . \tag{4.120}$$

In this context, note that

$$\boxed{\dot{x}^\mu = u^\mu \ , \quad \dot{g}_{\mu\nu} = u^\rho g_{\mu\nu,\rho}} \ . \tag{4.121}$$

We still have

$$\boxed{x^\mu{}_{,i}\xi^i{}_{,\nu} + \dot{x}^\mu\tau_{,\nu} = \delta^\mu{}_\nu} \ , \tag{4.122}$$

which is basically just the chain rule applied to $\partial x^\mu/\partial x^\nu$ under the assumption that the transformation to the label coordinates (ξ, τ) is invertible. Likewise by the chain rule,

$$\boxed{u^\mu{}_{,i}\xi^i{}_{,\nu} + \dot{u}^\mu\tau_{,\nu} = u^\mu{}_{,\nu}} \ . \tag{4.123}$$

The chain rule still gives the four relations

$$\boxed{\xi^i{}_{,\mu}x^\mu{}_{,j} = \delta^i{}_j \ , \quad \xi^i{}_{,\mu}\dot{x}^\mu = 0 \ , \quad \tau_{,\mu}x^\mu{}_{,i} = 0 \ , \quad \tau_{,\mu}\dot{x}^\mu = 1} \ . \tag{4.124}$$

Finally,

$$\begin{aligned}P_{\mu\nu}\dot{x}^\nu{}_{,i} &= P_{\mu\nu}\frac{\partial^2 x^\nu}{\partial\tau\partial\xi^i} = P_{\mu\nu}u^\nu{}_{,i} \\ &= g_{\mu\nu}u^\nu{}_{,i} + u_\mu u_\nu u^\nu{}_{,i} \\ &= u_{\mu,i} - g_{\mu\nu,i}u^\nu + u_\mu u_\nu u^\nu{}_{,i} \ ,\end{aligned}$$

whence

$$\boxed{P_{\mu\nu}\dot{x}^{\nu}{}_{,i} = u_{\mu,i} - g_{\mu\nu,i}u^{\nu} - \frac{1}{2}u_{\mu}g_{\nu\rho,i}u^{\rho}u^{\nu}}\,. \tag{4.125}$$

Let us now see what becomes of the argument (4.113) on p. 182 when it is transposed to a curved spacetime.

The first step is

$$\begin{aligned} r_{\mu\nu} &:= \dot{\gamma}_{ij}\xi^{i}{}_{,\mu}\xi^{j}{}_{,\nu} \\ &= \big(\dot{P}_{\sigma\tau}x^{\sigma}{}_{,i}x^{\tau}{}_{,j} + P_{\sigma\tau}\dot{x}^{\sigma}{}_{,i}x^{\tau}{}_{,j} + P_{\sigma\tau}x^{\sigma}{}_{,i}\dot{x}^{\tau}{}_{,j}\big)\xi^{i}{}_{,\mu}\xi^{j}{}_{,\nu} \\ &= \big(\dot{g}_{\sigma\tau} + \dot{u}_{\sigma}u_{\tau} + u_{\sigma}\dot{u}_{\tau}\big)\big(\delta^{\sigma}{}_{\mu} - u^{\sigma}\tau_{,\mu}\big)\big(\delta^{\tau}{}_{\nu} - u^{\tau}\tau_{,\nu}\big) \\ &\quad + \big(g_{\sigma\tau} + u_{\sigma}u_{\tau}\big)\big(u^{\sigma}{}_{,\mu} - \dot{u}^{\sigma}\tau_{,\mu}\big)\big(\delta^{\tau}{}_{\nu} - u^{\tau}\tau_{,\nu}\big) \\ &\quad + \big(g_{\sigma\tau} + u_{\sigma}u_{\tau}\big)\big(\delta^{\sigma}{}_{\mu} - u^{\sigma}\tau_{,\mu}\big)\big(u^{\tau}{}_{,\nu} - \dot{u}^{\tau}\tau_{,\nu}\big)\,, \end{aligned}$$

where $\dot{g}_{\sigma\tau} := u^{\rho}g_{\sigma\tau,\rho}$, $\dot{u}_{\sigma} := u^{\rho}u_{\sigma,\rho}$, and so on. Note immediately that

$$\big(g_{\sigma\tau} + u_{\sigma}u_{\tau}\big)\big(\delta^{\tau}{}_{\nu} - u^{\tau}\tau_{,\nu}\big) = g_{\sigma\nu} + u_{\sigma}u_{\nu}\,,$$

since $u^{\tau}u_{\tau} = -1$, so $u^{\tau}(g_{\sigma\tau} + u_{\sigma}u_{\tau}) = 0$. Recalling (4.117), we thus write

$$\begin{aligned} r_{\mu\nu} = &\Big(u^{\rho}g_{\sigma\tau,\rho} + \overset{\odot}{u}_{\sigma}u_{\tau} + u_{\sigma}\overset{\odot}{u}_{\tau} + \Gamma^{\rho}_{\sigma\phi}u^{\phi}u_{\rho}u_{\tau} + \Gamma^{\rho}_{\tau\phi}u^{\phi}u_{\rho}u_{\sigma}\Big) \\ &\qquad\qquad\qquad\qquad \times\big(\delta^{\sigma}{}_{\mu} - u^{\sigma}\tau_{,\mu}\big)\big(\delta^{\tau}{}_{\nu} - u^{\tau}\tau_{,\nu}\big) \\ &+\Big[\big(g_{\sigma\nu} + u_{\sigma}u_{\nu}\big)\big(u^{\sigma}{}_{,\mu} - \overset{\odot}{u}{}^{\sigma}\tau_{,\mu} + \Gamma^{\sigma}_{\rho\phi}u^{\rho}u^{\phi}\tau_{,\mu}\big) + (\mu\leftrightarrow\nu)\Big]\,. \end{aligned}$$

Recalling that $u^{\sigma}\overset{\odot}{u}_{\sigma} = 0$, the bottom line is

$$\boxed{g_{\sigma\nu}u^{\sigma}{}_{,\mu} - \overset{\odot}{u}_{\nu}\tau_{,\mu} + g_{\sigma\nu}\Gamma^{\sigma}_{\rho\phi}u^{\rho}u^{\phi}\tau_{,\mu} + u_{\nu}u_{\sigma}u^{\sigma}{}_{,\mu} + u_{\nu}u_{\sigma}\Gamma^{\sigma}_{\rho\phi}u^{\rho}u^{\phi}\tau_{,\mu} + (\mu\leftrightarrow\nu)}\,. \tag{4.126}$$

Regarding the top line, we have first

$$\begin{aligned} &\Big(u^{\rho}g_{\sigma\tau,\rho} + \overset{\odot}{u}_{\sigma}u_{\tau} + u_{\sigma}\overset{\odot}{u}_{\tau} + \Gamma^{\rho}_{\sigma\phi}u^{\phi}u_{\rho}u_{\tau} + \Gamma^{\rho}_{\tau\phi}u^{\phi}u_{\rho}u_{\sigma}\Big)\big(\delta^{\sigma}{}_{\mu} - u^{\sigma}\tau_{,\mu}\big) \\ &\qquad = u^{\rho}g_{\mu\tau,\rho} + \overset{\odot}{u}_{\mu}u_{\tau} + u_{\mu}\overset{\odot}{u}_{\tau} + \Gamma^{\rho}_{\mu\phi}u^{\phi}u_{\rho}u_{\tau} + \Gamma^{\rho}_{\tau\phi}u^{\phi}u_{\rho}u_{\mu} \\ &\qquad\quad - u^{\rho}g_{\sigma\tau,\rho}u^{\sigma}\tau_{,\mu} + \overset{\odot}{u}_{\tau}\tau_{,\mu} - \Gamma^{\rho}_{\sigma\phi}u^{\phi}u_{\rho}u_{\tau}u^{\sigma}\tau_{,\mu} + \Gamma^{\rho}_{\tau\phi}u^{\phi}u_{\rho}\tau_{,\mu}\,. \end{aligned}$$

Multiplying this now by $\big(\delta^{\tau}{}_{\nu} - u^{\tau}\tau_{,\nu}\big)$, we obtain the terms

$$\boxed{\begin{gathered}u^\rho g_{\mu\nu,\rho} + \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + \Gamma^\rho_{\mu\phi} u^\phi u_\rho u_\nu + \Gamma^\rho_{\nu\phi} u^\phi u_\rho u_\mu - u^\rho g_{\sigma\nu,\rho} u^\sigma \tau_{,\mu} \\ + \overset{\odot}{u}_\nu \tau_{,\mu} - \Gamma^\rho_{\sigma\phi} u^\phi u_\rho u_\nu u^\sigma \tau_{,\mu} + \Gamma^\rho_{\nu\phi} u^\phi u_\rho \tau_{,\mu} - u^\tau \tau_{,\nu} u^\rho g_{\mu\tau,\rho} \\ + \tau_{,\nu} \overset{\odot}{u}_\mu + \tau_{,\nu} \Gamma^\rho_{\mu\phi} u^\phi u_\rho - \Gamma^\rho_{\tau\phi} u^\phi u_\rho u_\mu u^\tau \tau_{,\nu} + u^\rho g_{\sigma\tau,\rho} u^\sigma \tau_{,\mu} u^\tau \tau_{,\nu} \\ - \Gamma^\rho_{\sigma\phi} u^\phi u_\rho u^\sigma \tau_{,\mu} \tau_{,\nu} - u^\tau \tau_{,\nu} \Gamma^\rho_{\tau\phi} u^\phi u_\rho \tau_{,\mu}\end{gathered}} \qquad (4.127)$$

So $r_{\mu\nu}$ is the sum of all the terms in (4.126) and (4.127). The coefficient of all terms containing $\tau_{,\mu}\tau_{,\nu}$ is

$$\begin{aligned} u^\rho g_{\sigma\tau,\rho} u^\sigma u^\tau - \Gamma^\rho_{\sigma\phi} u^\phi u_\rho u^\sigma - u^\tau \Gamma^\rho_{\tau\phi} u^\phi u_\rho &= \left(g_{\sigma\tau,\rho} - 2\Gamma_{\sigma\tau\rho}\right) u^\sigma u^\tau u^\rho \\ &= \left[g_{\sigma\tau,\rho} - (g_{\rho\sigma,\tau} + g_{\rho\tau,\sigma} - g_{\sigma\tau,\rho})\right] u^\sigma u^\tau u^\rho \\ &= 0\,, \end{aligned}$$

where we have used (4.30) on p. 160, assuming a torsion free connection. The coefficient of $\tau_{,\mu}$ in the sum of all the terms in (4.126) and (4.127) is

$$\begin{gathered} -\overset{\odot}{u}_\nu + \Gamma_{\nu\rho\phi} u^\rho u^\phi + u_\nu \Gamma_{\sigma\rho\phi} u^\sigma u^\rho u^\phi - u^\rho u^\sigma g_{\sigma\nu,\rho} + \overset{\odot}{u}_\nu \\ - \Gamma_{\rho\sigma\phi} u^\phi u^\rho u^\sigma u_\nu + \Gamma_{\rho\nu\phi} u^\phi u^\rho\ . \end{gathered}$$

This in turn is equal to

$$u^\rho u^\sigma \left(\Gamma_{\nu\rho\sigma} + \Gamma_{\rho\nu\sigma} - g_{\sigma\nu,\rho}\right) = 0\,,$$

after another application of (4.30) on p. 160. By symmetry, the coefficient of $\tau_{,\nu}$ in the sum of all the terms in (4.126) and (4.127) is also zero.

We thus have

$$\begin{aligned} r_{\mu\nu} &= g_{\sigma\nu} u^\sigma{}_{,\mu} + u_\nu u_\sigma u^\sigma{}_{,\mu} + g_{\sigma\mu} u^\sigma{}_{,\nu} + u_\mu u_\sigma u^\sigma{}_{,\nu} \\ &\quad + u^\rho g_{\mu\nu,\rho} + \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + \Gamma_{\rho\mu\phi} u^\phi u^\rho u_\nu + \Gamma_{\rho\nu\phi} u^\phi u^\rho u_\mu \\ &= \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + (g_{\sigma\nu} + u_\sigma u_\nu) u^\sigma{}_{,\mu} + (g_{\sigma\mu} + u_\sigma u_\mu) u^\sigma{}_{,\nu} \\ &\quad + u^\rho \left(g_{\mu\nu,\rho} + \Gamma_{\rho\mu\phi} u^\phi u_\nu + \Gamma_{\rho\nu\phi} u^\phi u_\mu\right)\,, \end{aligned}$$

where

$$\begin{aligned} (g_{\sigma\nu} + u_\sigma u_\nu) u^\sigma{}_{,\mu} &= \left[(g_{\sigma\nu} + u_\sigma u_\nu) u^\sigma\right]_{,\mu} - (g_{\sigma\nu} + u_\sigma u_\nu)_{,\mu} u^\sigma \\ &= -g_{\sigma\nu,\mu} u^\sigma - u_{\sigma,\mu} u^\sigma u_\nu + u_{\nu,\mu} \end{aligned}$$

and

$$(g_{\sigma\mu} + u_\sigma u_\mu) u^\sigma{}_{,\nu} = -g_{\sigma\mu,\nu} u^\sigma - u_{\sigma,\nu} u^\sigma u_\mu + u_{\mu,\nu}\,,$$

whence

$$
\begin{aligned}
r_{\mu\nu} &= \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu - g_{\sigma\nu,\mu}u^\sigma - g_{\sigma\mu,\nu}u^\sigma + u^\rho g_{\mu\nu,\rho} \\
&\quad - u_{\sigma,\mu}u^\sigma u_\nu + u_{\nu,\mu} - u_{\sigma,\nu}u^\sigma u_\mu + u_{\mu,\nu} + u^\rho \left(\Gamma_{\rho\mu\phi}u^\phi u_\nu + \Gamma_{\rho\nu\phi}u^\phi u_\mu\right) .
\end{aligned}
$$

But

$$
\begin{aligned}
-g_{\sigma\nu,\mu}u^\sigma - g_{\sigma\mu,\nu}u^\sigma + u^\rho g_{\mu\nu,\rho} &= -u^\sigma \left(g_{\sigma\nu,\mu} + g_{\sigma\mu,\nu} - g_{\mu\nu,\sigma}\right) \\
&= -2\Gamma_{\sigma\mu\nu}u^\sigma ,
\end{aligned}
$$

and in addition,

$$
-u_{\sigma,\mu}u^\sigma u_\nu - u_{\sigma,\nu}u^\sigma u_\mu + u^\rho \left(\Gamma_{\rho\mu\phi}u^\phi u_\nu + \Gamma_{\rho\nu\phi}u^\phi u_\mu\right) = 0 ,
$$

because

$$
\begin{aligned}
-u_{\sigma,\mu}u^\sigma + \Gamma_{\rho\mu\phi}u^\phi u^\rho &= -u^\sigma \left(u_{\sigma,\mu} - \Gamma^\rho_{\mu\sigma}u_\rho\right) \\
&= -u^\sigma u_{\sigma;\mu} \\
&= -\frac{1}{2}(u^\sigma u_\sigma)_{;\mu} = 0 .
\end{aligned}
$$

The expression for the rate of strain tensor has now boiled down to

$$
r_{\mu\nu} = \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + u_{\nu,\mu} + u_{\mu,\nu} - 2\Gamma_{\sigma\mu\nu}u^\sigma ,
$$

and hence,

$$
r_{\mu\nu} = \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + u_{\nu;\mu} + u_{\mu;\nu} .
$$

But of course,

$$
\begin{aligned}
P_\mu{}^\sigma P_\nu{}^\tau (u_{\sigma;\tau} + u_{\tau;\sigma}) &= (\delta_\mu{}^\sigma + u_\mu u^\sigma)(\delta_\nu{}^\tau + u_\nu u^\tau)(u_{\sigma;\tau} + u_{\tau;\sigma}) \\
&= u_{\mu;\nu} + u_{\nu;\mu} + u_{\mu;\tau}u_\nu u^\tau + u_\nu u^\tau u_{\tau;\mu} + u_{\sigma;\nu}u_\mu u^\sigma \\
&\quad + u_{\nu;\sigma}u_\mu u^\sigma + u_\mu u^\sigma u_\nu u^\tau u_{\sigma;\tau} + u_\mu u^\sigma u_\nu u^\tau u_{\tau;\sigma} \\
&= \overset{\odot}{u}_\mu u_\nu + u_\mu \overset{\odot}{u}_\nu + u_{\nu;\mu} + u_{\mu;\nu} ,
\end{aligned}
$$

which completes the second proof of (4.115) back on p. 182.

The second proof is given here to show the value of the trick used in the first proof, which basically means that, at any preselected spacetime event, we can drop all the terms in the second proof that depend on first coordinate derivatives of the metric or on the connection coefficients, provided we assume a torsion-free connection. As mentioned, this in turn happens because the weak equivalence principle is built into the manifold model of spacetime, in the sense that, at any event, we can find coordinates in some neighbourhood of that event such that the metric is very close to the Minkowski form (in the sense made precise earlier).

Summary

To sum up then, the result

$$r_{\mu\nu} := \dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu} = P_\mu{}^\sigma P_\nu{}^\tau(u_{\sigma;\tau} + u_{\tau;\sigma}) \tag{4.128}$$

expresses the rate of strain tensor in terms of coordinate derivatives of the four-velocity field of the medium. We now characterise relativistic rigid motion by

$$r_{\mu\nu} = 0\,, \qquad \dot{\gamma}_{ij} = 0\,. \tag{4.129}$$

Once again, we observe that the criterion for rigid motion, viz., $r_{\mu\nu} = 0$, is independent of the coordinates, because $r_{\mu\nu}$ is a tensor, even in a curved spacetime.

Furthermore, up to a factor of 1/2, the rate of strain tensor $r_{\mu\nu}$ is just the expansion tensor $\Theta_{\mu\nu}$ as defined by (4.74) on p. 172. So we may conclude that what we called the expansion in Sect. 4.3.13 will be zero if and only if the motion of the timelike congruence is rigid.

4.3.17 Stationary and Static Spacetimes

A stationary spacetime is one with a globally defined, timelike Killing vector field K. We already mentioned Killing vector fields (KVF) in Sect. 2.4.8 since they have been taken to play a role in the physical interpretation of general relativistic spacetimes. In fact, each KVF is associated with a one-parameter isometry. The mathematical condition for K to be a KVF is that the Lie derivative of the metric tensor along K should be zero:

$$\boxed{L_K g = 0}\,. \tag{4.130}$$

In terms of components relative to arbitrary coordinates, using the definition of the Lie derivative of a type (2,0) tensor, this becomes

$$K_{\mu;\nu} + K_{\nu;\mu} = 0\,. \tag{4.131}$$

We can then choose coordinates in such a way that the components of the metric are independent of the time coordinate. Such coordinates can be obtained by taking a parameter along the flow lines of the KVF as time coordinate and completing it by a suitable set of spatial coordinates. The idea then is that, along a flow line of the Killing vector field, the geometry of spacetime 'looks the same' at all times.

The converse is also true here. If there exist coordinates such that the metric components are independent of the time coordinate, then the temporal coordinate curves (curves such that only x^0 changes, while x^1, x^2, and x^3 are held constant) constitute the flow curves of a KVF and the spacetime is stationary (up to the problem of fixing this up globally). It is instructive to prove this. Let $K := \partial_0 := \partial_{x^0}$, whence it has components $K^0 = 1$, $K^1 = K^2 = K^3 = 0$. The hypothesis is that

$$g_{\mu\nu,0} := \partial_0 g_{\mu\nu} = 0 \;, \quad \forall \mu, \nu \in \{0,1,2,3\} \;,$$

and we would like to show that

$$K_{\mu;\nu} + K_{\nu;\mu} = 0 \;,$$

that is,

$$K_{\mu,\nu} + K_{\nu,\mu} - 2\Gamma^{\rho}_{\mu\nu} K_{\rho} = 0 \;.$$

Now

$$K_{\mu,\nu} = (g_{\mu\rho} K^{\rho})_{,\nu} = g_{\mu 0,\nu} \;, \quad K_{\nu\mu} = g_{\nu 0,\mu} \;.$$

But

$$\begin{aligned} 2\Gamma^{\rho}_{\mu\nu} K_{\rho} &= (g_{\rho\mu,\nu} + g_{\rho\nu,\mu} - g_{\mu\nu,\rho}) K^{\rho} \\ &= g_{0\mu,\nu} + g_{0\nu,\mu} - g_{\mu\nu,0} \\ &= g_{0\mu,\nu} + g_{0\nu,\mu} \;, \end{aligned}$$

which effectively completes the proof.

In general the Killing vector field will not be orthogonal to a continuum of spacelike hypersurfaces, e.g., the hypersurfaces obtained by holding x^0 constant and allowing only the three spacelike coordinates x^1, x^2, and x^3 to vary. It will in general possess a non-vanishing spatial component relative to such coordinates. In terms of the metric, when it is expressed relative to such coordinates, there will be nonzero components of the form g_{0i} and g_{i0} for $i = 1,2,3$.

If the Killing field is also orthogonal to a family of hypersurfaces, which are then automatically spacelike, the spacetime is said to be static. In terms of the components of the metric tensor this means that coordinates can be chosen such that the metric has block diagonal form with $g_{0i} = 0 = g_{i0}$ for $i = 1,2,3$, and the components are all independent of the time coordinate. Naturally, every static spacetime is also stationary.

In a stationary spacetime, the integral curves of the associated timelike Killing vector field K form a timelike congruence of the kind we have been investigating. Of course, K might not do for the vector field X in previous sections, because X is supposed to be normalised, and we do not know whether K is normalised. Suppose that K does happen to be normalised, so that we can set $X \equiv K$. Now the Killing equation is

$$X_{\mu;\nu} + X_{\nu;\mu} = 0 \;,$$

and interestingly this means precisely that the expansion is zero, and hence also that any medium whose fluid elements have this motion is moving rigidly. Unfortunately, the vorticity is not necessarily zero. If it had been, from the results in Sect. 4.3.14, we would have automatically obtained the other requirement for the spacetime to be static, namely, the existence of a family of orthogonal hypersurfaces.

To investigate further, one might introduce the scalar function κ with the property that $X = \kappa K$. We then have

$$-1 = (X|X) = \kappa^2 (K|K) , \qquad \kappa = \left[-\frac{1}{(K|K)} \right]^{1/2} ,$$

assuming that $(K|K)$ is never zero. (This is part of the assumption that the Killing vector field is timelike, hence non-null.) We then have

$$X_{\mu;\nu} = \kappa_{,\nu} K_\mu + \kappa K_{\mu;\nu} ,$$

and therefore

$$X_{\mu;\nu} + X_{\nu;\mu} = \kappa_{,\nu} K_\mu + \kappa_{,\mu} K_\nu , \quad X_{\mu;\nu} - X_{\nu;\mu} = \kappa_{,\nu} K_\mu - \kappa_{,\mu} K_\nu + 2\kappa K_{\mu;\nu} .$$

The vorticity is half the projection of the second of these onto the directions perpendicular to the congruence. If there is a family of 3D hypersurfaces orthogonal to the congruence (or equivalently, orthogonal to the Killing vector field), then the vorticity has to be zero, from what we have said above. This gives another equation on the Killing vector field that must be satisfied if it makes the spacetime static:

$$\left(\kappa_{,\nu} K_\mu - \kappa_{,\mu} K_\nu + 2\kappa K_{\mu;\nu} \right) h^\mu_\rho h^\nu_\sigma = 0 . \tag{4.132}$$

4.3.18 Transformation Properties of the Tetrad Quantities

Let us return to the earlier discussion of general timelike congruences. The question addressed here is: how do the different quantities defined transform under change of tetrad, i.e., under local Lorentz transformation? This is highly relevant if we are to determine whether a quantity such as the vorticity or the expansion really describes the congruence or whether it describes the congruence and the tetrad together. These quantities might not transform as Lorentz tensors. In particular, they only have 9 components as defined above! Are the other components set to zero, and would they stay that way under change of tetrad?

We can immediately answer the last point by referring to (4.73) and (4.74) on p. 172 for the expansion and (4.83) and (4.84) on p. 173 for the vorticity. We have to extend the definitions like this:

$$\Theta_{\hat{\alpha}\hat{\beta}} := \Theta_{\mu\nu} n^\mu_{\hat{\alpha}} n^\nu_{\hat{\beta}} , \qquad \Theta_{\mu\nu} := X_{(\rho;\sigma)} h^\rho_\mu h^\sigma_\nu = \frac{1}{2} \left(X_{\rho;\sigma} + X_{\sigma;\rho} \right) h^\rho_\mu h^\sigma_\nu , \tag{4.133}$$

$$\omega_{\hat{\alpha}\hat{\beta}} := \omega_{\mu\nu} n^\mu_{\hat{\alpha}} n^\nu_{\hat{\beta}} , \qquad \omega_{\mu\nu} := X_{[\rho;\sigma]} h^\rho_\mu h^\sigma_\nu = \frac{1}{2} \left(X_{\rho;\sigma} - X_{\sigma;\rho} \right) h^\rho_\mu h^\sigma_\nu . \tag{4.134}$$

These are just the tetrad components of well defined tensors. Note that this means

$$\Theta_{\hat{\alpha}\hat{0}} = 0 = \Theta_{\hat{0}\hat{\alpha}} , \quad \forall\, \hat{\alpha} \in \{0, 1, 2, 3\} . \tag{4.135}$$

This follows because we have projected $X_{\rho;\sigma} + X_{\sigma;\rho}$ onto a space orthogonal to the worldlines. Explicitly, we have

$$h^{\rho}_{\mu} n^{\mu}_{\hat{\alpha}} = n^{\hat{b}}_{\mu} n^{\rho}_{\hat{b}} n^{\mu}_{\hat{\alpha}} = n^{\rho}_{\hat{b}} \delta^{\hat{b}}_{\hat{\alpha}} = \begin{cases} n^{\rho}_{\hat{\alpha}} & \text{for } \hat{\alpha} \in \{1,2,3\}\,, \\ 0 & \text{for } \hat{\alpha} = 0\,, \end{cases}$$

and hence,

$$\begin{aligned} \Theta_{\hat{\alpha}\hat{\beta}} &= \frac{1}{2}\big(X_{\rho;\sigma} + X_{\sigma;\rho}\big) h^{\rho}_{\mu} h^{\sigma}_{\nu} n^{\mu}_{\hat{\alpha}} n^{\nu}_{\hat{\beta}} \\ &= \begin{cases} \dfrac{1}{2}\big(X_{\rho;\sigma} + X_{\sigma;\rho}\big) n^{\rho}_{\hat{\alpha}} n^{\sigma}_{\hat{\beta}} & \text{for } \hat{\alpha}, \hat{\beta} \in \{1,2,3\}\,, \\ 0 & \text{for } \hat{\alpha} = 0 \text{ or } \hat{\beta} = 0\,. \end{cases} \end{aligned}$$

Likewise,

$$\omega_{\hat{\alpha}\hat{0}} = 0 = \omega_{\hat{0}\hat{\alpha}}\,, \quad \forall\, \hat{\alpha} \in \{0,1,2,3\}\,. \tag{4.136}$$

So if we now make a *local* Lorentz transformation

$$n'_{\hat{\alpha}} = L_{\hat{\alpha}}{}^{\hat{\beta}} n_{\hat{\beta}}\,, \tag{4.137}$$

i.e., one that may vary from point to point in spacetime, we find that $\Theta_{\hat{\alpha}\hat{\beta}}$ and $\omega_{\hat{\alpha}\hat{\beta}}$ transform according to

$$\Theta'_{\hat{\alpha}\hat{\beta}} = L_{\hat{\alpha}}{}^{\hat{\gamma}} L_{\hat{\beta}}{}^{\hat{\delta}} \Theta_{\hat{\gamma}\hat{\delta}}\,, \qquad \omega'_{\hat{\alpha}\hat{\beta}} = L_{\hat{\alpha}}{}^{\hat{\gamma}} L_{\hat{\beta}}{}^{\hat{\delta}} \omega_{\hat{\gamma}\hat{\delta}}\,. \tag{4.138}$$

Of course, any Lorentz transformation that does not fix $n_{\hat{0}}$ will spoil the fact that the tetrad is adapted to this particular congruence. It will also mean that $\Theta'_{\hat{\alpha}0}$ and $\omega'_{\hat{\alpha}0}$ are not generally zero. This is why we limit the discussion of these quantities to the local rotation group, i.e., local Lorentz transformations that do not alter $n_{\hat{0}}$ at any point in spacetime. In fact, these are precisely the Lorentz transformations $L_{\hat{\alpha}}{}^{\hat{\beta}}$ for which $L_{\hat{0}}{}^{\hat{0}} = 1$ and otherwise both $\hat{\alpha}$ and $\hat{\beta}$ must be in $\{1,2,3\}$ to obtain a nonzero value.

Transformation of Structure Constants

Let us establish the relation between $C_{\hat{\rho}\hat{\sigma}\hat{\tau}}$ and $C'_{\hat{\alpha}\hat{\beta}\hat{\gamma}}$, referring to the general transformation (4.14) for structure constants given on p. 159:

$$C'^{\hat{\varepsilon}}{}_{\hat{\beta}\hat{\gamma}} = L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\delta}} C^{\hat{\delta}}{}_{\hat{\rho}\hat{\sigma}} + 2 L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} n_{[\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\rho}]}\,, \tag{4.139}$$

whence

$$C'_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = \eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\delta}} C^{\hat{\delta}}{}_{\hat{\rho}\hat{\sigma}} + 2\eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} n_{[\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\rho}]}\,. \tag{4.140}$$

For the record, it should be remembered that these relations are designed to ensure that

$$[n_{\hat\beta}, n_{\hat\gamma}] = C^{\hat\alpha}{}_{\hat\beta\hat\gamma} n_{\hat\alpha} \quad \Longrightarrow \quad [n'_{\hat\beta}, n'_{\hat\gamma}] = C'^{\hat\alpha}{}_{\hat\beta\hat\gamma} n'_{\hat\alpha} \, .$$

Consider the first term on the right-hand side of (4.140), viz.,

$$\eta_{\hat\alpha\hat\varepsilon} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} L^{-1\hat\varepsilon}{}_{\hat\delta} C^{\hat\delta}{}_{\hat\rho\hat\sigma} = \eta_{\hat\alpha\hat\varepsilon} L^{-1\hat\varepsilon}{}_{\hat\delta} \eta^{\hat\delta\hat\tau} C_{\hat\tau\hat\rho\hat\sigma} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \, .$$

Now in matrix language, we have $L\eta L^{\mathrm{T}} = \eta$, whence $L = \eta L^{-1\mathrm{T}} \eta^{-1}$ and

$$\eta_{\hat\alpha\hat\varepsilon} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} L^{-1\hat\varepsilon}{}_{\hat\delta} C^{\hat\delta}{}_{\hat\rho\hat\sigma} = L_{\hat\alpha}{}^{\hat\tau} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} C_{\hat\tau\hat\rho\hat\sigma} \, ,$$

which is a neat enough form for this term, and of course, quite unsurprising.

Consider now the second term on the right-hand side of (4.140). We have manipulations of the kind

$$\begin{aligned} L_{\hat\beta}{}^{\hat\rho} n_{\hat\sigma} \big(L^{-1\hat\varepsilon}{}_{\hat\rho} \big) &= L_{\hat\beta}{}^{\hat\rho} n_{\hat\sigma}{}^{\mu} \partial_\mu \big(L^{-1\hat\varepsilon}{}_{\hat\rho} \big) \\ &= n_{\hat\sigma}{}^{\mu} \partial_\mu \big(L_{\hat\beta}{}^{\hat\rho} L^{-1\hat\varepsilon}{}_{\hat\rho} \big) - \big(n_{\hat\sigma}{}^{\mu} \partial_\mu L_{\hat\beta}{}^{\hat\rho} \big) L^{-1\hat\varepsilon}{}_{\hat\rho} \\ &= -\big(\partial_{\hat\sigma} L_{\hat\beta}{}^{\hat\rho} \big) L^{-1\hat\varepsilon}{}_{\hat\rho} \, , \end{aligned}$$

using the notation

$$\partial_{\hat\sigma} := n_{\hat\sigma}{}^{\mu} \partial_\mu \, .$$

So the second term on the right-hand side of (4.140) is

$$\begin{aligned} 2\eta_{\hat\alpha\hat\varepsilon} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} n_{[\hat\sigma} L^{-1\hat\varepsilon}{}_{\hat\rho]} &= \eta_{\hat\alpha\hat\varepsilon} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \Big[n_{\hat\sigma} \big(L^{-1\hat\varepsilon}{}_{\hat\rho} \big) - n_{\hat\rho} \big(L^{-1\hat\varepsilon}{}_{\hat\sigma} \big) \Big] \\ &= \eta_{\hat\alpha\hat\varepsilon} \Big[-L_{\hat\gamma}{}^{\hat\sigma} \big(\partial_{\hat\sigma} L_{\hat\beta}{}^{\hat\rho} \big) L^{-1\hat\varepsilon}{}_{\hat\rho} + L_{\hat\beta}{}^{\hat\rho} \big(\partial_{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \big) L^{-1\hat\varepsilon}{}_{\hat\sigma} \Big] \\ &= -L_{\hat\alpha}{}^{\hat\tau} \eta_{\hat\tau\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \big(\partial_{\hat\sigma} L_{\hat\beta}{}^{\hat\rho} \big) + L_{\hat\alpha}{}^{\hat\tau} \eta_{\hat\tau\hat\sigma} L_{\hat\beta}{}^{\hat\rho} \big(\partial_{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \big) \\ &= \eta_{\hat\rho\hat\sigma} \Big[L_{\hat\alpha}{}^{\hat\rho} L_{\hat\beta}{}^{\hat\tau} \big(\partial_{\hat\tau} L_{\hat\gamma}{}^{\hat\sigma} \big) - L_{\hat\alpha}{}^{\hat\sigma} L_{\hat\gamma}{}^{\hat\tau} \big(\partial_{\hat\tau} L_{\hat\beta}{}^{\hat\rho} \big) \Big] \, , \end{aligned}$$

where we have used $\eta L^{-1\mathrm{T}} = L\eta$ in the third step and changed dummies in the last. We are free to swap the dummies $\hat\rho$ and $\hat\sigma$ inside the square bracket, due to the symmetry of $\eta_{\hat\rho\hat\sigma}$. We now have

$$C'_{\hat\alpha\hat\beta\hat\gamma} = L_{\hat\alpha}{}^{\hat\tau} L_{\hat\beta}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} C_{\hat\tau\hat\rho\hat\sigma} + \eta_{\hat\rho\hat\sigma} \Big[L_{\hat\alpha}{}^{\hat\rho} L_{\hat\beta}{}^{\hat\tau} \big(\partial_{\hat\tau} L_{\hat\gamma}{}^{\hat\sigma} \big) - L_{\hat\alpha}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\tau} \big(\partial_{\hat\tau} L_{\hat\beta}{}^{\hat\sigma} \big) \Big] \, . \quad (4.141)$$

But since $L\eta L^{\mathrm{T}} = \eta$, we also have results like

$$\begin{aligned} \eta_{\hat\rho\hat\sigma} L_{\hat\alpha}{}^{\hat\rho} \partial_{\hat\tau} L_{\hat\gamma}{}^{\hat\sigma} &= \partial_{\hat\tau} \big(\eta_{\hat\rho\hat\sigma} L_{\hat\alpha}{}^{\hat\rho} L_{\hat\gamma}{}^{\hat\sigma} \big) - \eta_{\hat\rho\hat\sigma} L_{\hat\gamma}{}^{\hat\sigma} \partial_{\hat\tau} L_{\hat\alpha}{}^{\hat\rho} \\ &= -\eta_{\hat\rho\hat\sigma} L_{\hat\gamma}{}^{\hat\sigma} \partial_{\hat\tau} L_{\hat\alpha}{}^{\hat\rho} \, , \end{aligned}$$

and likewise,

$$\eta_{\hat\rho\hat\sigma} L_{\hat\alpha}{}^{\hat\rho} \partial_{\hat\tau} L_{\hat\beta}{}^{\hat\sigma} = -\eta_{\hat\rho\hat\sigma} L_{\hat\beta}{}^{\hat\sigma} \partial_{\hat\tau} L_{\hat\alpha}{}^{\hat\rho} \, .$$

Hence we finally arrive at the slightly tidier result

$$\boxed{C'_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = L_{\hat{\alpha}}{}^{\hat{\tau}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} C_{\hat{\tau}\hat{\rho}\hat{\sigma}} + 2\eta_{\hat{\rho}\hat{\sigma}} L_{[\hat{\beta}}{}^{\hat{\sigma}} L_{\hat{\gamma}]}{}^{\hat{\tau}} \partial_{\hat{\tau}} L_{\hat{\alpha}}{}^{\hat{\rho}} \ .} \tag{4.142}$$

Transformation of Connection Coefficients

For the transformation of $\Gamma_{\hat{a}\hat{b}\hat{c}}$, we refer to the formula (4.15) on p. 159, viz.,

$$\Gamma'^{\hat{\varepsilon}}{}_{\hat{\beta}\hat{\gamma}} = L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\delta}} \Gamma^{\hat{\delta}}{}_{\hat{\rho}\hat{\sigma}} - L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} n_{\hat{\sigma}} \left(L^{-1\hat{\varepsilon}}{}_{\hat{\rho}} \right) , \tag{4.143}$$

whence

$$\Gamma'_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = \eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\delta}} \Gamma^{\hat{\delta}}{}_{\hat{\rho}\hat{\sigma}} - \eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} n_{\hat{\sigma}} \left(L^{-1\hat{\varepsilon}}{}_{\hat{\rho}} \right) . \tag{4.144}$$

As for the structure constants, we have

$$\eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} L^{-1\hat{\varepsilon}}{}_{\hat{\delta}} \Gamma^{\hat{\delta}}{}_{\hat{\rho}\hat{\sigma}} = L_{\hat{\alpha}}{}^{\hat{\sigma}} L_{\hat{\beta}}{}^{\hat{\tau}} L_{\hat{\gamma}}{}^{\hat{\rho}} \Gamma_{\hat{\sigma}\hat{\tau}\hat{\rho}} \ .$$

The manipulations with the other term are not identical, but they are very similar. We have

$$\begin{aligned} -\eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\beta}}{}^{\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} n_{\hat{\sigma}} \left(L^{-1\hat{\varepsilon}}{}_{\hat{\rho}} \right) &= \eta_{\hat{\alpha}\hat{\varepsilon}} L_{\hat{\gamma}}{}^{\hat{\sigma}} \left(\partial_{\hat{\sigma}} L_{\hat{\beta}}{}^{\hat{\rho}} \right) L^{-1\hat{\varepsilon}}{}_{\hat{\rho}} \\ &= L_{\hat{\alpha}}{}^{\hat{\tau}} \eta_{\hat{\tau}\hat{\rho}} L_{\hat{\gamma}}{}^{\hat{\sigma}} \partial_{\hat{\sigma}} L_{\hat{\beta}}{}^{\hat{\rho}} \\ &= -\eta_{\hat{\tau}\hat{\rho}} \left(\partial_{\hat{\sigma}} L_{\hat{\alpha}}{}^{\hat{\tau}} \right) L_{\hat{\gamma}}{}^{\hat{\sigma}} L_{\hat{\beta}}{}^{\hat{\rho}} \\ &= -\eta_{\hat{\rho}\hat{\sigma}} \left(\partial_{\hat{\tau}} L_{\hat{\alpha}}{}^{\hat{\rho}} \right) L_{\hat{\gamma}}{}^{\hat{\tau}} L_{\hat{\beta}}{}^{\hat{\sigma}} , \end{aligned}$$

by the same kind of manipulations as before. Putting the results together, we have

$$\boxed{\Gamma'_{\hat{\alpha}\hat{\beta}\hat{\gamma}} = L_{\hat{\alpha}}{}^{\hat{\rho}} L_{\hat{\beta}}{}^{\hat{\sigma}} L_{\hat{\gamma}}{}^{\hat{\tau}} \Gamma_{\hat{\rho}\hat{\sigma}\hat{\tau}} - \eta_{\hat{\rho}\hat{\sigma}} L_{\hat{\beta}}{}^{\hat{\sigma}} L_{\hat{\gamma}}{}^{\hat{\tau}} \partial_{\hat{\tau}} L_{\hat{\alpha}}{}^{\hat{\rho}}} \ . \tag{4.145}$$

Transformation of Fermi Rotation Coefficients

We are only concerned here with Lorentz transformations that fix the tangent to the congruence everywhere, i.e., rotations of the spacelike triad. This means that $L_{\hat{0}}{}^{\hat{\tau}} = \delta_{\hat{0}}^{\hat{\tau}}$ for $\hat{\tau} \in \{0,1,2,3\}$ and $L_{\hat{a}}{}^{\hat{0}} = 0$ for $\hat{a} \in \{1,2,3\}$ as usual. Using (4.145), we have the following deduction:

$$
\begin{aligned}
\Lambda'_{\hat{a}\hat{b}} &= -\Gamma'_{\hat{a}\hat{b}\hat{0}} \\
&= -L_{\hat{a}}{}^{\hat{\rho}} L_{\hat{b}}{}^{\hat{\sigma}} L_{\hat{0}}{}^{\hat{\tau}} \Gamma_{\hat{\rho}\hat{\sigma}\hat{\tau}} + \eta_{\hat{\rho}\hat{\sigma}} L_{\hat{b}}{}^{\hat{\sigma}} L_{\hat{0}}{}^{\hat{\tau}} \partial_{\hat{\tau}} L_{\hat{a}}{}^{\hat{\rho}} \\
&= -L_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}} \Gamma_{\hat{r}\hat{s}\hat{0}} + \eta_{\hat{r}\hat{s}} L_{\hat{b}}{}^{\hat{s}} \partial_{\hat{0}} L_{\hat{a}}{}^{\hat{r}} \\
&= L_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}} \Lambda_{\hat{r}\hat{s}} + \delta_{\hat{r}\hat{s}} \dot{L}_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}} \,,
\end{aligned}
$$

or

$$
\boxed{\Lambda'_{\hat{a}\hat{b}} = L_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}} \Lambda_{\hat{r}\hat{s}} + \delta_{\hat{r}\hat{s}} \dot{L}_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}}}\,, \tag{4.146}
$$

where

$$
\dot{L}_{\hat{a}}{}^{\hat{r}} := \partial_{\hat{0}} L_{\hat{a}}{}^{\hat{r}} = X^{\mu} \frac{\partial}{\partial x^{\mu}} L_{\hat{a}}{}^{\hat{r}} \,. \tag{4.147}
$$

This shows that the Fermi rotation coefficients do not transform as Lorentz tensors, even under these purely rotational Lorentz transformations. We observe, however, that we can always solve the equation $\Lambda'_{\hat{a}\hat{b}} = 0$ by suitable choice of rotation $L_{\hat{a}}{}^{\hat{r}}$. So there is always a Fermi–Walker transported tetrad for any congruence.

The coefficients $\Gamma_{\hat{a}\hat{b}\hat{c}}$ do not transform as Lorentz tensors under space rotations either. By (4.145), we have

$$
\Gamma'_{\hat{a}\hat{b}\hat{c}} = L_{\hat{a}}{}^{\hat{r}} L_{\hat{b}}{}^{\hat{s}} L_{\hat{c}}{}^{\hat{t}} \Gamma_{\hat{r}\hat{s}\hat{t}} - \delta_{\hat{r}\hat{s}} L_{\hat{b}}{}^{\hat{s}} L_{\hat{c}}{}^{\hat{t}} \partial_{\hat{t}} L_{\hat{a}}{}^{\hat{r}} \,. \tag{4.148}
$$

The rotations $L_{\hat{b}}{}^{\hat{a}}$ have to satisfy

$$
\partial_{\hat{c}} L_{\hat{b}}{}^{\hat{a}} := n^{\mu}_{\hat{c}} \frac{\partial}{\partial x^{\mu}} L_{\hat{b}}{}^{\hat{a}} = 0 \,, \quad \hat{a}, \hat{b}, \hat{c} \in \{1,2,3\} \,, \tag{4.149}
$$

everywhere in order for $\Gamma_{\hat{a}\hat{b}\hat{c}}$ to behave as Lorentz tensors under the rotations.

4.4 Energy–Momentum Tensor of a Conservative Continuous Medium

It is interesting to see some of the above machinery in action for a crucial problem in general relativity, namely the description of energy density and the dynamics of a continuous fluid. This section is adapted from [14].

The energy–momentum density plays the role of source for the gravitational field. Indeed, matter in *all* its forms is coupled to the gravitational field. If we wish to find the gravitational field produced by a certain material system for which we do not have an action functional, we then need a general description of the system and its dynamical behavior to keep track of its energy and momentum content without necessarily knowing all fundamental aspects of its structure.

Here we consider briefly a phenomenological treatment of a conservative continuous medium, i.e., one in which there are no irreversible dissipative processes. We use the notation developed to discuss rigid motions of continuous media in Sects. 2.3

and 4.3.15, but we add two new elements. First of all, we postulate an orthonormal triad field $n_a{}^\mu$ defined *throughout* the medium and satisfying everywhere

$$n_a \cdot n_b = \delta_{ab} \;, \qquad n_a \cdot u = 0 \;, \qquad u^\mu = \frac{\partial}{\partial \tau} x^\mu(\xi, \tau) \;,$$

as in Sect. 4.3. Secondly, we postulate a scalar field w_0 equal at each point to the *proper energy density* at that point, i.e., the density of total energy (rest mass as well as internal energy) as viewed in the local rest frame of the medium defined by the $n_a{}^\mu$ at that point. We do not impose any additional conditions on the $n_a{}^\mu$, e.g., Fermi–Walker transport, beyond their orthonormality and orthogonality to u^μ. As before, τ is the proper time of the particle labelled by ξ and u^μ is the 4-velocity of that particle. We recall that ξ, τ can be used as coordinates for the region of spacetime occupied by the medium.

We assume that the dynamical behavior of the medium is determined solely by its proper energy density and its internal stresses, to be described shortly. The density w_0 is defined in a local Cartesian rest frame of the medium. The first task is to reexpress it relative to the arbitrary curvilinear coordinates x^μ of spacetime, and also the internal coordinate system provided by the labels ξ^i.

The transformations between the local Cartesian rest frame and the (in general curvilinear) frame of the ξ^i are described by the transformation coefficients

$$A_{ai} := n_{a\mu} x^\mu{}_{,i}$$

and their inverses

$$A^{-1i}{}_a = \xi^i{}_{,\mu} n_a{}^\mu \;.$$

That these are the inverses is proven as follows:

$$A^{-1i}{}_a A_{aj} = \xi^i{}_{,\mu} n_a{}^\mu n_{a\nu} x^\nu{}_{,j} = \xi^i{}_{,\mu}(\delta^\mu{}_\nu + u^\mu u_\nu) x^\nu{}_{,j} = \delta^i{}_j \;, \tag{4.150}$$

$$A_{ai} A^{-1i}{}_b = n_{a\mu} x^\mu{}_{,i} \xi^i{}_{,\nu} n_b{}^\nu = n_{a\mu}(\delta^\mu{}_\nu - \dot{x}^\mu \tau_{,\nu}) n_b{}^\nu = \delta_{ab} \;. \tag{4.151}$$

A key fact used here is that

$$n_{a\mu} n_{a\nu} = g_{\mu\nu} + u_\mu u_\nu =: P_{\mu\nu} \;, \tag{4.152}$$

the relativistic projection operator onto the instantaneous hyperplane of simultaneity of an observer moving with the fluid element at τ, ξ^i. This is basically the first relation of (4.41) on p.165. We note that

$$A_{ai} A_{aj} = n_{a\mu} x^\mu{}_{,i} n_{a\nu} x^\nu{}_{,j} = P_{\mu\nu} x^\mu{}_{,i} x^\nu{}_{,j} = \gamma_{ij} \;, \tag{4.153}$$

where γ_{ij} is the proper metric of the medium defined in (4.101) on p. 178. Hence, assuming as we may that the ξ axes have the same relative orientation as the vectors $n_a{}^\mu$,

$$\det(A_{ai}) = \gamma^{1/2} \;, \quad \text{where } \gamma = \det(\gamma_{ij}) \;.$$

As discussed on p. 201, it will then turn out that the proper energy density in the ξ coordinate system can be written

$$w_\xi := \det(A_{ai})w_0 = \gamma^{1/2} w_0 \ . \tag{4.154}$$

4.4.1 Three Reference Frames

It is a good thing to be clear about the three reference frames available here:

- We have a general coordinate system $\{x^\mu\}$ with associated frame $\{\partial/\partial x^\mu\}$.
- We also have a coordinate system $\{\tau, \xi^i\}$, where ξ^i, $i = 1,2,3$, label particles, and $x^\mu(\tau, \xi^i)$ is the worldline of the particle or fluid element labelled by ξ^i, with τ its proper time. The frame associated with the label coordinates $\{\tau, \xi^i\}$ is of course $\{\partial/\partial\tau, \partial/\partial\xi^i\}$. Now $u^\mu = \partial x^\mu/\partial\tau$ is the 4-velocity of the particle or fluid element labelled by ξ^i, fixed in working out the partial derivative $\partial x^\mu/\partial\tau$. Hence $u = \partial/\partial\tau$ and this frame is $\{u, \partial/\partial\xi^i\}$. Note that $\partial/\partial\xi^i$ has components $\partial x^\mu/\partial\xi^i = x^\mu{}_{,i}$ in the general coordinate frame.
- We have one other frame, that is not a coordinate frame, viz., $\{u, n_a\}$, which shares the timelike vector u with the label frame and is a tetrad frame.

Let us relate the two frames $\{u, n_a\}$ and $\{u, \partial/\partial\xi^i\}$. First observe that

$$\begin{aligned}
(A_{ai}n_a)^\nu &= n_a{}^\nu n_{a\mu} x^\mu{}_{,i} \\
&= (\delta^\nu{}_\mu + u^\nu u_\mu) x^\mu{}_{,i} \\
&= \frac{\partial x^\nu}{\partial\xi^i} + u^\nu u_\mu \frac{\partial x^\mu}{\partial\xi^i} \ ,
\end{aligned}$$

using the fact that

$$n_a{}^\nu n_{a\mu} = \delta^\nu{}_\mu + u^\nu u_\mu =: P^\nu{}_\mu \ ,$$

whence

$$\frac{\partial}{\partial\xi^i} + \left(u \cdot \frac{\partial}{\partial\xi^i}\right) u = A_{ai} n_a \ .$$

But

$$A_{ai} A^{-1i}{}_b = \delta_{ab} \ ,$$

so we have

$$A_{ai} n_a A^{-1i}{}_b = n_b \ ,$$

and hence

$$n_a = A^{-1i}{}_a \left[\frac{\partial}{\partial\xi^i} + \left(u \cdot \frac{\partial}{\partial\xi^i}\right) u\right] \ .$$

Finally, we have the result

$$\boxed{n_a = A^{-1i}{}_a \frac{\partial}{\partial \xi^i} + \left(u \cdot \frac{\partial}{\partial \xi^i}\right) A^{-1i}{}_a u}\,, \tag{4.155}$$

expressing n_a in terms of the 4 frame vectors u, $\partial/\partial\xi^i$. We observe that some u is required even though $n_a \cdot u = 0$, because $u \cdot \partial/\partial\xi^i$ is not generally zero. Indeed it follows immediately from (4.155) that $n_a \cdot u = 0$.

Let us now obtain $\partial/\partial\xi^i$ in terms of n_a, u. We start with

$$\begin{aligned} A^{-1i}{}_a \frac{\partial}{\partial \xi^i} &= n_a{}^\mu \frac{\partial \xi^i}{\partial x^\mu} \frac{\partial}{\partial \xi^i} \\ &= n_a{}^\mu \left(\frac{\partial}{\partial x^\mu} - \frac{\partial \tau}{\partial x^\mu} \frac{\partial}{\partial \tau} \right) \\ &= n_a - \tau_{,\mu} n_a{}^\mu u\,. \end{aligned}$$

But

$$A_{aj} A^{-1i}{}_a \frac{\partial}{\partial \xi^i} = \delta^i{}_j \frac{\partial}{\partial \xi^i} = \frac{\partial}{\partial \xi^j}\,,$$

so

$$\boxed{\frac{\partial}{\partial \xi^i} = A_{ai} n_a - A_{ai} \tau_{,\mu} n_a{}^\mu u}\,. \tag{4.156}$$

Once again, some u is required in the mix.

So far, A is a 3×3 matrix. Although we shall not need it in what follows, it can be extended naturally to a 4×4 matrix with components

$$\left(A^\alpha{}_\phi\right) = \begin{pmatrix} 1 & -u \cdot \partial/\partial\xi^1 & -u \cdot \partial/\partial\xi^2 & -u \cdot \partial/\partial\xi^2 \\ 0 & & & \\ 0 & & (A^a{}_i) & \\ 0 & & & \end{pmatrix}\,, \tag{4.157}$$

where $\alpha \in \{0,1,2,3\}$ labels rows and $\phi \in \{0,1,2,3\}$ labels columns, and we have raised the a on A_{ai} with impunity using η, so in fact $A^a{}_i = A_{ai}$, for $a, i \in \{1,2,3\}$. Then extending the tetrad frame notation so that $n_0 := u$ and using a Greek index α to cover all four members of the tetrad frame, we have

$$u = n_\alpha A^\alpha{}_0\,, \quad \frac{\partial}{\partial \xi^i} = n_\alpha A^\alpha{}_i\,. \tag{4.158}$$

Then A is just the matrix that transforms from the tetrad frame to the label coordinate frame. The first relation of (4.158) reads $u = n_0$, since $A^a{}_0 = 0$, for $a = 1,2,3$, while the second relation reads

$$\begin{aligned}\frac{\partial}{\partial\xi^i} &= uA^0{}_i + n_a A^a{}_i = -\left(u\cdot\frac{\partial}{\partial\xi^i}\right)u + n_a n_{a\mu}\frac{\partial x^\mu}{\partial\xi^i} \\ &= -\left(u\cdot\frac{\partial}{\partial\xi^i}\right)u + \left(n_a\cdot\frac{\partial}{\partial\xi^i}\right)n_a\ ,\end{aligned}$$

which is just the expansion of $\partial/\partial\xi^i$ in terms of the tetrad basis $\{u, n_a\}_{a=1,2,3}$.

It is easy to check that the above matrix A has inverse

$$\left(A^{-1\phi}{}_\alpha\right) = \begin{pmatrix} 1 & \tau_{,\mu}n_1{}^\mu & \tau_{,\mu}n_2{}^\mu & \tau_{,\mu}n_3{}^\mu \\ 0 & & & \\ 0 & & (A^{-1i}{}_a) & \\ 0 & & & \end{pmatrix}\ , \tag{4.159}$$

where $\phi \in \{0,1,2,3\}$ now labels rows and $\alpha \in \{0,1,2,3\}$ labels columns. Indeed, it is obvious that $A^0{}_\phi A^{-1\phi}{}_0 = 1$, while

$$\begin{aligned}A^0{}_\phi A^{-1\phi}{}_a &= \tau_{,\mu}n_a{}^\mu - \left(u\cdot\frac{\partial}{\partial\xi^i}\right)\xi^i{}_{,\mu}n_a{}^\mu \\ &= \tau_{,\mu}n_a{}^\mu - u\cdot\left(\frac{\partial\xi^i}{\partial x^\mu}\frac{\partial}{\partial\xi^i}\right)n_a{}^\mu \\ &= \tau_{,\mu}n_a{}^\mu - u\cdot\left(\frac{\partial}{\partial x^\mu} - \frac{\partial\tau}{\partial x^\mu}\frac{\partial}{\partial\tau}\right)n_a{}^\mu \\ &= \tau_{,\mu}n_a{}^\mu - u_\mu n_a{}^\mu + (u\cdot u)\tau_{,\mu}n_a{}^\mu \\ &= 0\ ,\quad \text{since } u\cdot n_a = 0 \text{ and } u^2 = -1\ .\end{aligned}$$

Finally, $A^a{}_\phi A^{-1\phi}{}_b = A^a{}_i A^{-1i}{}_b = \delta^a_b$ by (4.151). So A is a left inverse for A^{-1} and hence also a right inverse.

4.4.2 Useful Identities

Returning now to the 3×3 matrices $A^{-1i}{}_a$ $(:= \xi^i{}_{,\mu}n_a{}^\mu)$ and A_{ai} $(:= n_{a\mu}x^\mu{}_{,i})$, we have the following identities:

$$\boxed{A^{-1i}{}_a A^{-1j}{}_a = \gamma^{ij}\ ,\qquad \gamma_{ij}A^{-1i}{}_a A^{-1j}{}_b = \delta_{ab}\ ,\qquad \gamma^{ij}A_{ai}A_{bj} = \delta_{ab}}\ , \tag{4.160}$$

where γ^{ij} is the contravariant proper metric tensor, inverse to γ_{ij}. As proof, consider

$$\gamma_{ik}A^{-1k}{}_a A^{-1j}{}_a = A_{bi}A_{bk}A^{-1k}{}_a A^{-1j}{}_a = A_{ai}A^{-1j}{}_a = \delta_i{}^j\ ,$$

$$\gamma_{ij}A^{-1i}{}_a A^{-1j}{}_b = A_{ci}A_{cj}A^{-1i}{}_a A^{-1j}{}_b = \delta_{ca}\delta_{cb} = \delta_{ab}\ ,$$

$$\gamma^{ij}A_{ai}A_{bj} = A^{-1i}{}_c A^{-1j}{}_c A_{ai}A_{bj} = \delta_{ca}\delta_{cb} = \delta_{ab} \ ,$$

as required. ■

If we use the notation

$$\begin{pmatrix} u^\mu \\ n_a{}^\mu \end{pmatrix} := \begin{pmatrix} u^0 & u^1 & u^2 & u^3 \\ n_1{}^0 & n_1{}^1 & n_1{}^2 & n_1{}^3 \\ n_2{}^0 & n_2{}^1 & n_2{}^2 & n_2{}^3 \\ n_3{}^0 & n_3{}^1 & n_3{}^2 & n_3{}^3 \end{pmatrix} ,$$

then we have

$$\begin{pmatrix} -u^\mu \\ n_a{}^\mu \end{pmatrix}^{\mathrm{tr}} \begin{pmatrix} u^\nu \\ n_a{}^\nu \end{pmatrix} = (-u^\mu u^\nu + n_a{}^\mu n_a{}^\nu) = (g^{\mu\nu}) \ , \tag{4.161}$$

by (4.152), and this implies

$$-\left[\det \begin{pmatrix} u^\mu \\ n_a{}^\mu \end{pmatrix}\right]^2 = \det(g^{\mu\nu}) = -g^{-1} \ ,$$

where we define $g := -\det g_{\mu\nu}$. Hence, assuming u^μ, $n_1{}^\mu$, $n_2{}^\mu$, $n_3{}^\mu$ to have respectively the same relative orientation as positive displacements along the x^0, x^1, x^2, x^3 axes,

$$\det \begin{pmatrix} u^\mu \\ n_a{}^\mu \end{pmatrix} = g^{-1/2} \ . \tag{4.162}$$

From this and the fact that

$$\begin{pmatrix} -u_\mu \\ n_{a\mu} \end{pmatrix}^{\mathrm{tr}} = \begin{pmatrix} u^\mu \\ n_a{}^\mu \end{pmatrix}^{-1} ,$$

because

$$\begin{pmatrix} -u_\mu \\ n_{a\mu} \end{pmatrix}^{\mathrm{tr}} \begin{pmatrix} u^\nu \\ n_a{}^\nu \end{pmatrix} = \left(-u_\mu u^\nu + n_{a\mu} n_a{}^\nu\right) = (\delta_\mu{}^\nu) \ ,$$

by (4.152) once again, it follows quite unsurprisingly that

$$\det \begin{pmatrix} -u_\mu \\ n_{a\mu} \end{pmatrix} = g^{1/2} \ . \tag{4.163}$$

Now let us show that

$$\boxed{\varepsilon_{abc} n_{a\mu} n_{b\nu} n_{c\sigma} = {}^{-1}\varepsilon_{\tau\mu\nu\sigma} g^{1/2} u^\tau} \ , \tag{4.164}$$

where the -1 on the 4D permutation symbol ${}^{-1}\varepsilon_{\tau\mu\nu\sigma}$ reminds us that it is a tensor density of weight -1, in the sense that it transforms according to

$$ {}^{-1}\overline{\varepsilon}_{\mu_1\dots\mu_n} = \frac{\partial(\overline{x})}{\partial(x)}\frac{\partial x^{\nu_1}}{\partial\overline{x}^{\mu_1}}\cdots\frac{\partial x^{\nu_n}}{\partial\overline{x}^{\mu_n}}{}^{-1}\varepsilon_{\nu_1\dots\nu_n} = {}^{-1}\varepsilon_{\mu_1\dots\mu_n} \ . $$

As a preamble to the proof, it is interesting to demonstrate (4.152) in the following way. Recall that

$$ g_{\mu\nu}u^\mu u^\nu = -1 \ , \qquad n_a^\mu n_{b\mu} = \delta_{ab} \ , \qquad n_{a\mu}u^\mu = 0 \ , $$

and from this we wish to deduce that

$$ n_a^\mu n_a^\nu = g^{\mu\nu} + u^\mu u^\nu \ . \tag{4.165} $$

A simple proof is found by choosing, for any event z in spacetime, a coordinate frame at z in which $g(z) = \eta$ and

$$ u^\mu = (1000) \ , \quad n_1^\mu = (0100) \ , \quad n_2^\mu = (0010) \ , \quad n_3^\mu = (0001) \ . \tag{4.166} $$

Now, at z, $n_a^\mu n_a^\mu$ can be written as the matrix

$$ n_a^\mu n_a^\nu = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \ , $$

with μ the row index and ν the column index. In the same way,

$$ g^{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \ , \qquad u^\mu u^\nu = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \ . $$

We then find that (4.165) holds as a matrix relation in this frame, and since it is a tensorial relation, it must hold in all frames.

We can prove that

$$ \varepsilon_{\mu\nu\rho\sigma}u^\sigma = -g^{-1/2}(z)\varepsilon_{abc}n_{a\mu}n_{b\nu}n_{c\rho} \tag{4.167} $$

in a similar way. We first check the result in the above frame. $\varepsilon_{\mu\nu\rho\sigma}u^\sigma$ is equal to zero unless $\{\mu,\nu,\rho\} = \{1,2,3\}$ (6 possibilities). The same is true of $\varepsilon_{abc}n_{a\mu}n_{b\nu}n_{c\rho}$ because $n_{a0} = 0$. Furthermore, $\varepsilon_{1230}u^0 = -1$ and

$$ \varepsilon_{abc}n_{a1}n_{b2}n_{c3} = \det\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} = 1 \ . $$

We also note that $g^{-1/2}(z) = 1$ in this frame. The result (4.167) therefore holds in this frame. Now in some primed frame,

$$\begin{aligned}
-g'^{-1/2}(z)\varepsilon_{abc}n'_{a\mu'}n'_{b\nu'}n'_{c\rho'} &= -\det\frac{\partial z'}{\partial z}g^{-1/2}\varepsilon_{abc}\frac{\partial z^{\varepsilon}}{\partial z^{\mu'}}n_{a\varepsilon}\frac{\partial z^{\phi}}{\partial z^{\nu'}}n_{c\phi}\frac{\partial z^{\lambda}}{\partial z^{\rho'}}n_{c\lambda} \\
&= -\left(\det\frac{\partial z'}{\partial z}\right)\frac{\partial z^{\varepsilon}}{\partial z^{\mu'}}\frac{\partial z^{\phi}}{\partial z^{\nu'}}\frac{\partial z^{\lambda}}{\partial z^{\rho'}}g^{-1/2}\varepsilon_{abc}n_{a\varepsilon}n_{b\phi}n_{c\lambda} \\
&= \left(\det\frac{\partial z'}{\partial z}\right)\frac{\partial z^{\varepsilon}}{\partial z^{\mu'}}\frac{\partial z^{\phi}}{\partial z^{\nu'}}\frac{\partial z^{\lambda}}{\partial z^{\rho'}}\varepsilon_{\varepsilon\phi\lambda\iota}u^{\iota} \\
&= \left(\det\frac{\partial z'}{\partial z}\right)\frac{\partial z^{\varepsilon}}{\partial z^{\mu'}}\frac{\partial z^{\phi}}{\partial z^{\nu'}}\frac{\partial z^{\lambda}}{\partial z^{\rho'}}\frac{\partial z^{\iota}}{\partial z^{\sigma'}}\varepsilon_{\varepsilon\phi\lambda\iota}u^{\sigma'} \\
&= \det\frac{\partial z'}{\partial z}\det\frac{\partial z}{\partial z'}\varepsilon_{\mu'\nu'\rho'\sigma'}u^{\sigma'} \\
&= \varepsilon_{\mu'\nu'\rho'\sigma'}u^{\sigma'}\ ,
\end{aligned}$$

which shows that the result (4.167) is tensorial and must hold in every frame if it holds in one. ■

We can now use (4.164) to make the deduction

$$\begin{aligned}
\varepsilon_{ijk}\gamma^{1/2} &= \varepsilon_{ijk}\det(A_{ab}) = \varepsilon_{abc}A_{ai}A_{bj}A_{ck} \\
&= \varepsilon_{abc}n_{a\mu}n_{b\nu}n_{c\sigma}x^{\mu}{}_{,i}x^{\nu}{}_{,j}x^{\sigma}{}_{,k} \\
&= {}^{-1}\varepsilon_{\tau\mu\nu\sigma}g^{1/2}u^{\tau}x^{\mu}{}_{,i}x^{\nu}{}_{,j}x^{\sigma}{}_{,k} \\
&= \varepsilon_{ijk}g^{1/2}\frac{\partial(x)}{\partial(\tau,\xi)}\ ,
\end{aligned}$$

so that

$$\boxed{\gamma^{1/2} = g^{1/2}\frac{\partial(x)}{\partial(\tau,\xi)}}\ . \tag{4.168}$$

4.4.3 Energy Density

The last relation enables us to write

$$w_{\xi} = \gamma^{1/2}w_0 = \frac{\partial(x)}{\partial(\tau,\xi)}w\ , \tag{4.169}$$

where w is the proper energy density of the medium relative to the coordinates x^{μ}, defined by

$$\boxed{w := g^{1/2}w_0}\ . \tag{4.170}$$

Note that w_0 is a scalar under both transformations of the ξ (relabelling) and transformations of the x^μ, whereas w_ξ is a scalar under transformations of the x^μ but transforms as a density of unit weight under transformations of the ξ. The quantity w is a scalar under transformations of the ξ but transforms as a density of unit weight under transformations of the x^μ.

A few comments are perhaps in order. Consider first the physical interpretation of w_0, w_ξ, and w. Let $\mathrm{d}^3\xi$ be an infinitesimal coordinate volume in the label coordinates, for a volume element at some (τ_0, ξ_0), keeping τ constant through the region, i.e., the volume element is

$$\mathscr{E}_{(\tau_0,\xi_0)} = \Big\{(\tau,\xi) : \tau = \tau_0\ , \quad \xi^i \in \big(\xi_0^i - \delta^i, \xi_0^i + \delta^i\big)\ ,\ i = 1,2,3\Big\}\ ,$$

where $2\delta^i = \mathrm{d}\xi^i$. An observer moving with the particle labelled by ξ_0 considers the element to be more or less at rest, because nearby particles move with infinitesimally close values of their 4-velocity.

The first claim here is that $w_\xi \mathrm{d}^3\xi = \gamma^{1/2} w_0 \mathrm{d}^3\xi$ is the rest mass energy plus internal energy contained in the element. But of course $\gamma^{1/2}\mathrm{d}^3\xi$ is the proper volume of this element. This was precisely the idea behind the definition $\gamma_{ij} := P_{\mu\nu} x^\mu{}_{,i} x^\nu{}_{,j}$ on p. 178. Although the observer moving with ξ_0 considers the element to be at rest, $\mathrm{d}^3\xi$ does not give its proper 3-volume for this observer. We know that the quantity $\mathrm{d}s^2 = \gamma_{ij}\mathrm{d}\xi^i \mathrm{d}\xi^j$ gives the distance $\mathrm{d}s$ by which the two particles ξ and $\xi + \mathrm{d}\xi$ appear to be separated in the instantaneous rest frame of either. This means that $\gamma := \det \gamma_{ij}$ is the volume factor converting the coordinate volume $\mathrm{d}^3\xi$ to a proper volume in the rest frame of ξ_0.

If we multiply this proper volume by the scalar field w_0 equal at each point to the *proper energy density* at that point, i.e., the density of total energy (rest mass as well as internal energy) as viewed in the local rest frame defined by the $n_a{}^\mu$ at that point, we obtain the rest mass plus internal energy contained in the element. Note that the existence of a smooth label coordinate system does suggest no random motions within the volume element and hence no internal energy, but just rest mass. On the other hand, one may just be labelling fluid elements, within which there are random motions, so that the quantity delivered here does include internal energy. This needs to be borne in mind when the theory is applied.

The quantity $w := g^{1/2} w_0$ is interpreted as the proper energy density of the medium relative to the coordinates x^μ. But what does this mean? Is the idea that $w\mathrm{d}^3 x$ should be the energy (rest mass plus internal) in a volume element throughout which x^0 is kept constant, like the one above for the label coordinates? Such an interpretation could not be valid, because $g^{1/2}$ delivers volumes of spacetime via $g^{1/2}\mathrm{d}^4 x$, and not volumes of space in spacelike hypersurfaces, in the general case it is supposed to represent. Clearly, there is no simple physical interpretation of this object w for general coordinates, something that should be borne in mind when it is put to use.

Note, however, that the definition $w := g^{1/2} w_0$ is at least consistent in some sense with the definition $w_\xi = \gamma^{1/2} w_0$, for we may take our general coordinates $\{x^\mu\}$ to be the label coordinates $\{\tau, \xi^i\}$ and use the result (4.168) proven above, viz.,

$$\gamma^{1/2} = g^{1/2} \frac{\partial(x)}{\partial(\tau, \xi)} ,$$

to conclude that, when the general coordinates are just the label coordinates, we have $\gamma^{1/2} = g^{1/2}$. This is just saying that $\det \gamma_{ij}$ is the determinant of the metric in the label coordinate system, up to a sign. And of course in this case $w = w_\xi$ as we have defined it above. So there are some specific cases in which w has a physical interpretation.

Concerning the claims that w_0 is a scalar under both transformations of the ξ (relabelling) and transformations of the x^μ, whereas w_ξ is a scalar under transformations of the x^μ but transforms as a density of unit weight under transformations of the ξ :

- w_0 is a scalar because it is defined in the unique (up to space rotation) rest frame.
- w_ξ is scalar under transformations of the x^μ because γ_{ij} is a scalar field on the region of spacetime occupied by the medium.
- w_ξ transforms as a density of unit weight under transformations of the ξ for the following reason. If $\{\tau, \xi_i'\}$ are new label coordinates, then

$$\gamma'^{1/2} = \frac{\partial(\tau, \xi^i)}{\partial(\tau, \xi'^i)} \gamma^{1/2} ,$$

whence

$$w_{\xi'} = \gamma'^{1/2} w_0 = \frac{\partial(\tau, \xi^i)}{\partial(\tau, \xi'^i)} \gamma^{1/2} w_0 = \frac{\partial(\tau, \xi)}{\partial(\tau, \xi')} w_\xi ,$$

which is indeed the transformation rule for a density of weight 1.

Concerning the claims that the quantity w is a scalar under transformations of the ξ but transforms as a density of unit weight under transformations of the x^μ :

- w is a scalar under transformations of the ξ for the simple reason that its definition $w := g^{1/2} w_0$ makes no reference to the label coordinates.
- w transforms as a density of unit weight under general coordinate transformations because w_0 is a scalar and $g^{1/2}$ transforms as a density of unit weight under these coordinate transformations.

4.4.4 Stress Tensor

We now ask how the proper energy density varies with time. The argument in this section is a fleshed out version of the one given in [14, Chap. 10]. If the medium is *conservative*, which means that energy does not flow around by dissipative mechanisms, w_0 can vary only as a result of the action of forces on the component parts of the medium. These forces can be described phenomenologically by means of a *stress tensor*.

Let us suppose to begin with that the coordinates x^μ have been chosen to be canonical at a certain point x, oriented in such a way that

$$\begin{pmatrix} u^\mu \\ n_a{}^\mu \end{pmatrix}$$

becomes the unit matrix at x [see (4.166) on p. 199], and adjusted in the neighbourhood of x so that the derivatives of the metric tensor vanish at x. Then the coordinates x^μ may be regarded as an extension of the local Minkowski frame, which strictly speaking has mathematical existence only in the tangent space, to a small neighborhood of x.

So basically, given x, we can choose coordinates $\{x^\mu\}$ such that

$$u\big|_x = \frac{\partial}{\partial x^0}\bigg|_x \,, \qquad n_1\big|_x = \frac{\partial}{\partial x^1}\bigg|_x \,, \qquad n_2\big|_x = \frac{\partial}{\partial x^2}\bigg|_x \,, \qquad n_3\big|_x = \frac{\partial}{\partial x^3}\bigg|_x \,,$$

and we can also arrange these coordinates so that the derivatives of the metric tensor are zero at x, whence the Levi-Civita connection coefficients are also zero at x. Bear in mind, however, that the derivatives of the metric and Levi-Civita connection coefficients can only be made exactly equal to zero on a whole neighbourhood of x if the spacetime is flat there.

Now let $\mathrm{d}\Sigma_a$ be a directed 2D surface element in this frame. Then one would intuitively expect the material on the side of $\mathrm{d}\Sigma_a$ away from the direction in which $\mathrm{d}\Sigma_a$ points to exert on the material on the opposite side a force that depends linearly on $\mathrm{d}\Sigma_a$:

$$\boxed{\mathrm{d}F_a = t_{ab}\mathrm{d}\Sigma_b} \,. \tag{4.171}$$

The coefficients t_{ab} of the linear dependence are the components of the stress tensor in the local Minkowski rest frame.

Of course, this looks nice as long as one does not think too much about it! However, it reveals something about the way physics is often done. One has an intuition for what the right formulation should be, from simpler situations, e.g., a flat or even non-relativistic spacetime, and one does not worry too much about the actual physical interpretation in the more sophisticated context (here, a curved spacetime). The definition (4.171) immediately raises the question as to whether t_{ab} will be symmetric in its indices, on the basis of another intuition. Keep reading for the answer.

The force $\mathrm{d}F_a$ is a contact force and as such is expected to respect the law of action and reaction. This means that the material on the side of $\mathrm{d}\Sigma_a$ *toward* which $\mathrm{d}\Sigma_a$ points must exert a force $-\mathrm{d}F_a$ across $\mathrm{d}\Sigma_a$. As a consequence the total force experienced by a small volume V of the medium, as a result of the action of the surrounding medium, is given by

$$F_a = -\int_\Sigma \mathrm{d}F_a = -\int_\Sigma t_{ab}\mathrm{d}\Sigma_b = -\int_V t_{ab,b}\mathrm{d}^3x \,, \tag{4.172}$$

where Σ is the surface of V. Here V is assumed to contain the point x and the derivative in the final integrand is taken with respect to the extended local coordinates.

According to DeWitt [14], because V is otherwise arbitrary, it is evident that the internal stresses which the tensor t_{ab} describes give rise to a net force density in the immediate vicinity of x given by

$$\boxed{f_a = -t_{ab,b}}\ . \tag{4.173}$$

This is not actually quite correct, as we shall see shortly. However, it will not raise problems for the rest of DeWitt's argument, exposed hereafter.

Suppose the origin of the coordinates x^μ is taken at the point x. Then, lowering the spatial indices on the x^μ, we may express the torque about x, exerted on V by the surrounding medium, in the form

$$\begin{aligned} T_a &:= -\int_\Sigma \varepsilon_{abc} x_b \mathrm{d}F_c = -\varepsilon_{abc}\int_\Sigma x_b t_{cd}\mathrm{d}\Sigma_d \\ &= -\varepsilon_{abc}\int_V (x_b t_{cd})_{,d}\mathrm{d}^3x = T_a^{\mathrm{I}} + T_a^{\mathrm{II}}\ , \end{aligned} \tag{4.174}$$

where

$$T_a^{\mathrm{I}} := \varepsilon_{abc}\int_V x_b f_c \mathrm{d}^3x\ , \qquad T_a^{\mathrm{II}} := \varepsilon_{abc}\int_V t_{bc}\mathrm{d}^3x\ . \tag{4.175}$$

The value of T_a^{I} depends on the location of the origin. It is what one would expect to get for the torque using the force density f_a. The value of T_a^{II}, on the other hand, is independent of the location of the origin, and is in fact an unwanted residual. We argue that it must be zero by the following argument. In the limit $V \to 0$, it may be expressed simply as

$$T_a^{\mathrm{II}} = V\varepsilon_{abc}t_{bc}\ .$$

On the other hand, the moment of inertia of V is of the order

$$I \sim w_0 V^{5/3}\ .$$

To understand this estimate, consider the moment of inertia $\sim mr^2$ of a solid sphere of mass m and radius r. The mass m is proportional to the volume, which goes as r^3, so the moment of inertia goes as r^5, hence as $V^{5/3}$. The residual torque therefore imparts a contribution to the angular acceleration of V given by

$$\dot{w}_a^{\mathrm{II}} = \frac{T_a^{\mathrm{II}}}{I} \sim V^{-2/3}w_0^{-1}\varepsilon_{abc}t_{bc}\ .$$

But unless $\varepsilon_{abc}t_{bc} = 0$, this becomes infinite as $V \to 0$, which is absurd, and we thus conclude that

$$\varepsilon_{abc}t_{bc} = 0\ ,$$

or, alternatively,

$$t_{ab} - t_{ba} = (\delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc})t_{cd} = \varepsilon_{abe}\varepsilon_{ecd}t_{cd} = 0\ .$$

That is, the stress tensor is necessarily symmetric.

The symmetry of the stress tensor may be illustrated in the particularly simple case of a gas at equilibrium, where we obviously have

$$t_{ab} = p\delta_{ab} \;, \tag{4.176}$$

p being the pressure. We note that p, like w_0, is a scalar field.

So far we have obtained the stress tensor in the local Minkowski rest frame, but like the energy density, it can also be expressed in the ξ coordinate system and in the general coordinate system x^μ. When viewed in an arbitrary coordinate system, unlike the energy density, it must be regarded as a tensor density, known as the *stress density*. The relevant definitions are then

$$\boxed{t^{ij} := \gamma^{1/2} A^{-1i}{}_a A^{-1j}{}_b t_{ab} \;, \qquad t^{\mu\nu} := g^{1/2} n_a{}^\mu n_b{}^\nu t_{ab}} \;. \tag{4.177}$$

We note that

$$t^{\mu\nu} u_\nu = 0 \;.$$

Once again, these are just definitions, but the real question is, what can we do with them? The definition of t^{ij} can be motivated as follows:

$$\begin{aligned} t_{ab} n_a \otimes n_b &= t_{ab} \left[A^{-1i}{}_a \frac{\partial}{\partial \xi^i} + O_a u \right] \otimes \left[A^{-1j}{}_b \frac{\partial}{\partial \xi^j} + O_b u \right] \\ &= t_{ab} A^{-1i}{}_a A^{-1j}{}_b \frac{\partial}{\partial \xi^i} \otimes \frac{\partial}{\partial \xi^j} + \text{terms in } \frac{\partial}{\partial \xi^i} \otimes u,\; u \otimes \frac{\partial}{\partial \xi^i}, \text{ and } u \otimes u \;, \end{aligned}$$

where O_a is a shorthand for

$$O_a := \left(u \cdot \frac{\partial}{\partial \xi^i} \right) A^{-1i}{}_a \;,$$

and we have used (4.155) on p. 196. So $t^{ij} \partial/\partial\xi^i \otimes \partial/\partial\xi^j$ is the piece of $t_{ab} n_a \otimes n_b$ in the $\partial/\partial\xi^i \otimes \partial/\partial\xi^j$ direction, so to speak, but with a factor of $\gamma^{1/2}$ inserted, which reminds us that this is a density.

Regarding $t^{\mu\nu}$, note that

$$t_{ab} n_a \otimes n_b = t_{ab} n_a{}^\mu n_b{}^\nu \frac{\partial}{\partial x^\mu} \otimes \frac{\partial}{\partial x^\nu} = g^{-1/2} t^{\mu\nu} \frac{\partial}{\partial x^\mu} \otimes \frac{\partial}{\partial x^\nu} \;.$$

Once again, the factor of $g^{1/2}$ in the definition of $t^{\mu\nu}$ reminds us that it is a density.

There does remain the question of what these mean physically and what one can do with them, but for the moment, let us show that

$$\boxed{t^{ij} = \frac{\partial(x)}{\partial(\tau,\xi)} \xi^i{}_{,\mu} \xi^j{}_{,\nu} t^{\mu\nu} \;, \qquad t^{\mu\nu} = \frac{\partial(\tau,\xi)}{\partial(x)} P^\mu{}_\sigma P^\nu{}_\tau x^\sigma{}_{,i} x^\tau{}_{,j} t^{ij}} \;. \tag{4.178}$$

We have

$$\begin{aligned} t^{ij} &= \gamma^{1/2} A^{-1i}{}_a A^{-1j}{}_b t_{ab} = \gamma^{1/2} \xi^i{}_{,\mu} n_a{}^\mu \xi^j{}_{,\nu} n_b{}^\nu t_{ab} \\ &= \gamma^{1/2} g^{-1/2} \xi^i{}_{,\mu} \xi^j{}_{,\nu} t^{\mu\nu} = \frac{\partial(x)}{\partial(\tau,\xi)} \xi^i{}_{,\mu} \xi^j{}_{,\nu} t^{\mu\nu} , \end{aligned}$$

using (4.168) on p. 200, and

$$t_{ab} = \gamma^{-1/2} A_{ai} A_{bj} t^{ij} = \gamma^{-1/2} n_{a\mu} x^\mu{}_{,i} n_{b\nu} x^\nu{}_{,j} t^{ij} ,$$

whence

$$\begin{aligned} t^{\mu\nu} &= g^{1/2} n_a{}^\mu n_b{}^\nu t_{ab} = g^{1/2} n_a{}^\mu n_b{}^\nu \gamma^{-1/2} n_{a\sigma} x^\sigma{}_{,i} n_{b\tau} x^\tau{}_{,j} t^{ij} \\ &= \frac{\partial(\tau,\xi)}{\partial(x)} P^\mu{}_\sigma P^\nu{}_\tau x^\sigma{}_{,i} x^\tau{}_{,j} t^{ij} , \end{aligned}$$

as required.

In the two relations of (4.178), we basically transform from coordinates $\{x^\mu\}$ to coordinates $\{\tau, \xi^i\}$ and back again. The Jacobian factors show that we have a density of weight 1. Note also that there are 16 components of $t^{\mu\nu}$ and only 9 components of t^{ij}. The projectors $P^\mu{}_\sigma$ and $P^\nu{}_\tau$ in the second relation ensure that

$$u_\mu t^{\mu\nu} = 0 = t^{\mu\nu} u_\nu . \tag{4.179}$$

The point here has been to eliminate the quantities t_{ab} which expressed the stresses relative to a very special local Minkowski rest frame at the chosen event x. We are still waiting to see what can be done with the quantities $t^{\mu\nu}$ and t^{ij}.

4.4.5 Dynamics. Accounting for the Energy Balance

Consider now three nonparallel infinitesimal displacements $\delta_i \xi^j$, $i = 1,2,3$, that are fixed in the medium and have the same orientation as the vectors $n_a{}^\mu$. Relative to the local Minkowski rest frame, these become

$$\boxed{\delta_i x_a = A_{aj} \delta_i \xi^j , \quad i = 1,2,3} . \tag{4.180}$$

Let us understand this better.

Note first that $\{x_a\}$, $a = 1,2,3$, are the local Minkowski rest frame coordinates of p. 203. We have to select our event x in spacetime first and carry out the calculation in the vicinity of that event. The displacements $\delta_1 \xi^i$, $\delta_2 \xi^i$, and $\delta_3 \xi^i$, each with three components relative to the label coordinates ξ^i, link neighbouring material points with the same proper time coordinate τ. Now

$$A_{aj} \delta_i \xi^j = n_{a\mu} x^\mu{}_{,j} \delta_i \xi^j = n_{a\mu} \frac{\partial x^\mu}{\partial \xi^j} \delta_i \xi^j = n_{a\mu} \delta_i x^\mu ,$$

and in this coordinate system (see p. 203),

$$n_a\big|_x = \frac{\partial}{\partial x^a}\bigg|_x ,$$

so

$$n_a{}^\mu\big|_x = \frac{\partial x^\mu}{\partial x^a}\bigg|_x = \delta_a{}^\mu ,$$

whence

$$n_{a\mu}\big|_x = g_{\mu\nu}(x) n_a{}^\nu\big|_x = \eta_{\mu\nu} n_a{}^\nu\big|_x = \eta_{\mu a} , \tag{4.181}$$

and finally,

$$A_{aj}\delta_i\xi^j = n_{a\mu}\delta_i x^\mu = \delta_i x_a ,$$

as claimed.

The three displacements define an infinitesimal parallelipiped whose volume is

$$\delta V = \det(\delta_i x_a) = \det(A_{ak})\det(\delta_i\xi^j) = \gamma^{1/2}\det(\delta_i\xi^j) . \tag{4.182}$$

The first step, viz., $\delta V = \det(\delta_i x_a)$, follows because $g^{1/2} = 1$ at x in these coordinates. Note also that one would expect $\delta V = \gamma^{1/2}\det(\delta_i\xi^j)$ from the result that the instantaneous proper distance δs between particles with neighbouring values of their labels as viewed in the instantaneous rest frame of either is given by

$$(\delta s)^2 = (\delta x)^2 = \gamma_{ij}\delta\xi^i\delta\xi^j ,$$

as explained earlier, where the proper metric γ_{ij} of the medium is in fact defined by this relation.

The surface elements of the three pairs of opposite faces of this parallelipiped are $\pm\delta_i\Sigma_a$, where

$$\begin{aligned}
\delta_i\Sigma_a &= \frac{1}{2}\varepsilon_{ijk}\varepsilon_{abc}\delta_j x_b\delta_k x_c \\
&= \frac{1}{2}\varepsilon_{ijk}\varepsilon_{abc}A_{bm}A_{cn}\delta_j\xi^m\delta_k\xi^n \\
&= \frac{1}{2}\gamma^{1/2}\varepsilon_{ijk}\varepsilon_{lmn}A^{-1l}{}_a\delta_j\xi^m\delta_k\xi^n .
\end{aligned}$$

On the object $\delta_i\Sigma_a$, the index i numbers the pairs and a gives the 3 spatial components of the surface element numbered by i. Further, $\delta_1\Sigma_a$ is the surface element defined by $\delta_2 x$ and $\delta_3 x$, in that order, $\delta_2\Sigma_a$ is the surface element defined by $\delta_3 x$ and $\delta_1 x$, in that order, and $\delta_3\Sigma_a$ is the surface element defined by $\delta_1 x$ and $\delta_2 x$, in that order. The factor of 1/2 in the last calculation appears because the sum over j and k gives two equal terms.

The forces exerted on these faces by the surrounding medium are $\pm\delta_i F_a$, as given by (4.171), so

$$\begin{aligned}\delta_i F_a &= -t_{ab}\delta_i \Sigma_b = -\frac{1}{2}\gamma^{1/2}\varepsilon_{ijk}\varepsilon_{lmn}A^{-1l}{}_b t_{ab}\delta_j \xi^m \delta_k \xi^n \\ &= -\frac{1}{2}\varepsilon_{ijk}\varepsilon_{lmn}A_{ar}t^{rl}\delta_j \xi^m \delta_k \xi^n \ ,\end{aligned} \tag{4.183}$$

using the first definition of (4.177) in the last step. During an increment $\mathrm{d}\tau$ of proper time, the faces of the parallelipiped will suffer displacements relative to its center given by

$$\pm\frac{1}{2}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\delta_i x_a\right)\mathrm{d}\tau = \pm\frac{1}{2}\dot{A}_{aj}\delta_i \xi^j \mathrm{d}\tau \ , \tag{4.184}$$

according to (4.180). Note that A_{aj} was defined as $A_{aj} = n_{a\mu}x^{\mu}{}_{,j}$. This is a function of the event in spacetime, hence of the coordinates $\{\tau, \xi^i\}$. The quantity $\dot{A}_{aj}$ is then the rate of change of A_{aj} with respect to τ, keeping the ξ^i fixed.

The rate of change of the energy density w_ξ with proper time may be computed by taking into account the work done by the forces $\pm\delta_i F_a$ on the faces of the parallelipiped as a result of these displacements. Recall that w_ξ is that quantity such that $w_\xi \mathrm{d}^3\xi$ is the rest mass energy plus internal energy contained in the label coordinate volume element $\mathrm{d}^3\xi$ (see p. 201). But $\det(\delta_i \xi^j)$ is the coordinate volume of that element, and $\dot{w}_\xi \det(\delta_i \xi^j)$ is therefore the proper time rate of change of rest mass energy plus internal energy in the element. Hence we are interested in

$$\dot{w}_\xi \det(\delta_i \xi^j) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left[w_\xi \det(\delta_i \xi^j)\right] = \frac{\mathrm{d}}{\mathrm{d}\tau}(w_0 \delta V) \ . \tag{4.185}$$

In the second equality, we have used the fact that

$$w_\xi \det(\delta_i \xi^j) = w_0 \delta V \ ,$$

which follows from the fact that $w_\xi = \gamma^{1/2} w_0$ [see (4.154) on p. 195] and the result $\delta V = \gamma^{1/2}\det(\delta_i \xi^j)$ [see (4.182) on p. 207]. One can interpret $\mathrm{d}(w_0\delta V)/\mathrm{d}\tau$ as the proper time rate of change of energy in δV in the local Cartesian rest frame moving with δV, which corroborates the interpretation of $\dot{w}_\xi \det(\delta_i \xi^j)$ just made.

All the real physics, i.e., not just interpretation, but physical law, now occurs in the step

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\tau}(w_0 \delta V) = \left(\frac{\mathrm{d}}{\mathrm{d}\tau}\delta_i x_a\right)\delta_i F_a} \ . \tag{4.186}$$

The proper time rate of energy change in the material element is equal to the proper time rate at which the forces on the surface do work. The right-hand side is a sum over i of 3D Cartesian scalar products (sums over a) of each 3-force element with each rate of displacement 3-vector in the local Minkowski frame. In actual fact, the relevant rate of displacement vectors are

$$\pm\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\tau}(\delta_i x_a) \ ,$$

relative to the center of the volume element, but the result is the same because the forces are in pairs $\pm\delta_i F_a$ (each i labels one of the pairs of opposite faces).

This brings us to the conclusion that

$$\begin{aligned}\dot{w}_\xi \det(\delta_i \xi^j) &= \left(\frac{\mathrm{d}}{\mathrm{d}\tau}\delta_i x_a\right)\delta_i F_a = \dot{A}_{as}\delta_i\xi^s\delta_i F_a \\ &= -\frac{1}{2}\varepsilon_{ijk}\varepsilon_{lmn}\dot{A}_{as}A_{ar}t^{rl}\delta_i\xi^s\delta_j\xi^m\delta_k\xi^n ,\end{aligned}$$

using (4.184) to get the second equality and (4.183) to get the third. Factoring out the determinant, we get

$$\dot{w}_\xi = -\dot{A}_{al}A_{ar}t^{rl} , \tag{4.187}$$

because

$$\varepsilon_{ijk}\delta_i\xi^s\delta_j\xi^m\delta_k\xi^n = \varepsilon^{smn}\det(\delta_i\xi^j) ,$$

and then

$$\varepsilon^{smn}\varepsilon_{lmn} = 2\delta_l{}^s .$$

Hence, we now have

$$\begin{aligned}\dot{w}_\xi &= -\dot{A}_{al}A_{ar}t^{rl} \\ &= -\frac{1}{2}(\dot{A}_{ai}A_{aj} + A_{ai}\dot{A}_{aj})t^{ij} \quad (t^{ij} \text{ is symmetric}) \\ &= -\frac{1}{2}\dot{\gamma}_{ij}t^{ij} \quad [\text{see (4.153) on p. 194}] \\ &= -\frac{1}{2}\frac{\partial(x)}{\partial(\tau,\xi)}\dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu}t^{\mu\nu} \quad [\text{see (4.178) on p. 205}] \\ &= -\frac{1}{2}\frac{\partial(x)}{\partial(\tau,\xi)}r_{\mu\nu}t^{\mu\nu} = -\frac{\partial(x)}{\partial(\tau,\xi)}u_{\mu;\nu}t^{\mu\nu} ,\end{aligned} \tag{4.188}$$

where $r_{\mu\nu}$ is the rate-of-strain tensor discussed in Sect. 4.3.16. This tensor is defined by

$$r_{\mu\nu} := \dot{\gamma}_{ij}\xi^i{}_{,\mu}\xi^j{}_{,\nu} ,$$

from which it was shown that

$$r_{\mu\nu} = P_\mu{}^\sigma P_\nu{}^\tau(u_{\sigma;\tau} + u_{\tau;\sigma}) .$$

It was the latter relation that justified calling it the rate of strain tensor. Now

$$\begin{aligned}r_{\mu\nu}t^{\mu\nu} &= (u_{\sigma;\tau} + u_{\tau;\sigma})t^{\mu\nu}P_\mu{}^\sigma P_\nu{}^\tau \\ &= 2u_{\sigma;\tau}t^{\mu\nu}P_\mu{}^\sigma P_\nu{}^\tau ,\end{aligned}$$

because $t^{\mu\nu} = t^{\nu\mu}$. By the second relation of (4.178) on p. 205,

$$t^{\mu\nu} = \frac{\partial(\tau,\xi)}{\partial(x)} P^{\mu}{}_{\rho} P^{\nu}{}_{\phi} x^{\rho}{}_{,i} x^{\phi}{}_{,j} t^{ij} \; ,$$

and because P is a projection operator, we have

$$P^{\mu}{}_{\rho} P_{\mu}{}^{\sigma} = P^{\sigma}{}_{\rho} \; , \qquad P^{\nu}{}_{\phi} P_{\nu}{}^{\tau} = P^{\tau}{}_{\phi} \; ,$$

so it follows that

$$r_{\mu\nu} t^{\mu\nu} = 2u_{\sigma;\tau} t^{\sigma\tau} \; ,$$

as claimed above.

We now observe that

$$\begin{aligned}
\dot{w}_{\xi} &= \frac{\partial}{\partial\tau}\left[\frac{\partial(x)}{\partial(\tau,\xi)} w\right] \quad \text{[keeping } \xi \text{ constant and using (4.169) on p. 200]} \\
&= \frac{\partial(x)}{\partial(\tau,\xi)}\left[\left(\frac{\partial\tau}{\partial x^{\mu}}\frac{\partial^2 x^{\mu}}{\partial\tau^2} + \frac{\partial\xi^i}{\partial x^{\mu}}\frac{\partial^2 x^{\mu}}{\partial\xi^i\partial\tau}\right) w + w_{,\mu}\dot{x}^{\mu}\right] \\
&= \frac{\partial(x)}{\partial(\tau,\xi)}\left[w\frac{\partial}{\partial x^{\mu}}\left(\dot{x}^{\mu}\right) + w_{,\mu}\dot{x}^{\mu}\right] = \frac{\partial(x)}{\partial(\tau,\xi)}(wu^{\mu})_{;\mu} \; .
\end{aligned} \tag{4.189}$$

In the second step, it is not so obvious why

$$\frac{\partial}{\partial\tau}\frac{\partial(x)}{\partial(\tau,\xi)} = \frac{\partial(x)}{\partial(\tau,\xi)}\frac{\partial\xi^m}{\partial x^{\mu}}\frac{\partial^2 x^{\mu}}{\partial\xi^m\partial\tau} \; ,$$

where we define $\xi^0 := \tau$ and m is summed over 0, 1, 2, and 3. Here is a proof. We have

$$\varepsilon^{\alpha\beta\gamma\delta}\frac{\partial(x)}{\partial(\tau,\xi)} = \varepsilon_{ijkl}\frac{\partial x^{\alpha}}{\partial\xi^i}\frac{\partial x^{\beta}}{\partial\xi^j}\frac{\partial x^{\gamma}}{\partial\xi^k}\frac{\partial x^{\delta}}{\partial\xi^l} \; ,$$

so

$$\begin{aligned}
\varepsilon^{\alpha\beta\gamma\delta}\frac{\partial}{\partial\tau}\frac{\partial(x)}{\partial(\tau,\xi)} &= \varepsilon_{ijkl}\left(\frac{\partial}{\partial\tau}\frac{\partial x^{\alpha}}{\partial\xi^i}\right)\frac{\partial x^{\beta}}{\partial\xi^j}\frac{\partial x^{\gamma}}{\partial\xi^k}\frac{\partial x^{\delta}}{\partial\xi^l} + 3 \text{ similar terms} \\
&= \varepsilon_{ijkl}\frac{\partial^2 x^{\alpha}}{\partial\xi^m\partial\tau}\frac{\partial\xi^m}{\partial x^{\varepsilon}}\frac{\partial x^{\varepsilon}}{\partial\xi^i}\frac{\partial x^{\beta}}{\partial\xi^j}\frac{\partial x^{\gamma}}{\partial\xi^k}\frac{\partial x^{\delta}}{\partial\xi^l} + 3 \text{ similar terms} \\
&= \frac{\partial^2 x^{\alpha}}{\partial\xi^m\partial\tau}\frac{\partial\xi^m}{\partial x^{\varepsilon}}\varepsilon^{\varepsilon\beta\gamma\delta}\frac{\partial(x)}{\partial(\tau,\xi)} + 3 \text{ similar terms} \\
&= \frac{\partial(x)}{\partial(\tau,\xi)}\left[\varepsilon^{\varepsilon\beta\gamma\delta}\frac{\partial^2 x^{\alpha}}{\partial\xi^m\partial\tau} + \varepsilon^{\alpha\varepsilon\gamma\delta}\frac{\partial^2 x^{\beta}}{\partial\xi^m\partial\tau}\right. \\
&\qquad \left. + \varepsilon^{\alpha\beta\varepsilon\delta}\frac{\partial^2 x^{\gamma}}{\partial\xi^m\partial\tau} + \varepsilon^{\alpha\beta\gamma\varepsilon}\frac{\partial^2 x^{\delta}}{\partial\xi^m\partial\tau}\right]\frac{\partial\xi^m}{\partial x^{\varepsilon}} \; .
\end{aligned}$$

Put $(\alpha\beta\gamma\delta) = (0123)$. The object in square brackets multiplied by $\partial\xi^m/\partial x^{\varepsilon}$ is then

$$\frac{\partial \xi^m}{\partial x^\mu}\frac{\partial^2 x^\mu}{\partial \xi^m \partial \tau} ,$$

as claimed.

There is one more detail concerning the calculation (4.189). At the very end we have

$$w\frac{\partial}{\partial x^\mu}\left(\dot{x}^\mu\right) + w_{,\mu}\dot{x}^\mu = (wu^\mu)_{;\mu} .$$

A priori we only know that

$$w\frac{\partial}{\partial x^\mu}\left(\dot{x}^\mu\right) + w_{,\mu}\dot{x}^\mu = (wu^\mu)_{,\mu} .$$

However,

$$(wu^\mu)_{,\mu} = (wu^\mu)_{;\mu} ,$$

because the ordinary coordinate divergence of a contravector density of unit weight is always equal to its covariant divergence, something that is not hard to prove.

Hence, finally, putting together (4.188) and (4.189), we deduce the neat result

$$\boxed{(wu^\mu)_{;\mu} + u_{\mu;\nu}t^{\mu\nu} = 0} , \tag{4.190}$$

or, alternatively,

$$\boxed{u_\mu(wu^\mu u^\nu + t^{\mu\nu})_{;\nu} = 0} . \tag{4.191}$$

Let us just check that the relation (4.191) follows from its predecessor (4.190). We have

$$\begin{aligned} u_\mu\left(wu^\mu u^\nu + t^{\mu\nu}\right)_{;\nu} &= u_\mu u^\mu (wu^\nu)_{;\nu} + u_\mu wu^\nu u^\mu{}_{;\nu} + u_\mu t^{\mu\nu}{}_{;\nu} \\ &= u_{\mu;\nu}t^{\mu\nu} + u_\mu t^{\mu\nu}{}_{;\nu} \quad \text{[by (4.190)]} \\ &= (u_\mu t^{\mu\nu})_{;\nu} = (u_\mu t^{\mu\nu})_{,\nu} = 0 . \end{aligned}$$

The second term on the right-hand side of the first line is zero because

$$u_\mu u^\mu{}_{;\nu} = \frac{1}{2}(u_\mu u^\mu)_{;\nu} = \frac{1}{2}(u_\mu u^\mu)_{,\nu} ,$$

and $u_\mu u^\nu = -1$. We have

$$(u_\mu t^{\mu\nu})_{;\nu} = (u_\mu t^{\mu\nu})_{,\nu}$$

because $u_\mu t^{\mu\nu}$ is a contravector of unit weight again, so its covariant divergence reduces to its coordinate divergence. Finally, $(u_\mu t^{\mu\nu})_{,\nu} = 0$ because $u_\mu t^{\mu\nu} = 0$.

4.4.6 Dynamics. Accounting for the Momentum Balance

This accounts for the energy balance in the medium. Let us now account for the momentum balance, a task that turns out to be much easier. Consider again the parallelipiped of volume δV. Its 4-momentum is

$$p^\mu = w_0 u^\mu \delta V \ ,$$

because δV is the proper volume measured in a frame moving with the material and w_0 is the energy density in that frame. In the local (instantaneous) rest frame of the parallelipiped, the time rate of change of this momentum is equal to

$$\boxed{n_{a\mu}\dot{p}^\mu = w_0 n_{a\mu}\dot{u}^\mu \delta V = w_0 a_a \delta V} \ , \tag{4.192}$$

where the dot denotes the covariant proper time derivative and the a_a are the rest-frame components of the absolute acceleration of δV. Note that the term containing the proper time rate of change of the rest frame energy $w_0\delta V$ of the material element as it moves is projected out by the presence of $n_{a\mu}$, i.e.,

$$n_{a\mu}\dot{p}^\mu = n_{a\mu}\frac{\mathrm{d}}{\mathrm{d}\tau}(w_0\delta V)u^\mu + n_{a\mu}\dot{u}^\mu w_0 \delta V \ ,$$

and the first term is zero because $n_{a\mu}u^\mu = 0$.

This change of momentum can only be caused by the stresses which are

$$\boxed{F_a = -g^{-1/2} n_{a\mu} t^{\mu\nu}{}_{;\nu}\delta V} \ . \tag{4.193}$$

Equation (4.193) looks plausible, but it is not that obvious. On p. 204, we had (4.173), viz.,

$$F_a = -t_{ab,b}\delta V \ ,$$

and to get (4.193) from this, we would have to show the plausible result

$$t_{ab,b} = g^{-1/2} n_{a\mu} t^{\mu\nu}{}_{;\nu} \ , \tag{4.194}$$

using the definition (4.177) on p. 205, viz.,

$$t^{\mu\nu} := g^{1/2} n_a{}^\mu n_b{}^\nu t_{ab} \ . \tag{4.195}$$

Note that both sides of (4.194) are scalars under general coordinate change, because the t_{ab} are defined in a fixed frame and hence scalars, and both sides of (4.195) are type (2,0) tensor densities of weight 1 under general coordinate change, for the same reason. This means that it would suffice to show (4.194) when the right-hand side is expressed in a particular, well chosen coordinate frame.

For example, we could choose the coordinates on p. 203, i.e., the local Minkowskian frame fitted to u^μ, $n_a{}^\mu$ at some event x, which are the very coordinates used to express the left-hand side of (4.194). In (4.181) on p. 207, we noted that

$$n_{a\mu}\big|_x = \eta_{a\mu} \ .$$

Furthermore, the first derivatives of the metric tensor are zero at x, whence

$$t^{\mu\nu}{}_{;\nu}\big|_x = t^{\mu\nu}{}_{,\nu}\big|_x$$

and $g^{1/2}(x) = 1$. We would thus be down to showing that

$$t_{ab,b} = \eta_{a\mu} t^{\mu\nu}{}_{,\nu}\big|_x \ ,$$

in these coordinates, using

$$t^{\mu\nu}\big|_x = \delta_a{}^\mu \delta_b{}^\nu t_{ab} \ ,$$

and it would be nice just to say that this is obvious! But there is a problem, because we cannot get $t^{\mu\nu}{}_{,\nu}\big|_x$ from the defining relation (4.195) if we have restricted it to x. We have to consider

$$t^{\mu\nu}{}_{;\nu} = (g^{1/2} n_a{}^\mu n_b{}^\nu)_{;\nu} t_{ab} + g^{1/2} n_a{}^\mu n_b{}^\nu t_{ab,\nu} \ ,$$

recalling that t_{ab} is a scalar for each a,b. If the first term is zero, we have

$$n_{a\mu} t^{\mu\nu}{}_{;\nu}\big|_x = \eta_{a\mu} \delta_c{}^\mu \delta_b{}^\nu t_{cb,\nu} = t_{ab,b} \ ,$$

as required. Now $(g^{1/2})_{;\nu} = 0$, so it would be sufficient to show that

$$(n_a{}^\mu n_b{}^\nu)_{;\nu} = 0 \ .$$

In the special coordinates at x, the covariant derivative is the coordinate derivative, and this would amount to showing that

$$(n_a{}^\mu n_b{}^\nu)_{,\nu}\big|_x = 0 \ .$$

However, it is clear that there is no general reason why this should be so.

As a matter of fact, as hinted earlier, the problem is actually with the relation (4.173) on p. 204. We can see directly that the correct relation (4.193) does not actually imply that $F_a = -t_{ab,b}\delta V$ in canonical coordinates at x. We may write

$$\begin{aligned} F_a &= -g^{-1/2} n_{a\mu} t^{\mu\nu}{}_{;\nu} \delta V \\ &= -g^{-1/2} n_a{}^\mu g^{\nu\sigma} t_{\mu\sigma;\nu} \delta V \ , \end{aligned}$$

and in the canonical coordinates at x,

$$t_{\mu\sigma;\nu} = t_{\mu\sigma,\nu} \ , \quad n_a{}^\mu = \delta_a{}^\mu \ , \quad g^{\nu\sigma} = \eta^{\nu\sigma} \ , \quad g^{-1/2} = 1 \ ,$$

provided we look only at x. It thus follows that

$$F_a = -\eta^{\nu\sigma} t_{a\nu,\sigma} \delta V = -t_{ab,b} \delta V + t_{a0,0} \delta V \ , \quad \text{at } x \ .$$

This may look good if we are aiming to prove (4.173), because we would only have to show that $t_{a0,0} = 0$, and we know that $t_{\mu\nu} u^\nu = 0$, according to (4.195), and in the canonical coordinate frame at x, u has components $(1,0,0,0)$, so at x, $t_{\mu 0} = 0$. But of course what we need here is the derivative of t_{a0} along the worldline of the material through x, i.e., along u, and we just do not know that t_{a0} will remain zero along this worldline.

In a specific case, we might be able to arrange for this, but note that, if we adopt the proper time along the relevant worldline as time coordinate as we did in Sect. 3.2 when setting up our normal frame, then we no longer get all the connection coefficients equal to zero at x, unless the fluid element passing through event x happens to have zero four-acceleration there [see (3.4) on p. 151].

Of course, it would suffice to have $u \propto \partial / \partial x^0$ in order to obtain $t_{a0} = 0$ along the worldline and hence $t_{a0,0} = 0$ at x, but this flexibility is unlikely to allow us to waive the consequence $\Gamma^i_{00} \neq 0$ in general since it only gives us one degree of freedom with which we have to get rid of three quantities, as i takes values in $\{1,2,3\}$. But in any case, DeWitt implied that (4.173) was valid for any choice of canonical coordinates at x and this contradicts the implications of the correct relation (4.193).

We can see what goes wrong in (4.172) back on p. 203, and it is a good illustration of the complications that arise in relativistic theories. When we talk about 2D surface elements in this discussion, those elements are near x, but not at x. Naturally, they should be taken *simultaneous* with x according to the choice of coordinates, for mathematical reasons, in order to derive (4.173). But the simultaneity dictated by the canonical coordinates is largely arbitrary physically speaking as far as the relevant fluid element is concerned. When we come to do physics, as in the present section, the 'force density' derived there proves to be inadequate.

So why is (4.193) the right relation? Ultimately, the answer is that it will lead to the right formulation of the energy–momentum–stress density in the next section.

Now we come to the physics. Equating F_a in (4.193) and $n_{a\mu} \dot{p}^\mu$ in (4.192), and recalling that $w := g^{1/2} w_0$ according to (4.170), we get

$$\begin{aligned} 0 &= n_{a\mu} (w \dot{u}^\mu + t^{\mu\nu}{}_{;\nu}) \\ &= n_{a\mu} (w u^\mu{}_{;\nu} u^\nu + t^{\mu\nu}{}_{;\nu}) \\ &= n_{a\mu} (w u^\mu u^\nu + t^{\mu\nu})_{;\nu} \ , \end{aligned}$$

using the fact that $n_a \cdot u = 0$ to get the last step. The conclusion here is thus that

$$\boxed{n_{a\mu} (w u^\mu u^\nu + t^{\mu\nu})_{;\nu} = 0} \ . \tag{4.196}$$

4.4.7 Energy–Momentum–Stress Density

Equation (4.196) may be combined with the energy balance equation (4.191) to yield finally

$$T^{\mu\nu}{}_{;\nu} = 0 \,, \tag{4.197}$$

where $T^{\mu\nu}$ is the energy–momentum–stress density defined by

$$\boxed{T^{\mu\nu} := wu^\mu u^\nu + t^{\mu\nu}} \,. \tag{4.198}$$

As an example, consider the well known case of a gas at equilibrium. Recall from (4.176) on p. 205 that $t_{ab} = p\delta_{ab}$ in this case. Then we have

$$t^{\mu\nu} := g^{1/2} n_a{}^\mu n_b{}^\nu t_{ab} = g^{1/2} n_a{}^\mu n_a{}^\nu p = g^{1/2} P^{\mu\nu} p \,,$$

and hence,

$$T^{\mu\nu} = g^{1/2}\left(w_0 u^\mu u^\nu + P^{\mu\nu} p\right) \quad \text{(gas at equilibrium)} \,. \tag{4.199}$$

Let us also examine $T^{\mu\nu}$ in canonical coordinates in the case of flat spacetime in which one has the strictly conserved quantities

$$P^\mu = \int_\Sigma T^{\mu\nu} \mathrm{d}\Sigma_\nu \,,$$

where Σ is any spacelike hypersurface and $\mathrm{d}\Sigma$ the usual measure. Separating P^μ into its energy and momentum components and choosing for Σ the hypersurface $x^0 =$ constant, we have

$$P^0 = \int T^{00} \mathrm{d}^3 x \,, \qquad P^i = \int T^{i0} \mathrm{d}^3 x \,.$$

These expressions, together with the differential identities

$$T^{00}{}_{,0} + T^{0i}{}_{,i} = 0 \,, \qquad T^{i0}{}_{,0} + T^{ij}{}_{,j} = 0 \,,$$

allow one to make the standard identifications

- $T^{00} =$ energy density,
- $T^{i0} = T^{0i} =$ momentum density = energy flux density,
- $T^{ij} =$ momentum flux density.

In the case of the conservative medium, we have

$$T^{\mu\nu} := wu^\mu u^\nu + t^{\mu\nu} \,,$$

and, since $g^{1/2} = 1$ in canonical coordinates, and also $t^{\mu\nu} u_\nu = 0$ from (4.179) on p. 206, whence

$$t^{00} u_0 = t^{0i} u_i \,, \quad t^{i0} u_0 = t^{ij} u_j \,,$$

and consequently, defining $v_i := u_i/u_0$,

$$t^{00} = t^{0i}v_i \,, \quad t^{i0} = t^{ij}v_j \,, \quad t^{00} = t^{ij}v_iv_j \,,$$

it follows that

$$T^{00} = w_0u^0u^0 + v_iv_jt^{ij} \,, \qquad T^{i0} = w_0u^0u^i + t^{ij}v_j \,. \tag{4.200}$$

The first terms on the right-hand sides of these equations are easy to understand. Because of Lorentz contraction, the proper energy density, i.e., the total energy density of the medium in the local Minkowski rest frame, becomes w_0u^0 in an arbitrary Lorentz frame, and these terms evidently give the contributions to the densities of energy and momentum arising from the bulk motion of the matter. The remaining terms are described by DeWitt as curious residuals arising from the internal stresses. These residuals are by no means unimportant and DeWitt gives a simple mechanical example [14, Chap. 10].

4.4.8 Summary

The point of exposing this analysis was to illustrate the way the label coordinate frame based on coordinates $\{\tau, \xi^i\}$ can be put to use to build a heuristic picture of what is happening in a continuous fluid, even given the complexities that arise in curved spacetimes.

We began by defining the proper energy density w_0 of the medium, and variants w_ξ and w in Sect. 4.4.3. The quantity $w_\xi \mathrm{d}^3\xi$ was the rest mass energy plus internal energy contained in the fluid element $\mathrm{d}^3\xi$ 'comoving' with the fluid, where 'comoving' was determined by some synchronization of the proper times of the particles labelled by ξ^i. But we also saw there that it is not always possible to give a physical interpretation of a quantity like w when it relates to arbitrary coordinates, a general issue for all physical quantities when expressed relative to arbitrary coordinates.

In Sect. 4.4.4, we introduced what we called canonical coordinates at some preselected spacetime event x. These were just the normal coordinates of Sect. 3.1 for the given choice of tetrad $\{u, n_a\}_{a=1,2,3}$ along the fluid worldline through x. We then defined a stress tensor density t_{ab} somewhat heuristically by its components at x, showing in (4.171) on p. 203 how it determines the force $\mathrm{d}F_a = t_{ab}\mathrm{d}\Sigma_b$ on a surface element described by a 3-vector $\mathrm{d}\Sigma_b$ relative to the canonical coordinates at x. We criticised the naive conclusion in (4.173) that $f_a = -t_{ab,b}$ could be interpreted as the force density at x.

By considering turning forces, it was shown in the remainder of Sect. 4.4.4 that t_{ab} had to be symmetric in a and b, and it was shown in (4.177) on p. 205, viz.,

$$t^{ij} := \gamma^{1/2}A^{-1i}{}_aA^{-1j}{}_bt_{ab} \,, \qquad t^{\mu\nu} := g^{1/2}n_a{}^\mu n_b{}^\nu t_{ab} \,, \tag{4.201}$$

how to express the stress tensor density relative to the label coordinates and arbitrary coordinates, respectively. The transformation from arbitrary to label coordinates and back was then specified by (4.178) on p. 205, viz.,

$$t^{ij} = \frac{\partial(x)}{\partial(\tau,\xi)}\xi^i{}_{,\mu}\xi^j{}_{,\nu}t^{\mu\nu}\,, \qquad t^{\mu\nu} = \frac{\partial(\tau,\xi)}{\partial(x)}P^\mu{}_\sigma P^\nu{}_\tau x^\sigma{}_{,i}x^\tau{}_{,j}t^{ij}\,. \tag{4.202}$$

The quantities t_{ab} which depended on a specific choice of coordinates for each x were thereby eliminated.

In Sect. 4.4.5, we tackled the dynamics of the fluid by considering an infinitesimal volume element in the form of a parallelipiped specified by three linearly independent infinitesimal displacements $\delta_1\xi^i$, $\delta_2\xi^i$, and $\delta_3\xi^i$ (each with three components) fixed in the medium in the sense that they link neighbouring material points at the same value of their proper time. This assumes some kind of sensible synchronisation of the proper times of the fluid elements. We also made use once more of the local canonical coordinates at a preselected spacetime event x.

The reader should be quite clear as to how these coordinate systems served as an intermediary allowing us to formulate the forces $\delta_i F_a$, $i = 1,2,3$, acting on the six faces $\pm\delta_i\Sigma_a$ of the parallelipiped in (4.183) on p. 208. We were then able to formulate the physical law (4.186) on p. 208, which states that the proper time rate of energy change in the material element is equal to the proper time rate at which the forces on the surface do work.

We thus found in (4.188) on p. 209 that

$$\dot{w}_\xi = -\frac{\partial(x)}{\partial(\tau,\xi)}u_{\mu;\nu}t^{\mu\nu}\,, \tag{4.203}$$

a completely covariant formulation relative to arbitrary coordinates. But w_ξ was related to w by (4.169) on p. 200, viz.,

$$w_\xi = \frac{\partial(x)}{\partial(\tau,\xi)}w\,, \tag{4.204}$$

whence we obtained (4.189) on p. 210, viz.,

$$\frac{\partial(x)}{\partial(\tau,\xi)}(wu^\mu)_{;\mu}\,, \tag{4.205}$$

which is another covariant expression for $\dot{w}_\xi$ in terms of arbitrary coordinates, to be equated with the one in (4.203). Finally, we obtained (4.191) on p. 211, viz.,

$$u_\mu(wu^\mu u^\nu + t^{\mu\nu})_{;\nu} = 0\,. \tag{4.206}$$

Once again, what is important is to see how we obtain a generally covariant result like this by first formulating the physics in special coordinate systems adapted to the physical medium.

In Sect. 4.4.6, by considering the time rate of change (4.192) of the momentum of the tiny parallelipiped in the local instantaneous rest frame of the parallelipiped, viz.,

$$n_{a\mu}\dot{p}^{\mu} = w_0 n_{a\mu}\dot{u}^{\mu}\delta V = w_0 a_a \delta V \ , \tag{4.207}$$

and equating it to the force on the parallelipiped given by (4.193), viz.,

$$F_a = -g^{-1/2} n_{a\mu} t^{\mu\nu}{}_{;\nu} \delta V \ , \tag{4.208}$$

we obtained (4.196) on p. 214, viz.,

$$n_{a\mu}(w u^{\mu} u^{\nu} + t^{\mu\nu})_{;\nu} = 0 \ . \tag{4.209}$$

Defining the energy–momentum–stress density $T^{\mu\nu}$ by

$$T^{\mu\nu} := w u^{\mu} u^{\nu} + t^{\mu\nu} \ , \tag{4.210}$$

the relations (4.206) and (4.209) implied its covariant conservation, i.e.,

$$T^{\mu\nu}{}_{;\nu} = 0 \ . \tag{4.211}$$

Of course, with this approach, it remains to interpret $T^{\mu\nu}$ physically, and this is done for a very simple case in Sect. 4.4.7. What is achieved here physically is that we understand how the conservation of this quantity expresses Newton's laws, or their generalisation to general relativity, operating within the continuous medium. The standard interpretation of T is then that $T_{\mu\nu}v^{\mu}v^{\nu}$ is the energy density that would be measured by an observer with instantaneous four-velocity v using standard techniques, whatever standard techniques may mean.

Chapter 5
Weak Locality Hypothesis

Let us go back to flat spacetimes and the frames that may be set up by accelerating observers, i.e., frames adapted to accelerating worldlines, using the formalism of Friedman and Scarr [23], as discussed in Sects. 2.4–2.7.

So far we have just done something entirely theoretical. We have found new coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ for a region of spacetime, such that the accelerating observer sits permanently at $\mathbf{y} = 0$. They have other properties that may seem nice theoretically. In fact, they have all the properties we achieve for any semi-Euclidean (SE) coordinate system, plus the rigidity property (superhelical rigidity if there is rotational acceleration), as summarized in Sect. 2.8.

But as we learn from general relativity, coordinates are just coordinates. The whole physical problem is to relate some given coordinates to things like lengths and times that would be measured physically. A case in point is the time dilation discussed by Friedman and Scarr [23], and which we commented on in Sect. 2.5. The quantity

$$\tilde{\gamma} = \frac{\mathrm{d}\tau}{\mathrm{d}\tau_{\mathrm{p}}} ,$$

defined in (2.164) on p. 63, is supposed to be the time dilation factor between a clock or physical process ticking along with the particle in question and a clock carried by the observer. But this claim contains no physics at all for the moment because it does not explain how these quantities could be compared.

There is a danger here of talking about time running at different rates in different frames. This would be a mistake, not because it is inherently wrong when interpreted correctly, but because it reverses the logic of the situation. When we talk about time in a given frame, thinking of it as a coordinate, this is just a way of comparing the rates of different physical processes, and it is better to talk about the rates of different physical processes than the rate of a construct from that, viz., the time coordinate. This gives logical precedence to the physical process, and other physical processes (e.g., in clocks) used to measure their rates. This is the dynamical approach to relativity theory discussed by Brown [6], and advocated here.

The accelerating observer has specified hyperplanes of simultaneity by borrowing them from the instantaneously comoving inertial frames. Of course, the observer must choose some way of separating out space and time. At least that is the way, perhaps naively, that we conceive of our environment. And from a mathematical point of view, it seems a natural thing to borrow from an ICIF, but it is nevertheless physically arbitrary, like all aspects of non-inertial coordinate frame construction.

When her proper time changes from τ to $\tau + \mathrm{d}\tau$, she can then ask the clock or physical process ticking along with the particle how much τ_p has changed by between the events where the two hyperplanes of simultaneity $\mathrm{HOS}(\tau)$ and $\mathrm{HOS}(\tau + \mathrm{d}\tau)$ intersect the worldline of the particle. The whole problem of relating theoretical coordinates to physical measurements still remains, however, because the observer must be able to establish the events where the two borrowed hyperplanes of simultaneity intersect the particle worldline. The way her time coordinate is spread out over the neighbouring region of spacetime would indeed appear to be the trickiest physical aspect of relating coordinates to real measurements.

Let us see what Friedman and Scarr have to say about all this [23]. They begin as follows. K' is the uniformly accelerated frame determined by the initial position $\hat{x}(0)$ of the observer, the initial comoving frame $\hat{\lambda}$, and the acceleration tensor A. The worldline $\hat{x}(\tau)$ of the observer is known, given $\hat{x}(0)$, $u(0)$, and A, and the comoving frame matrix $\lambda(\tau)$ can also be given explicitly in terms of $\hat{\lambda}$ and A [see (2.83) on p. 41].

For each τ, they define X_τ to be the collection of all events simultaneous with $\hat{x}(\tau)$ in K_τ, which is the ICIF at τ. This is what we refer to above as borrowing the HOS from an instantaneously comoving inertial observer. Now X_τ is orthogonal to $u(\tau)$, so there exists a neighbourhood of τ and a spatial neighbourhood of the observer in which the X_τ are pairwise disjoint. So spacetime can be split locally into disjoint spacelike hyperplanes. The splitting does not depend on the choice of the original laboratory inertial frame K, and it ensures that, at least locally, the same event does not occur at two different coordinate (proper) times for the observer setting up these coordinates. So we can define local coordinates for the observer in an unambiguous way.

This is of course entirely theoretical. The whole physical problem now consists in specifying those spacelike hypersurfaces X_τ in the real world. At this point, Friedman and Scarr introduce what they call the *weak hypothesis of locality*:

> Let K' be a uniformly accelerated frame for an accelerated observer with worldline $\hat{x}(\tau)$. For any fixed time τ, the rates of the clock of the accelerated observer and the clock of the comoving inertial observer at time τ are the same, and for events simultaneous to $\hat{x}(\tau)$ in the instantaneously comoving inertial frame K_τ, the comoving inertial observer and accelerated observer measure the same spatial lengths.

Hidden in here are what are often called the *clock hypothesis* and the *ruler hypothesis*. The length aspect is more problematic than the time aspect, although both are problematic. Regarding the proper time of the accelerating observer, she merely has to carry a good clock and look at it when necessary. But what is a good clock? One that satisfies the condition of always running at the same rate as an instantaneously coincident clock in the ICIF.

This ignores the fundamental fact that maybe no clock could ever do this. Acceleration may well have universal effects. That would presumably take us beyond the relativity theories as we know them, since there is no hint in our present relativity theories of how those universal effects could be taken into account. We are assuming here precisely that there are none. That is perhaps the sole *theoretical* import of the weak locality hypothesis with regard to time. The clock hypothesis refers to a specific putative clock and assumes that it fits the bill.

Regarding lengths, we do not specify how the accelerating observer measures them. In a rigid frame like this one, a rigid ruler would be needed. This is one that satisfies the ruler hypothesis: however it is moving, it is always instantaneously ready to measure length in the instantaneously comoving inertial frame. As with the clock, this may never be possible due to universal effects of acceleration that are not taken into account by our relativity theories as they stand. Presumably the weak locality hypothesis says that this is at least *theoretically* possible, in the sense that it rules out universal effects due to acceleration, but the question still remains as to whether such rulers actually exist. The ruler hypothesis refers to a specific putative ruler and assumes that it fits the bill.

We shall return to the clock and ruler hypotheses in the following chapters. But what about using light signals somehow? If we assume Maxwell's equations, the theory tells us where light beams will go. Then on the basis of that theoretical assumption, we still need to be able to time emission and reception, so from a practical standpoint, we still need the assumptions about the clocks. At this point it is interesting to consider a very general discussion of the question of ascertaining proper lengths between neighbouring worldlines, so let us refer to Ryder's account in his excellent *Introduction to General Relativity* [50].

For a moment then, we consider an argument that will be valid also in general relativity. Ryder merely refers to two points A and B. The analysis shows that he is discussing the proper separation, defined in some way to be specified, between two timelike worldlines, and in particular two worldlines determined by fixing space coordinates. We thus consider the worldlines obtained by fixing (x^1, x^2, x^3) for A and $(x^1 + \mathrm{d}x^1, x^2 + \mathrm{d}x^2, x^3 + \mathrm{d}x^3)$ for B, and then allowing only the time coordinate $x^0 = ct$ to vary.

Ryder specifies an *operational definition* for the proper distance between these two worldlines. This operational definition uses light signals. What we shall do here is to compare the resulting formula with a theoretical definition that declares their proper separation to be the proper distance orthogonal to either worldline, where orthogonality is defined in the relativistic sense by orthogonality to the 4-velocity of either worldline.

Let us begin with the operational definition. We send a light signal from worldline A to worldline B, where it is reflected and returns to A. The distance between the two worldlines is then defined from the total proper time elapsing at A between emission and reception of the light signal. Let us say that the light leaves A at coordinate time $x^0 + \mathrm{d}x^{0(2)}$, reaches B at coordinate time x^0, and returns to A at coordinate time $x^0 + \mathrm{d}x^{0(1)}$, where $\mathrm{d}x^{0(1)}$ will be positive and $\mathrm{d}x^{0(2)}$ negative.

Now the intervals

- from $\left(x^0+\mathrm{d}x^{0(2)},x^1,x^2,x^3\right)$ to $\left(x^0,x^1+\mathrm{d}x^1,x^2+\mathrm{d}x^2,x^3+\mathrm{d}x^3\right)$

and

- from $\left(x^0,x^1+\mathrm{d}x^1,x^2+\mathrm{d}x^2,x^3+\mathrm{d}x^3\right)$ to $\left(x^0+\mathrm{d}x^{0(1)},x^1,x^2,x^3\right)$

are both null. We conclude that

$$0=g_{00}(\mathrm{d}x^0)^2+2g_{0i}\mathrm{d}x^0\mathrm{d}x^i+g_{ik}\mathrm{d}x^i\mathrm{d}x^k\ , \tag{5.1}$$

where $\mathrm{d}x^0$ stands in for either $\mathrm{d}x^{0(1)}$ or $\mathrm{d}x^{0(2)}$. Note that this result holds for the completely general case, i.e., any coordinates in any spacetime, even a curved one. Now (5.1) is effectively a quadratic equation for $\mathrm{d}x^0$, with two solutions

$$\mathrm{d}x^0=\frac{1}{g_{00}}\left[-g_{0i}\mathrm{d}x^i\pm\sqrt{(g_{0i}g_{0k}-g_{00}g_{ik})\mathrm{d}x^i\mathrm{d}x^k}\right]\ , \tag{5.2}$$

where we use the convention that Latin indices run over $\{1,2,3\}$. Naturally, there may be no solutions or only one, but these cases correspond to situations in which observers could not sit on worldlines A and B and exchange light signals, for some reason which will not concern us here.

Bearing in mind that $\mathrm{d}x^{0(1)}$ will be positive and $\mathrm{d}x^{0(2)}$ negative, and also that $g_{00}<0$, we deduce that the coordinate time Δx^0 elapsed between emission and reception of the signal at A is given by the positive difference between the two solutions in (5.2), viz.,

$$\Delta x^0=\mathrm{d}x^{0(1)}-\mathrm{d}x^{0(2)}=-\frac{2}{g_{00}}\sqrt{(g_{0i}g_{0k}-g_{00}g_{ik})\mathrm{d}x^i\mathrm{d}x^k}\ . \tag{5.3}$$

The corresponding proper time interval between these events is then given by

$$c\Delta\tau=\sqrt{-g_{00}}\Delta x^0\ ,$$

and the operational proper distance $\mathrm{d}l$ by

$$\mathrm{d}l=\frac{c}{2}\Delta\tau\ ,$$

whence finally,

$$\boxed{\mathrm{d}l^2=\left(g_{ik}-\frac{g_{0i}g_{0k}}{g_{00}}\right)\mathrm{d}x^i\mathrm{d}x^k}\ . \tag{5.4}$$

The aim now will be to derive this same result as the orthogonal proper distance between the two worldlines A and B, which means that it is the proper distance between the two worldlines as measured by either observer in their instantaneous rest frame, i.e., the instantaneously comoving inertial frame.

We are considering worldlines $x^\mu(\tau)$ at fixed space coordinates, whence

$$\dot{x}^\mu:=\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}=\left(\frac{\mathrm{d}x^0}{\mathrm{d}\tau},0,0,0\right)\ ,$$

whence

$$\dot{x}^\mu = \frac{1}{\sqrt{-g_{00}}}(1,0,0,0) = \frac{1}{\sqrt{-g_{00}}}\delta^\mu{}_0 \,, \tag{5.5}$$

which is of course the unit tangent vector to such a worldline.

We now introduce the projection $P^\mu{}_\nu$ onto the normal to the worldline, given as usual by

$$P^\mu{}_\nu = \delta^\mu{}_\nu + \dot{x}^\mu \dot{x}_\nu \,, \tag{5.6}$$

where

$$\dot{x}_\nu := g_{\nu\sigma}\dot{x}^\sigma = \frac{1}{\sqrt{-g_{00}}}(g_{00}, g_{01}, g_{02}, g_{03}) \,. \tag{5.7}$$

In particular,

$$\dot{x}_i = \frac{1}{\sqrt{-g_{00}}} g_{0i} \,. \tag{5.8}$$

It is straightforward to show that, for this kind of worldline with fixed space coordinates, this takes the form

$$P^\mu{}_\nu = -\frac{1}{g_{00}} \begin{pmatrix} 0 & g_{10} & g_{20} & g_{30} \\ 0 & -g_{00} & 0 & 0 \\ 0 & 0 & -g_{00} & 0 \\ 0 & 0 & 0 & -g_{00} \end{pmatrix} \,. \tag{5.9}$$

Now if we were using coordinate simultaneity for these coordinates, the separation of the worldlines in hyperplanes of constant x^0 would be

$$\mathrm{d}x^\mu = (0, \mathrm{d}x^1, \mathrm{d}x^2, \mathrm{d}x^3) \,.$$

The idea then is to project this spacelike separation 4-vector to one that is actually orthogonal to worldline A at each instant of proper time of A. We now obtain an object

$$\begin{aligned} \delta x_\perp := P^\mu{}_\nu \mathrm{d}x^\nu &= -\frac{1}{g_{00}} \begin{pmatrix} 0 & g_{10} & g_{20} & g_{30} \\ 0 & -g_{00} & 0 & 0 \\ 0 & 0 & -g_{00} & 0 \\ 0 & 0 & 0 & -g_{00} \end{pmatrix} \begin{pmatrix} 0 \\ \mathrm{d}x^1 \\ \mathrm{d}x^2 \\ \mathrm{d}x^3 \end{pmatrix} \\ &= -\frac{1}{g_{00}} \begin{pmatrix} g_{0i}\mathrm{d}x^i \\ -g_{00}\mathrm{d}x^1 \\ -g_{00}\mathrm{d}x^2 \\ -g_{00}\mathrm{d}x^3 \end{pmatrix} \,, \end{aligned} \tag{5.10}$$

and it is the length of this 4-vector that we intend to show is equal to Ryder's $\mathrm{d}l$. We thus consider

$$\begin{aligned}
\delta x_{\perp}^2 &= g_{\mu\nu} P^{\mu}{}_{i} \mathrm{d}x^i P^{\nu}{}_{j} \mathrm{d}x^j \\
&= g_{\mu\nu} (\delta^{\mu}{}_{i} + \dot{x}^{\mu} \dot{x}_i)(\delta^{\nu}{}_{j} + \dot{x}^{\nu} \dot{x}_j) \mathrm{d}x^i \mathrm{d}x^j \\
&= g_{\mu\nu} \left(\delta^{\mu}{}_{i} - \frac{1}{g_{00}} \delta^{\mu}{}_{0} g_{i0} \right) \left(\delta^{\nu}{}_{j} - \frac{1}{g_{00}} \delta^{\nu}{}_{0} g_{j0} \right) \mathrm{d}x^i \mathrm{d}x^j \\
&= \left(g_{ij} - \frac{g_{\mu\nu} \delta^{\mu}{}_{0} g_{i0} \delta^{\nu}{}_{j}}{g_{00}} - \frac{g_{\mu\nu} \delta^{\nu}{}_{0} g_{j0} \delta^{\mu}{}_{i}}{g_{00}} + \frac{g_{\mu\nu} \delta^{\mu}{}_{0} \delta^{\nu}{}_{0} g_{i0} g_{j0}}{g_{00}^2} \right) \mathrm{d}x^i \mathrm{d}x^j \\
&= \left(g_{ij} - \frac{g_{0i} g_{0j}}{g_{00}} \right) \mathrm{d}x^i \mathrm{d}x^j \,,
\end{aligned}$$

which is indeed $\mathrm{d}l^2$ as given by (5.4).

The above discussion refers to *neighbouring* points, whereas [23] concerns rather the proper distances between any two points in a given HOS of the accelerating observer. However, it does establish that, at least to a good approximation, one natural way of determining proper distance, viz., using light signals, does deliver a proper distance in the hyperplane orthogonal to the observer worldline.

Note, however, that for a more remote particle worldline under observation, we have not considered a light signal measurement and we do not actually know whether a spacelike hyperplane in inertial coordinates would be the most natural spacelike hypersurface to use to measure the proper distance. Furthermore, we do not even know what spacelike curve in the chosen spacelike hypersurface is the one whose length should be measured to give the most natural proper length as it would naturally be measured by the accelerating observer.

But is naturalness important? If not, then it does not matter what spacelike curve we use to specify proper length between two non-neighbouring events, and proper length becomes somewhat arbitrary. But if we are only trying to set up coordinates such that the spatial coordinates are proper lengths according to some choice of spacelike curves, then that arbitrariness is just part of the arbitrariness we know to be inherent in all coordinate systems.

What is different for an inertially moving observer is that we do have a natural choice for these things, which naturally leads to inertial coordinates in flat spacetimes, and locally inertial coordinates in general spacetimes. In fact, what guides us ultimately in those cases is the simplicity of our field theories of matter, a point we shall develop later on. And if there are only arbitrary coordinate systems, and no fundamentally natural ones, for the accelerating observer, then we have to admit that the theoretical usefulness of the role of such observers is reduced to nothing.

Before moving on, an interesting detail in Ryder's calculation with the light signals is the claim that the proper distance between the neighbouring worldlines should be determined from the proper time interval $\Delta\tau$ between emission and reception by the simple relation $\mathrm{d}l = c\Delta\tau/2$. But why should the observer use her proper time interval and not an interval of some other time coordinate to get the proper distance by this formula? After all, we do not specify what time system is used when we say that light travels c meters per unit time. Presumably, the only guide here is, in some unspecified sense, the naturalness of the idea.

Returning to the weak locality hypothesis, we are interested in this because it is clearly the point where the purely theoretical construction of the coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ is related to what might actually be measured physically. This comes close to a point that will be criticised in detail later: how do we interpret quantities expressed relative to non-inertial coordinate frames? In the end, what really interests us here is to see the interface between the theory and the real world.

We construct the coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ theoretically in terms of some inertial coordinates $\{x^{\mu}\}_{\mu=0,1,2,3}$ using the relation (2.97) on p. 44, viz.,

$$x^{\mu} = \hat{x}^{\mu}(\tau) + y^{(i)}\lambda_{(i)}(\tau) \,. \tag{5.11}$$

This is done in a purely mathematical way, and the whole physical problem then consists in relating those new coordinates to things like lengths and times that would be measured physically.

Of course, since we think we know how to relate inertial coordinates to measured lengths and times, we could just establish the inertial coordinate system physically in this way and then use (5.11) to establish the coordinate system $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ physically. But our idea is to try to relate the latter coordinates directly to what an accelerating observer might measure. And the aim of the weak locality hypothesis (WLH) is to claim that we can achieve the same relation between the coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ and physically measured lengths and times as we would obtain if we first related the latter to inertial coordinates and then used the conversion (5.11), or rather its inverse.

It assumes first that, at any proper time τ of the accelerating person, the rate of her clock will be instantaneously the same as that of the instantaneously comoving inertial observer. Let us see what would be involved in such a claim. First of all, we define τ in terms of the inertial time t in the laboratory inertial frame K to be such that

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma(v) \,, \tag{5.12}$$

where v is the instantaneous speed of the observer relative to K. Theoretically, we make that definition so that the 4-velocity of the observer $\mathrm{d}x^{\mu}/\mathrm{d}\tau$ has unit pseudolength.

Now for an instantaneously comoving inertial observer (ICIO) at proper time τ, moving therefore with constant velocity v relative to the laboratory frame K,

$$\frac{\mathrm{d}t}{\mathrm{d}t'} = \gamma(v) \,, \tag{5.13}$$

where t' is the time coordinate of that ICIO. If this ICIO carried an ordinary clock, we consider it to be established by the special theory of relativity (SR) that the time t' she would measure with that clock would be related to inertial time t in K by this last relation.

So the *theoretical* definition of proper time for the accelerating observer in (5.12) has it precisely that its rate of change relative to the inertial time t in K will be the same as the rate of change of the proper time of the ICIO given in (5.13). This means that, if real clocks actually do always measure proper time as defined by (5.12), we do not need the first assumption of WLH. But is this likely to be borne out in reality? After all, there is a major difference between the ICIO and the AO: the latter is accelerating.

The question thus remains as to whether the accelerating observer would actually measure this proper time if she carried an ordinary clock with her. In other words, would a real clock, or a real physical process, really go at the rate specified in (5.12) if it were being accelerated? Clearly, the first statement of WLH concerning time measurement is the assumption that some specified clock will run at this rate.

Later we shall discuss the Unruh–DeWitt (UD) detector in quantum field theory (see Chap. 14). We shall see that this detector will detect nothing if it moves at constant velocity through the usual Minkowski space vacuum of the quantum field, whereas it will excite if it accelerates in any way. One can say that it is designed not to excite when sitting still in the usual QFT vacuum, and it then follows that it will not excite when moving at any constant velocity. This boils down to a consequence of Lorentz symmetry. The QFT vacuum looks the same in any inertial frame and its relation with the detector is the same. But something changes when the detector accelerates.

There is a general point here about any accelerating detector. Imagine designing two different detectors to measure the same physical quantity. Whenever they are moving inertially in the same physical context, we expect them to deliver the same value for whatever quantity it is they are supposed to measure. This is because all our field theories of matter, which govern both the internal constitution of the detectors and the environment of the detectors, are Lorentz symmetric (or locally Lorentz symmetric in GR). But what can we say when they are accelerating? Will they always deliver the same result for the given physical quantity? After all, there is no corresponding acceleration symmetry in our field theories of matter.

It is this lack of acceleration symmetry that shows that we can never be sure that a standard detector of something, designed to detect that thing in a certain way whenever it is moving inertially, will still be a standard detector of that thing when it is accelerating. We could of course try to define whatever it does register when accelerating to be the same quantity for a comoving accelerating observer, but even that is problematic given the point made in the last paragraph, because our definitions may become detector dependent.

Another example here is the Mould EM radiation detector, also discussed in Chaps. 11 and 14. According to its inventor, this measures no EM radiation when comoving with an accelerating charge, even though the latter would be considered to radiate by any inertially moving observer. We could define whatever it measures to be radiation for whoever is carrying it, and that may or may not be a useful definition. But what if there is another perfectly good design of radiation detector, which always delivers the same values as the Mould detector when moving inertially, but delivers different values to the Mould detector when accelerating? Which detector should we then use to define radiation for the accelerating observer?

In the present situation, we have to consider time detectors, i.e., clocks. We could consider a hydrogen atom, counting orbits of its electron. In a now classic paper [2], Bell put forward an argument using classical electromagnetism which shows that, if the acceleration is not too great in relation to certain parameters determining the structure of the atom, then this time detector will work rather well for the AO. However, there will be some error, even on the purely theoretical level, in the sense that it will never exactly register the proper time of the AO due to changes in its constitution that arise because it is accelerating, or rather, due to specific, acceleration-related delays to such changes.

What we are saying is that the atom will satisfy the clock hypothesis to a good approximation, depending on the way it is accelerated, but never exactly. Another kind of clock will also satisfy the clock hypothesis to some level of approximation, the accuracy of which will depend on the way it is made to accelerate, but in the detail it is very unlikely to give the same readings as the atom. On the other hand, we have a theoretical definition of proper time and we can design better and better clocks for reading it when those clocks have the given acceleration. In other words, we can make specific designs for each kind of acceleration.

And if we are just after measuring proper time for this curve, and we know what the proper time should be theoretically from the definition (5.12), maybe there is no point in such an exercise. After all, if the Bell-type argument allows us to estimate the error in the reading from the hydrogen atom, then we can just recalibrate the reading from the hydrogen atom to know what proper time has elapsed.

It should be noted that the heart of the Bell-type argument is the standard assumption that the relevant field theories of matter are Lorentz symmetric. But that is nevertheless an assumption, which amounts to assuming that the special theory of relativity is right. What would happen here if there were in fact universal effects due to acceleration in the sense that there are universal effects due to velocity? This would presumably imply that WLH could never be fully borne out in the finest details.

So what is the aim of WLH? The idea of the first statement that the rates of the clock carried by the AO and one carried by the ICIO are the same is to say that AO would *naturally* measure the theoretical quantity we call her proper time τ, i.e., that τ is somehow her *natural* time coordinate. The unspoken implication is also that the hyperplanes of simultaneity of the ICIOs are somehow the HOSs that AO should naturally adopt to spread her natural time coordinate over the spacetime neighbourhood of her worldline. This assumption should perhaps feature more explicitly in WLH, where it is only implicit.

Is that spreading of time a natural one? As we saw above, the light signal determination of proper distances between neighbouring worldlines does pick out the HOS of the ICIO, at least to a first approximation, and this is the hyperplane instantaneously orthogonal to the AO worldline. What other spreading is available? Whether this is natural or not, any other spreading is likely to be less natural. What we are talking about here is naturalness of coordinate construction *in reality*, because in theory any coordinates are as good as any other coordinates.

The other claim in WLH says that, for events simultaneous with $\hat{x}(\tau)$ in the ICIO K_τ, the comoving and accelerating observers measure the same spatial lengths. This assumes that we know what events should count as simultaneous for an accelerating observer, as mentioned in the last paragraph. The natural choice may well be the HOS of the ICIO, but we are still just making an operational definition on the basis of naturalness, for what that is worth. In fact, it is the worth of these naturalness claims that constitutes the main subject of the discussion.

And what are the grounds for thinking that AO would measure the same spatial components as the ICIO, viz., $y^{(i)}$, $i = 1,2,3$? As implied above, there is a sense in which AO can measure what she wants by suitably designing and refining her length detector, i.e., her ruler. But the suggestion here is once again that she would *naturally* measure the length intervals between events in the given HOS, and that she would *naturally* measure the length intervals found by intervals in the values of the $y^{(i)}$, $i = 1,2,3$, within that HOS.

But no mention is made of the methods she could use to do that. Would she just use a rigid ruler? How would it have to be moving to deliver the right values? Could AO get it moving like that? Is that how it would naturally move? Or would AO use light signals somehow? Does the theory of null structure in the spacetime show that this would work?

And what is the aim of the claim that these coordinates are what AO would naturally measure? We are asking, even if that were the case, so what? It is almost as though the aim were to somehow validate this particular coordinate choice. But why should we need to do that? After all, coordinates are just coordinates (unless they are inertial coordinates, which have some privileges, because these are the ones in which all our field theories of matter take on their simplest forms).

We have to be careful that there is no hidden agenda here. We shall see later how some authors try to pretend that the coordinates $\{y^{(\mu)}\}_{\mu=0,1,2,3}$ in (5.11) are inertial coordinates, so that they may interpret EM or other fields expressed relative to this coordinate frame as though they were the EM or other fields we know and love from our school days (see Chap. 11).

But even if we could license this kind of somewhat naive interpretation of fields by whatever arguments may support WLH, what are these interpretations for? Are they just to make us feel more comfortable about these spacetime theories? The fact is that this hypothesis, if it is worth anything at all, must be a hypothesis about specific instruments and methods for measuring lengths and times under specific conditions of acceleration. Its theoretical content is only to rule out (if it is valid) universal effects of acceleration.

Chapter 6
Extending Bell's Approach to General Relativity

In his paper *How to Teach Special Relativity* [2], Bell considers the nucleus of an atom as an accelerating point charge for which the exact electromagnetic potential (the Liénard–Wiechert potential) is known from Maxwell's theory, then proposes to calculate the exact orbit of the electron in this field as the nucleus accelerates. In principle, this would give a perfect (pre-quantum) description of the way the length of the atom would change in the direction of acceleration, as measured in the original inertial frame. One can also imagine calculating the way the period of the electron orbit will change, as measured in the original inertial frame. The result is a dynamical explanation of the FitzGerald contraction and time dilation, using as premise Maxwell's theory of electromagnetism.

It is worth noting at the outset that such a dynamical explanation of these phenomena does not circumvent the theoretical need to posit a spacetime with the usual attributes, as first described so elegantly by Minkowski, viz., a metric and a coordinate description in which the metric takes the usual diagonal form $\eta = \mathrm{diag}\,(1,-1,-1,-1)$. Otherwise one could not even apply Maxwell's theory. The view adopted here is that dynamical explanations like this live happily alongside the standard philosophy and add to our understanding of what is going on (see also Chap. 9).

It is natural to ask if one could not do something similar for an atom sitting in a curved spacetime that was supposed to represent some gravitating configuration. Here we shall sketch briefly how this can be done for three spacetimes:

- the de Sitter empty universe,
- the Schwarzschild spacetime,
- the static homogeneous gravitational field.

This may seem a strange choice. The point about the first spacetime is that one has an empty world, but in which space itself is supposed to be expanding. This does raise the question of what one means by space expanding, and we use the Bell atom to try to find out. Regarding the second choice, it will be an opportunity to present a mild criticism of the way this spacetime is often presented in textbooks. One proposes coordinates that are suggestively represented by r, θ, ϕ for a spherically symmetric manifold and solves Einstein's equations for the metric, but the obvious question is not always directly addressed: what do the coordinates mean in terms of real life? Once again, we can use the Bell atom to carry out an investigation. Finally, the highly idealistic static homogeneous gravitational field (SHGF) brings us back to the issue originally addressed by Bell, because the metric usually used here is precisely the semi-Euclidean metric generally proposed for a uniformly accelerating observer in a Minkowski spacetime with no gravitational effects, not even non-tidal ones.

The heart of the matter here is the strong equivalence principle (SEP), without which one could not even do electromagnetism in curved spacetime, at least, without which one would have to invent some other way of formulating non-gravitational bits of physics within general relativity. Throughout the discussion, we use absolutely standard general relativity with zero torsion, so that the connection is the well-known metric connection with coefficients

$$\Gamma^{\mu}_{\nu\sigma} := \frac{1}{2} g^{\mu\tau}\left(g_{\nu\tau,\sigma} + g_{\tau\sigma,\nu} - g_{\nu\sigma,\tau}\right) . \tag{6.1}$$

As mentioned earlier, this already guarantees a weak form of equivalence principle (WEP), viz., at any spacetime event P, there exists a choice of coordinates in some neighbourhood of that event for which

$$g_{\mu\nu}\big|_P = \eta_{\mu\nu} , \qquad \Gamma^{\mu}_{\nu\sigma}\big|_P = 0 , \tag{6.2}$$

where η is the usual Minkowski metric. In fact, for any original set of coordinates $\{x^a\}$, a new set with the property of having zero connection coefficients at the chosen event is

$$x'^{\mu} := x^{\mu} - x_0^{\mu} + \frac{1}{2}\Gamma^{\mu}_{\nu\rho}\big|_0 \left(x^{\nu} - x_0^{\nu}\right)\left(x^{\rho} - x_0^{\rho}\right) , \tag{6.3}$$

using the subscript 0 to denote the chosen event and arranging for this event to lie at the origin of the new coordinates. One then gets the metric components into Minkowski form by a linear coordinate transformation, which preserves the fact that the connection coefficients are zero at the chosen event.

The idea here is that the spacetime looks roughly Minkowskian at the chosen event, and even in a small neighbourhood of it, when we use these coordinates. Because the connection coefficients are zero at P in the primed system, the first coordinate derivatives of the metric components are also zero at P in the primed system, so the metric components are not going to change quickly from the Minkowski values when we move away from P. However, the connection coefficients themselves will generally change when we move away from P, unless the spacetime is flat. Any primed coordinates with the property (6.2) will be called normal coordinates. This is a cheap version of the more elegant ploy in Sect. 3.1.

The strong equivalence principle (SEP) is then the hypothesis that other theories of physics not concerned with gravity itself will look roughly as they do in flat spacetime relative to such a coordinate system. This is not a very precise formulation. We implement it more precisely by saying that any inertial coordinate derivatives in the dynamical equations of the flat spacetime theory are to be replaced by covariant derivatives with respect to the metric connection to get the minimal extension of the theory in the curved spacetime. Non-minimal extensions are not considered here.

So what can the Bell atom do for us in a curved spacetime? We have to look again at what it was doing in special relativity, where we said it could explain time dilation and FitzGerald contraction. If our atom is moving inertially, we now make the definition that it can be used as a clock and a ruler in the inertial frame moving with it, by taking its spatiotemporal dimensions to gauge intervals of space and time in that frame. What we find is that the length and time values it delivers agree with what the Minkowski metric components give according to the usual interpretation of the metric components in relativity.

One can well imagine that this works because Maxwell's theory is Lorentz covariant. One can also imagine that the spatiotemporal dimensions of any system governed by a Lorentz covariant theory could thus be used as a clock and ruler that would predict time dilation and FitzGerald contraction. One is giving logical precedence to the Lorentz covariance of one's theories and then understanding the physical effects of setting one's clocks and rulers in motion. Then the class of coordinate systems relative to which one's theories are Lorentz covariant is defined as the class of inertial coordinate systems.

In general relativity, there are not usually any inertial coordinate systems, but we have SEP, so theories that work in special relativity can be carried over in a perfectly defined way to general relativity (although it remains to be seen whether they are still good theories when so transformed). When we are given some coordinates and the components of a metric for those coordinates, we do not a priori know how to associate those coordinates with real events in the world we are trying to model. On the standard account, it is the metric that tells us how to convert a coordinate interval into a real time or length as it would be measured. It is then effectively *assumed* that, wherever they are and whatever they are doing, our clocks and rulers will deliver the time or space intervals predicted by the metric (if interpreted correctly). One might also conclude that, wherever it it is and whatever it is doing, an atom could be used as a clock or ruler that would deliver the time or space intervals predicted by the metric.

The present idea is to turn this around. We start with the theories carried over by SEP from the theories in special relativity that govern our clocks and rulers, we see what happens to our clocks and rulers in the relevant context in our curved spacetime, according to the theory provided by SEP, and we make the operational *definition* that they can be used to measure space and time intervals. The aim is then to show why the metric in such a construction delivers the lengths and times that would be measured by our clocks and rulers. Naturally, we focus on one atom because it provides the simplest possible measuring instrument.

The program described here is too ambitious to be carried out to the letter. For one thing, it is not obvious what theory to use for the atom because the minimum extension of Maxwell electromagnetism to the curved spacetime actually predicts that the electron will radiate electromagnetic waves and crash into the nucleus. And even if there were a stable orbit for the electron, it would be difficult to carry out exact calculations, so one would have to resort to numerical approximation, as happened even in the special relativistic case when the atom was accelerating. The obvious theory to use would be a quantum theory, and the problem of obtaining an exact result would surely be made worse. On the other hand, it is possible to carry out approximate calculations. The strong equivalence principle provides a very simple way of doing this.

The idea actually put into practice in each of the examples below is as follows. Given some event in the spacetime, through which the nucleus of our atom is assumed to pass, one finds normal coordinates at that event such that the nucleus is instantaneously stationary when it passes through it. One then assumes a very simple form for the electron orbit, viz., circular with a well-defined period, relative to these normal coordinates. The actual theory proposed above, carried over from special relativity, might well give some other spatiotemporal characteristics relative to the normal coordinates, but whatever it did give, one could then transform the characteristics of the resulting atom to the original coordinate system, in which one is trying to interpret the metric components. With the operational definition of space and time intervals, viz., wherever it is and whatever it is doing, an atom can be used as a clock or ruler that will deliver the time or space intervals (if it is used correctly, i.e., provided that the atom is instantaneously stationary in the relevant coordinate system), one then obtains the usual interpretation of the metric components.

Since the details of the atom have completely disappeared here — in the end we make a completely arbitrary decision about its structure, with no regard for theory — one must ask what remains of Bell's original idea. The key to understanding what the metric components do for us in some arbitrary coordinate interval now lies entirely in the conversion of a normal coordinate interval to the interval relative to the relevant coordinates. Obviously, this is something that could be examined without any reference to atoms or the theories governing their structure. Of course, what remains is the justification for this procedure. It is the idea that our clocks and rulers measure normal coordinates (when used correctly), and this because the physics of these objects is governed by theories that carry over from Lorentz covariant theories in special relativity by means of the strong equivalence principle.

In fact, what remains is the understanding that, without the strong equivalence principle, not only could one not even do non-gravitational physics in general relativity (without the help of some other such principle), but one could not even *understand* how the coordinates should be related to what is actually observed, because one must simply *postulate* that the metric provides time and length intervals from coordinate intervals.

But there is another point. This idea of carrying out an exact calculation for the physical structure of some real physical entity, and the possibility of doing so even approximately, means that one can in principle test another hypothesis, the clock hypothesis. Or rather, one can assess it theoretically in specific instances, i.e., for specific clocks under specific acceleration conditions, whence it is itself no longer required as a further hypothesis. This is achieved on the basis of our Lorentz covariant theories in special relativity, where the problem already existed, and those same theories assisted by SEP in general relativity, since the problem carries over identically to curved spacetimes, precisely via the strong equivalence principle.

6.1 De Sitter Empty Universe

It would be easy to present this spacetime by saying that we have a metric of the form

$$(g_{\mu\nu}) = \mathrm{diag}\left(c^2, -\mathrm{e}^{2Ht}, -\mathrm{e}^{2Ht}, -\mathrm{e}^{2Ht}\right), \tag{6.4}$$

where H is a constant. What we mean in fact, and it is logically rather different, is that there are coordinates $\{t,x,y,z\}$ such that the metric takes the above form. But since we now begin by mentioning coordinates, this in turn raises the question as to what the coordinates correspond to physically. That is, how would we set up such a coordinate system, or even recognise the coordinates corresponding to a given spacetime event, in the real world around us? This question ought to be raised for any curved spacetime of general relativity, not just this one which happens to be empty and thereby exacerbates the problem.

We shall not be concerned with the dynamics of this spacetime, which was never proposed as a serious cosmological model. In fact, it would have to be empty of matter and energy and pumped up by a cosmological constant Λ related to the constant H in the above equation by [40]

$$\Lambda = \frac{3H^2}{c^2}.$$

The test particles considered here are all assumed to have negligible effect on this geometry.

The aim here is to make a critical analysis of what coordinates mean in this cosmological context. It is usual to say by examination of the above metric that the de Sitter universe is expanding since the proper distance between particles sitting at fixed spatial coordinates is increasing. Of course, the fact that the spatial coordinates of the particle are unchanging means nothing in itself, because coordinates are just a relative thing, i.e., not intrinsic to the spacetime. Note, however, that such particles would then be following geodesics of the metric, as can be shown by considering the connection coefficients (given below). This is taken to mean that there are no forces on them, in the sense that the only things that can be affecting a particle following a timelike geodesic are gravity and the cosmic repulsion given by Λ. And it is clear that the cosmic repulsion is at work here, since what is usually called the proper distance between two such particles, on the usual interpretation of the metric, is actually increasing exponentially with the time coordinate t, which happens to be the proper time of a particle sitting at fixed spacetime coordinates, on the usual interpretation of the metric.

In fact, we set up our notion of proper distance in a coordinate dependent way. It depends on getting the metric into a form with $g_{0\mu} = 0$ for $\mu = 1, 2, 3$. This is the ploy motivating the Weyl postulate [40], a starting assumption of cosmological models which says roughly speaking that the worldlines of galaxies designated as fundamental observers form a 3-bundle of non-intersecting geodesics orthogonal to a series of spacelike hypersurfaces. The 0-coordinate is then called cosmic time and hypersurfaces $t =$ constant are spacelike with an obvious positive-definite metric induced from g. The latter determines proper distance. In the case of the de Sitter metric displayed above, this means that the proper distance between test particles sitting at constant values of (x,y,z) increases exponentially with the cosmic time t.

But what are x,y,z, and t? Does the construction in the last paragraph really answer that question? Does it answer that question in an empty universe? Of course, if one disallows the usual postulate about the geometric interpretation of the metric, there is no hope of linking these coordinates with reality without the strong equivalence principle (SEP). This is what we shall invoke here in the context of a pre-quantum atom, governed by Maxwell's theory and its minimal extension via SEP to the curved spacetime. The fact that the atom is not understood here via quantum mechanics is irrelevant, as is the fact that we ignore the radiation disaster for such an atom, which means that we are even ignoring one of the consequences of the Maxwell theory that carries over directly via SEP to its minimal extension. In fact, even the details of Maxwell's theory will turn out to be irrelevant. What will be relevant will just be the fact that something we understand in flat spacetime carries over to a good approximation to the curved spacetime.

We consider an atom at the spatial origin for some $t = T$. There is no loss of generality here, since the de Sitter universe is spatially homogeneous and isotropic. Letting the subscript 0 denote this point $(T, 0, 0, 0)$, we use the normal coordinates specified in (6.3), viz.,

$$x'^{\mu} = x^{\mu} - x_0^{\mu} + \frac{1}{2}\,\Gamma^{\mu}_{\nu\rho}\big|_0 (x^{\nu} - x_0^{\nu})(x^{\rho} - x_0^{\rho}) \,.$$

Note that the primed coordinates are adjusted so that the origin is $(0, 0, 0, 0)$ in these coordinates. They are not quite the coordinates we want. The first derivatives of the metric are zero at our origin, so that the connection coefficients are also zero at the origin, but we have not necessarily got the metric into the Minkowski form there. This is done by a making a further linear transformation of coordinates. We shall see that it is a simple matter in the present case, because the metric remains in diagonal form.

The connection coefficients are easily calculated from the metric using (6.1), and evaluated at the chosen origin (the last step just means putting $t = T$). They can be presented most conveniently in the form of four matrices:

$$\Gamma^{0}_{\mu\nu}\big|_0 = -2H\mathrm{e}^{2HT}\begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}\,, \quad \Gamma^{1}_{\mu\nu}\big|_0 = 2H\begin{pmatrix} 0 & 1 & & \\ 1 & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}\,,$$

$$\Gamma^{2}_{\mu\nu}\big|_0 = 2H\begin{pmatrix} & & 1 & 0 \\ & & 0 & 0 \\ 1 & 0 & & \\ 0 & 0 & & \end{pmatrix}\,, \quad \Gamma^{3}_{\mu\nu}\big|_0 = 2H\begin{pmatrix} & & 0 & 1 \\ & & 0 & 0 \\ 0 & 0 & & \\ 1 & 0 & & \end{pmatrix}\,.$$

We can now write down the coordinates which make the connection zero at the origin. We use (y^μ) rather than the prime notation:

$$\begin{aligned}
y^0 &= t - T - H\mathrm{e}^{2HT}\left[(x^1)^2 + (x^2)^2 + (x^3)^2\right] ,\\
y^1 &= x^1 + 2H(t-T)x^1 ,\\
y^2 &= x^2 + 2H(t-T)x^2 ,\\
y^3 &= x^3 + 2H(t-T)x^3 .
\end{aligned}$$

Note that $t = x^0$. We must recalculate the metric in the y coordinates using the formula

$$g'^{\mu\nu} = \frac{\partial x'^\mu}{\partial x^\sigma}\frac{\partial x'^\nu}{\partial x^\tau} g^{\sigma\tau} .$$

It is easier to transform the contravariant tensor, because we do not have the inverse coordinate transformation. We are only interested in the metric at the origin. At this point, the two matrices of partial derivatives occurring in the last formula are just the identity, a general feature of the way we defined the coordinate change. We thus have

$$g'^{\mu\nu}(0) = \mathrm{diag}\left(c^{-2}, -\mathrm{e}^{-2Ht}, -\mathrm{e}^{-2Ht}, -\mathrm{e}^{-2Ht}\right) .$$

We now apply a linear map $y^\mu \to z^\mu$ which gets this into the Minkowski form $\mathrm{diag}(c^{-2}, -1, -1, -1)$. We require

$$\frac{\partial z^\mu}{\partial y^\nu} = \begin{pmatrix} 1 & & & \\ & \mathrm{e}^{HT} & & \\ & & \mathrm{e}^{HT} & \\ & & & \mathrm{e}^{HT} \end{pmatrix} .$$

Hence, keeping the origin at $(0, 0, 0, 0)$ in the z coordinates,

$$z^0 = y^0 , \quad z^1 = \mathrm{e}^{HT} y^1 , \quad z^2 = \mathrm{e}^{HT} y^2 , \quad z^3 = \mathrm{e}^{HT} y^3 ,$$

so that finally,

$$\begin{aligned}
z^0 &= t - T - H\mathrm{e}^{2HT}\left[(x^1)^2 + (x^2)^2 + (x^3)^2\right] ,\\
z^1 &= \mathrm{e}^{HT} x^1\left[1 + 2H(t-T)\right] ,\\
z^2 &= \mathrm{e}^{HT} x^2\left[1 + 2H(t-T)\right] ,\\
z^3 &= \mathrm{e}^{HT} x^3\left[1 + 2H(t-T)\right] .
\end{aligned}$$

These are the required normal coordinates at the chosen event.

We now introduce electron orbits about the nucleus at the origin. The first thing to note is that the nucleus is stationary in the z coordinates. Indeed, its path in these coordinates is just $(t - T, 0, 0, 0)$, using t as the parameter. The electron orbits are thus the usual flat space paths for non-moving nuclei, relative to the z coordinates. The reason this works out so neatly is just that the nucleus is following a geodesic, i.e., it is in free fall. This will slightly complicate the analysis in Schwarzschild spacetime, but in an interesting and important way. The problem is that we shall consider a supported atom, i.e., the nucleus will be sitting at fixed Schwarzschild space coordinates r, θ, ϕ and hence will not be following a geodesic. It will therefore have motion relative to the normal coordinates, even if we arrange for it to be instantaneously stationary.

Let us assume following Bell that there is a circular orbit relative to these z coordinates. It can be chosen in the plane $z^3 = 0$, giving $z^1(z^0)$ and $z^2(z^0)$ such that

$$[z^1(z^0)]^2 + [z^2(z^0)]^2 = r_{\mathrm{atom}}^2 .$$

The idea is to take the atomic radius as a unit of proper distance here. We are making the operational definition that the atom can be used to measure proper lengths, wherever it is, whatever it is doing (with the proviso about being at least instantaneously stationary relative to the relevant coordinate system). This is what happens in special relativity, where the inertially moving observer continues to use the same standard atom even though it may appear to have shrunk to another observer. We are now going to see how this unit of proper distance will look in the original de Sitter coordinates, something we usually deduce from the metric without further explanation.

We shall go further, however. We shall take

$$z^1(z^0) = r_{\mathrm{atom}} \cos \frac{2\pi z^0}{P} , \quad z^2(z^0) = r_{\mathrm{atom}} \sin \frac{2\pi z^0}{P} .$$

We are going to take P, the period of the electronic orbit, as a unit of proper time for the atomic nucleus (not for the electron, of course). In other words, we make the operational definition that this physical process defines proper time. We can then look at the period of the atom, the unit of proper time, when referred to the original de Sitter coordinates.

Note in passing a rather pleasant idea, which illustrates perfectly why the details of Maxwell's theory are irrelevant here. We could take our time-defining process to be a planetary orbit. On this cosmological scale, the planetary orbit would still be infinitesimal. Some thought must be given to the physical theory used to define the orbit, however, since we are now outside Maxwellian electromagnetism. We need a reason why the planetary orbit should look like a flat space orbit in the normal coordinates, and this reason is the strong equivalence principle.

What does our electron orbit look like in x coordinates? The easiest way to get $z^3 = 0$ is just $x^3 = 0$. The other equations are now

$$\mathrm{e}^{HT} x^1 \left[1 + 2H(t-T)\right] = r_{\mathrm{atom}} \cos \frac{2\pi}{P} \left\{ t - T - H\mathrm{e}^{2HT} \left[(x^1)^2 + (x^2)^2 \right] \right\} ,$$

$$\mathrm{e}^{HT} x^2 \left[1 + 2H(t-T)\right] = r_{\mathrm{atom}} \sin \frac{2\pi}{P} \left\{ t - T - H\mathrm{e}^{2HT} \left[(x^1)^2 + (x^2)^2 \right] \right\} .$$

Note that

$$\frac{x^2}{x^1} = \tan \frac{2\pi}{P} \left\{ t - T - H\mathrm{e}^{2HT} \left[(x^1)^2 + (x^2)^2 \right] \right\} .$$

The term $He^{2HT}\left[(x^1)^2+(x^2)^2\right]$ is roughly constant over the orbit. In fact, we have a roughly circular orbit in the de Sitter coordinates, with the same period P and radius

$$e^{-HT} r_{\text{atom}} \ .$$

We are neglecting the terms $H(t-T)$ since H is assumed very small and $t \approx T$ during the orbit. This is the result we were looking for. The coordinate t measures proper time, but we must multiply dr by an exponential factor e^{HT} to get the proper distance. Orbits shrink by e^{-HT} with increasing T when viewed in de Sitter coordinates, so that their proper radius can remain the same.

Before continuing with a similar analysis of the Schwarzschild spacetime, let us ask what could be achieved by an exact calculation here. In other words, what could we discover if we took our best theory of electron orbits (or wave functions) in flat spacetime (presumably a quantum theory) and replaced all inertial coordinate derivatives by covariant derivatives with respect to the metric connection for the de Sitter spacetime, then worked out exactly what the electron orbit (or wave function) looked like relative to the above de Sitter coordinates? Naturally, we are not concerned about whether there is such a thing as an orbit in the given theory. We could take any spatial and temporal characteristics of the electron dynamics that our theory could provide us with. These could then be used to find an exact relation between the coordinates and the proper times and distances that this atom would measure for us between neighbouring events, *if we took this as the operational definition for these quantities*.

Of course, no measurements are exact in practice. When space and time intervals themselves are being measured, by placing atoms side by side or counting electron orbits over some period of time, it is easy to see how error will be introduced. But one does at least have a practical understanding of how to link a coordinatised manifold with the real world.

6.2 Schwarzschild Spacetime

The aim here is to study the shape and period of electron orbits around a nucleus placed in the context of Schwarzschild spacetime. Once again, although we assume here that Maxwell's theory of electromagnetism extends successfully by application of the strong principle of equivalence, it will make no difference which theory we actually extend in this way. Maxwell's theory just provides a vocabulary for talking about the spatiotemporal dimensions of the system, i.e., the atom. So we expect the electron orbits to look roughly as they would in flat spacetime when referred to normal coordinates.

Relative to the usual pseudopolar coordinates (t,r,θ,ϕ) for the Schwarzschild spacetime, whatever they are supposed to correspond to in the real world, the metric is

$$g_{\mu\nu} = \begin{pmatrix} \left(1-\dfrac{A}{r}\right)c^2 & & & \\ & -\left(1-\dfrac{A}{r}\right)^{-1} & & \\ & & -r^2 & \\ & & & -r^2\sin^2\theta \end{pmatrix} ,$$

where $A = 2GM/c^2$. This is usually derived directly from Einstein's equation without further comment about what the coordinates mean or how the metric should be interpreted. It is only by asking such a question explicitly when examining some of the applications that the reader can get an idea of what is happening. However, this idea does not come by understanding what justifies the interpretation. It comes rather by guesswork from what is being claimed of the coordinates, in particular, when authors discuss what are called the experimental tests of general relativity, e.g., gravitational redshift, perihelion precession of Mercury, deflection of light, and so on.

The deflection of light is a good example of the didactic problems that are raised when talking about spacetime. The reader is challenged to find a textbook which explains relative to what the light is being deflected. The excellent book [50] by Ryder illustrates the consequences of the 'deflection' by means of a diagram. Nearby stars close to the limb of the Sun (itself momentarily concealed by the Moon during an eclipse) appear to shift away from the Sun as viewed relative to the distant background of celestial objects. One might just be tempted to say that the light from these stars is deflected from the path it would have had if the Sun had not been there. But there is no such path in the spacetime one happens to occupy. We are being asked to compare a path in one spacetime with a path in a different one that does not exist.

Let us just look at two typical claims in the literature of the Schwarzschild solution, to illustrate the problem with the didactic aspects of these discussions. We find this in an elementary course on general relativity [3]:

> To get a more quantitative feel for the distortion of the geometry produced by the gravitational field of a star, consider a long stick lying radially in the gravitational field, with its endpoints at the [Schwarzschild] coordinate values $r_1 > r_2$. To compute its length L, we have to evaluate
>
> $$L = \int_{r_2}^{r_1} \mathrm{d}r \, (1-2m/r)^{-1/2} \, .$$

Here m stands in for GM/c^2. Since this set of points lies in a hyperplane of simultaneity for the Schwarzschild coordinates, a Schwarzschild observer would call this the proper distance between the two endpoints, using the standard geometric interpretation of the metric. But is it really the length of a stick? What would we have to do to get a stick to do this? For example, none of the points of it are in free fall, so they all have some kind of 4-acceleration, and in fact, they all have different 4-accelerations. It is clear that if real rods do behave like this, there must be some physical reason for it.

In fact it is interesting to see how that account continues with regard to the related question of proper distance [3]:

Note that the [increment in the] proper radius R of a two-sphere [centered on the singularity], obtained from the spatial line element by setting $\theta = \text{const.}$, $\phi = \text{const.}$, is

$$\mathrm{d}R = (1-2m/r)^{-1/2}\mathrm{d}r > \mathrm{d}r\ . \tag{6.5}$$

In other words, the proper distance between spheres of radius r and radius $r+\mathrm{d}r$ is $\mathrm{d}R > \mathrm{d}r$, and hence larger than in flat space.

It is intriguing to wonder what the last comment means. For this is not really a comparison with any spheres in flat space. The coordinate interval $\mathrm{d}r$ need not be at a point where the spacetime is even approximately flat. The so-called proper distance is something that is related to the coordinate r in this way, according to the usual geometric interpretation of the metric. In fact, the quoted relation (6.5) is telling us how to understand the coordinates, although it is not telling us why our rulers should be apt for this.

As an aside, we have the same kind of pedagogical difficulty in the following, still in the context of the Schwarzschild metric [3]:

Let us consider proper time for a stationary observer, i.e., an observer at rest at fixed values of r, θ, ϕ. Proper time is related to coordinate time by

$$\mathrm{d}\tau = (1-2m/r)^{1/2}\mathrm{d}t < \mathrm{d}t\ . \tag{6.6}$$

Thus clocks go slower in a gravitational field.

But they go slower than what? Of course, this is a neat inequality and very simple. But does it really tell us that the clock is going slower than the same clock in flat spacetime? After all, $\mathrm{d}t$ is a coordinate change at a place where $r \neq \infty$ and spacetime is not flat. The above relation tells us how to understand the coordinate t at the relevant point, provided that we understand how to interpret proper time as given by the metric.

Anyway, the aim here is indeed to go directly to a justification for the way the metric is interpreted, and the main claim is that one must appeal to the strong equivalence principle to get this. So let us return to the present issue. It is convenient to display the metric connection coefficients relative to these coordinates in the form of four 4×4 matrices:

$$\Gamma^0_{\mu\nu} = \begin{pmatrix} 0 & \dfrac{A}{2r^2}B(r)^{-1} & 0 & 0 \\ \dfrac{A}{2r^2}B(r)^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\ ,$$

$$\Gamma^1_{\mu\nu} = \begin{pmatrix} \frac{Ac^2}{2r^2}B(r) & 0 & 0 & 0 \\ 0 & -\frac{A}{2r^2}B(r)^{-1} & \frac{1}{r} & 0 \\ 0 & \frac{1}{r} & -rB(r) & 0 \\ 0 & 0 & 0 & -rB(r)\sin^2\theta \end{pmatrix} ,$$

$$\Gamma^2_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sin\theta\cos\theta \end{pmatrix} , \quad \Gamma^3_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{r} \\ 0 & 0 & 0 & \cot\theta \\ 0 & \frac{1}{r} & \cot\theta & 0 \end{pmatrix} ,$$

where the definition $B(r) := 1 - A/r$ is just a device to get the second matrix onto one line.

We must now choose a point at which to place the atom. Since $\cot\theta = 0$ when $\theta = \pi/2$, we shall take this value, together with $\phi = 0$ and $r = R$. We may as well choose $t = 0$. These choices are based on the symmetries of the metric, e.g., it is independent of t and ϕ, so one may as well choose the simplest values. In fact, the coordinate curves associated with t and ϕ, and even θ, are the flows of Killing vector fields [30, Sect. 16.4].

Normal coordinates are then given in the neighbourhood of this point by the standard formula (6.3) on p. 230:

$$\begin{aligned} y^0 &= t + \frac{A}{2R^2}\left(1 - \frac{A}{R}\right)^{-1}(r-R)t \, , \\ y^1 &= r - R + \frac{Ac^2}{4R^2}\left(1-\frac{A}{R}\right)t^2 - \frac{A}{4R^2}\left(1-\frac{A}{R}\right)^{-1}(r-R)^2 \\ &\quad -\frac{1}{2}R\left(1-\frac{A}{R}\right)\left(\theta - \frac{\pi}{2}\right)^2 + \frac{1}{R}(r-R)\left(\theta-\frac{\pi}{2}\right) - \frac{1}{2}R\left(1-\frac{A}{R}\right)\phi^2 \, , \\ y^2 &= \theta - \frac{\pi}{2} \, , \\ y^3 &= \phi + \frac{1}{R}(r-R)\phi = \frac{r\phi}{R} \, . \end{aligned}$$

Note the simplification for y^3. Note also that these coordinates are adjusted so that their origin is exactly at the point we chose. The second order terms in t, r, θ and ϕ will not be of much interest to us, because the atomic radius and period are assumed more or less infinitesimal here. This is indeed an assumption inherent in their use as a gauge for length and time.

The next step is to calculate the metric components in the new coordinates so that we can proceed to put it into Minkowski form, by a linear transformation $y^\mu \to z^\mu$. We use the contravariant metric components because we have y^μ in terms of x^μ:

$$g'^{\mu\nu} = \frac{\partial y^\mu}{\partial x^\sigma}\frac{\partial y^\nu}{\partial x^\tau} g^{\sigma\tau} ,$$

to be evaluated at the point $x = (0, R, \pi/2, 0)$. Now the matrices of partial derivatives are just the identity at this point, by the way we have defined the normal coordinates. Hence,

$$g'^{\mu\nu}\big|_{\text{origin}} = \begin{pmatrix} \frac{1}{c^2}\left(1-\frac{A}{R}\right)^{-1} & & & \\ & 1-\frac{A}{R} & & \\ & & \frac{1}{R^2} & \\ & & & \frac{1}{R^2} \end{pmatrix} .$$

Relative to y coordinates, the z coordinates must have derivative matrix

$$\frac{\partial z}{\partial y} = \begin{pmatrix} \left(1-\frac{A}{R}\right)^{1/2} & & & \\ & \left(1-\frac{A}{R}\right)^{-1/2} & & \\ & & R & \\ & & & R \end{pmatrix} .$$

We can now write down the normal coordinates in the form required:

$$\begin{aligned}
z^0 &= \left(1-\frac{A}{R}\right)^{1/2} t \left[1 + \frac{A}{2R^2}\left(1-\frac{A}{R}\right)^{-1}(r-R)\right] ,\\
z^1 &= \left(1-\frac{A}{R}\right)^{-1/2}(r-R)\\
&\quad + \left(1-\frac{A}{R}\right)^{-1/2}\left[\frac{Ac^2}{4R^2}\left(1-\frac{A}{R}\right)t^2 - \frac{A}{4R^2}\left(1-\frac{A}{R}\right)^{-1}(r-R)^2\right.\\
&\quad \left. -\frac{1}{2}R\left(1-\frac{A}{R}\right)\left(\theta-\frac{\pi}{2}\right)^2 + \frac{1}{R}(r-R)\left(\theta-\frac{\pi}{2}\right) - \frac{1}{2}R\left(1-\frac{A}{R}\right)\phi^2\right] ,\\
z^2 &= R\left(\theta - \frac{\pi}{2}\right) ,\\
z^3 &= r\phi .
\end{aligned}$$

We shall consider two electron orbits: one occurring at a more or less fixed value of r and one occurring over a small range of r values. However, we must first examine the motion of the nucleus relative to the normal coordinates. Indeed, if it were moving in the sense of having a coordinate velocity relative to the normal coordinates, SEP tells us that we should have to consider Lorentz contraction effects. And even if it had zero coordinate velocity at the relevant event, it might have nonzero coordinate acceleration, and then we would have to take into account another often unspoken assumption which lies at the very heart of this discussion, namely the clock hypothesis. At this point, let us just spell out the problem as clearly as possible, and return to the clock hypothesis at the end of the section.

The worldline of the nucleus is in fact $(t, R, \pi/2, 0)$ in Schwarzschild coordinates, where only t varies. In z coordinates, this becomes

$$\left(\left[1-\frac{A}{R}\right]^{1/2} t,\ \frac{Ac^2}{4R^2}\left[1-\frac{A}{R}\right]^{1/2} t^2, 0, 0\right) ,$$

where t can be taken as a parameter. So the nucleus is going to move, but at present finds itself at standstill in the normal coordinates. In the de Sitter case, our nucleus was following a geodesic and it came as no surprise to find that it had no motion whatever relative to the normal coordinates. Indeed, not only was it not moving, but it was not going to move either, having no coordinate acceleration or higher order time derivatives of its spatial position in these coordinates. [Its path in the normal coordinates was just $(t-T, 0, 0, 0)$.] The situation now is rather different, because we have a supported nucleus. Although it is not moving in the sense that its speed is zero relative to the normal coordinates, it is going to move, i.e., it has nonzero acceleration relative to the normal coordinates.

The t^2 dependence of z^1 implies that z^1 goes as $(z^0)^2$ for small z^0. Hence, it has zero velocity in these coordinates, even though its acceleration is nonzero. This is just what one would expect. The acceleration relative to these coordinates is the acceleration that any object must have in order not to be in free fall, but rather to just stay put relative to the Schwarzschild system. In fact,

$$\frac{\mathrm{d}z^1}{\mathrm{d}z^0} = \frac{Ac^2}{2R^2(1-A/R)^{1/2}} z^0 . \tag{6.7}$$

We shall return to this point at the end of the section.

But for the time being let us consider our first electron orbit, assuming it to be circular (see Fig. 6.1). Taking the x axis along $\theta = \pi/2$ and $\phi = 0$, and the y axis along $\theta = \pi/2$ and $\phi = \pi/2$, our origin occurs on the x axis at $x = R$, the y and z coordinates being zero. At this point, a small change in z is given by $R\mathrm{d}\theta$. But this is just $\mathrm{d}z^2$. In addition, when $\mathrm{d}r = 0$, $\mathrm{d}z^3 = R\mathrm{d}\phi = \mathrm{d}y$. If we consider an orbit lying in a plane parallel to the (y, z) plane, we expect no distortion or dilatation, but the period should be changed by an R dependent factor. The orbit will be the circular orbit given by $z^1 = 0$ and

$$z^2(z^0) = r_{\text{atom}} \cos \frac{2\pi}{P} z^0 ,$$
$$z^3(z^0) = r_{\text{atom}} \sin \frac{2\pi}{P} z^0 .$$

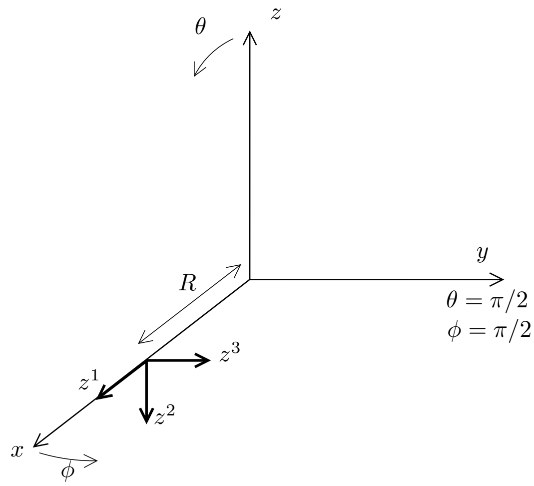

Fig. 6.1 Normal coordinates z^1, z^2, z^3

Now $z^1 = 0$ implies $r \approx R$ remains roughly constant. We are basically ignoring second order terms in the coordinate transformations. Hence,

$$\frac{2\pi}{P} z^0 \approx \frac{2\pi}{P}\left(1-\frac{A}{R}\right)^{1/2} t\ .$$

When t changes by $P/(1-A/R)^{1/2}$, τ has changed by P, because we defined the changes in proper time by the atomic process. This means

$$\mathrm{d}\tau = \left(1-\frac{A}{R}\right)^{1/2} \mathrm{d}t\ .$$

This is exactly what we would conclude from the usual interpretation of the metric.

What about the spatial aspect of the orbit in Schwarzschild coordinates? We have

$$R\left[\theta(t)-\frac{\pi}{2}\right] \approx r_{\mathrm{atom}} \cos\frac{2\pi}{P}\left(1-\frac{A}{R}\right)^{1/2} t\ ,$$
$$R\phi(t) \approx r_{\mathrm{atom}} \sin\frac{2\pi}{P}\left(1-\frac{A}{R}\right)^{1/2} t\ .$$

From the quasi-Euclidean interpretation of the coordinates on the left-hand side, we see that we have a roughly circular orbit relative to Schwarzschild coordinates, with the same radius relative to these coordinates but slowed down by the R-dependent factor. The latter claim means that t must change by slightly more than P for the electron to return to its starting point. This does not imply that any observer will notice anything! If her clock measures proper time, the atom will be behaving normally. All the processes in the observer's body will be slowed down likewise. This slowing down is relative to the coordinate t, which would only be the proper time of an observer at infinity.

We now consider an orbit in the (z^1, z^3) plane, with $z^2 = 0$. According to the analysis above, this lies in the plane $\theta = \pi/2$. We shall take

$$z^1(z^0) = r_{\text{atom}} \cos \frac{2\pi}{P} z^0 ,$$
$$z^3(z^0) = r_{\text{atom}} \sin \frac{2\pi}{P} z^0 .$$

We now find the orbit in Schwarzschild coordinates:

$$\left(1 - \frac{A}{R}\right)^{-1/2} (r - R) \approx r_{\text{atom}} \cos \frac{2\pi}{P} \left(1 - \frac{A}{R}\right)^{1/2} t ,$$

and also

$$r\phi = r_{\text{atom}} \sin \frac{2\pi}{P} \left(1 - \frac{A}{R}\right)^{1/2} t .$$

Here we view r and ϕ as functions of the parameter t. Now $r - R$ is roughly dx and $r\phi$ is roughly dy for these tiny (atomic scale) values of $r - R$ and ϕ. We have an ellipse of width $2r_{\text{atom}}$ in the y direction and $2(1 - A/R)^{1/2} r_{\text{atom}}$ in the x direction.

We conclude that the orbital motion of the atom is slowed down relative to the t coordinate, by just the same factor as in the first orbit. In addition, the orbit is now squashed up in the x direction, which is the radial direction, relative to the r coordinates. A rigid rod comprising a row of juxtaposed atoms lying along a small interval of the x axis will take up less dr than it did at infinity. In fact,

$$\mathrm{d}r = 2\left(1 - \frac{A}{R}\right)^{1/2} r_{\text{atom}}$$

corresponds to proper distance $2r_{\text{atom}}$, so that proper distance is given by

$$\left(1 - \frac{A}{R}\right)^{-1/2} \mathrm{d}r ,$$

just as one would predict from the usual interpretation of the metric.

So what about the fact that our atomic nucleus is about to move relative to the normal coordinates? We said at the outset that we were applying SEP, so that anything happening relative to such coordinates must look roughly as it would in an inertial frame in flat spacetime. (Not forgetting that SEP does much better than this statement containing the word 'roughly', because it provides a way of carrying physical theories over exactly.) This means that the problem of an accelerating atom that happens to be stationary at the relevant spacetime event, referring everything to normal coordinates in the curved spacetime, reduces to precisely the same problem but referring everything to inertial coordinates in a flat spacetime.

But this was (part of) the problem discussed by Bell in his paper [2] and also in Chap. 5. It is often treated in special relativity by making the following assumption, generally called the clock hypothesis [6]:

> When a clock is accelerating, the effect of motion on the rate of the clock is no more than that associated with its instantaneous velocity, i.e., the acceleration changes nothing.

There is a related statement for the effect of motion on the length of a measuring rod, which one might call the ruler hypothesis. These hypotheses are discussed in the book by Brown [6]. The point being made in the present section is of course that one does not need to assume this, because SEP must deliver it via the Bell argument. If Maxwell's theory were the final word in our atomic theory today, we could use an atom as both clock and ruler and calculate exactly what would happen to its shape and period under acceleration, at least in principle. In other words, one could find out whether the above hypotheses were justified for the atom as clock and ruler and for the particular acceleration being considered.

Naturally, if they were not justified in the case of the atom, one might seek some other clock or ruler. The situation gets complicated, because one might not have a good enough theory to test the hypothesis. Indeed, coming back to the atom, which one would nowadays model with quantum theory, one might not be able, even in theory, to test the clock hypothesis. There is room for some more debate here, but it does look as though the key to not introducing any further principles or hypotheses into relativity might be found in Bell's approach, which Brown calls the dynamical perspective in the title of his book [6] and which is being advocated here. And once one has achieved this in special relativity, it is claimed here that SEP immediately transposes the result to general relativity in the way described above for the de Sitter and Schwarzschild spacetimes.

As a final point, note that we might consider the case of a freely falling atom in the Schwarzschild spacetime. To simplify, we could assume that the atom is stationary relative to the Schwarzschild coordinates at the relevant spacetime event of our analysis, like the supported atom in the above analysis. This atom would then follow a timelike geodesic and would sit at the spatial origin of our normal coordinates. Its coordinate acceleration relative to the normal coordinates would be zero and we could apply the strong principle of equivalence and draw the same conclusions about measuring times and lengths with either supported or freely falling atoms, provided that the freely falling atoms are instantaneously stationary when used for measurement.

What we are saying with regard to the clock and ruler hypotheses is therefore this. We are still assuming them here, in our rather rough demonstration that, if atoms are used as clocks and rulers, then they deliver the usual geometric interpretation of the metric. The claim is that exact calculations with our Lorentz covariant theories in special relativity, and exact calculations with the transposition of these theories to general relativity, could in principle decide one way or the other whether these 'hypotheses' are justified for a given clock or ruler. In brief, they are consequences in this approach, rather than assumptions.

6.3 Static Homogeneous Gravitational Field

This is a spacetime with coordinates (y^0, y^1, y^2, y^3) such that the metric components take the form

$$g_{\mu\nu} = \begin{pmatrix} \left(1+\dfrac{gy^3}{c^2}\right)^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{6.8}$$

It is well worth discussing the physical interpretation of this spacetime in some detail before beginning our analysis of the Bell atom in this context. We should say immediately that, when $y^3 = -c^2/g$, this metric is degenerate because $g_{00} = 0$ and g^{00} is not defined. However, we shall soon find that this problem is due to the choice of coordinates, rather like the singularity at the event horizon in the usual representation of the Schwarzschild metric. Such singularities can be removed by a better choice of coordinates.

Of course, we ought to say a word about the coordinates (y^0, y^1, y^2, y^3) in which the metric assumes this representation, in particular, about the way they might be related to observations of the real world. With some pessimism, the reader is invited to search some standard textbooks for an answer.

Intuitively, the idea is that there is a parallel gravitational field in the negative y^3 direction. This field is declared to be static because the metric components do not depend on y^0, which is clearly a time coordinate of some kind. But what about homogeneity? Well, it is natural to seek a locally inertial frame, e.g., at the spatiotemporal origin of these coordinates, and this will in fact explain why this spacetime is said to describe a homogeneous spacetime. Consider the coordinates (t, x, y, z) defined by

$$t = \frac{c}{g}\sinh\frac{gy^0}{c^2} + \frac{y^3}{c}\sinh\frac{gy^0}{c^2} , \tag{6.9}$$

$$x = y^1 , \qquad y = y^2 , \tag{6.10}$$

$$z = \frac{c^2}{g}\left(\cosh\frac{gy^0}{c^2} - 1\right) + y^3\cosh\frac{gy^0}{c^2} . \tag{6.11}$$

Note that the spatiotemporal origins of the two systems coincide. For the record, the inverse transformation is

$$y^0 = \frac{c^2}{g}\tanh^{-1}\frac{gt/c}{1+gz/c^2} , \tag{6.12}$$

$$y^1 = x , \qquad y^2 = y , \tag{6.13}$$

$$y^3 = \left[\left(z+\frac{c^2}{g}\right)^2 - c^2t^2\right]^{1/2} - \frac{c^2}{g} . \tag{6.14}$$

If we evaluate the metric components relative to the coordinates (t,x,y,z), we find that the metric takes the Minkowski form everywhere in the spacetime. These coordinates are not therefore merely locally inertial, with Minkowski form at the event in question and zero connection coefficients at that same event. They are in fact globally inertial. The spacetime we are considering has zero curvature everywhere and is therefore identical to the Minkowski spacetime, apart from the limited range of validity of the original coordinates.

This means that (y^0,y^1,y^2,y^3) constitute some non-inertial coordinatisation of a flat spacetime, in which there are therefore no tidal effects (since curvature describes precisely the tidal effects). One can immediately deduce that the origin $(x,y,z) = (0,0,0)$ of the spatial hypersurfaces $\{t = \text{constant}\}$ in the new coordinate system is in free fall, since it obviously satisfies the geodesic equation. According to the usual rules of general relativity, and in particular SEP, the coordinates (t,x,y,z) are the ones that would be set up in a freely falling frame, in which the connection coefficients are zero and freely falling bodies, i.e., not subjected to any forces, will follow straight lines. (In general relativity, one adopts the idea that free fall is not due to a force. This is a linguistic adjustment which not everyone accepts, made to accord with the idea that forces and accelerations are intimately linked. In general relativity, freely falling bodies follow geodesics, and the geodesic equation says precisely that their four-acceleration is zero.)

For the freely falling observer who uses the (t,x,y,z) coordinate system, the origin $(y^1,y^2,y^3) = (0,0,0)$ of the spatial hypersurfaces $\{y^0 = \text{constant}\}$ of the (y^i) coordinate system is moving with uniform acceleration. The worldline comprising these events is given, with parameter σ, by

$$y^0 = \sigma\,, \quad y^1 = y^2 = y^3 = 0\,. \tag{6.15}$$

In the inertial coordinates, we have

$$t = \frac{c}{g}\sinh\frac{g\sigma}{c^2}\,, \quad x = 0 = y\,, \quad z = \frac{c^2}{g}\left(\cosh\frac{g\sigma}{c^2} - 1\right)\,. \tag{6.16}$$

Eliminating the parameter, this gives the worldline

$$x = 0 = y\,, \quad z = \frac{c^2}{g}\left[\left(1 + \frac{g^2t^2}{c^2}\right)^{1/2} - 1\right]\,. \tag{6.17}$$

This is the worldline of a particle whose 4-acceleration a^μ has constant squared magnitude $a^2 = \eta_{\mu\nu}a^\mu a^\nu = -g^2$ (see Sect. 2.9).

The above change of coordinates (6.9)–(6.11) seems to be somewhat miraculous, in the sense that we did not explain how it was obtained. However, one could find it in this way. Starting with a Minkowski spacetime and inertial coordinates like (t,x,y,z), we can envisage an accelerating observer, following the worldline (6.17). We can set up semi-Euclidean coordinates for the accelerating observer, using the process described earlier in this book, i.e., at each event on her worldline, the observer borrows the space coordinates of an instantaneously comoving inertial observer for all events in the instantaneous hyperplane of simultaneity, and attributes her own proper time to all events in that hyperplane of simultaneity. These semi-Euclidean coordinates are precisely the coordinates (y^0, y^1, y^2, y^3) that we started out with here, and indeed the component form of the Minkowski metric in (6.8), with Euclidean geometry on the spatial hypersurfaces $\{y^0 = \text{constant}\}$, is precisely the component form of the Minkowski metric we obtained for a TUA observer in (2.231) on p. 79. Such coordinates are also called Rindler coordinates.

The idea put forward in standard presentations is that a uniformly accelerating observer in a flat spacetime without even non-tidal gravitational effects would adopt these coordinates (y^i), but this is a debatable point as already discussed, if only because, in general relativity, an observer is free to adopt any coordinates whatever. One way to motivate the classic view is to consider the notion of rigidity given by Rindler [48]. The idea is that, when a rod is made to move along its axis, i.e., accelerated, different elements of it will be moving at different speeds, but the rod is declared to be rigid if each element instantaneously adopts the appropriate FitzGerald contracted length for that speed. One does not worry about the microphysical explanation of this process, or even whether it is possible for any object to behave like this.

In fact it is very easy to see that, if the uniformly accelerating observer (UAO) carried measuring rods that were rigid in Rindler's sense, laying them along the axis of motion, they would always mark out the semi-Euclidean coordinate length. Put another way, they would always have constant semi-Euclidean coordinate length. Rindler rigidity corresponds exactly to the more general notion of rigid motion we have been discussing in this book [31]. But when the semi-Euclidean coordinates are set up, UAO adopts precisely the spatial coordinates of an instantaneously comoving inertial observer (ICIO). One comes straight back to the ruler hypothesis, and also to Bell's idea here. The point is that these Rindler-rigid rods exactly obey the ruler hypothesis. But if the rods were Bell atoms, or strings of Bell atoms laid side by side and held together by electromagnetic forces, one could in principle do an exact calculation with Maxwell's theory to find out whether the Bell ruler would serve as a Rindler ruler. That is, one should be able to test, by calculation, the hypothesis that when a ruler is accelerating, the effect of motion on the length of the ruler is no more than that associated with its instantaneous velocity, i.e., that the acceleration changes nothing.

Let us return to the metric (6.8) at the beginning of this section, what would the general relativist deduce from the discovery of a global inertial frame? There would appear to be several possibilities, all of which are equivalent as far as general relativity is concerned. One might say that one was in a perfectly flat region of spacetime far from any source of curvature (gravity), and that the coordinates $\{y^\mu\}$ were simply those adopted by a uniformly accelerating observer clever enough to set up a system so well adapted to her motion. This is what we have just been considering. But one might also say that the coordinates (t, x, y, z) are those of an observer freely falling in a static and homogeneous gravitational field (SHGF). The fact that the field is static and homogeneous is what leads to there being no tidal effects, i.e., zero curvature.

Presumably, in the second case, if the observer looked around, she would find some source, i.e., some distribution of energy (e.g., mass energy) to which one could attribute the presence of this SHGF. It is hard to imagine what it might be. However, a practically-minded observer with knowledge of physics would probably just say that, in reality, the curvature is not quite zero but that this is a good approximation over some region of spacetime, of the kind usually made in Earth-based laboratories, given the accuracy with which measurements can be made.

But the Earth-based observer would in fact have a considerable advantage over one who was simply presented with the interval associated with this metric, viz., (6.18) below, and could not see the source. In the laboratory, one can measure the acceleration relative to the floor, for example, assuming that it is at rest relative to the gravitational source. At some risk of confusion, let us call this the acceleration due to gravity, but bearing in mind that there is no acceleration due to gravity in general relativity (there is only acceleration when one is not allowed to free-fall). Without sight of the gravitational source or other reference point, our relativist in possession of the appropriate interval, viz.,

$$\mathrm{d}s^2 = \left(1+\frac{gy^3}{c^2}\right)^2 (\mathrm{d}y^0)^2 - (\mathrm{d}y^1)^2 - (\mathrm{d}y^2)^2 - (\mathrm{d}y^3)^2 \,, \tag{6.18}$$

could not even determine the acceleration due to gravity. Looking at the expression in (6.18), one might want to say that it was equal to g, because the interval does indeed single out this value. But in a certain sense, the value g is just an artefact of the choice of coordinates.

One should have no doubt about this. Starting with the Minkowski frame, one could have chosen any uniformly accelerating observer, with value g' say, and obtained new coordinates $\{y'^{\mu}\}$ such that the interval takes the form

$$\mathrm{d}s^2 = \left(1+\frac{g'y'^3}{c^2}\right)^2 (\mathrm{d}y'^0)^2 - (\mathrm{d}y'^1)^2 - (\mathrm{d}y'^2)^2 - (\mathrm{d}y'^3)^2 \,. \tag{6.19}$$

The fact is, of course, that in this specific context, one cannot say how much of the acceleration is due to gravity and how much is due to some other effect, e.g., the kind of accelerating effect we were presumably imagining when we considered a uniformly accelerating body in a flat spacetime (with no gravity).

Perhaps enough has been said to show how tricky relativity can be when one begins to think about coordinates and the way they might be related to the real world. Let us turn to the kind of analysis we have done with an ideal atom in the other two spacetimes. Starting in the proposed SHGF laboratory frame given by the coordinates $\{y^{\mu}\}$, we seek normal coordinates at some point. Since the metric components (6.8) depend on y^3, we take inertial coordinates about some point $\{y^{\mu}\} = (0,0,0,Z)$. We now seek coordinates $\{x^{\mu}\}$ with inertial metric and the correspondence $(0,0,0,Z) \longleftrightarrow (0,0,0,0)$. It is also essential that an atom sitting motionless at the spatial point $(0,0,Z)$ of the $\{y^{\mu}\}$ system should also be sitting motionless at the spatial origin of the $\{x^{\mu}\}$ system, at least at the event $(x^{\mu}) = (0,0,0,0)$. (We shall soon show that it must start moving relative to the inertial system, but it is at least instantaneously stationary there.) For this, we have to find the worldline of the spatial point $(0,0,Z)$ of the $\{y^\mu\}$ coordinate system as y^0 goes by and work out its motion relative to the inertial coordinates.

We propose the coordinates

$$x^0 = \left(\frac{c^2}{g} + y^3\right) \sinh \frac{gy^0}{c^2} , \quad x^3 = \left(\frac{c^2}{g} + y^3\right) \cosh \frac{gy^0}{c^2} - \frac{c^2}{g} - Z , \tag{6.20}$$

with the usual $x^1 = y^1$ and $x^2 = y^2$. Note that the event $(0,0,0,Z)$ in the $\{y^\mu\}$ system corresponds to the origin of the new system. The transformation is easily inverted. Since,

$$\tanh \frac{gy^0}{c^2} = \frac{x^0}{x^3 + c^2/g + Z} , \tag{6.21}$$

we have

$$y^0 = \frac{c^2}{g} \tanh^{-1} \frac{x^0}{x^3 + c^2/g + Z} , \tag{6.22}$$

whence, after a little manipulation,

$$y^3 = \left[\left(x^3 + \frac{c^2}{g} + Z\right)^2 - (x^0)^2\right]^{1/2} - \frac{c^2}{g} . \tag{6.23}$$

What we have done here is to use the previous inertial coordinates, whose origin coincided with the event $(y^\mu) = (0,0,0,0)$, and translate the origin a distance Z up the third axis. It is quite obvious that these coordinates must be inertial too, i.e., if we worked out the metric, it would have the Minkowski form.

What about the speed of the spatial point $(0,0,Z)$ of the $\{y^\mu\}$ system as viewed relative to these inertial coordinates? As the time y^0 goes by, this point is given by

$$x^0(\sigma) = \left(\frac{c^2}{g} + Z\right) \sinh \frac{g\sigma}{c^2} , \quad x^3(\sigma) = \left(\frac{c^2}{g} + Z\right) \left(\cosh \frac{g\sigma}{c^2} - 1\right) , \tag{6.24}$$

where we have renamed the parameter y^0 as σ. Naturally, $x^1(\sigma) = 0$ and $x^2(\sigma) = 0$. We now eliminate the parameter to obtain the worldline in the form

$$x^1 = 0 = x^2 , \quad x^3(x^0) = \left[\left(\frac{c^2}{g} + Z\right)^2 + (x^0)^2\right]^{1/2} - \left(\frac{c^2}{g} + Z\right) . \tag{6.25}$$

This is similar to (6.17) on p. 247, which is retrieved by putting $Z = 0$. The speed as viewed in the inertial frame is

$$v(x^0) = c\frac{\mathrm{d}x^3}{\mathrm{d}x^0} = \frac{cx^0}{\left[\left(\dfrac{c^2}{g} + Z\right)^2 + (x^0)^2\right]^{1/2}} , \tag{6.26}$$

and this is zero when $x^0 = 0$. It is interesting to note that this worldline also has uniform acceleration a^μ. In fact, it turns out that

$$a_\mu a^\mu = -\frac{g^2}{\left(1+\dfrac{gZ}{c^2}\right)^2} , \tag{6.27}$$

so the pseudolength of the four-acceleration of the worldline through $(y^1, y^2, y^3) = (0,0,Z)$ decreases as Z increases, something already established in Sect. 2.9. This shows that the Rindler ruler mentioned above, carried by a uniformly accelerating observer, must undergo different accelerations at different points along its length.

We have a similar situation to the one we had for the Schwarzschild spacetime. An atomic nucleus with worldline $(y^0, 0, 0, Z)$ as y^0 varies is instantaneously stationary relative to the global inertial coordinate system $\{x^\mu\}$ when $y^0 = 0$, but it is accelerating, so it is going to move, and indeed it was moving prior to $y^0 = 0$. We are now ready to examine the electron orbit. As before, we shall assume it to have the form

$$x^1(x^0) = r_{\text{atom}} \cos\frac{2\pi x^0}{P} , \quad x^3(x^0) = r_{\text{atom}} \sin\frac{2\pi x^0}{P} , \tag{6.28}$$

on the grounds that we expect Maxwell's theory to apply in this frame and the nucleus is instantaneously at rest there. But we shall bear in mind that, even though it is an inertial frame, and not just a locally inertial frame, there are still errors here due to the fact that the nucleus was moving a short while ago, and it is its state a short while ago which determines its fields out at the radius of the electron orbit. For the moment we may consider that we are making the approximation that the electron orbit is small and fast enough for us to be able to ignore this fact for the present purposes. In this view of things, the clock and ruler hypotheses (as applied to the atom) are taken as stating that this approximation is adequate for the present purposes, so they could be viewed as assumptions here. However, it should be borne in mind that what is being advocated here is the idea that they are not independent hypotheses, but claims that could be justified by theory, using a better theory for the atom and a more accurate calculation.

We reexpress the electron orbit in terms of the $\{y^\mu\}$ coordinates. The result is

$$y^1 = r_{\text{atom}} \cos\left[\frac{2\pi}{P}\left(\frac{c^2}{g}+y^3\right)\sinh\frac{gy^0}{c^2}\right] , \tag{6.29}$$

$$\left(\frac{c^2}{g}+y^3\right)\cosh\frac{gy^0}{c^2} - \left(\frac{c^2}{g}+Z\right) = r_{\text{atom}} \sin\left[\frac{2\pi}{P}\left(\frac{c^2}{g}+y^3\right)\sinh\frac{gy^0}{c^2}\right] . \tag{6.30}$$

We must now approximate. The period will be very short compared with times we would expect to measure. Throughout the short duration of the orbit, the coordinate y^0 will barely change, remaining close to its initial value of zero. To a first approximation, we can take

$$\sinh\frac{gy^0}{c^2}\approx\frac{gy^0}{c^2}\quad\text{and}\quad\cosh\frac{gy^0}{c^2}\approx 1\ . \tag{6.31}$$

Now the orbit is given by

$$y^1\approx r_{\text{atom}}\cos\frac{2\pi\left(1+\dfrac{gy^3}{c^2}\right)y^0}{P}\ , \tag{6.32}$$

$$y^3-Z\approx r_{\text{atom}}\sin\frac{2\pi\left(1+\dfrac{gy^3}{c^2}\right)y^0}{P}\ . \tag{6.33}$$

Since the orbit occurs near the point $(0,0,0,Z)$ of the $\{y^\mu\}$ coordinates, y^3 also remains close to Z, as attested by (6.33), and we insert $y^3\approx Z$. It thus takes the form

$$y^1\approx r_{\text{atom}}\cos\frac{2\pi\left(1+\dfrac{gZ}{c^2}\right)y^0}{P}\ , \tag{6.34}$$

$$y^3-Z\approx r_{\text{atom}}\sin\frac{2\pi\left(1+\dfrac{gZ}{c^2}\right)y^0}{P}\ . \tag{6.35}$$

Our electron orbit will therefore have the same circular shape with unchanged radius in the $\{y^\mu\}$ picture, but its period will now be

$$P_{\text{lab}}=P\left(1+\frac{gZ}{c^2}\right)^{-1}\approx P\left(1-\frac{gZ}{c^2}\right)\ , \tag{6.36}$$

relative to the y^0 coordinate. Note that an orbit in the (x^2,x^3) or (x^1,x^2) planes would give the same results, i.e., the same radius and a period changed in the above way.

We can now express proper time and proper distance in terms of the $\{y^\mu\}$ coordinates, something usually taken directly from the metric. The atomic radius is r_{atom} in the $\{y^\mu\}$ coordinate system, so the coordinates y^1,y^2,y^3 give the proper distance directly. This agrees with what we would deduce from the metric, which has $(-1,-1,-1)$ down the leading diagonal of the spatial sector in these coordinates. However, it is now

$$\left(1+\frac{gZ}{c^2}\right)\mathrm{d}y^0$$

that measures the proper time interval when y^0 changes by $\mathrm{d}y^0$. Once again, this is what the metric suggests in the usual interpretation.

6.4 Taking Stock

The last example has brought us back to the subject of Bell's paper [2], because the characteristics of this spacetime allow one to interpret it as a spacetime in which there are no gravitational effects, not even non-tidal ones. The worldline $(y^0, 0, 0, 0)$ with varying y^0 can be viewed as the worldline of an observer with uniform acceleration in such a spacetime and the coordinates $\{y^\mu\}$ as semi-Euclidean coordinates this observer might adopt to describe events. Alternatively one can view the worldline $(y^0, 0, 0, 0)$ with varying y^0 as the worldline of an observer supported, i.e., prevented from free fall, in an SHGF with the appropriate 'acceleration due to gravity'. Whatever the interpretation, our idea is to use an atom following the worldline $(y^0, 0, 0, Z)$ to define proper length and time intervals in the neighbourhood of this event: its radius indicates proper length and its period indicates proper time, by definition. Likewise for the other two spacetimes.

The strong equivalence principle (SEP) is used to relate the proper lengths and times as defined operationally in this way with normal coordinates in the vicinity of the chosen events. Approximations inevitably enter here. For one thing, normal coordinates for a given event are not unique for various reasons, one being that the connection coefficients are only required to be zero at the chosen event and in fact there is no restriction on them away from this event. Another is that the atom has spatial extent, and its period extends in time, so it cannot be said to probe just one event. Yet another is that we have not gone into the details of the theory governing the structure of the atom, and even if we did, we would only have approximate theoretical solutions. But we can still claim that the approximation is good enough to demonstrate the worth of this operational definition, because we are able to reproduce the usual interpretation of the metric. In fact, the three calculations have been completely concerned with this last issue, which is really the question of how the normal coordinates relate to whatever coordinates we started with, relative to which the metric components were originally specified. So the problem has fallen into two parts:

- The first asks how we measure normal coordinates.
- The second asks how we relate normal coordinates to the coordinates we started out with, and which we are trying to understand physically, particularly with regard to the metric components as specified relative to that original coordinate system.

The calculation part of our discussions deals with the second point and is largely trivial because it is indeed a quite general result: intervals of normal coordinates are intervals of proper time or length for the relevant observer as they are defined under the standard interpretation of the metric. The interesting part of the discussion is whether atoms can be used to measure normal coordinates, and if so, why. The claim here is just that, if we decree proper length and time to be what atoms measure, by definition, we can then use SEP and whatever theories we have for the atom to justify the claim that normal coordinates give a good approximation to proper length and time under this definition. At this point we deduce that the metric can be used to give proper length and time in the usual way. In this approach, we understand why the metric components can be used in this way. So we are defining proper time and length by what the atom measures, then using SEP to say that normal coordinates will give this to a good approximation.

This does turn things round compared with a standard view in which we posit Minkowski spacetime in SR or a curved metric manifold in GR, with metric fields η or g, respectively, and *postulate* that the metric field corresponds in the usual way to lengths and times we measure by the usual techniques. And it shows that, for this classic approach to work, we are *compelled* to build Lorentz symmetric or locally Lorentz symmetric field theories of matter, otherwise the devices we traditionally use to measure lengths and times could not always agree, even approximately, with the dictates of the metric. So this kind of postulate regarding the metric is not innocent with regard to our field theories of matter, a point that may be overlooked in anti-constructivist accounts such as [44] (see also Chap. 9). The converse view discussed and advocated in this chapter, whilst stopping short of constructivist pretensions, claims that, if we build Lorentz symmetric or locally Lorentz symmetric field theories of matter, then we can deduce the usual interpretation of the metric field.

Let us just focus once more on the problem that led us to consider the clock and ruler hypotheses. In two of our investigations with the Bell atom, we consider an observer who is not following a geodesic, viz., an observer sitting at fixed Schwarzschild space coordinates and another sitting at fixed semi-Euclidean space coordinates in a flat spacetime. One can arrange normal coordinates at any given event on these worldlines in such a way that an atom moving with these observers is instantaneously stationary relative to the coordinates, but one cannot avoid the atom beginning to move or having had motion relative to these coordinates in some neighbourhood of the chosen event. Our definition of proper length and time for these observers tells us that these accelerating atoms will indicate these quantities by their spatiotemporal dimensions. In the process we use to deduce what the atoms will look like relative to these normal coordinates, we find that we have a problem that was already encountered in special relativity when an atom is accelerating, viz., the fact that the atom is instantaneously stationary does not guarantee that its spatiotemporal aspects will be identical to those of a coincident atom that is permanently stationary.

In fact, Bell's analysis indicates that this is only likely to work as an approximation. Using some exact theory in special relativity, one can in principle say exactly what the accelerating atom will look like. One can therefore justify the hypothesis that it will look sufficiently like the inertial atom, should that be the case. In other words, the clock and ruler hypotheses that make the similarity claim for the two atoms in special relativity are not fundamental assumptions required by the theory. They are just assumptions about measurement accuracy and they can be justified or rejected on Bell's approach, although one might prefer to reject the atom as measuring device, a point we come to below. In any case, this problem is carried over identically by SEP to general relativity and solved in the same way there.

We could for example look at things like this. Suppose the accelerating observers we have been talking about in Schwarzschild spacetime or in an SHGF were to use freely falling atoms to measure proper length and time, since this is in a sense what the clock and ruler hypotheses would appear to recommend. It is clear from our definition of proper length and time by the spatiotemporal characteristics of atoms that these freely falling atoms are not a priori indicating the proper length and time of the accelerating observers, although it might turn out by chance that they were, or indeed that they gave a good approximation. Relative to the normal coordinates at the chosen events, one can arrange for the freely falling atoms to remain stationary at the origin of these normal coordinate systems, and this would mean that they had slightly different spatiotemporal characteristics to the atoms that accelerate relative to these coordinate systems. This is exactly the question raised by the clock and ruler hypotheses. The question in any particular circumstance is just whether the expected difference would be big enough given the accuracies elsewhere in the calculation or the measurement.

Brown says this about the clock hypothesis (extending the quote on p. 245):

> This is the claim that when a clock is accelerating, the effect of motion on the rate of the clock is no more than that associated with its instantaneous velocity — the acceleration adds nothing. This allows for the identification of the integration of the metric along an arbitrary timelike curve — not just a geodesic — with the proper time. This hypothesis is no less required in general relativity than it is in the special theory. The justification of the hypothesis inevitably brings in dynamical considerations, in which forces internal and external to the clock (rod) have to be compared. Once again, such considerations ultimately depend on the quantum theory of the fundamental non-gravitational interactions involved in material structure.

So these ‘hypotheses’ follow from the minimal extensions of our theories of physics from special to general relativity with the help of the strong equivalence principle, provided we are ready to apply Bell’s approach. What we can do (at least in principle) when we have the minimally extended theories, and when we are ready to apply Bell’s reasoning, is to find the exact motion of the electron around the nucleus in arbitrary coordinates for an arbitrary spacetime, or in quantum theoretical terms, the exact shape of the electron wave function as a 4D region of spacetime. By the way these theories are set up via the strong equivalence principle, we know that, in normal coordinates in which the atom is instantaneously at rest at the origin, we will get a close approximation to the behaviour of the same atom but stationary in a flat spacetime. We thus decree this as the operational definition of clock and ruler, accepting of course that there will be a small error due to the curvature and the non-inertial motion of the atom. The standard, purely theoretical relation between the normal coordinates and whatever coordinates we are trying to understand (along with their metric) then tells us how to interpret the metric components in those other coordinates.

Along the way, the clock and rod hypotheses will be justified by the fact that the minimally extended theories and this Bell approach give practically the same answers for cases that are actually sure to differ slightly. For example, one can look at the number of revolutions of the electron (as given exactly by the minimal extension of Maxwell's theory MEME) for an atom that is accelerating in a flat spacetime and for another which, at some event, has instantaneously the same speed (and maybe is not accelerating, or has a different acceleration). There is an inevitable approximation due to the fact that the period of the electron is finite, even though small, so that, in order to count a few periods of the two electrons, one cannot avoid the two atoms moving to different spacetime events. The point is that the main effect here is due to the speed and that any effect due to curvature or acceleration is simply written off by the certitude that one is condemned to make an approximation.

But if one found oneself in a situation where the clock or ruler hypotheses were not valid at the given level of accuracy, would this not suggest that one's measuring instrument were not adequate to that situation? Would this be a criterion for seeking some more refined measuring tool? And how far could one go on using quantum wave functions to gauge lengths and times before spacetime so defined became impossible to manage? How far could one go in imposing the spacetime continuum of a differentiable manifold on length and time intervals obtained in this way?

6.5 Linking Theory to Measurement. Interpretation of the Metric

Friedman and Scarr's weak locality hypothesis (WLH) is inspired by Mashhoon's locality hypothesis (LH), as reviewed in [36, 37]. The aim in the next chapter will be to identify Mashhoon's motivations and confirm the conclusions of this and the last chapter. In Mashhoon's own terms, LH is the assumption that, at each instant along her worldline, an accelerated observer is physically equivalent to an otherwise identical momentarily comoving inertial observer, or more briefly, accelerated observers are always pointwise inertial. We shall need to understand what is meant by *physically equivalent*, or indeed by *being pointwise inertial*.

But before commenting on the above papers, let us attempt to make a clear statement about how to link the manifold and metric structure of GR to observations of the real world. To begin with, we may postulate a weak equivalence principle (WEP) which states that the locally inertial frames we know to exist in the mathematical structure of the theory correspond to freely falling, non-rotating laboratory frames in the following sense: if we *naturally* set up coordinates in such laboratories by using standard techniques (whatever they may be!), then we obtain locally inertial coordinates. There are two pressing questions:

- Why should this be a good starting point?
- How do we know whether our laboratory is non-rotating?

The techniques we use to measure lengths and times and hence set up real coordinates involve real measuring equipment, so they indirectly involve all our field theories of physics. To justify this WEP as starting point, we thus need the strong equivalence principle (SEP) that tells us how to ship our non-gravitational theories of physics into the curved spacetime context. The usual way of doing this is to say that, in a locally inertial frame (LIF), our non-gravitational field theories should look roughly as they do in flat spacetime. This can be achieved in a minimal way by replacing coordinate derivatives by covariant derivatives for the metric connection, for example. So if the freely falling, non-rotating laboratory really did correspond to a LIF, one would expect to just use standard measuring techniques in the normal way, as though one were in a flat spacetime.

Regarding the second question, the answer is probably that we do not know whether the laboratory is non-rotating. We can think of it as a specific hypothesis about the specific situation we are in. If we succeed in setting up a correspondence between theory and measurement that reveals no contradictions, we could suppose the hypothesis valid. If we fail, it may be this hypothesis that is invalid and we must then start comparing options. A worse situation would be one in which we obtain what appears to be a good correspondence but this hypothesis along with one of the others, such as WEP, is invalid. On the other hand, a good metaphysical hypothesis here is perhaps that such collusion would be improbable. This is then a typical question of global consistency between theory and measurement. The hypothesis that laboratories with fixed orientations relative to remote objects in the Universe are non-rotating seems successful within present accuracies, although we do not know why.

This is actually the end of the story when it comes to linking the manifold and metric structure of GR to observations of the real world. We can then consider non-gravitational phenomena in the LIF and transform to any other coordinates or frame we choose. This exploits SEP and the hypothesis that the representation of all physical quantities by tensorial objects in the mathematical picture gives a good correspondence with reality.

This is where the idea of interpreting the metric field comes in, along with the possibility of assessing the suitability of putative clocks and rulers, i.e., real time and length measuring devices, under specific conditions of acceleration using our non-gravitational field theories of physics. The main claim is that one can actually prove that the metric field in GR can be interpreted as it usually is in terms of lengths and times as they are normally measured. In Bell's approach, we ask what happens to a clock or ruler when it accelerates. We use an idealistic classical atom to fulfill both functions. We make the operational definition that, wherever it is and whatever it is doing, the radius and period of the atom indicate lengths and times, respectively, for an observer moving with it. The minimum extension of electromagnetism (MEME) to the curved spacetime, or a better theory, can be used to say exactly what happens to the atom when carried by an observer following some arbitrary worldline. We first describe the atom in normal, i.e., locally inertial coordinates, then convert to any coordinates we like, e.g., coordinates adapted to the observer in some sense.

We find that to a good approximation the atom indicates lengths and times as we would usually calculate them using the metric. We can also test the clock and ruler hypotheses for this particular clock and ruler, viz., the atom, under whatever conditions are specified by the worldline of the nucleus. That is, we can actually assess the extent to which the atom resembles an instantaneously comoving inertial atom, i.e., an identical atom in free fall at the same event, with the same four-velocity there. Put another way, if the aim is to show that the metric is good for defining lengths and times, we can calculate the expected error using the atom as ruler and clock, and hence actually determine the extent to which the atom fails as clock or ruler for the given accelerating worldline.

But could all this also *prove* that the metric field delivers the lengths and times that would be measured by standard techniques by observers out there in the real world? A counterargument is that this been assumed anyway in the first step of this discussion, where we set up locally inertial coordinates in a real laboratory. We said earlier that, if we naturally set up coordinates in freely falling, non-rotating laboratories by using standard techniques, whatever they may be, then we obtain locally inertial coordinates. But such coordinates are intimately related to the metric, because they are coordinates such that the metric is assumed to take on the Minkowski diagonalised form $\mathrm{diag}(1,-1,-1,-1)$. So saying that standard techniques naturally measure lengths and times in the freely falling, non-rotating laboratory does seem implicitly to *assume* that the metric field indicates those lengths and times the way we usually take it to.

In this picture then, or rather in this attempt at a logical account of the link between the manifold and metric structure of GR and observations of the real world, the metric field would be *defined* as specifying lengths and times, but statements like the clock and ruler hypotheses, or more generally the Mashhoon locality hypothesis (see Chap. 7), could be tested by calculations using SEP for the specific measuring equipment proposed, and estimates made for the errors they involve.

There is nevertheless a sense in which, without SEP, one would have difficulty even interpreting the metric field, although this may seem unlikely on the above account, where we *define* the metric field as being the thing that specifies lengths and times in LIFs. Is there any vestige of truth in such a claim? Here we argue that there is, since we do need to invoke SEP, which tells us how to ship our non-gravitational theories of physics into the curved spacetime context, precisely so that we can justify carrying out measurements as though we were in a flat spacetime. Indeed, SEP says that, in a locally inertial frame, our non-gravitational field theories should look roughly as they do in flat spacetime, and it is these theories that govern our measuring devices.

Another line of thinking would be to decree that the metric indicates lengths and times once we have coordinates, but the problem is then to know what coordinates we have set up. If we could also decree that coordinates naturally set up in freely falling, non-rotating laboratories are inertial, without trying to justify that claim, it seems that we would logically succeed in simply imposing the meaning of the metric by edict, but then we lose the benefits of the justification described above. And as mentioned in Sect. 6.4, we are then forced to build Lorentz symmetric field theories of matter in order to understand why what we usually refer to as clocks and rulers do in fact deliver times and lengths, so the edict is not innocent with respect to our dynamical theories.

So far we have considered two possible views of the metric field:

- Definable as what specifies lengths and times.
- Provable as what specifies lengths and times.

The discussion in Brown's paper entitled *The behaviour of rods and clocks in general relativity, and the meaning of the metric field* [7] focuses on the following opposing views:

- The metric field is fundamentally of *geometrical* significance and regarded as being an intrinsic property of spacetime, specifying the very fabric of spacetime in some sense.
- The metric field is just like any other field over spacetime, but happens to specify lengths and times as we would usually measure them (at least approximately) by our rulers and clocks for a deeper reason related directly to the fact that all our non-gravitational field theories of matter governing the internal constitution and interactions of our rulers and clocks are locally Lorentz symmetric.

The above discussion is clearly related to this through the attempt to specify the exact role of SEP and the dynamical theories that are absolutely necessary in every case when trying to decide whether the metric does in fact yield the length or time measured by a real ruler or clock.

The fact that we always require several non-gravitational field theories to understand the relationship between a real length or time measurement and the metric field suggests that the latter is not primarily or intrinsically concerned with lengths or times. On the other hand, the fact that the metric field can nevertheless always yield a value for the length or time actually measured suggests that it is somehow intimately related to our notions of length and time. Brown suggests that this intimate relation is due to the deeper fact that, for some mysterious reason, all our non-gravitational field theories of matter are locally Lorentz symmetric.

He shows that (global) Lorentz symmetry of our field theories of matter is also the most fundamental hypothesis of special relativity. In his account, it is better to begin with this than principles like the relativity principle which can be deduced easily from it, or talk of light signals that implicitly depend on a theory of light which needs to be Lorentz symmetric, while at the same time one must deduce anyway at the end of the day that *all* field theories are actually Lorentz symmetric.

The opposing idea that the metric field is fundamentally of *geometrical* significance and should be regarded as an intrinsic property of spacetime, specifying the very fabric of spacetime in some sense, is best put by Brown's quote from Will [52]:

> The property that all non-gravitational fields should couple in the same manner to a single gravitational field is sometimes called universal coupling. Because of it, one can discuss the metric as a property of spacetime itself rather than as a field over spacetime. This is because its properties may be measured and studied using a variety of different experimental devices, composed of different non-gravitational fields and particles, and the results will be independent of the device. Thus for instance the proper time between two events is characteristic of spacetime and of the location of the events, not of the clocks used to measure it.

But the interpretation of the mathematically defined quantity proper time as the extent to which physical processes of all kinds will have proceeded in systems following the chosen worldline depends heavily on the non-gravitational field theories at work in the system and the only reason why different systems deliver the same (or closely similar) estimates for what we usually call time is that we assume SEP, which itself assumes local Lorentz symmetry of those non-gravitational field theories. Put another way, the fact that we *can* discuss the metric as a property of spacetime itself rather than as a field over spacetime distracts us from the fact that this depends entirely on the unexplained and often unmentioned hypotheses of universal coupling and local Lorentz symmetry that license this.

So there is a risk of missing the fundamental role of these hypotheses, which are themselves mysterious in Brown's sense of being explanation-seeking, if we simply decree the metric as being part of the very fabric of spacetime, or merely declare it to specify the geometry of spacetime. In addition, Brown's picture reminds us that we must actually set up coordinates in the real world, which we expect to correspond somehow to the theoretical coordinates, and here we come to the idea put forward above that we can achieve this by identifying freely falling, non-rotating laboratories through a notion of global consistency of construction and theory, always assuming that the theory is good enough to describe reality.

Naturally, the issue of which account is most logical has been widely debated, particularly among philosophers. Even in pure mathematics, there are different logical accounts of any given theory. These go down as different axiomatisations. And axiomatisation serves two specific purposes in the context of pure mathematics:

- By filtering an axiomatisation, i.e., by keeping a subset of the axioms and dropping others, a given mathematical theory can be generalised.
- Generalisation, but also the modification of axiomatisations, leads to an understanding of the mathematical theory, by showing what aspects of it depend on what other aspects, rather in the manner of analysis (extracting axioms) and synthesis (producing theorems).

But although pure mathematics is *always* inspired by something in the real world (since we are part of that world), it is never held accountable by that world in the way that a theory of physics would be. It is held accountable to some extent, because it has to be logically consistent, but it is not required to correspond in detail to things that happen around us in the way that a theory of physics must, particularly through its predictions.

So on top of the axioms of the theory, if one may refer to the logical grounding of a physical theory in those terms (and why not, since it is only mathematics?), there is another major ingredient in physics: the correspondence with reality. All the more reason then to try to get a better understanding of physics by doing something like axiomatisation, but taking into account the extra ingredient. And the purpose in this case is presumably only the second of the above, namely to obtain a better understanding, by generating several explanations for what we are doing which happily live together without conflict.

Since there is so much debate, sometimes conflictual, between human beings who prefer one way of putting things rather than another, instead of viewing the existence of mutually consistent explanations as a bonus, perhaps it is worth remembering that all these explanations, even what we call the laws of nature, are only *our* way of understanding the world, no matter how well they seem to us to predict what is going on in that world. The present view is that the dynamical explanations for special relativistic effects are worth having, and we shall return to this in detail in Chap. 9, since these explanations turn out to be closely connected with the phenomenon of acceleration.

Chapter 7
Mashhoon's Locality Hypothesis

Let us now turn specifically to Mashhoon's locality hypothesis, as reviewed in [36,37]. He motivates the whole discussion by the observation that the basic laws of microphysics have been formulated with respect to ideal inertial observers, whereas in fact all actual observers are accelerated. So, for example, we originally obtained Maxwell's equations of electromagnetism, describing these effects in inertial frames, and now we have the problem of describing what a real, i.e., accelerating observer, will consider to be these effects.

For Mashhoon, it is clear that we are talking about a *real* problem, namely the problem of relating what real, probably accelerating observers observe to the fundamental theory. This problem concerns observers in gravity-free contexts just as much as observers in situations where there is gravity. His aim then is to establish a connection between actual and inertial observers, achieved in SR by the locality hypothesis, viz., the assumption that an accelerated observer at each instant along her worldline is physically equivalent to an otherwise identical momentarily comoving inertial observer. He considers a non-inertial observer as passing through, in some sense, a continuous infinity of hypothetical momentarily comoving inertial observers.

It is important to understand what is meant in practice by the term *physically equivalent* and this is the aim of the following sections of this chapter. We shall see (e.g., on p. 269) that it means that every measurement of any physical quantity by the accelerating observer or her measuring device will give the same measurement as would have been obtained by the instantaneously comoving inertial observer, to within measurement accuracy.

So here is Mashhoon's picture. The special theory of relativity, i.e., the standard relativistic physics of Minkowski spacetime, is primarily based on a fundamental symmetry in nature, namely, Lorentz symmetry, which allows us to relate the physical measurements of inertial observers at rest in one inertial frame to those at rest in another inertial frame. Of course, we have to be able to identify the inertial frames, and this is done on the level of global consistency of theory and measurement: the (real) inertial frames are ones in which measurements show the theory to take on its simplest form.

The basic laws of microphysics are initially formulated with respect to inertial observers, i.e., in inertial coordinate frames, but all actual observers are accelerating in some way. The term 'observer' is used in an extended sense to include any measuring device. In order to interpret experimental results then, we need to establish a *physical* connection between accelerated and inertial observers. The assumption that is supposed to achieve this in the standard theory of relativity is the hypothesis of locality:

> An accelerated observer (measuring device) along its worldline is at each instant *physically equivalent* to a hypothetical inertial observer (measuring device) that is otherwise identical and instantaneously comoving with the accelerated observer (measuring device).

Then Lorentz symmetry and the locality hypothesis together form the physical basis for the special theory of relativity in this view. He notes that Lorentz invariance is consistent with quantum theory, but that this is not the case with his locality hypothesis (LH). The aim of Mashhoon's nonlocal formulation of special relativity [37] (not discussed here) is to correct this situation. He then considers LH to be an extension of Lorentz symmetry, in some sense.

The inspiration is Newtonian point particle mechanics: the accelerated observer and the otherwise identical instantaneously comoving inertial observer have the same position and velocity, and hence share the same state, whatever that means. He claims that they are thus pointwise physically identical in classical mechanics. But this seems an odd suggestion since, even in Newtonian mechanics, accelerating observers have to imagine the so-called inertial forces in order to pretend that they are moving inertially! Such a motivation therefore looks rather weak.

In Mashhoon's picture, in SR, the accelerated observer is envisaged through LH as a continuous infinity of hypothetical momentarily comoving inertial observers, and interestingly enough, this is strongly reminiscent of the construction of semi-Euclidean (SE) coordinate systems. Mashhoon mentions that Lorentz first introduced such an assumption in his theory of electrons when he conceived of an electron as a small ball of charge that would always be exactly FitzGerald contracted along its direction of motion, an assumption still commonly made when carrying out self-force calculations (see Sect. 9.3 and [32]). This is indeed precisely the assumption that the electron is rigid, or undergoing rigid motion, in the sense described earlier. Lorentz viewed this as an approximation. Physically, it amounted to assuming that the electron velocity would change over a much longer time scale than the period of its internal oscillations.

But are we talking about subjects of physical investigation like electrons, or observers and the coordinate systems or other frames they might set up? There is certainly a similarity in the assumptions, but what remains crucial is what Mashhoon means by *physically equivalent* in his statement about accelerating observers in LH. Actually, this highlights the difference between accelerating observers and accelerating measuring devices in that statement. The measuring device does not need to set up coordinates and so is more like Lorentz's electron!

Mashhoon also mentions Einstein's understanding of rods and clocks, claiming that the hypothesis of locality underlies Einstein's development of the theory of relativity. He sees the locality assumption as fitting perfectly together with Einstein's local principle of equivalence to ensure that *every observer in a gravitational field is pointwise inertial*. On the other hand, only freely falling observers are usually considered to be inertial, so it is not obvious how to interpret such a claim.

He goes on to say that, in order to preserve the operational significance of Einstein's heuristic principle of equivalence, which he takes to be the presumed local equivalence of an observer in a gravitational field with an accelerated observer in Minkowski spacetime, whatever equivalence may mean here, it must be coupled with a statement regarding what accelerated observers actually measure. Taken from this angle, it looks like it is the measuring device that really matters in the statement of LH, rather than people with preferences about coordinate systems.

When coupled with the hypothesis of locality, then, Mashhoon considers that Einstein's principle of equivalence provides a physical basis for a field theory of gravitation that is consistent with (local) Lorentz invariance. But how do we formulate that notion of consistency? The idea is perhaps that the instantaneously comoving (locally) inertial, i.e., freely falling, measuring devices we invoke in LH are somehow consistent with (local) Lorentz invariance because physics 'looks like' flat spacetime physics to observers comoving with such devices. However, that looks rather like making a mountain out of a molehill, or taking LH to be a fundamental principle rather than a purely pragmatic one.

The earlier of the two papers [36] gets more quickly down to the nitty gritty of the approximations involved. The idea is that, if all physical phenomena were somehow reducible to pointlike events, then LH would be exactly valid. The problem thus arises, in this view, because things like EM waves involve intrinsic length and time scales, viz., their wavelength λ and period λ/c. To measure the frequency of a wave would require observation of several oscillations, but during this time, the observer or measuring device will have changed its Newtonian state, i.e., changed its velocity.

If the change in velocity occurring over a few periods of the wave can be disregarded, then LH is considered to be vindicated. This idea, or approximation, is quantified by defining an acceleration length $\mathscr{L}$ for the observer and saying that LH amounts to $\lambda/\mathscr{L} \ll 1$ for the case of the EM wave, for example. The observer will also have an acceleration time $\mathscr{L}/c$. For translational acceleration a, we could take $\mathscr{L} = c^2/a$, and for rotation at angular speed Ω, we could take $\mathscr{L} = c/\Omega$.

One can also consider the locality hypothesis as an adiabaticity assumption analogous to the one for sufficiently slow processes in thermodynamics. This would therefore be expected to be a good approximation only up to some acceleration, for a given measuring device and measured phenomenon. The whole issue here is one of estimating the validity of approximations in real physical measurement situations. In itself, it contains no profound features.

If we consider an accelerated measuring device in Minkowski spacetime, the internal dynamics of the device is then subject to inertial effects that consist of the inertial forces of classical mechanics together with electromagnetic and quantum effects. If the net influence of these inertial effects integrates over the relevant length and time scales of a measurement to perturbations that do not appreciably disturb the result of the measurement and can therefore be neglected, then LH is valid and the device can be considered acceptable for that particular measurement.

The response of measuring devices to acceleration, i.e., the influence of inertial effects on their operation, should eventually be determined on the basis of a proper theory of accelerated systems. This is precisely what is done in Bell's approach to the one-electron atom in flat spacetime, which is subject to Maxwell's equations, but can also be done in curved spacetimes using MEME, as we saw in Chap. 6. The added difficulty in GR is that one needs to specify how to set up coordinates, e.g., locally inertial coordinates, and this is a heavily theory-laden process. In SR, we think we know how to set up inertial coordinates, and SEP licenses the idea of doing exactly the same thing operationally speaking when freely falling in the curved spacetime. We expect this to set up locally inertial coordinates in the latter case.

As Mashhoon notes, in an Earth-based laboratory, the translational acceleration length would be $\mathscr{L}_{\text{trans}} = c^2/g \sim 1$ light-year, with g as the acceleration due to gravity at the Earth's surface, while the rotational acceleration length would be $\mathscr{L}_{\text{rot}} = c/\Omega \sim 28$ A.U., with Ω the angular frequency of rotation of the Earth about its axis. These acceleration effects are thus likely to be negligible for most measurement purposes.

Mashhoon also mentions the decay time of muons in storage rings as investigated by Eisele under the title *On the behaviour of an accelerated clock* [20], also cited and discussed by Brown in [7]. The hypothesis of locality implies that $\tau_\mu = \gamma\tau_\mu^0$, where τ_μ^0 is the lifetime of the muon when it is at rest, and γ is the Lorentz factor corresponding to the circular motion of muons in the storage ring. But one can avoid the locality hypothesis by using a model in which the muon decays from a high-energy Landau level in a constant magnetic field. The muon decay is then susceptible to quantum calculations which give

$$\tau_\mu \approx \gamma\tau_\mu^0 \left[1 + \frac{2}{3}\left(\frac{\lambda}{\mathscr{L}}\right)^2\right] , \tag{7.1}$$

where $\lambda = \hbar/mc$ is the Compton wavelength of the muon and $\mathscr{L}$ is its effective acceleration length, i.e., $\mathscr{L} = c^2/a$, where $a = \gamma^2 v^2/r \sim 10^{18} g$ is the effective centripetal acceleration of the muon in the storage ring. In practice, the nonlocal correction term turns out to be very small.

Brown says that, in all such experiments, the clock retardation as gauged by muon decay is calculated in conformity with the clock hypothesis, which is the instance of LH we are concerned with here, whence the effect is due to root mean square velocities or integration over instantaneous velocities of the clocks. The instantaneous accelerations themselves are assumed to contribute nothing to the effect. The point to note here, however, is that, for any given clock, no matter how ideal its behaviour when moving inertially, there will in principle be an acceleration such that, in order to achieve that acceleration, the external force acting on the clock will disrupt its inner workings sufficiently to make a significant difference.

Mashhoon does not mention the possibility, not encoded within present theories, that there may also be universal effects due to acceleration, just as there are universal effects due to velocity. The word 'universal' here means that we refer to effects that do not depend on the specific clock. Such effects would presumably signal some kind of acceleration symmetry that should be included in our field theories of matter.

In Brown's account, we have more on the Eisele calculation. The muon orbits are described in terms of a Landau level with high quantum number, and perturbation techniques in the theory of the weak interactions are used to estimate the muon lifetime theoretically. Even for these high accelerations $a \sim 10^{18} g$, the correction to the clock hypothesis estimate would only be of the order of 1 part in 10^{25}, far too small for the original experiment to detect. But note that there is nothing to prevent a more accurate experiment detecting this one day.

No clock is perfect, but with a better clock, we should always be able to measure the inaccuracies of the less perfect one. A better clock is one which, by definition, conforms better to the requirements of the clock hypothesis in the given circumstance, i.e., for the given worldline. This is because proper time is *defined* within the theory, but that definition is justified by the fact that it is relevant to the way real clocks behave. As Eisele says, the most interesting part of the calculation lies not in potential applications, but rather in the possibility of checking the clock hypothesis in this special case with the help of an accepted physical theory. There is one clock hypothesis for each system, and we can in principle check it.

Brown puts the issue particularly clearly. In contrast to time dilation induced by uniform motion, which is usually understood to be independent of the constitution of the clock because it is universal in the sense mentioned above, the effects of the acceleration *will* depend on the constitution of the clock. The reason why time dilation induced by uniform motion appears to be independent of the clock constitution boils down to the fact that all the field theories of matter governing the functioning of the clock are Lorentz symmetric. Of course, we expect acceleration effects that depend on the clock constitution, and there may in principle be other acceleration effects that are universal because they arise from an as yet unknown acceleration symmetry of the field theories of matter.

So Eisele's calculations tell us why the muon can be used as a microscopic way-wiser for timelike curves in relativistic spacetime, even when the curves are non-geodesic and involve enormous accelerations of the order of $10^{18} g$. It is simply because the process they instantiate nevertheless runs at the rate decreed by the proper time of the curve.

Mashhoon goes on to discuss Lorentz's models of the electron as spatially extended charge distributions with the usual rigidity assumption, which amounts to another instance of LH, indeed, what is usually called the ruler hypothesis. He also mentions Einstein's remark that one *must* assume that the behaviour of rods and clocks depends only upon velocities and not upon accelerations, which as we have seen is not strictly true. This is in fact just a reasonable assumption in many cases. The real issue here is whether specific instances of rods and clocks are up to the task in the given situation, i.e., for the given worldline.

Oddly, Mashhoon also seems to miss this point. He says that the modern experimental foundation of Einstein's theory of gravitation requires an extension of this assumption to all standard measuring devices, whence LH supersedes the clock and ruler hypotheses. But it would be better to say that, in each instance, i.e., for each measuring device and each worldline, we must apply our fundamental theories to estimate what these devices will actually register. We can then compare with what would be expected if LH were valid and determine whether the accuracy of the device requires us to take the acceleration or gravitational field into account.

So LH is not a fundamental addition to our theoretical paraphernalia, as Mashhoon seems to imply at times. It is merely an approximation that we can in principle always justify, or reject. He notes for example that LH rests upon the possibility of defining instantaneous inertial rest frames along the worldline of an arbitrary point particle, and that Minkowski raised this possibility, and hence LH likewise, to the level of a fundamental axiom. The present view is that this confuses theory and practice. Theoretically, we can always construct instantaneously comoving inertial frames in the usual manifold context. Making this an axiom in some other approach to theoretical construction of the spacetime changes nothing in the usual theoretical picture.

The real question here is whether we can construct something in the real world that corresponds to what we construct, or lay down axiomatically in theory. That is another matter. The fact that one can construct the theory by laying such a condition down as an axiom does not help us in any way. In practice, we have to assume at some point that what we are actually doing out there with clocks and rulers can be taken to construct a real physical counterpart to the theoretical structure. Theory can tell us what is likely to work as a good clock or ruler in the given circumstance. Then every experiment serves to check the overall consistency of the relevant theoretical considerations and practical constructions. When disagreements are found, we first question the practical construction, and if there is no reason for doubt there, we finally obtain criticism of the theory itself.

7.1 Mashhoon on Length Measurement by Rotating Observers

This section is based on the discussion in [36]. Observers A and B move round a circle of radius r about the origin in the (x,y) plane of an inertial frame L in flat spacetime. At $t=0$, A is at azimuthal angle $\varphi_A = 0$ and B is at $\varphi_B = \Delta$. For $t>0$, they have motions specified by the same angular frequency $\hat{\Omega}_0(t) > 0$, so that their positions are given by

$$\varphi_A(t) = \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' \,, \qquad \varphi_B(t) = \Delta + \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' \,. \tag{7.2}$$

For an observer in L, the angular separation $\Delta\varphi$ of A and B is constant at any time $t > 0$, viz.,

$$\Delta\varphi = \varphi_B(t) - \varphi_A(t) = \Delta \ .$$

The spatial separation along the arc is

$$l(t) = r\Delta\varphi = r\Delta \ .$$

We now consider a set $\mathscr{O}$ of observers O covering all points on the arc from A to B, and moving exactly as A and B, with angular frequency $\hat{\Omega}_0(t)$. For $t > 0$, the observer in L considers all members of $\mathscr{O}$ to be at rest in a rotating coordinate system (x', y', z') obtained from (x, y, z) by rotation about the z axis with frequency $\hat{\Omega}_0(t)$.

We now ask what length the arc from A to B should have as assessed by the observers in $\mathscr{O}$. The locality hypothesis suggests (or rather hopes) that, at any time $t > 0$ in L, each observer O is instantaneously equivalent to a comoving inertial observer O', in the sense that both O and O' would register exactly the same values for all physical quantities they try to measure by applying exactly the same measuring device and method to the task. This is what is meant by the term *physically equivalent* in Mashhoon's general statement of LH on p. 264.

In this case, the hypothesis hopes that O will register the same infinitesimal arc length as O', and we know that to be $\delta l'(t) = \hat{\gamma}\delta l(t)$, where $\hat{\gamma}(t)$ is the Lorentz factor corresponding to the speed $\hat{v}(t) = r\hat{\Omega}_0(t)$, and $\delta l(t)$ is the arc length in question as gauged in L. If we now define

$$\boxed{l' := \sum \delta l' = \hat{\gamma}(t) \sum \delta l = \hat{\gamma}(t) l} \ , \tag{7.3}$$

this could be called the arc length as measured by the observers $\mathscr{O}$. However, each $\delta l'$ is an infinitesimal length at rest in a different inertial frame.

Mashhoon claims that the same result is obtained if length is measured using light travel time over infinitesimal distances between observers in $\mathscr{O}$. We know theoretically from Chap. 5 that measurements of infinitesimal proper distances using light travel times should deliver the answer $\delta l'(t) = \hat{\gamma}\delta l(t)$, at least to a certain level of accuracy. The problem with (7.3), if it is a problem, is the sum, because we add up quantities measured in very different inertial frames.

Mashhoon says that (7.3) is not a proper geometric definition of length, and even wonders whether it is physically reasonable. But would the situation be improved, as he suggests, by somehow combining the infinite number of disjoint inertial frames into one continuous accelerated frame of reference? The word 'combining', although not appropriate here, immediately makes us think of the SE frames. Mashhoon's proposal, the most natural in his view, is to choose one of the noninertial observers on the arc, say A, and set up a geodesic coordinate system along its worldline.

He claims that the measure of interval along the worldline (proper time) and away from it (proper length) would be determined by LH. But does this not reveal a misunderstanding of his own idea? Surely LH does not determine anything, only proposes that such and such a measuring device will to a sufficiently good approximation (better than measurement accuracy) measure the same thing as an instantaneously comoving inertial measuring device of the same kind, whence theory tells us what it will measure, at least to within measurement accuracy.

The misunderstanding seems to be confirmed by his statement:

> [...] at any instant of proper time the rules of Euclidean geometry are applicable, as the accelerated observer is instantaneously inertial.

But it would be hard to get a clear understanding by putting things in this way. An accelerated observer is *never* inertial, even instantaneously. That is the whole problem. LH just seems to embody the idea that we can sometimes pretend it is, to within measurement accuracy. So for the moment, let us not make further claims about the role of LH in understanding the continuous accelerated frame of reference. Let us just follow Mashhoon's account of how to construct such a frame both theoretically and pragmatically, and assess its role for ourselves.

The aim then is to find the proper length of the arc from A to B in what Mashhoon calls a geodesic coordinate system adapted to the worldline of A. Of course, this will deliver a different answer to (7.3). But so what? Where is the surprise? And why use this coordinate system? If we want to measure the result in (7.3) or the result we are about to calculate, we will of course have to do something different in each case. How will that throw light on LH? Let us establish how this new definition of the arc length might be measured physically and see how it elucidates this issue.

7.2 Calculating a Length in the Rotating Tetrad Coordinate System

Here we shall actually investigate two ways of calculating the arc length occupied by the continuous set $\mathscr{O}$ of observers O lying between the two observers A and B, as described in Sect. 7.1. We refer here to the formalism set up in Sect. 2.11. The two approaches are as follows:

- The most natural way to do this is to intersect the cylinder of worldlines by a HOS of the rotating tetrad coordinate system associated with A, i.e., by a hyperplane of constant T, and then integrate $\mathrm{d}L := (\mathrm{d}X^2 + \mathrm{d}Y^2)^{1/2}$ along the appropriate interval corresponding to the arc from A to B. This is straightforward, apart from establishing the endpoint corresponding to the last observer B.
- Mashhoon integrates $\mathrm{d}L := (\mathrm{d}X^2 + \mathrm{d}Y^2)^{1/2}$ along a different curve in spacetime, namely, the intersection of the cylinder of worldlines with a HOS of the inertial coordinate system, i.e., a hyperplane of constant t. This approach is less natural, because it mixes views from two different frames, but it does evaluate a length for *the same curve* in spacetime as (7.3). It is also more difficult, because T varies from one observer to the next, and one must identify exactly how it varies.

Note that $\mathrm{d}L$ is worked out as though the (X,Y) plane had Euclidean geometry, and it does in a hyperplane of constant T [see the metric (2.340) on p. 105], a point in favour of the first calculation.

7.2.1 Calculating the Length in a Constant T Hyperplane

The worldline of A is given in inertial coordinates by

$$x_A(t) = \big(t, r\cos\varphi_A(t), r\sin\varphi_A(t), 0\big) , \tag{7.4}$$

as t varies, where

$$\varphi_A(t) = \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' . \tag{7.5}$$

The worldline of B is given in inertial coordinates by

$$x_B(t) = \big(t, r\cos\varphi_B(t), r\sin\varphi_B(t), 0\big) , \tag{7.6}$$

as t varies, where

$$\varphi_B(t) = \Delta + \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' , \tag{7.7}$$

for a constant Δ which effectively specifies B. And finally, the worldline of O is given in inertial coordinates by

$$x(t) = \big(t, r\cos\varphi(t), r\sin\varphi(t), 0\big) , \tag{7.8}$$

as t varies, where

$$\varphi(t) = \delta + \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' , \tag{7.9}$$

for a constant δ which specifies O.

If $\Delta = 2\pi$, all these observer worldlines form a cylinder in spacetime as viewed in the inertial frame (if we drop the z dimension), specified by

$$x^2 + y^2 = r^2 . \tag{7.10}$$

Since from (2.331) on p. 104, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} X + r \\ \gamma Y \end{pmatrix} , \tag{7.11}$$

it follows that

$$\begin{pmatrix} X + r \\ \gamma Y \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} . \tag{7.12}$$

It then follows immediately that

$$(X + r)^2 + \gamma^2 Y^2 = x^2 + y^2 , \tag{7.13}$$

so the above cylinder (7.10) is expressed by

$$\frac{(X + r)^2}{r^2} + \frac{Y^2}{(r\gamma^{-1})^2} = 1 . \tag{7.14}$$

Dropping the Z dimension, this is an elliptical cylinder in spacetime, not surprisingly. The observer A at the space origin of the (T,X,Y,Z) coordinate system is instantaneously borrowing the HOS of an ICIO, and we expect this to cut the circular cylinder in an ellipse.

The direction of motion of A is the Y direction, tangential to the circle in the (x,y) plane, as can be seen from (2.330) on p. 103 [see in particular $\lambda_{(1)}^{\mu}$], so the ellipse is squashed by a factor of $\gamma^{-1} < 1$ in the Y direction, where γ corresponds to the instantaneous speed of A. The general formula for an ellipse centered at the point $(X,Y) = (-r,0)$ is

$$\frac{(X+r)^2}{a^2} + \frac{Y^2}{b^2} = 1 \ . \tag{7.15}$$

This has eccentricity

$$e := \frac{\sqrt{a^2 - b^2}}{a} \ . \tag{7.16}$$

In the present case, $a = r$ and $b = r\gamma^{-1}$, so it turns out that the eccentricity here is $e = v$. The semi-major axis is r, since the ellipse passes through the space origin of the (X,Y) plane (where A sits), and the semi-minor axis is $r\gamma^{-1}$.

So we have a picture of the cylinder of all the observer worldlines as it would be recorded by A relative to the (T,X,Y,Z) coordinates in constant T snapshots, ignoring the Z dimension. In such a constant T hyperplane, it is convenient to replace X and Y by another parameter, in fact, a single parameter θ specified by

$$X + r = r\cos\theta \ , \quad Y = r\gamma^{-1}\sin\theta \ , \tag{7.17}$$

which runs from 0 to 2π as one moves around the ellipse from A to A. In a constant T hyperplane, and systematically ignoring Z in the following, the event on the elliptical cylinder specified by θ has inertial coordinates

$$t = F(T) + \gamma(T)v(T)Y = F(T) + rv(T)\sin\theta \ , \tag{7.18}$$

$$\begin{aligned} x &= (X+r)\cos\phi(T) - \gamma(T)Y\sin\phi(T) \\ &= r\cos\theta\cos\phi(T) - r\sin\theta\sin\phi(T) \\ &= r\cos\big[\theta + \phi(T)\big] \ , \end{aligned} \tag{7.19}$$

and

$$\begin{aligned} t &= (X+r)\sin\phi(T) + \gamma(T)Y\cos\phi(T) \\ &= r\cos\theta\sin\phi(T) - r\sin\theta\cos\phi(T) \\ &= r\sin\big[\theta + \phi(T)\big] \ . \end{aligned} \tag{7.20}$$

We can now say which observer $O \in \mathscr{O}$ happens to be at this event, since we have (7.8), and comparison dictates that

$$\varphi(t) = \theta + \phi(T) , \tag{7.21}$$

recalling that $\varphi(t) = \delta + \varphi_A(t)$ specifies an observer through the value of δ. The relation (7.18) is important here because it tells us the value of t we should be using to understand (7.21).

At the risk of becoming tedious, it is worth formulating this a little more closely since there are so many parameters around. The aim is to integrate the quantity $\mathrm{d}L := \sqrt{\mathrm{d}X^2 + \mathrm{d}Y^2}$ around part of the above ellipse in a hyperplane of fixed $T = T_A$ chosen by A at the outset. The relevant part of the ellipse is specified by some value Θ of θ which specifies the event where the worldline of observer B intersects the hyperplane $T = T_A$. This value of θ will generally depend on the chosen value T_A of T.

Now from (7.17) specifying X and Y in terms of θ, we have

$$\mathrm{d}X = -r\sin\theta\mathrm{d}\theta , \quad \mathrm{d}Y = r\gamma^{-1}\cos\theta\mathrm{d}\theta , \tag{7.22}$$

so

$$\begin{aligned}\mathrm{d}L &= r(\sin^2\theta + \gamma^{-2}\cos^2\theta)^{1/2}\mathrm{d}\theta \\ &= r\big[1 - \cos^2\theta(1-\gamma^{-2})\big]^{1/2}\mathrm{d}\theta \\ &= r(1 - v^2\cos^2\theta)^{1/2}\mathrm{d}\theta .\end{aligned} \tag{7.23}$$

The aim is therefore to find $\Theta(T_A)$ and calculate

$$L = r\int_0^{\Theta}(1 - v^2\cos^2\theta)^{1/2}\mathrm{d}\theta . \tag{7.24}$$

Note immediately that v in the integrand does not depend on θ, but is just the speed $v(T_A)$ of A at the chosen time T_A (which is a proper time of A, in fact). This means that the integral itself is rather easily found from tables of elliptic integrals.

An incomplete elliptic integral of the second kind is

$$E(k,\alpha) := \int_0^{\alpha}\sqrt{1 - k^2\sin^2\theta}\,\mathrm{d}\theta , \qquad 0 < k < 1 . \tag{7.25}$$

This is also called Legendre's form for the elliptic integral of the second kind [51]. In the integral for L, the constant v satisfies $0 < v < 1$, but we have a cosine rather than a sine function. However,

$$\cos\theta = \sin(\pi/2 - \theta) ,$$

so

$$\int_0^{\Theta} \sqrt{1-v^2\cos^2\theta}\,\mathrm{d}\theta = \int_{\pi/2-\Theta}^{\pi/2} \sqrt{1-v^2\sin^2\theta}\,\mathrm{d}\theta$$
$$= \int_0^{\pi/2} \sqrt{1-v^2\sin^2\theta}\,\mathrm{d}\theta - \int_0^{\pi/2-\Theta} \sqrt{1-v^2\sin^2\theta}\,\mathrm{d}\theta\ .$$

Then we have

$$\boxed{L = r\Big[E(v,\pi/2) - E(v,\pi/2-\Theta)\Big]}\ . \tag{7.26}$$

We still need to determine $\Theta(T_A)$ specifying the position of B on the ellipse formed by intersecting the HOS of A at its proper time T_A with the cylinder of worldlines described earlier. What equation do we need to solve to obtain $\Theta(T_A)$?

Given T_A and Δ specifying B via its angular position

$$\varphi_B(t) = \Delta + \int_0^t \hat{\Omega}_0(t')\mathrm{d}t' = \Delta + \varphi_A(t)\ , \tag{7.27}$$

we have to find where the hyperplane H_{T_A} of constant $T = T_A$ cuts the worldline W_B of B. The latter is given in inertial coordinates by

$$x_B(t) = \big(t, r\cos\varphi_B(t), r\sin\varphi_B(t), 0\big)$$
$$= \Big(t, r\cos\big[\Delta + \varphi_A(t)\big], r\sin\big[\Delta + \varphi_A(t)\big], 0\Big)\ .$$

The former is given by

$$H_{T_A} = \Bigg\{\Big(F(T_A) + \gamma(T_A)v(T_A)Y, (X+r)\cos\phi(T_A) - \gamma(T_A)Y\sin\phi(T_A),$$
$$(X+r)\sin\phi(T_A) + \gamma(T_A)Y\cos\phi(T_A), Z\Big) : X,\,Y,\,Z \text{ alone vary}\Bigg\}\ .$$

The ellipse lies on this plane and we know there will be a value of θ depending on Δ such that the intersection of W_B and H_{T_A} occurs for that θ. We see here why this value Θ of θ should be expected to depend on T_A.

As noted earlier [see (7.18), but we now have a more specific notation], the inertial time $t(T_A)$ for the intersection we seek is given by

$$t(T_A) = F(T_A) + \gamma(T_A)v(T_A)Y = F(T_A) + rv(T_A)\sin\Theta(T_A)\ . \tag{7.28}$$

We also have the value of the inertial coordinate x at the intersection, viz.,

$$r\cos\varphi_B\big(t(T_A)\big) = (X+r)\cos\phi(T_A) - \gamma(T_A)Y\sin\phi(T_A)$$
$$= r\cos\Big[\Theta(T_A) + \phi(T_A)\Big]\ ,$$

mirroring (7.19), and the value of the inertial coordinate y at the intersection, viz.,

$$
\begin{aligned}
r\sin\varphi_B\big(t(T_A)\big) &= (X+r)\sin\phi(T_A) + \gamma(T_A)Y\cos\phi(T_A) \\
&= r\sin\Big[\Theta(T_A)+\phi(T_A)\Big] ,
\end{aligned}
$$

so we only need one other relation to describe the event in spacetime that concerns observer B for the purposes of this calculation, viz.,

$$
\varphi_B\big(t(T_A)\big) = \Theta(T_A) + \phi(T_A) . \tag{7.29}
$$

Feeding in (7.28), we can obtain $\Theta(T_A)$ by solving

$$
\varphi_B\Big(F(T_A) + rv(T_A)\sin\Theta(T_A)\Big) = \Theta(T_A) + \phi(T_A) . \tag{7.30}
$$

Note that, by (7.27), this can be written

$$
\varphi_A\Big(F(T_A) + rv(T_A)\sin\Theta(T_A)\Big) + \Delta = \Theta(T_A) + \phi(T_A) . \tag{7.31}
$$

Recall also from the definition of ϕ just prior to (2.272) on p. 90, that $\phi(T_A)$ is equal to $\varphi_A(t_{\text{something}})$ for some value $t_{\text{something}}$ of the inertial time corresponding to the proper time T_A of A. However, the two occurrences of the function φ_A appearing in (7.31) are evaluated at *different* values of the inertial time! Otherwise we might have made the hasty deduction that $\Delta = \Theta(T_A)$ just after (7.21) on p. 273.

To be precise, the time $t(T_A) = F(T_A) + rv(T_A)\sin\Theta(T_A)$ at which φ_A is evaluated on the left-hand side is the inertial time when B is considered by the ICIO for A at its proper time T_A to be simultaneous with A, whereas the time $t_{\text{something}}$ is found from (7.18) on p. 272 with $T = T_A$ and $Y = 0$, whence

$$
t_{\text{something}} = F(T_A) \neq F(T_A) + rv(T_A)\sin\Theta(T_A) . \tag{7.32}
$$

Equation (7.31) could thus be written

$$
\boxed{\varphi_A\Big(F(T_A) + rv(T_A)\sin\Theta(T_A)\Big) + \Delta = \Theta(T_A) + \varphi_A\big(F(T_A)\big)} . \tag{7.33}
$$

Note that the relations (7.31) and (7.33) generalise immediately to any of the intermediately placed observers $O \in \mathscr{O}$ specified by $\delta \in [0,\Delta]$. We would then obtain a generally different value of $\Theta(T_A)$ for the given value of T_A, so we ought to use a symbol like $\theta(\delta, T_A)$. The general relations are thus

$$
\varphi_A\Big(F(T_A) + rv(T_A)\sin\Theta(T_A)\Big) + \delta = \theta(\delta, T_A) + \phi(T_A) \tag{7.34}
$$

or

$$
\boxed{\varphi_A\Big(F(T_A) + rv(T_A)\sin\Theta(T_A)\Big) + \delta = \theta(\delta, T_A) + \varphi_A\big(F(T_A)\big)} . \tag{7.35}
$$

For the present purposes then, we use (7.26) together with (7.33) to calculate the length between observers A and B in the most natural way if A records data using the coordinate system (T,X,Y,Z). It would be miraculous if that delivered the value for the length obtained in (7.3) on p. 269.

But so what? Even inertial observers with different motions get different lengths. The difference when an observer has acceleration is that there is no single agreed definition for a length, even given the motion of the observer, and this ultimately is because our field theories of matter have no acceleration symmetry to match their velocity (Lorentz) symmetry.

There will not even be a natural measured length, since different operational definitions will deliver different values, and this too is ultimately because our field theories of matter have no acceleration symmetry to match their velocity (Lorentz) symmetry.

7.2.2 Calculating the Length in a Constant t Hyperplane

Let us now calculate $\mathrm{d}L := \sqrt{\mathrm{d}X^2 + \mathrm{d}Y^2}$ on a circular arc of $x^2 + y^2 = r^2$ in a constant t slice, even though this is still somewhat unnatural, as pointed out previously, since it adds up lengths calculated in one frame but at different times in that frame, whereas in the first approach, the one leading to (7.3), we added up lengths calculated in different frames, and in the approach described in the last section we added up lengths calculated in one frame and at one time as specified by that frame. The calculation in this section is the one advocated by Mashhoon.

We can use the relation (7.18) on p. 272, viz.,

$$t = F(T) + \gamma(T)v(T)Y = F(T) + rv(T)\sin\theta\ , \tag{7.36}$$

which gives the inertial time of the event on the cylinder of worldlines specified by θ and T. The point is that T will vary now as we change θ, whereas we previously held T constant. Differentiating (7.36), we have

$$0 = \frac{\mathrm{d}t}{\mathrm{d}T} = \frac{\mathrm{d}F(T)}{\mathrm{d}T} + r\dot{v}(T)\sin\theta(T) + rv(T)\cos\theta(T)\frac{\mathrm{d}\theta}{\mathrm{d}T}\ ,$$

and we know from (2.271) on p. 90 that

$$\frac{\mathrm{d}F(T)}{\mathrm{d}T} = \gamma(T)\ .$$

Hence,

$$\boxed{\frac{\mathrm{d}T}{\mathrm{d}\theta} = -\frac{rv\cos\theta}{\gamma + r\dot{v}\sin\theta}}\ . \tag{7.37}$$

Now $X = r\cos\theta - r$, so

$$\frac{\mathrm{d}X}{\mathrm{d}\theta} = -r\sin\theta\ ,$$

and $Y = r\sqrt{1-v^2}\sin\theta$, with v dependent on T and hence on θ, so

$$\frac{\mathrm{d}Y}{\mathrm{d}\theta} = -rv\dot{v}(1-v^2)^{-1/2}\sin\theta\frac{\mathrm{d}T}{\mathrm{d}\theta} + r\gamma^{-1}\cos\theta\ .$$

Hence,

$$\begin{aligned}\frac{1}{r^2}\left[\left(\frac{\mathrm{d}X}{\mathrm{d}\theta}\right)^2 + \left(\frac{\mathrm{d}Y}{\mathrm{d}\theta}\right)^2\right] &= \sin^2\theta + \left(\frac{\gamma r v^2\dot{v}\sin\theta\cos\theta}{\gamma + r\dot{v}\sin\theta} + \gamma^{-1}\cos\theta\right)^2\\ &= 1-\cos^2\theta\left[1-\left(\frac{\gamma r v^2\dot{v}\sin\theta}{\gamma + r\dot{v}\sin\theta} + \frac{1}{\gamma}\right)^2\right]\ .\end{aligned}$$

In square brackets, we have

$$\begin{aligned}1-\frac{1}{\gamma^2}-\frac{2rv^2\dot{v}\sin\theta}{\gamma + r\dot{v}\sin\theta}-\frac{\gamma^2 r^2 v^4\dot{v}^2\sin^2\theta}{(\gamma + r\dot{v}\sin\theta)^2} &= v^2\left[1-\frac{2r\dot{v}\sin\theta}{\gamma + r\dot{v}\sin\theta}-\frac{\gamma^2 r^2 v^2\dot{v}^2\sin^2\theta}{(\gamma + r\dot{v}\sin\theta)^2}\right]\\ &= v^2\gamma^2\frac{1-r^2\dot{v}^2\sin^2\theta}{(\gamma + r\dot{v}\sin\theta)^2}\ ,\end{aligned}$$

after a short calculation.

So finally,

$$\mathrm{d}L = \sqrt{\left(\frac{\mathrm{d}X}{\mathrm{d}\theta}\right)^2 + \left(\frac{\mathrm{d}Y}{\mathrm{d}\theta}\right)^2} = r\sqrt{1-v^2W\cos^2\theta}\ , \tag{7.38}$$

where

$$W := \gamma^2\frac{1-r^2\dot{v}^2\sin^2\theta}{(\gamma + r\dot{v}\sin\theta)^2}\ , \tag{7.39}$$

and

$$\boxed{L = r\int_0^{\Theta}\sqrt{1-v^2W\cos^2\theta}\,\mathrm{d}\theta}\ . \tag{7.40}$$

The upper limit of the integral is somehow determined by the worldline of B. Indeed, it specifies the event at the intersection of the worldline of B with the hyperplane of constant t.

Now for the given value of t, observer B is located at

$$x_B(t) = \big(t, r\cos\varphi_B(t), r\sin\varphi_B(t), 0\big)\ , \tag{7.41}$$

where $\varphi_B(t) = \Delta + \varphi_A(t)$. This event has coordinates (T_B, X_B, Y_B, Z_B), where $Z_B = 0$ and

$$\boxed{t = F(T_B) + \gamma(T_B)V(T_B)Y_B = F(T_B) + rv(T_B)\sin\theta_B}\ , \tag{7.42}$$

with θ_B such that

$$X_B + r = r\cos\theta_B\ , \quad Y_B = r\gamma^{-1}(T_B)\sin\theta_B\ . \tag{7.43}$$

By (7.19) and (7.20) on p. 272, this leads to

$$r\cos\varphi_B(t) = r\cos\left[\theta_B + \phi(T_B)\right]\ , \quad r\sin\varphi_B(t) = r\sin\left[\theta_B + \phi(T_B)\right]\ ,$$

which imply

$$\theta_B + \phi(T_B) = \varphi_B(t)\ , \tag{7.44}$$

or

$$\boxed{\theta_B + \phi(T_B) = \Delta + \varphi_A(t)}\ . \tag{7.45}$$

The key relations for determining the two unknowns $\Theta := \theta_B$ and T_B given t are (7.42) and (7.45). Note that $\phi(T_B) = \varphi_A(t_{\text{something}})$ by the definition just prior to (2.272) on p. 90, where $t_{\text{something}}$ is the inertial time corresponding to a proper time T_B at A, i.e., $t_{\text{something}} = F(T_B)$, since A is at $Y = 0$. So the final equations to be solved to find T_B and eventually $\Theta := \theta_B$ are

$$\boxed{\begin{cases} t = F(T_B) + rv(T_B)\sin\theta_B\ , \\ \Delta + \varphi_A(t) = \theta_B + \varphi_A\big(F(T_B)\big)\ . \end{cases}} \tag{7.46}$$

Naturally, this would be impossible to do in all generality, and difficult even in many specific cases.

Mashhoon proposes to solve the problem in the case where all observers move at constant angular speed, so that $\dot{v} = 0$. Then his function W in (7.39) is equal to unity. We return to the case of an elliptic integral for L in (7.40) and the upper limit $\Theta := \theta_B$ is found as follows. We have $F(T_B) = \gamma T_B$, where γ is constant, so

$$t = \gamma T_B + rv\sin\theta_B\ .$$

Furthermore, $\varphi_A(t) = vt/r$ and

$$\varphi_A\big(F(T_B)\big) = \frac{v}{r}F(T_B) = \frac{v}{r}\gamma T_B\ .$$

Hence,

$$\Delta + \frac{v}{r}t = \theta_B + \frac{v}{r}\gamma T_B\ .$$

Eliminating T_B, we obtain

$$\Delta + \frac{v}{r}t = \theta_B + \frac{v}{r}(t - rv\sin\theta_B)\ ,$$

and finally,

$$\Delta = \theta_B - v^2\sin\theta_B\ . \tag{7.47}$$

Mashhoon finds once again that L is not generally equal to the length obtained in (7.3) on p. 269. Once again, where is the surprise?

He comments as follows. The proper acceleration length of the uniformly rotating observer A is given by $\mathscr{L} = 1/\gamma\Omega_0$, since c is taken as unity here (it would normally be c/Ω_0, according to Mashhoon's earlier remarks). How does the γ factor come in? This factor ranges from unity for $v = 0$ to infinity as $v \to c$, so it makes $\mathscr{L}$ smaller, and that is important, because $\mathscr{L}$ is supposed to be a large number. This problem aside, Mashhoon has proven that $L \neq l'$ in (7.3) on p. 269, but that, irrespective of the magnitude of Δ,

$$\frac{L}{l'} \longrightarrow 1 \quad \text{as} \quad \frac{r}{\mathscr{L}} = v\gamma \longrightarrow 0\,.$$

So provided that r is much smaller than $\mathscr{L}$, which amounts to requiring v to be small of course, the two ways of defining the length between A and B, although known theoretically to give different answers, will give very similar answers.

He also notes that, for $\Delta \to 0$, we have $L/l' \to 1$ regardless of the value of v. So the two ways of defining the length between A and B tend to give very similar answers if the two observers are very close. Mashhoon remarks that 'consistency can be achieved' only if the length under consideration is negligibly small compared to the acceleration length of the observer. The implication is that, in other cases, the acceleration might be too great and thereby mess up the agreement, depending on one's measurement accuracy.

7.3 Conclusion

There is is a final discussion section to the paper [36]. He reiterates that the locality hypothesis is an essential element of the theories of special and general relativity. This is supposed to be true because all real observers must be accelerating to some extent, and presumably can be assumed not to know this in many situations. They may thus be going through the motions of some standard length measurement procedure as though they had set up inertial coordinates comoving with them, and incur errors through this ignorance. The locality hypothesis then steps in to rescue them, or not in some cases.

This seems a dubious way of viewing the situation. After all, the *theories* of special and general relativity do not require us to start talking about how we relate them to the real world, even though they would be of no use to us without doing so. The problem with length is one of making operational definitions and then seeing what the theories have to say about the values we should expect to get. For circular motion the theories deliver values like L and l', with different values for L depending on the theoretical definition, as we have discussed. One must then decide whether one's operational definition corresponds best to this or that theoretical calculation. Surely that is the end of the story?

Mashhoon claims that relativistic measurement theory must take the basic assumption he refers to as the locality hypothesis into account along with its limitations. But when one does this, it does not have the status of an assumption at all! It is just an analysis of a typical situation at the interface between theory and reality, in which one asks whether it might not be possible to pretend that accelerating coordinates are inertial coordinates. But since there is no obligation to treat accelerating coordinates in this way, there would be no real problem in situations where that turned out to introduce inaccuracies, especially since our theories are apparently capable of dealing with such situations.

As far as time measurement is concerned, there is no doubt that we can make realistic estimates of how acceleration is likely to affect time measurements when defined operationally in some way or other for real situations encountered today in experimental work, e.g., on board Earth-orbiting satellites. But the situation for distance measurements is rather different precisely because there is a significant lack of uniqueness in the operational definitions of lengths, as attested by the calculations in this section. On the other hand, we are not always just intent on measuring proper times along worldlines, but sometimes also times we would like to attribute to remote events according to some operational procedure for setting up coordinates. When we measure proper times along worldlines, the only question is: what kind of clock are we using, and how will its working be affected by the acceleration? In the case of length measurements, we are always forced to choose some procedure as well as the measuring instrument.

Mashhoon's view is that 'this kind of problem can be resolved' for distance measurements, by which he means that one can sometimes nevertheless ignore the effects of acceleration in the sense of pretending that one's coordinate system is an inertial system. This works, for example, when the distance to be measured is much smaller than the relevant acceleration length of the observer. On the other hand, when that condition is not fulfilled, there seems to be no problem anyway, since the theory tells us how to deal with such situations.

Interestingly, he concedes that 'from a basic standpoint', which presumably means 'on the most fundamental level', the significance of non-inertial frames is rather limited. This is quite obvious. They are just coordinate systems, and we know very well that physical significance should not depend on choice of coordinates. This is the whole issue of covariance. On the other hand, any specific measurement will depend on how it is made, and we do always require some coordinate system or another, and some kind of operational definitions for measurements. What is new about this?

It seems then that there is nothing deep here, but just a purely practical problem such as: can we treat our operationally defined length and time measurements in an Earth-based laboratory as though the laboratory were moving inertially, even though it is rotating about the Earth's axis of rotation? Mashhoon has estimated that there could be an error of about 10^{-2} cm in a measurement of the Earth's equatorial circumference, due to this effect. Here is a real physical application of these considerations. But the very fact that we can estimate the error means that we *do not require* the locality hypothesis!

Chapter 8
Acceleration, Self-Force, and Inertia

The aim here will be to revive an old idea that has been around now for over 100 years, since before the advent of relativity theory. The subjects will be:

- inertial mass,
- Newton's second law $\mathbf{F} = m\mathbf{a}$,
- the dichotomy between spatially extended and point particles,

and

- the idea of classical mass renormalisation as introduced by Dirac in 1938.

All the claims in this chapter are justified in detail in [32].

8.1 Electromagnetic Mass and Newton's Laws

We begin with a simplistic model of the electron as a uniform spherical shell of charge of radius a. When the charge shell is sitting still in an inertial frame, we can calculate the energy in its Coulomb fields, and we obtain the value

$$\text{Energy in EM fields} = \frac{e^2}{2a} ,$$

where e is the charge distributed over the shell. Today we would automatically associate a mass with this energy by dividing by c^2:

$$\text{Mass associated with EM field energy} = \frac{e^2}{2ac^2} . \tag{8.1}$$

We can call this the energy-derived mass of the fields.

Now there is an obvious problem here if we let a tend to zero, since both this energy and the associated mass tend to infinity. And this raises the question as to whether it is ever justified to take the point particle limit for a charged particle.

We can also calculate the three-momentum of the EM fields produced by the charge shell. When it is sitting still in an inertial frame, there is no momentum in the fields, but when it is moving with constant velocity **v**, we find

$$\mathbf{p} = \frac{4}{3}\frac{e^2}{2ac^2}\mathbf{v} .$$

This is the momentum of a particle of mass

$$\text{Momentum-derived mass of EM fields} = \frac{4}{3}\frac{e^2}{2ac^2} . \tag{8.2}$$

Note that the momentum-derived mass of the fields of the charged shell is equal to 4/3 of the energy-derived mass, a point we shall return to later.

If the charge shell is moving very fast, but still at constant velocity, in an inertial frame, we can do a relativistic calculation of the three-momentum in the EM fields and we obtain exactly the same thing as before except that now there is a relativistic factor $\gamma(v)$, which is an increasing function of the speed v:

$$\mathbf{p} = \frac{4}{3}\frac{e^2}{2ac^2}\gamma(v)\mathbf{v} , \quad \gamma(v) := \frac{1}{\sqrt{1-v^2}} .$$

This is the momentum of a particle of mass

$$\text{Momentum-derived mass of EM fields} = \frac{4}{3}\frac{e^2}{2ac^2}\gamma(v) .$$

So what we discover here is that the momentum-derived mass of the EM fields due to this charged shell increases with speed v in exactly the way one would expect the inertial mass of a particle to increase with speed in relativistic dynamics (see Sect. 9.3 for further discussion of this kind of explanation).

Note that we have to assume that the charge shell FitzGerald contracts in the direction of motion and this reminds us that there must be some binding forces in the system. Indeed the various elements of negative charge distributed over the spherical surface will tend to repel one another and some other force will be needed to hold these elements in place. The actual shape of the charge distribution will depend on the balance between the repulsive EM forces between the charge elements and these binding forces.

So far we have talked about the energy and momentum in the fields of a charged shell and the associated masses, but how do we actually predict the resistance that something will show to being accelerated, as quantified by its inertial mass?

In modern particle physics there are basically two kinds of particle:

- Truly elementary particles like quarks and leptons (the electron is an example of a lepton) which are not considered to be made of smaller particles.
- A whole host of bound state particles.

A good example of the latter would be the proton, a bound state of three quarks, according to modern theory. Now if we had to estimate the mass of a proton, we would certainly want to include the rest masses of the constituent quarks, but we would also want to include the kinetic energy of those quarks, not to mention any strong, weak, or electromagnetic binding energy involved in the system, with the energies being suitably divided by c^2. Here, of course, we are making ample use of the celebrated relation $E = mc^2$.

But what about the electron? If it really is a point particle, as often assumed, we cannot make a model for its inertia that is intrinsic to its structure in order to predict its inertial mass. So for the truly elementary particles like quarks and leptons, we invent a field called the Higgs field, and we arrange for these particles to interact with that field in such a way that moving through it is rather like moving through honey, according to one analogy. Put another way, the elementary particles get their inertia from the outside.

However, we still need to renormalise the electron mass in quantum electrodynamics (QED), and we need to renormalise the masses of the other elementary particles likewise in the sophisticated quantum field theories appropriate to them. This suggests that we should not treat the electron as a point particle. At least that would save us the trouble we noted for the charge shell model of the electron in the classical case, as discussed above. So for the purposes of this discussion, let us assume that there are in fact no elementary particles, i.e., that all particles do in fact have some structure, and make the rather radical bootstrap hypothesis that it is the very structure of each particle that causes it to resist being accelerated.

It should be noted that at the present time we use a hybrid model for bound state particles. For example, the quarks in the proton get their mass by interacting with the Higgs field, but most of the mass of the proton comes from its internal structure. It should be noted in this context that ab initio determinations of the light hadron masses are now possible using lattice quantum chromodynamics (QCD) [19]. The hadrons are strongly interacting particles like the proton and the neutron, and QCD is our best theory for the strong force. Using very sophisticated computer simulations, these calculations do indeed confirm that most of the mass of the proton and other similar bound state particles comes from their internal structure.

With the above bootstrap hypothesis in mind, could it be then that all the mass of the electron comes somehow from the electrodynamic effect, i.e., from the mass associated with its EM fields? If this were to be true, we would have to have

$$m_{\mathrm{e}} \sim \frac{e^2}{ac^2} \, ,$$

where m_{e} is the electron mass, e the electron charge, c the speed of light, and a the linear dimension of the charge distribution being used to model the electron. This can be rearranged to estimate the latter:

$$a \sim \frac{e^2}{m_{\mathrm{e}} c^2} =: r_{\text{classical}} \, ,$$

which is called the classical electron radius. The latter can be calculated from the measured values of the constants in the above expression, whence

$$r_{\text{classical}} = 2.82 \times 10^{-15}\ \text{m}\,.$$

Unfortunately this is much too big. Experiment suggests that the linear dimension of the electron cannot be greater than 10^{-18} m, so it would have to be at least a thousand times more massive for this to work.

Another problem is the strange discrepancy between the energy-derived and momentum-derived EM masses, i.e., the factor of 3/4 in the relation

$$m_{\text{EM}}^{\text{EDM}} = \frac{3}{4} m_{\text{EM}}^{\text{MDM}}\,,$$

as can be seen from (8.1) and (8.2). This has caused a long controversy, still underway in some quarters. However, the basic explanation was pointed out by Poincaré over a hundred years ago! It is due to leaving out the binding forces in the system. The point is that the EM energy–momentum tensor for the charge shell system is not conserved everywhere, and one cannot obtain a Lorentz covariant four-momentum by integrating a non-conserved energy–momentum tensor over spacelike hypersurfaces.

One approach here is to redefine the EM energy–momentum tensor, or the energy–momentum of the EM fields, in an ad hoc way so that things work out. But another is to say that the discrepancy *should* be there in general until we include *all* the forces involved in the particle. Then we obtain a total energy–momentum tensor which is conserved, and we can integrate that over spacelike hypersurfaces to obtain an energy–momentum four-vector which is Lorentz covariant.

So whatever happened to electromagnetic mass? Feynman thought that as soon as one had to introduce unknown binding forces into the model for the electron, that made the whole idea too complicated to be worth bothering about [21]. And then of course quantum electrodynamics came along and that deals with the electron in a very different way, although as mentioned earlier, it does leave us with some of the same problems, in particular, the problem of mass renormalisation.

And what about a mechanism here? Why should these electromagnetic effects cause a spatially extended charge distribution to resist being accelerated? So far we have talked about the EM fields around the charge shell, found their energy, and divided it by c^2 to associate a mass with that. And we have found the momentum of the EM fields, but not the momentum of the charge shell itself. However, we may ask what brings about this momentum.

Remember that, when the charge shell is sitting still in an inertial frame, there is no momentum in the fields. But if we push it for a while and get it moving, we find that there is some momentum in the fields, and this suggests that we must have supplied some force in addition to the one required by the mechanical inertial of the electron, by which we mean any inertia from other origins than the electromagnetic effects. But that in turn suggests that there must have been a corresponding extra force acting back on the accelerating agent.

And this is indeed the case. Whenever we try to accelerate a spatially extended charge distribution, it will exert an EM force on itself, in fact, an EM self-force. Here is the formula Feynman gives in [21]:

$$F_{\text{self}} = -\alpha \frac{e^2}{ac^2}\ddot{x} + \frac{2}{3}\frac{e^2}{c^3}\dddot{x} + \beta \frac{e^2 a}{c^4}\ddddot{x} + O(a^2) . \tag{8.3}$$

What we have here is the EM self-force for a rather arbitrarily shaped spatially extended charge distribution or charge blob of linear dimension a, moving in an arbitrary way in one dimension. The symbol x denotes the position of some preselected point within the charge blob, a function of the proper time τ along the worldline of that point. Dots over the x denote proper time derivatives, so $\dot{x}$ is the speed, $\ddot{x}$ is the acceleration, and so on.

The self-force has been expanded as a power series in a, with a leading order term that goes as a^{-1}, then a term that is independent of a, then a whole infinite sum of terms going as a, a^2, and so on. The total charge on the charge blob is e, and α and β are constants that depend more or less only on the shape of the charge distribution (although not quite, as we shall see). For example, α has the value 2/3 for the spherical charge shell.

The first thing to note is that, if we try to take a point particle limit by letting a tend to zero, that leading order term goes to infinity, so the EM self-force is always infinite in the point particle limit.

The second thing to note is that the leading order term, going as a^{-1}, is proportional to the acceleration $\ddot{x}$. This will be very important when we come to consider mass renormalisation in a moment. The coefficient of the acceleration in this leading order term has units of mass and we may call it the self-force-derived mass of the charge blob:

$$m_{\text{EM}}^{\text{SFDM}} = \alpha \frac{e^2}{ac^2} . \tag{8.4}$$

This can be compared with the energy-derived and momentum-derived masses in (8.1) and (8.2).

We can think of the self-force as expressing a breakdown of Newton's third law within the particle, in the sense that the sum of all the electromagnetic actions and reactions within the particle is not zero *when the particle is being accelerated*. This last condition is very important. The EM self-force effect makes a clear distinction between constant velocity motion, when the self-force is always zero, as can be seen by inserting $\dot{x} = \text{constant}$ in (8.3), and accelerated motion, when it is unlikely ever to be zero.

This is very important because it saves Newton's first law, which proclaims that no external force should be required to keep an object moving at constant velocity. This law would not be true for any spatially extended charge distribution, not even one that is overall electrically neutral, unless the self-force were zero for constant velocity motion.

We can put this in a rather amusing way. When Newton's first law applies, i.e., there is no external force and hence we have constant velocity motion, Newton's third law also applies, in the above sense that the sum of all the EM actions and reactions within the particle is then zero. But when Newton's first law fails, i.e., we have an external force and accelerated motion, then Newton's third law also fails.

Another important point to mention is that the momentum-derived and self-force-derived masses in (8.2) and (8.4) are always exactly equal, i.e.,

$$m_{\mathrm{EM}}^{\mathrm{MDM}} = m_{\mathrm{EM}}^{\mathrm{SFDM}} .$$

This confirms the idea that momentum gets into the fields by our having to overcome this leading order term in the self-force.

We should say a word about the shape of the accelerating charge blob. As mentioned earlier, its shape in any given situation results from the equilibrium between the binding forces and the repulsive EM forces between charge elements making up the blob. When it changes its velocity, it will be desperately trying to FitzGerald contract to suit its instantaneous velocity, but with some delay effects depending on exactly how it is accelerated.

In self-force calculations, we generally assume rigidity of the blob, in the relativistic sense, known as Born rigidity, much discussed earlier in this book. This basically amounts to assuming that the charge blob always has exactly the same shape in its instantaneous rest frame, that is, in the instantaneously comoving inertial frame. Achieving this physically will depend as much on how the blob is accelerated as on the nature of the binding forces.

Note, however, that more sophisticated self-force calculations are possible that do not assume rigidity. The reader is referred to very recent work by Gralla, Harte, and Wald in Chicago [26].

We should also comment on the second term in the EM self-force, viz.,

$$F_{\mathrm{self}}^{\mathrm{rad}} := \frac{2}{3}\frac{e^2}{c^3}\dddot{x} . \tag{8.5}$$

This term does not depend on the spatial dimensions a of the system, and it does not even depend on its shape, as can be seen from the above expression. This remarkable term is the radiation reaction force. The rate of doing work against this part of the bootstrap force is exactly the rate of energy emission by radiation as given by the Larmor formula.

Note that we would lose this explanation for the EM radiation by accelerating electrons if we treated them as point particles, so this is another bonus of our earlier bootstrap hypothesis, according to which there are no point particles.

8.2 Classical Mass Renormalisation

So we find that the EM self-force is in general a very complicated infinite sum of terms. Now obviously the expression (8.3) would be vastly simplified if we could set a equal to zero. Then we would get rid of all the terms going as a, a^2, and so on, and we would still have the radiation reaction term, which we would like to keep since it explains why accelerating charged particles radiate electromagnetic energy. But as noted earlier, the problem with this strategy is that the leading order term goes to infinity as the spatial extent of the charge distribution tends to zero.

Then in 1938, Dirac came along [16], and he said, let us suppose that the electron acts on itself only through the second self-force term and not through the first or any of the higher order terms. And that peculiar suggestion is the basis of classical mass renormalisation. Let us sketch briefly how that works.

We begin with Newton's second law, which says that force equals mass times acceleration. In this case the force is the sum of an external force causing the acceleration and the resulting EM self-force, and this is equal to the mass times the acceleration $\ddot{x}$, so we have

$$F_{\text{ext}} + F_{\text{self}} = m_{\text{bare}}\ddot{x} \, .$$

For the moment, let us call the mass the bare mass.

We now analyse the self-force into the infinite series (8.3), but drop all the terms going as a, a^2, and so on, since we intend to let a tend to zero at the end. We keep the radiation reaction term on the left of our new equation with the external force, but the clever thing here is that we group the potentially divergent leading order term with the bare mass times the acceleration. And the reason why this is useful is just that the leading order term is itself proportional to the acceleration. That is what makes this ploy work.

The result is a new version of Newton's second law, viz.,

$$F_{\text{ext}} + \frac{2}{3}\frac{e^2}{c^3}\dddot{x} = \left(m_{\text{bare}} + \alpha\frac{e^2}{ac^2}\right)\ddot{x} \, .$$

On the left we have the external force plus the radiation reaction force, and on the right a new mass term times the radiation. The renormalisation step in the argument consists in saying that everything in the round brackets on the right-hand side must just be the measured mass. We do not worry about the fact that part of it must tend to infinity in the limit $a \to \infty$. We thus define the measured or renormalised mass to be

$$m_{\text{ren}} := m_{\text{bare}} + \alpha\frac{e^2}{ac^2} \, .$$

This is the process of classical mass renormalisation.

8.3 Self-Force Calculations

The spherically symmetric charge shell is very nice because we can actually carry out the self-force calculation, but it is certainly not the simplest spatially extended charge distribution we can imagine. That must surely be the distribution shown in Fig. 8.1, i.e., two point charges A and B held some distance apart by some binding forces that we shall try not to think about.

Fig. 8.1 Charge dumbbell, or toy model for the electron, at rest in an inertial frame, as used to investigate EM self-force. There are some binding forces and the system is in equilibrium under the forces between *A* and *B*

How do we do a self-force calculation? We get the dumbbell moving in some arbitrary way in an inertial frame, describing the motion of each point charge, and then we use the remarkable Liénard–Wiechert formula which tells us the electric and magnetic fields **E** and **B** produced by each point charge and at every point in spacetime as a result of its specific motion. We can then calculate the fields created by *A* at *B* and vice versa, and hence the EM force exerted by *A* on *B*, and the EM force of *B* on *A*, and we simply add these together to estimate the total EM force the system exerts on itself.

Let us try to get an intuitive picture of how the self-force comes about by considering the charge dumbbell moving in some arbitrary way in one dimension, but perpendicular to its own axis, as shown in Fig. 8.2. Now *B* will produce some fields which will affect *A* slightly later due to retardation effects, and slightly later, *A* will have moved a little bit to the right. Of course, if *A* and *B* are like charges, they repel one another, so it looks as though the electric force of *B* on *A* will have a main component along the system axis *AB* and a small component in the direction of motion. Likewise it looks as though the electric force of *A* on *B* will have a main component along *AB* and a small component in the direction of motion. When we add these together, the components along the system axis will cancel exactly by symmetry, and it looks as though there will be two small components in the direction of motion that add up to give a net electric self-force in the direction of motion.

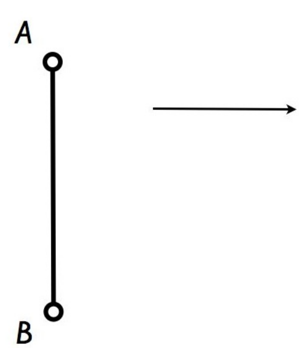

Fig. 8.2 Charge dumbbell moving with arbitrary motion in one dimension in an inertial frame, with direction of motion perpendicular to its own axis

However, that could not possibly be correct. If it were, we could just get the charge dumbbell moving at constant velocity in 1D perpendicular to its axis and it would begin to accelerate itself! This would be very nice, of course, because it would solve all today's energy problems. However, this does not happen. Recall that the EM self-force is zero for constant velocity motion, so the intuitive picture breaks down.

This is in fact due to a remarkable result from Maxwell's theory of electromagnetism: when a charged particle moves with constant velocity, the electric field it produces is radial, not from its retarded position, but from its current position. It is as though the fields corrected themselves to first order for retardation effects, and first order is sufficient for constant velocity motion. Here we should think about Newton's first law, which would not be true for the charge dumbbell with this kind of motion if it were not for this fundamental result from electromagnetic theory.

So the intuitive picture is not valid and we are stuck with carrying out a somewhat tedious calculation using the Liénard–Wiechert formulas for the fields. What do we find for the charge dumbbell moving in an arbitrary way but perpendicular to its own axis as shown in Fig. 8.2? In fact, the electric force of A on B does indeed have a small component in the direction of motion and a main component along the system axis, and likewise for the electric force of B on A. However, the small components in the direction of motion are always counteraligned with the acceleration for like charges, and pay no heed to the direction of the velocity.

When we add together these electric forces, the axial components (along AB) do indeed cancel, while the components counteraligned with the acceleration for like charges add up to produce a net electric self-force that is also opposed to the acceleration. The magnetic force of A on B lies along the system axis. Likewise for the magnetic force of B on A, and they cancel exactly.

Now here is an interesting thing. If the charges at A and B have opposite signs, then the self-force changes sign. So it will actually assist the acceleration! This may look somewhat suspicious, but it is exactly what one would expect for a charge dipole. Recall that the EM binding energy in a charge dipole is negative, and we expect any negative binding energy in a bound state particle to decrease its inertial mass.

So we expand the EM self-force in powers of d, the separation between A and B. We find that the Coulomb terms cancel so there is no term going as $1/d^2$. However, there is a residue going as $1/d$ and we find, as always, that the self-force diverges when d tends to zero. The EM self-force is in this case

$$F_{\text{self}} = -\frac{e^2}{4c^2 d}\gamma(v)^3 \ddot{x} + O(d^0) \; . \tag{8.6}$$

Here we see that the factor α in Feynman's general self-force expression (8.3) is equal to 1/4 for this shape of charge distribution with this motion.

Equation (8.6) is the result of a relativistic calculation and we see appearing the relativistic factor γ, function of the instantaneous speed. Note that it occurs as a cube, which may look awkward. However, this gamma factor is just right for renormalising the relativistic version of Newton's second law, thanks to the simple identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\big[\gamma(v)v\big] = \gamma(v)^3\ddot{x} \, .$$

So here we have a dynamical explanation as to why the inertial mass of an object should increase with speed as $\gamma(v)$. It is because the EM self-forces increase in the appropriate way, in this case as a cube of γ, and this in turn is presumably a consequence of the Lorentz symmetry of Maxwell's equations. One would expect the same explanation to work for contributions to the inertial mass from other fields operating within a bound state particle, if those fields satisfy Lorentz symmetric equations (see Sect. 9.3 for further discussion of this kind of explanation).

So for this system, we obtain the renormalised mass

$$m_{\text{ren}} := m_{\text{bare}} + \frac{1}{4}\frac{e^2}{c^2 d} \, .$$

If we calculate the next term in the series expansion of the self-force, i.e., the one independent of d, we do indeed get the correct radiation reaction to explain the EM energy radiation by this charge distribution when it moves in this way.

We can also consider our charge dumbbell moving in an arbitrary way in one dimension, but this time along its own axis, as shown in Fig. 8.3. Here we must make some assumption about the length, such as a rigidity assumption, for example. In this case there are no magnetic forces of either A on B or B on A, and the leading order term in the electric self-force acts along the system axis. Once again it is always counteraligned with the acceleration for like charges A and B and pays no heed to the direction of the velocity.

The EM self-force is now

$$F_{\text{self}} = -\frac{e^2}{2c^2 d}\gamma(v)^3\ddot{x} + O(d^0) \, . \tag{8.7}$$

Interestingly, the self-force-derived EM mass has changed. We now find

$$m_{\text{ren}} := m_{\text{bare}} + \frac{1}{2}\frac{e^2}{c^2 d} \, ,$$

so the factor α in Feynman's general self-force expression (8.3) is equal to 1/2 for this shape of charge distribution with this motion. Note that we still get the right γ factor in (8.7) to be able to renormalise the relativistic version of Newton's second law.

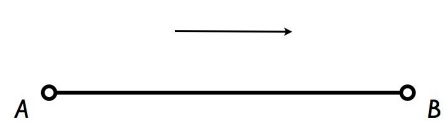

Fig. 8.3 Charge dumbbell moving with arbitrary motion in one dimension in an inertial frame, with direction of motion along its own axis

So what we discover here is that the EM mass of an object depends on which way it moves relative to its own geometric configuration. In the present case we have

$$m_{\mathrm{EM}}^{\mathrm{SFDM}}(\text{longitudinal}) = 2m_{\mathrm{EM}}^{\mathrm{SFDM}}(\text{transverse}) \ .$$

Such a dependence is hardly surprising when we consider what is going on within these bound state systems.

We can also consider the charge dumbbell rotating about a distant center as shown in Fig. 8.4. The distance R from the center of rotation should be considered much greater than the length d of the dumbbell. Now the acceleration is along the system axis and the velocity is perpendicular to it. But despite the very different configuration, we find once again that the leading order term in the self-force is counteraligned with the acceleration for like charges, i.e., it is radially outward from the center of rotation, and the relativistic factors work out perfectly to be able to renormalise the relativistic version of Newton's second law.

We can also consider the charge dumbbell rotating about a distant center as shown in Fig. 8.5. Once again the distance R from the center of rotation should be considered much greater than the length d of the dumbbell. This situation is quite different again. The acceleration is perpendicular to the system axis and the velocity lies along it, but once again we find that the leading order term in the self-force is counteraligned with the acceleration for like charges, i.e., it is still radially outward from the center of rotation, and the relativistic factors work out perfectly to be able to renormalise the relativistic version of Newton's second law.

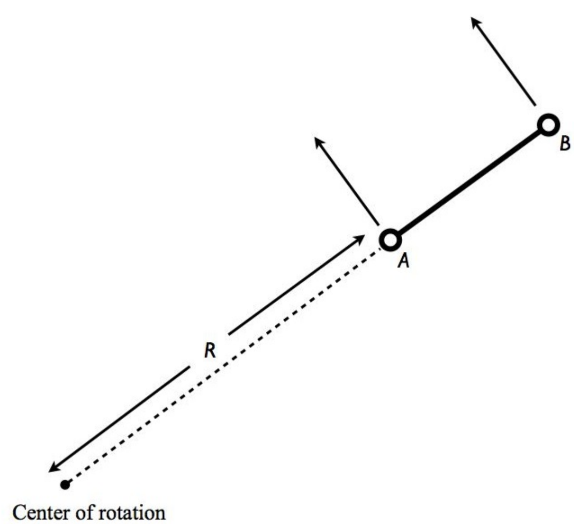

Fig. 8.4 Charge dumbbell rotating at constant angular speed about a center at distance R from A, in such a way that the system axis AB always points to the center of rotation

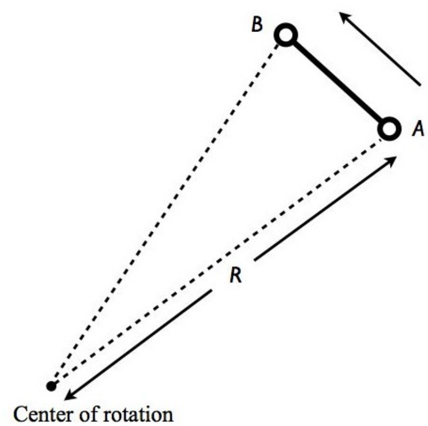

Fig. 8.5 Charge dumbbell rotating at constant angular speed about a center at distance R from A, in such a way that the system axis AB is always perpendicular to the line joining its midpoint to the center of rotation

This raises a question: is the leading order term in the EM self-force always aligned or counteraligned with the acceleration? The answer is negative. Despite the positive results just mentioned for the charge dumbbell, it turns out that these are exceptions. When the charge dumbbell moves (rigidly) along an arbitrary worldline, the leading order term in the self-force contains a contribution along the system axis, in addition to the contribution aligned or counteraligned with the acceleration which can be removed by mass renormalisation (see Chap. 10).

Note, however, that this unwanted term can be made to disappear by considering a spherically symmetric charge distribution. Indeed, the result for the charge dumbbell can be used in an integration to obtain the leading order term in the EM self-force for a spherically symmetric distribution, and one sees exactly how the unwanted contribution to this term drops out. This suggests that renormalisability in QFT may contain a hidden assumption of spherical symmetry.

So what is the connection between all this and bound state particles? In modern particle physics, energy and mass are the same thing because

$$E = mc^2 \ .$$

So this is why we include binding energy in the inertial mass of composite particles, and everything is very simple. However, this hides another dynamical explanation, albeit a pre-quantum theoretical explanation. Binding forces in composite particles lead to bootstrap effects, and that is why binding energy must be included in their inertial mass. So the message here is that we include binding energies because they reflect the related self-forces in those bound states (see Sect. 9.4 for further discussion of this kind of explanation).

8.4 Rewriting Newton's Second Law

So what are the benefits of our radical bootstrap hypothesis, made at the beginning of this discussion. First of all, Newton's second law has become much simpler. It now says just $F = 0$. However, F has become much, much more complicated. It is the sum of the external force causing acceleration and all the resulting self-forces due to fields operating within the bound state particle:

$$F_{\text{ext}} + \sum_{\text{fields}} F_{\text{self}} = 0 \,. \tag{8.8}$$

The usual form for Newton's second law is obtained by analysing the self-forces and moving the leading order terms, proportional to the acceleration, to the right-hand side.

And now we deduce results that were simply imposed before. For example, we get a numerical value for the mass in the $F = ma$ form of the above law. And since self-forces make a clear distinction between uniform velocity and changing velocity, we get a nice explanation of Newton's first law. We really understand why no external force is needed to keep a particle in constant velocity motion: it is because the particle then exerts no forces on itself.

At this point, one might wonder about the so-called Mach principle, which in one version suggests that a particle somehow (mysteriously, causally) gets its inertia from the overall distribution of matter and energy in the Universe. Under the present hypothesis, a particle gets its inertia solely and completely from within itself. What picks out the overall distribution of matter and energy in the Universe is that it seems to specify what pass for inertial frames, i.e., those frames in which our field theories of matter take on their simplest forms. But what seems more cogent would be the idea that the matter and energy in the Universe has evolved into its present distribution *because* of the existence of such frames, themselves a consequence of the underlying (Lorentz) symmetry of the field theories of matter.

8.5 Self-Force Effects in Modern Particle Physics

Let us now look briefly at what happens in particle physics today. As mentioned before, we have several truly elementary particles like leptons and quarks which are not considered to have any internal structure, and then we have a whole host of bound state particles like baryons and mesons which we try to organise into multiplets.

What is a multiplet? Mathematically, it is a vector space carrying a representation of a group. But physically, it is just a set of particles with similar properties. And one of the properties that has to be similar across a multiplet is the inertial mass. So one task of modern particle physics is to group bound state particles together into sets with similar masses, and another is to explain why the masses are not quite equal across a given multiplet, something known in the jargon as mass splitting.

A good example is provided by the neutron, with inertial mass 939.5 MeV/c^2, and the proton, with inertial mass 938.2 MeV/c^2. These have very similar masses, so they are ideal for putting together in a multiplet. We thus make the hypothesis that their quantum states carry an irreducible representation of the isospin SU(2) symmetry group:

$$|\mathrm{p}\rangle = |T = 1/2, T_3 = 1/2\rangle \;, \qquad |\mathrm{n}\rangle = |T = 1/2, T_3 = -1/2\rangle \;, \tag{8.9}$$

characterized by an isospin value of $T = 1/2$.

Another example is the pion isotriplet. Once again, the positively charged, neutral, and negatively charged pions have very similar masses so they are ideal for grouping together into a multiplet, and we make the hypothesis that their quantum states carry an irreducible representation of the isospin SU(2) symmetry group, but this time characterized by an isospin value of $T = 1$:

$$\begin{aligned} |\pi^+\rangle &= -|T = 1, T_3 = 1\rangle \;, \qquad 139.6\ \mathrm{MeV}/c^2 \;, \\ |\pi^0\rangle &= |T = 1, T_3 = 0\rangle \;, \qquad 135.0\ \mathrm{MeV}/c^2 \;, \\ |\pi^-\rangle &= |T = 1, T_3 = -1\rangle \;, \qquad 139.6\ \mathrm{MeV}/c^2 \;. \end{aligned} \tag{8.10}$$

Note the slightly smaller mass of the neutral pion, something we shall be able to explain shortly.

And of course we also have the quark flavour models for these particles. The neutron and proton are bound states of three quarks. The neutron is a bound state of one up quark and two down, while the proton is a bound state of one down and two up:

$$\mathrm{n} = \mathrm{udd} \;, \qquad \mathrm{p} = \mathrm{uud} \;.$$

The pions are mesons, i.e., quark–antiquark bound states:

$$\pi^- = \mathrm{d}\bar{\mathrm{u}} \;, \qquad \pi^0 = \mathrm{d}\bar{\mathrm{d}}\;,\; \mathrm{u}\bar{\mathrm{u}} \;, \qquad \pi^+ = \mathrm{u}\bar{\mathrm{d}} \;. \tag{8.11}$$

So, for example, the positively charged pion is a bound state of an up quark and an antidown quark and the neutral pion a superposition of down–antidown and up–antiup. And note that the up and down quarks themselves form a multiplet. We make the hypothesis that their quantum states carry an irreducible representation of the isospin SU(2) symmetry group characterized by an isospin value of $T = 1/2$.

Now these multiplets reflect a symmetry under the strong force. Recall that the strong interaction is at work in the above bound states, and indeed, they are largely held together by the strong interaction. In each case we model that by means of a strong interaction Hamiltonian H_{strong} which describes the energy of all the strong interactions going on within the bound state. Then the symmetry hypothesis is expressed by saying that this strong interaction Hamiltonian commutes with all the generators T_1, T_2, and T_3 of the isospin SU(2) symmetry group:

$$\left[H_{\text{strong}}, T_i\right] = 0 \,, \qquad i = 1,2,3 \,.$$

This in turn implies that the operator

$$S_{\text{strong}} = \exp \frac{\mathrm{i} H_{\text{strong}} t}{\hbar}$$

determining the time evolution of the bound state commutes with all generators and hence with all members of the isospin SU(2) group.

That is the mathematical statement of the symmetry assumption, but what does it mean physically? In fact it asserts that the strong interactions make no distinction between states in the multiplet. Put another way, the up quark and the down quark look the same as far as the strong force is concerned. However, electromagnetic interactions are also at work within these bound states, and these interactions do make a distinction here. The up quark has electric charge $+2/3$ and the down quark has electric charge $-1/3$, so these look very different to the EM interaction.

Now the mass of a state ψ in quantum theory is expressed as an expectation value of the relevant Hamiltonian in the given state, viz.,

$$M_\psi = \langle \psi | H | \psi \rangle \,, \tag{8.12}$$

where H is the Hamiltonian modelling all the energy sources in the system. In this case then, we expect to have

$$H = H_{\text{strong}} + H_{\text{EM}} \,.$$

Inserting this in the expression (8.12) for the mass, we expect the masses of these bound states to be a sum of a strong contribution, which is expected to largely dominate, and a much smaller EM contribution:

$$\begin{aligned} M_\psi &= \langle \psi | H_{\text{strong}} | \psi \rangle + \langle \psi | H_{\text{EM}} | \psi \rangle \\ &= M_\psi^{\text{strong}} + M_\psi^{\text{EM}} \,. \end{aligned}$$

And the upshot of the symmetry hypothesis is that the strong contributions to the masses of all the bound states within a given multiplet will all be exactly equal, if the symmetry assumption holds exactly, while the much smaller EM contributions will differ from one state to the next.

The symmetry hypothesis concerning the strong interactions therefore explains the nearly equal masses of all the bound states within a given multiplet, while the EM contributions explain, at least in part, the mass differences. For example, we expect

$$m_{\text{n}} = m_{\text{nucleon}}^{\text{strong}} + m_{\text{n}}^{\text{EM}} \,, \qquad m_{\text{p}} = m_{\text{nucleon}}^{\text{strong}} + m_{\text{p}}^{\text{EM}} \,,$$

so the strong contributions to the neutron and proton masses are expected to be exactly equal if the isospin symmetry holds exactly, while the EM contributions will explain, at least in part, the difference in mass between the two particles.

In modern particle physics then, the notion of electromagnetic mass is still there, but it is found in a very different way as the expectation value of the electromagnetic interaction Hamiltonian in the quantum state for the given bound state, e.g., for the proton and neutron,

$$m_{\mathrm{p}}^{\mathrm{EM}} = \langle \mathrm{p}|H_{\mathrm{EM}}|\mathrm{p}\rangle \ , \qquad m_{\mathrm{n}}^{\mathrm{EM}} = \langle \mathrm{n}|H_{\mathrm{EM}}|\mathrm{n}\rangle \ .$$

Before leaving the domain of particle physics, it is worth pointing out that the classical self-force model can be more sophisticated than just a charge shell or a charge dumbbell. This is fortunate because, as Feynman points out [21], if the proton were just a charged sphere and the neutron a neutral one, then from what was said earlier, we would expect the neutron to have the lower mass, and this is not the case [see (8.9)].

However, when we consider that the neutron and proton are each today considered to be bound states of three charged particles, there is absolutely no reason to think that, if we could actually carry out the classical self-force calculation, the neutron would turn out to have the lower EM mass simply on the grounds that it is overall electrically neutral.

And in fact the dumbbell model works rather well for the pions. As we can see from (8.11), the charged pions are composed of like charges, e.g., the positively charged pion $\pi^+ = \mathrm{u}\overline{\mathrm{d}}$ comprises an up quark with charge $+2/3$ and an antidown with charge $+1/3$, while the neutral pion is composed of opposite charges. So from what was said earlier, we would expect the neutral pion to have the smaller inertial mass, and indeed it does [see (8.10)].

We can even use the crude classical dumbbell model to estimate the length of a pion, obtaining a value $d \sim 10^{-16}$ m, which accords quite well with estimates of pion diameters from cross-section measurements. Note that the spherical shell model also works surprisingly well here.

So the conclusion from all this is that mass splittings in multiplets of bound state particles are explained today, at least in part, by a quantum theoretical version of the classical self-force idea. But this comes with a warning. Mass splittings in multiplets of quark bound states are complicated by the different masses of the different quark flavours.

In fact, the isospin SU(2) flavour symmetry is broken by the fact that the up and down quarks have very different masses. Worse, we cannot measure these masses directly because it has so far proven impossible to isolate an individual quark, so we can only infer their masses from other observations, and a lot remains to be understood yet.

8.6 An Interim Conclusion

What are the bonuses of our radical bootstrap hypothesis made at the beginning of this discussion, i.e., the assumption that there are no point particles and that the inertial mass of all particles results entirely from self-forces due to the various fundamental interactions operating within them? To begin with, we have a simpler version of Newton's second law, which now says just that the total force on any particle is always zero. And we have dynamical explanations for:

- Inertia and inertial mass, with the possibility of actually calculating the latter.
- The speed dependence of inertial mass as it is usually found in relativistic dynamics (see also Sect. 9.3).
- The inclusion of binding energies in inertial mass (see also Sect. 9.4).
- The EM energy radiation by accelerating charged particles.

But there is one more thing that we have not discussed in the above.

8.7 Passive Gravitational Mass and the Geodesic Principle

We have been talking about inertial mass, and it is well known that, according to very accurate measurements, the inertial mass of any object is exactly equal to its passive gravitational mass (PGM). Recall that the passive gravitational mass gauges the extent to which the particle is affected by gravity in Newtonian gravitational theory. Now it is sometimes said that general relativity explains why we should have this equality. The point is that any particle will follow a geodesic of the spacetime metric if there are no non-gravitational effects around to act on it. But that implies that any two particles, no matter what their inertial mass, will fall in the same way in the absence of any non-gravitational effects, i.e., free fall does not depend on the nature of the particle, and in particular on its inertial mass.

Actually, there are some provisos regarding this so-called geodesic principle, which we shall turn to in Sects. 8.7.2–8.7.5. But let us just note that the above explanation could be considered to turn things upside-down, since general relativity would not even be possible if it were not for the equality of inertial and passive gravitational mass, and it was the very discovery of this equality that led eventually to Einstein's formulation of the general theory of relativity (GR). Furthermore, there is a sense in which the above explanation lacks somewhat in impact, since it just seems to be a sophisticated *reformulation* of the equality of the PGM and the inertial mass, rather than providing any kind of mechanism. Let us try to do a little better.

8.7.1 Equality of Inertial Mass and Passive Gravitational Mass

If we place an object on our outstretched hand, we prevent it from free fall. In the GR picture there is only one force on the object, namely the force we exert upwards on it in order to push it off its geodesic. In the usual view of GR, gravity is not a force and there is no such thing as weight. This can be viewed as a linguistic adjustment, ensuring that forces are always associated with accelerations: the freely falling object has zero acceleration in the GR picture, while our supported object is being accelerated by the upward force from our hand.

However, we nevertheless feel the object pushing down on our hand, and it is interesting to wonder how it does that. One might say that this is just the reaction, according to Newton's third law, to the force we are exerting on the object. Indeed, by pushing on it, we slightly deform the microscopic structure of the object near its lower surface, and that structure will react to that. But there is another rather intriguing way of looking at this through the idea of the self-force. Let us examine how that works.

Just to set the scene in this GR view, consider a spacetime with coordinates

$$\left(y^0, y^1, y^2, y^3\right)$$

and a metric that only differs from the Minkowski form in the 00 component, which is the following function of one of the space coordinates:

$$g_{00} = \left(1 + \frac{gy^3}{c^2}\right)^2 ,$$

where c is the speed of light and g a constant with units of acceleration. These are supposed to be the coordinates one would set up in a laboratory held fixed relative to a distant gravitational source. The metric then describes a parallel gravitational field in the y^3 direction (see also Sect. 6.3).

By the weak equivalence principle (WEP), at any event in any curved spacetime there is a neighbourhood with coordinates such that the metric looks Minkowskian to a good approximation. However, for this particular metric, the neighbourhood can be the whole spacetime, because this spacetime happens to be flat. The curvature is zero and there are no tidal effects.

The Minkowski coordinates whose existence is guaranteed by WEP are supposed to be the coordinates that would naturally be used by a freely falling observer, in the sense that an observer sitting at the space origin of such coordinates would be following a geodesic. The coordinates $\{y^\mu\}$ can be viewed as coordinates that might be adopted by a uniformly accelerating observer, in the sense that an observer sitting at the space origin of the $\{y^\mu\}$ coordinates would have uniform acceleration.

Now consider a charge shell held fixed at the origin of the $\{y^\mu\}$ system, i.e., a sphere supported against the uniform gravitational field. Since the four-acceleration is nonzero, this requires a force. The sphere is being pushed off its geodesic. As viewed from the freely falling frame, the sphere will appear to be accelerating (and it has nonzero acceleration according to the GR definition of acceleration).

We now import the theory of electromagnetism using the strong principle of equivalence (SEP). This states that any theory of non-gravitational physics will look roughly as it does in flat spacetime when described in locally inertial coordinates. The principle is usually formulated by taking the flat spacetime field equations for the non-gravitational effect and replacing all (inertial) coordinate derivatives by covariant derivatives (minimal extension of the theory from flat to curved spacetime).

According to SEP, an *exactly* equivalent view in this case (because our spacetime is in fact flat) is of the charge shell accelerating uniformly in a flat spacetime without gravity. But then we know that the sphere will exert an EM force on itself that opposes the acceleration, i.e., that acts toward the gravitational source. Put another way, the EM self-force will oppose the supporting force of the laboratory table upon which the sphere sits. In fact, it contributes to its weight.

But what is weight in GR? We said above that GR effectively does away with this notion. However, there is a natural way to reinstate it. We simply define it to be the negative of the force required to support the object at a fixed distance from the source:

$$W = -F_{\mathrm{supp}} \, .$$

Then we need to reinstate passive gravitational mass, also rendered obsolete by general relativity. Since it is supposed to gauge the extent to which the object is affected by gravitational effects, the natural definition is to take it as (minus) the constant of proportionality between the weight and the four-acceleration A, i.e.,

$$W = -m_{\mathrm{PG}} A \, ,$$

whence

$$F_{\mathrm{supp}} = m_{\mathrm{PG}} A \, .$$

We now propose the new dynamical law (8.8), viz.,

$$F_{\mathrm{supp}} + F_{\mathrm{self}} = 0 \, , \tag{8.13}$$

but this time in the GR framework. Here we are assuming the radical bootstrap hypothesis that all the inertial mass of an object arises due to the leading order terms in self-forces. Since

$$F_{\mathrm{self}} = -m_{\mathrm{inertial}} A + \text{smaller terms} \, ,$$

at least for spherically symmetric charge distributions, we deduce that

$$m_{\mathrm{PG}} = m_{\mathrm{inertial}} \, . \tag{8.14}$$

Note that all this is just standard theory, and precisely the way we always naturally think about things today, although with a different, and classical, mechanism. Think, for example, about the weight of a proton or atomic nucleus: we automatically include the binding energy as part of the weight. In GR, energy is gravitationally attractable.

Amusingly, if some part of the binding energy of a particle is negative, we have an antigravity effect. Think, for example, of a charge dipole lying on the laboratory table. The negative EM binding energy in the system will slightly decrease its weight. And we can see why no particle or object has ever been seen to float up into the air under the Earth's gravity! It is simply because no particle or object we have ever observed has ever been found to have negative *inertial* mass.

Of course, the above argument leading to (8.14) is circular (see also Sect. 8.7.3). After all, we use this very same experimental result to formulate GR in the first place. But here perhaps we also have a mechanism, at least for bound state particles, or for the binding energy contributions to the inertial masses of bound state particles. Contrast with a naive special relativistic version of gravity in which gravity is just a force. The object supported by my outstretched hand is subject to two forces: the supporting force from my hand and its weight. But they exactly balance. There is no acceleration and there is no hope of an explanation of the kind just given.

In GR as it is usually presented, the supporting force is needed *because* the particle has nonzero four-acceleration. But we do not ask why a nonzero four-acceleration should require a supporting force, any more than we ask why an acceleration should require a force in Newtonian physics. The picture here, in the radical bootstrap hypothesis, is that all forces on an object must always exactly balance to give a zero resultant, as in (8.13).

And as mentioned earlier, self-forces make a clear distinction between uniform velocity and changing velocity. Here we understand why no supporting force is needed to keep a particle in free fall in GR. It is simply because it is not accelerating in the GR picture, so it does not exert any force on itself.

8.7.2 Geodesic Principle

This states that, when point particles are not acted upon by forces (apart from gravitational effects), their trajectories take the form

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}s^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\frac{\mathrm{d}x^\rho}{\mathrm{d}s} = 0\,, \tag{8.15}$$

where $x^\mu(s)$ gives the worldline as a function of the proper time s of the particle and $\Gamma^\mu_{\nu\rho}$ are the connection coefficients in the given coordinate system. In the literature, this is often derived from an action principle. One writes the worldline as a function $x^\mu(\lambda)$ of some arbitrary parameter λ, whence the appropriate action for the worldline between two points $P_1 = x(\lambda_1)$ and $P_2 = x(\lambda_2)$ of spacetime is

$$s(P_1,P_2) := \int_{\lambda_1}^{\lambda_2}\left(g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\right)^{1/2}\mathrm{d}\lambda = \int_{\lambda_1}^{\lambda_2} L\,\mathrm{d}\lambda = \int_{\lambda_1}^{\lambda_2}\mathrm{d}s\,, \tag{8.16}$$

with Lagrangian

$$L := \left(g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\right)^{1/2}\,. \tag{8.17}$$

The Euler–Lagrange equations extremising the action under variation of the worldline are

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) - \frac{\partial L}{\partial x^\mu} = 0\,, \tag{8.18}$$

with $\dot{x}^\mu := \mathrm{d}x^\mu/\mathrm{d}\lambda$, and these lead to the above geodesic equation (8.15).

The action for some particles labelled by a is

$$\mathscr{A} = -\sum_a c m_a \int \mathrm{d}s_a \ , \tag{8.19}$$

where m_a is the mass of particle a and s_a is its proper time. This is the action *because* variation of the worldline of particle a gives its equation of motion as

$$\frac{\mathrm{d}^2 a^\mu}{\mathrm{d}s_a^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\frac{\mathrm{d}a^\rho}{\mathrm{d}s_a} = 0 \ . \tag{8.20}$$

Likewise, if some of the particles are charged with charge e_a for particle a, and there are some EM fields $F_{\mu\nu}$, one declares the action to be

$$\mathscr{A} = -\sum_a c m_a \int \mathrm{d}s_a - \frac{1}{16\pi c}\int F_{\mu\nu}F^{\mu\nu}(-g)^{1/2}\mathrm{d}^4x - \sum_a \frac{e_a}{c}\int A_\mu \mathrm{d}a^\mu \ , \tag{8.21}$$

where A_μ is a 4-vector potential from which $F_{\mu\nu}$ derives, simply *because* variation of the worldline of particle a gives its equation of motion as

$$\frac{\mathrm{d}^2 a^\mu}{\mathrm{d}s_a^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\frac{\mathrm{d}a^\rho}{\mathrm{d}s_a} = \frac{e_a}{m_a}F^\mu{}_\nu\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a} \ , \tag{8.22}$$

the minimal generalisation of the Lorentz force law to a curved manifold, while variation of A_μ gives the EM field equations as the minimal extension of Maxwell's equations to the curved spacetime. Of course, these actions are designed to give appropriate field equations, and we are just decreeing here that the appropriate field equation for the particle labelled by a is a geodesic equation, or an equation like (8.22).

So what is the physical motivation for the geodesic principle? It is claimed here that the appropriate physical argument supporting (8.15) is an application of the strong principle of equivalence. This is where we discover exactly how we are to link what happens mathematically in a curved manifold with measurements in our own world. We start from an action, but at some point we must say what the point of contact would be with physical reality. Let us suppose we impose a strong principle of equivalence, that is, we say roughly speaking that any physical interaction other than gravitation behaves in a locally inertial frame as though gravity were absent. Relative to such a frame, any particle that is not subject to (non-gravitational) forces will then move in a straight line with uniform velocity, i.e., it will follow the trajectory described by

$$\frac{\mathrm{d}v^\mu}{\mathrm{d}s} = 0 \ , \tag{8.23}$$

where v^μ is its 4-velocity. This is expressed covariantly through (8.15). If there are non-gravitational forces, we start with $\mathbf{F} = m\mathbf{a}$ in the locally inertial frame and we find (8.15) with a force term on the right-hand side. It is quite clear that we still have a version of Newton's second law $\mathbf{F} = m\mathbf{a}$, so the present view is that we have not explained inertia and inertial effects by this ploy, but merely extended this equation of motion to the new theory.

It is worth looking more closely at the claim that (8.23) is expressed covariantly through (8.15). A cheap way is to set the connection coefficients equal to zero in (8.15). This is basically the observation that the two equations are the same relative to Cartesian coordinates in a flat spacetime. Such a claim misses out some of the machinery of the connection construction that lies at the heart of non-Euclidean geometry, but this is not the place to expose all that. A justification of sorts can be found in [30, Sect. 2.5].

It is not totally obvious from what has just been said that SEP is absolutely necessary here and some authors would claim that it is not. This will be discussed further in the following (see in particular Sect. 8.7.4).

8.7.3 Equality of Inertial and Passive Gravitational Mass Revisited

Equation (8.15) is thus taken as the equation of motion of a point particle upon which no forces are acting, unless one counts gravity as a force. We observe that there is no mention of any parameters characterising the point particle. In particular there is no mention of its inertial mass. Of course there is no mention of parameters describing its inner make-up. After all, it is supposed to be a point particle. In this chapter, we are considering what would happen to a slightly spatially extended particle, i.e., with a world tube that intersects spatial hypersurfaces in a small region rather than a single mathematical point. This object might be spinning in some sense, or contain a charge distribution, for example. We shall return to this point in a moment, but let us begin with the disappearance of the inertial mass since this is directly relevant to the discussion.

So where did the inertial mass of the particle go? If we look back to (8.22), viz.,

$$m_a^{\text{inertial}}\left(\frac{\mathrm{d}^2 a^\mu}{\mathrm{d}s_a^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\frac{\mathrm{d}a^\rho}{\mathrm{d}s_a}\right) = e_a F^\mu{}_\nu \frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\ , \tag{8.24}$$

we find the inertial mass m_a^{inertial} multiplying the acceleration term in the equation to give a force on the right that is determined by an external field $F_{\mu\nu}$ and a coupling constant e_a characterising the particle. This is a typical equation of motion when there is some non-gravitational force (in this case electromagnetic) acting on the particle. Now in Newtonian gravitational theory, if the external field happens to be a gravitational potential Φ, one gets an equation of motion like this:

$$m_a^{\text{inertial}}\left(\frac{\mathrm{d}^2 a^\mu}{\mathrm{d}s_a^2} + \overline{\Gamma}^\mu_{\nu\rho}\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\frac{\mathrm{d}a^\rho}{\mathrm{d}s_a}\right) = m_a^{\text{pg}} h^{\mu\nu}\Phi_{,\nu}\ , \tag{8.25}$$

where m_a^{pg} is the passive gravitational mass of particle a, $(h^{\mu\nu}) = \mathrm{diag}\,(0,1,1,1)$, and $\overline{\Gamma}^{\mu}_{\nu\rho}$ is the connection appropriate to Newtonian spacetime and relative to whatever coordinates we have chosen to describe it. But due to the observed equality of inertial mass and passive gravitational mass, viz., $m_a^{\mathrm{pg}} = m_a^{\mathrm{inertial}}$, the coupling factor on the right-hand side is just the same factor as we have on the left-hand side. The only relevant characteristic of our point particle thus cancels out.

This explains how the inertial mass disappears from the equation, precisely because of the observed equality of inertial mass and passive gravitational mass, but how do we get rid of the gravitational potential we have just introduced? Of course, we can absorb it into the connection, following the much more detailed account of all this in [22]. We now have a new connection

$$\Gamma^{\mu}_{\nu\rho} := \overline{\Gamma}^{\mu}_{\nu\rho} + h^{\mu\sigma}\Phi_{,\sigma} t_\nu t_\rho \;, \tag{8.26}$$

where $(t_\mu) := (1,0,0,0)$, so that $t_\mu = \partial t/\partial x^\mu$. Equation (8.25) becomes

$$\frac{\mathrm{d}^2 a^\mu}{\mathrm{d}s_a^2} + \Gamma^{\mu}_{\nu\rho}\frac{\mathrm{d}a^\nu}{\mathrm{d}s_a}\frac{\mathrm{d}a^\rho}{\mathrm{d}s_a} = 0 \;, \tag{8.27}$$

still in this Newtonian context. So the equality of inertial and passive gravitational mass allows us to treat the trajectories of particles subjected only to gravitational effects as geodesics of a non-flat connection, because we do expect this new connection in (8.26) to be non-flat in general.

As Friedman says in [22], the equality of inertial and passive gravitational mass implies the existence of a connection Γ such that freely falling objects follow geodesics of Γ. This does not work for other types of interaction, where the ratio of $m_a := m_a^{\mathrm{inertial}}$ to the coupling factor, e.g., e_a for a charged particle, is not the same for all bodies. The worldlines of charged particles in an EM field cannot be construed as the geodesics of any single connection, because m_a/e_a in (8.22) varies from one particle to another.

Put another way, the equality of inertial and passive gravitational mass must be true if any theory of gravitation like general relativity, in which gravitational interaction is explained by the dependence of a non-flat connection on the distribution of matter, is to be possible. Note in passing that general relativity is not the only theory of this type. Classical gravitational theory can also be formulated in this way by taking advantage of the very same equivalence of inertial and passive gravitational mass. Friedman's book [22] is recommended for anyone who thinks that Newtonian gravitational theory cannot be given a fully covariant and totally geometric treatment. The essential difference with general relativity is that, in this treatment of Newtonian gravity, there is a flat connection $\overline{\Gamma}$ living alongside the non-flat connection Γ of (8.26). The deep fact here is that, in general relativity, the non-flat connection is the *only* connection of the spacetime.

So what of the criticism mentioned earlier, namely that the very possibility of a theory like GR implies the equality of inertial and passive gravitational mass? Might this not undermine the result of Sect. 8.7.1 which used GR and SEP to show that at least the self-force contributions to inertial mass are likely to equal the self-force contributions to passive gravitational mass? Put like this, we appear to be assuming the result in order to demonstrate it.

Looking back at Sect. 8.7.1, what we proposed was a new law (8.13), viz.,

$$\boxed{\sum_{\text{fields}} F_{\text{self}} + F_{\text{supp}} = 0}\ , \tag{8.28}$$

which would replace Newton's second law $\mathbf{F} = m\mathbf{a}$ and its direct extensions to GR with the help of SEP. Newton's second law in its usual form follows from (8.28) by analysing the self-forces into some multiple of the four-acceleration, and as mentioned earlier, the whole problem of the research program suggested in this chapter is to show that this is always possible, not just for EM forces, but for the other forces too, and then to show that there is no other mechanical mass. So a dynamical law, viz., (8.28), is still necessary here, but from it we can deduce results that were merely imposed previously, at least in the case where the inertia is entirely due to self-force effects.

To repeat the arguments at the end of Sect. 8.7.1, we understand physically why a supporting force is needed, namely to balance self-forces. In GR as it is usually presented, the supporting force is needed because the particle has non-zero four-acceleration, but we do not know why a non-zero four-acceleration should require a (supporting) force any more than we know why an acceleration should require a force in Newtonian physics.

Another point is that self-forces make a distinction between uniform velocities and changing velocities. The self-force is zero when the particle has a uniform velocity, and only becomes non-zero when the particle velocity is changing. So we understand from (8.28) why no force F_{supp} is required on the particle to keep it in free fall. And we understand the contrast between Newton's first and second laws, in the same way as the self-force idea explains this contrast in Newtonian physics.

And finally, although the equality of inertial and passive gravitational mass was crucial to the very existence of a theory like GR, we do have a mechanism here to explain why this should be the case.

8.7.4 Do Einstein's Equations Explain Inertia?

It turns out that the geodesic principle is not a principle at all, and neither is it likely to be any better than an approximation for a real particle that cannot be treated as a mathematical point. For the fact is that the geodesic principle follows from Einstein's equations in general relativity, provided that we also have SEP and *provided that we can make suitable assumptions about the particle*. Here follows a proof of sorts.

Recall first that Einstein's equations can be written

$$G_{\mu\nu} = -\kappa T_{\mu\nu}\ , \tag{8.29}$$

where

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{8.30}$$

is the Einstein tensor expressed in terms of the Ricci tensor $R_{\mu\nu}$ and curvature scalar R, κ is a constant that turns out to be expressible as

$$\kappa = \frac{8\pi G}{c^4} , \tag{8.31}$$

and $T_{\mu\nu}$ is the energy–momentum tensor expressing the distribution of mass and energy in the spacetime.

Now the covariant divergence of the Einstein tensor is zero in many circumstances, in particular when the torsion is zero. But the torsion is indeed often zero. In fact, it is sourced by the spin currents of matter in such a way that, in contrast to curvature, it does not propagate in spacetime, so it could only be nonzero in regions where there is matter or energy with some rotational property. A very clear, though somewhat sophisticated account of all this can be found in [25, Chap. 5].

Anyway, in a region where there is no spinning matter, Einstein's equation (8.29) implies that the covariant divergence of the energy–momentum tensor is zero. This is what we shall now use to derive the geodesic 'principle'. A more sophisticated approach is given in Sect. 8.7.5. In both cases, it is crucial to note that *non-innocent assumptions are made about the particle*, and in particular about the energy–momentum tensor that is taken to describe it.

We consider an almost-pointlike particle. So when almost-point particles are not acted upon by forces (apart from gravitational effects), we would like to show that their trajectories take the form

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}s^2} + \Gamma^\mu_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}s} \frac{\mathrm{d}x^\rho}{\mathrm{d}s} = 0 . \tag{8.32}$$

We consider a small blob of dustlike (i.e., zero pressure) matter with density ρ and velocity field

$$v^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}s} . \tag{8.33}$$

This equation expresses the fact that we view each component dust particle as having its own worldline $x^\mu(s)$. The energy–momentum tensor for this matter is then

$$T^{\mu\nu} = \rho \frac{\mathrm{d}x^\mu}{\mathrm{d}s} \frac{\mathrm{d}x^\nu}{\mathrm{d}s} \tag{8.34}$$

and we are saying that Einstein's field equation (8.29) implies that

$$T^{\mu\nu}{}_{;\nu} = 0 . \tag{8.35}$$

We analyse (8.35) by inserting (8.34) and the result is the geodesic equation (8.15). For completeness, here is the argument. We have

$$\rho_{,\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\frac{\mathrm{d}x^\nu}{\mathrm{d}s}+\rho\left(\frac{\partial}{\partial x^\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}s}+\Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}s}\right)\frac{\mathrm{d}x^\nu}{\mathrm{d}s}+\rho\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\left(\frac{\partial}{\partial x^\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}s}+\Gamma^\nu_{\nu\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}s}\right)=0\;. \tag{8.36}$$

If we did not have the idea of a velocity field v^μ, it would be difficult to interpret partial derivatives of $\mathrm{d}x^\mu/\mathrm{d}s$ with respect to the coordinates. But as things are, we can say

$$\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\frac{\partial}{\partial x^\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}s}=\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\frac{\partial v^\mu}{\partial x^\nu}=\frac{\mathrm{d}v^\mu}{\mathrm{d}s}=\frac{\mathrm{d}^2x^\mu}{\mathrm{d}s^2}\;. \tag{8.37}$$

The terms in the second bracket of (8.36) are

$$\frac{\partial v^\nu}{\partial x^\nu}+\Gamma^\nu_{\nu\rho}v^\rho=\mathrm{div}\,v\;, \tag{8.38}$$

and the whole thing can now be expressed by

$$\mathrm{div}(\rho v)\frac{\mathrm{d}x^\mu}{\mathrm{d}s}+\rho\left(\frac{\mathrm{d}^2x^\mu}{\mathrm{d}s^2}+\Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}s}\frac{\mathrm{d}x^\rho}{\mathrm{d}s}\right)=0\;. \tag{8.39}$$

By mass conservation,

$$\mathrm{div}(\rho v)=0\;, \tag{8.40}$$

and the result follows.

This proof purports to show that each constitutive particle of the blob follows a geodesic. But then we did not allow these particles to jostle one another. For example, we have zero pressure, as attested by the form of the energy–momentum tensor in (8.34). And we did not allow the particles to generate any torsion by revolving about the center of energy of the blob. And neither did we endow them with electric charge. It is in this sense that the geodesic ‘principle’ is in fact just an approximation, unless the test particle is not a blob, but a mathematical point. But more importantly, we still assume a certain form for the appropriate energy–momentum tensor, viz., (8.34). Why should this be the right thing?

Some argue that inertia is explained in general relativity, precisely because of the above proof or better variants of it (such as the one in Sect. 8.7.5). This point of view is expressed in the philosophical study by Brown [6, p. 141]. He claims that GR is the first in the long line of dynamical theories, based on the Aristotelian distinction between natural and forced motions of bodies, that *explains* inertial motion. This is not the view taken here, for reasons to be explained shortly. However, other issues discussed in Brown’s book, in particular what he refers to as the dynamical approach to spacetime structure, are exactly in line with the approach advocated in the present book, and in particular with the issues discussed in Bell’s paper [2] and extensions of those points to GR (see Chap. 6).

Concerning the supposed explanation of inertial motion according to Brown [6, Sect. 9.3], the idea is that inertia in GR is just as much a consequence of the gravitational field equations as gravitational waves, i.e., inertial motion of test particles is just part of the gravitational dynamics. Here is an argument against that view.

Recall the discussion just after (8.22) on p. 301. It was pointed out that actions like (8.19) and (8.21) are designed to give appropriate field equations, and that the appropriate field equation for the particle labelled by a is a geodesic equation, or an equation like (8.22). Now in GR, one adds a gravitational part to the action, viz.,

$$\mathscr{A}_{\text{grav}} := \frac{c^3}{16\pi G} \int R(-g)^{1/2} \mathrm{d}^4 x \,. \tag{8.41}$$

Some textbooks motivate this as follows. When the metric is varied in $\mathscr{A}_{\text{grav}}$, a constant multiple of the Einstein tensor pops out. The point about this is the observation that, when the metric is varied in an action like (8.21), the energy–momentum tensor $T_{\mu\nu}$ pops out. One gets a sum of contributions to this tensor from the matter as encapsulated in the action term

$$-\sum_a c m_a \int \mathrm{d}s_a \,, \tag{8.42}$$

and from the EM fields as encapsulated in the action term

$$-\frac{1}{16\pi c} \int F_{\mu\nu} F^{\mu\nu} (-g)^{1/2} \mathrm{d}^4 x \,.$$

Setting the variation of the full action with respect to the metric equal to zero, one then obtains the Einstein equations, with the Einstein tensor on one side and the total non-gravitational energy–momentum on the other side.

Now the covariant divergence of the Einstein tensor is zero (assuming zero torsion) and this could in principle be worrying, because the Einstein equation then implies that the covariant divergence of the total energy–momentum is zero. It is interesting at this point to note that there is a general result according to which the energy–momentum tensor derived from a matter action of the form

$$\int L(-g)^{1/2} \mathrm{d}^4 x \tag{8.43}$$

by varying the metric always has zero covariant divergence on shell in the torsion-free case when L is a scalar, as explained very clearly in [14, Chap. 9] (see below). The expression 'on shell' means 'when the matter field equations are satisfied'. Another version of this theorem in [25, Sect. 6.5] shows that invariance of the matter action under the group of diffeomorphisms is sufficient to guarantee zero covariant divergence of the corresponding energy–momentum tensor on shell if the torsion is zero. As mentioned above, if the torsion is not zero, the covariant divergence of the Einstein tensor is not zero either. This case is not considered here.

Of course, the action $\mathscr{A}$ in (8.42) does not have the form (8.43), but one expects some general theorem to ensure that the resulting energy–momentum tensor will have zero covariant divergence on shell, i.e., when the field equations, that is, the geodesic equations, are satisfied. So here we have a very general *converse* of the idea that covariant conservation of the energy–momentum tensor implies the geodesic principle, a converse that only requires the matter action to be coordinate independent. We shall return to this general idea in Sect. 8.7.5.

The main point we would like to make is that the geodesic equations, and their variants with a force on one side, are built in by construction of the action. It is no surprise therefore that they should pop out again when we set the covariant divergence of the energy–momentum tensor equal to zero. Perhaps one should be more suspicious of arguments from actions. They are neat, and bring a level of unity in the sense that one can derive several dynamical equations from the same action by varying different items. On the other hand, we are only getting out what we put in somewhere else.

Actions like (8.16) at the beginning of Sect. 8.7.2, or (8.42) above, are a case in point. We saw that (8.16) was in fact designed to deliver the geodesic principle when extremised under variation of the worldline, so we have already fed in the idea that freely moving test particles follow (roughly) straight lines at constant velocity in (locally) inertial frames. Hence the conclusion at the end of Sect. 8.7.2 that the physical motivation for the geodesic principle must still pass by an application of the strong equivalence principle and the hypothesis, still a necessary assumption of the theory, that any particle not subject to forces in a flat spacetime will move with constant velocity.

A recent commentator [10] asserts that the motion of massive test particles is independent of SEP. This refers to the above idea that geodesic motion follows by conservation of energy–momentum, which in turn follows from Einstein's equations, whence the inertia of massive objects is supposed to be explained by the theory. So here we are arguing against both conclusions:

(i) independence from SEP, and
(ii) insofar as geodesic motion is a consequence of Einstein's equations, the claim that this explains inertia.

Note, however, that the paper [10] is recommended for its clear account of the idea advocated by Brown, and also in this book, that the metric tensor in relativity theories gets its geometric significance through detailed physical arguments.

Regarding (ii), the unwarranted claim that inertial motion drops out of Einstein's theory as a consequence of Einstein's equations for the gravitational field without further assumptions, we have just seen a counter-argument. When we wanted to deduce geodesic motion from Einstein's equations, we decreed that the energy–momentum tensor of the test particle was that of a very small cloud of dust, then reasoned heuristically and took a point-particle limit at the end of a short calculation. Further, we assumed that the component matter was not spinning in any way, and that the component particles carried no electric charge. Put like that, one sees how limited the proof is. No real particle with spatial extent could be like this, and indeed, no real particle would actually free fall along a geodesic, and nor even would its center of energy (see, for example, [14, Appendix A] for the analysis of spinning test particles and [13] for the case of a spatially extended charged particle interacting with its own fields). All this shows clearly that the motion of (real) test particles is indeed part of the dynamics of the relevant non-gravitational fields, *provided that we* *make sufficient assumptions about the nature of the test particles and provided that we have a principle like SEP to formulate those assumptions in the curved spacetime context*.

The point is then that one still needs to explain the choice of energy–momentum tensor. A more sophisticated proof of the geodesic principle from Einstein's equations using distributions will be sketched in Sect. 8.7.5. There one derives the energy–momentum tensor (distribution) for a point particle by varying the metric in the usual action $S = -m\int \mathrm{d}\tau$ for a point particle in relativity theory. One then shows that conservation implies the geodesic equation. But it would be a circular argument to say that this proves that the geodesic equation follows from Einstein's equation, because the action $S = -m\int \mathrm{d}\tau$ is designed to deliver the geodesic equation when one varies the particle worldline of which it is a functional. Once again, since the action S is invariant under coordinate changes and we assume zero torsion, this alone implies that the covariant divergence of the energy–momentum tensor derived from it by varying the metric will be zero if the particle follows a geodesic. But of course, as mentioned above, the action is designed to yield geodesic motion.

So far we have focused on massive test particles, but it is very instructive to ask why free photons should follow null geodesics. There is clearly more input here than just Einstein's equations for the gravitational field. We need to apply SEP in order to ship the flat spacetime situation into the locally inertial frames of curved spacetimes. But the flat spacetime situation can only be had by assuming Maxwell's equations. In other words, we do require the minimal extension of Maxwell's equations (MEME) to the curved spacetime in order to get the 'geodesic principle' for photons.

Interestingly, and probably significantly, the above discussion of material test particles assumes that there is no torsion, because it uses the Levi-Civita connection. But torsion is generated by spinning matter as mentioned above [25, Chap. 5]. This brings us to the claim (i) above that SEP would not be needed to show that test particles follow geodesics.

In the above demonstration of the geodesic principle, we require the particle to be moving in a region of spacetime where the torsion is zero, because this is a sufficient condition for the covariant divergence of the Einstein tensor to be zero. One often forgets torsion outright and just decrees the connection coefficients to be symmetric in their two lower indices. However, it is interesting to draw attention to torsion here since it is precisely spinning matter that generates torsion. As torsion does not propagate beyond its sources, one only needs to assume that the test particle (blob) moves in empty spacetime. But what if the blob is itself spinning?

Now it is known that a spinning blob of matter will not free fall along a geodesic. The spin angular momentum of the blob couples with the curvature and tweaks it off the geodesic [14, Appendix A]. How does one show this? One begins with a Lagrangian which treats the blob as an ensemble of particles, then expands everything about the center of energy. The best thing would be to include all the electromagnetic forces holding the particles together, but fortunately one can just make a quasi-rigidity assumption and go from there. The latter assumption avoids talking about, but nevertheless embodies, non-gravitational forces. It thus assumes, in a very hidden way admittedly, the strong principle of equivalence.

It is the center of energy of the blob that approximately (but not quite) follows a geodesic. It seems a remarkable achievement just to get this. But it is not so remarkable, because that Lagrangian mentioned in the last paragraph is precisely the one that is designed to deliver geodesic motion for the constituent particles of the blob, were they not constrained by quasi-rigidity.

In fact, Butterfield does specify that he is talking about non-rotating test particles in [10]. But the point remains that one really must ask what is meant by a test particle. It is supposed to be a mathematical point, but that is an approximation. And the fact is that all test particles are going to involve non-gravitational forces. Of course, it is precisely for the blob of dust that one gets one of the derivations, taking a limit in the end as the size of the blob goes to zero. On the other hand, any realistic test particle (even with a limit at the end) is going to involve non-gravitational forces, and will in general be spinning (not spinning would be a very special and improbable case). Even if not spinning, the general Lagrangian approach including EM forces is going to predict deviations from the geodesic.

Note once again that photons have to follow null curves because of Maxwell's equations [12, Chap. 7], and this brings in a need for SEP. Now it would be a strange thing in a way if SEP were required to show that photons have to follow null curves, but massive particles could get away without having to obey any vestige of the laws of any of the forces governing their make-up, and hence avoid any need for SEP. Put another way, if we say that there are no non-gravitational effects to be taken into account when considering our test particle, then since SEP deals only in non-gravitational effects, it cannot be needed to say *anything* about the test particle. This is pure logic. There is no physics at all in it.

8.7.5 Geodesic Principle from Einstein's Equations: Another 'Proof'

The argument to say that Einstein's equations explain inertia uses the idea that the geodesic principle is implied by Einstein's equations [6, Sect. 9.3]. But we need to postulate an energy–momentum tensor for the test particle. According to Einstein's equations, this is proportional to the Einstein tensor which has zero covariant divergence in the torsion-free case, so the energy–momentum tensor for the test particle must also have zero covariant divergence. Writing down the latter equation for the right energy–momentum tensor, the geodesic equation drops out for the test particle.

But what is the right energy–momentum tensor for the test particle? It is often derived from an action by varying the metric and applying a variational principle, so we have to ask where that action came from. We then find that the action was actually *designed* to give the geodesic principle anyway, by varying the particle worldline and applying another variational principle. That would make such a proof circular.

This will be illustrated here by the following approach adapted from results in [14]. For a point particle, it is commonplace to take the action functional to be

$$S = -m \int \mathrm{d}\tau \,. \tag{8.44}$$

We vary the metric and apply a variational principle in order to derive an energy–momentum tensor $T^{\mu\nu}$ from S, involving some distributions. Covariant conservation of $T^{\mu\nu}$ then implies the geodesic equation. But S was designed to deliver the geodesic equation when we vary the particle worldline and apply another variational principle.

The action functional of the free particle is given by (8.44). To obtain the functional derivative of this action with respect to the metric tensor $g_{\mu\nu}$, we must subject $g_{\mu\nu}$ to a variation $\delta g_{\mu\nu}$. If the worldline of the particle does not intersect the support of $\delta g_{\mu\nu}$, the action will remain unaffected. It is evident therefore that the functional derivative is going to involve a delta function $\delta(x,z)$ having as arguments the point x where the derivative is being taken and the location $z^{\alpha}(\lambda)$ of the particle. We use indices from the first part of the Greek alphabet to denote tensors taken at the point $z^{\alpha}(\lambda)$ and from the middle of the alphabet to denote tensors taken at the point x^{μ}. With this convention we may employ the abbreviations

$$g_{\mu\nu} = g_{\mu\nu}(x) \,, \qquad g_{\alpha\beta} = g_{\alpha\beta}\big(z(\lambda)\big) \,.$$

We also need the identity

$$\frac{\delta g_{\alpha\beta}}{\delta g_{\mu\nu}} = \delta^{\mu\nu}{}_{\alpha\beta} \,, \tag{8.45}$$

where

$$\delta^{\mu\nu}{}_{\alpha\beta} := \frac{1}{2}(\delta^{\mu}{}_{\sigma}\delta^{\nu}{}_{\tau} + \delta^{\nu}{}_{\sigma}\delta^{\mu}{}_{\tau})\delta(x,z)\Big|_{\sigma=\alpha,\tau=\beta} \,.$$

$\delta^{\mu\nu}{}_{\alpha\beta}$ is a bitensor density, of unit weight at the point x and zero weight at the point z. We show below that it satisfies

$$\delta^{\mu\nu}{}_{\alpha\beta;\nu} = -\frac{1}{2}(\delta^{\mu}{}_{\alpha;\beta} + \delta^{\mu}{}_{\beta;\alpha}) \,, \qquad \delta^{\mu}{}_{\alpha} := \delta^{\mu}{}_{\nu}\delta(x,z)\Big|_{\nu=\alpha} \,, \tag{8.46}$$

as may be verified by passing to a coordinate system in which the derivatives of $g_{\mu\nu}$ vanish at x. We record here two other properties of the functional derivative:

- Functional differentiation is commutative (like ordinary differentiation).
- Functional differentiation commutes with ordinary differentiation with respect to coordinates or worldline parameter λ. (It does *not* commute with covariant differentiation!)

Regarding the definition of $\delta^{\mu\nu}{}_{\alpha\beta}$, it seems reasonable to ask why we do not have

$$\frac{\delta g_{\alpha\beta}}{\delta g_{\mu\nu}} = \delta^{\mu}{}_{\sigma}\delta^{\nu}{}_{\tau}\delta(x,z)\Big|_{\sigma=\alpha,\tau=\beta} \,.$$

The problem is that there is a constraint on the functions $g_{\alpha\beta}$, viz., $g_{\alpha\beta} = g_{\beta\alpha}$. Indeed,

$$g_{\alpha\beta} = \frac{1}{2}(g_{\alpha\beta} + g_{\beta\alpha}) \ ,$$

and applying $\delta/\delta g_{\mu\nu}$ to this, we obtain the given result. Put another way, we expect $\delta g_{\alpha\beta}/\delta g_{\mu\nu}$ to be symmetric in α, β or in μ, ν, and the given result is obtained by symmetrising the above proposal.

However, the result (8.45) says that

$$\frac{\delta g_{\alpha=0,\beta=1}}{\delta g_{\mu=0,\nu=1}} = \frac{1}{2}\delta(x,z) \ ,$$

whereas one normally defines

$$\frac{\delta f(z)}{\delta f(x)} = \delta(x,z) \ .$$

The point here is that we can only account in this way for the interdependence (symmetry) of the $g_{\alpha\beta}$ if we insist on a sum over all indices, i.e., the given formula is only right when we sum over indices. Note that in the application (8.47), we do indeed sum over all α and β.

Another point here is the justification of the covariant derivative result in (8.46). By definition, the covariant derivative at x of an object with two contravariant indices at x and weight 1 at x is

$$\delta^{\mu\nu}{}_{\alpha\beta;\nu} := \delta^{\mu\nu}{}_{\alpha\beta,\nu} + \Gamma^{\mu}_{\rho\nu}\delta^{\rho\nu}{}_{\alpha\beta} + \Gamma^{\nu}_{\rho\nu}\delta^{\mu\rho}{}_{\alpha\beta} - \Gamma^{\rho}_{\rho\nu}\delta^{\mu\nu}{}_{\alpha\beta} \ ,$$

the last term being due to the weight at x. The last two terms cancel. Now

$$\delta^{\mu}{}_{\alpha} := \delta^{\mu}{}_{\nu}\delta(x,z)\Big|_{\nu=\alpha}$$

has weight 1 at x, hence weight 0 at z, so the covariant derivative at z is

$$\delta^{\mu}{}_{\alpha;\beta} := \delta^{\mu}{}_{\alpha,\beta} - \Gamma^{\gamma}_{\alpha\beta}\delta^{\mu}{}_{\gamma} \ .$$

(Here we use the convention for $\delta(x,z)$. One can attribute some weight at x and the rest at z, so that the total weight is 1, as required for this distribution.) Hence,

$$-\frac{1}{2}(\delta^{\mu}{}_{\alpha;\beta} + \delta^{\mu}{}_{\beta;\alpha}) = -\frac{1}{2}(\delta^{\mu}{}_{\alpha,\beta} + \delta^{\mu}{}_{\beta,\alpha}) + \Gamma^{\gamma}_{\alpha\beta}\delta^{\mu}{}_{\gamma} \ .$$

In a frame where

$$\Gamma^{\mu}_{\nu\rho}\Big|_{x} = 0 \ ,$$

we only have to show that

$$\delta^{\mu\nu}{}_{\alpha\beta,\nu} = -\frac{1}{2}(\delta^{\mu}{}_{\alpha,\beta} + \delta^{\mu}{}_{\beta,\alpha}) \ .$$

This amounts to showing that

$$(\delta^{\mu}{}_{\sigma}\delta^{\nu}{}_{\tau}+\delta^{\nu}{}_{\sigma}\delta^{\mu}{}_{\tau})\delta(x,z)_{,\nu}\Big|_{\sigma=\alpha,\tau=\beta} = -\delta^{\mu}{}_{\sigma}\delta(x,z)_{,\beta}\Big|_{\sigma=\alpha} - \delta^{\mu}{}_{\tau}\delta(x,z)_{,\alpha}\Big|_{\tau=\beta}$$
$$= \big[\delta^{\mu}{}_{\sigma}\delta(x,z)_{,\tau}+\delta^{\mu}{}_{\tau}\delta(x,z)_{,\sigma}\big]\Big|_{\sigma=\alpha,\tau=\beta} ,$$

which is true. Of course, we can also check that

$$\Gamma^{\mu}_{\rho\nu}\delta^{\rho\nu}{}_{\alpha\beta} = \Gamma^{\gamma}_{\alpha\beta}\delta^{\mu}{}_{\gamma} .$$

This says that

$$\frac{1}{2}\Gamma^{\mu}_{\rho\nu}(\delta^{\rho}{}_{\sigma}\delta^{\nu}{}_{\tau}+\delta^{\nu}{}_{\sigma}\delta^{\rho}{}_{\tau})\delta(x,z)\Big|_{\sigma=\alpha,\tau=\beta} = \delta^{\mu}{}_{\sigma}\Gamma^{\gamma}_{\alpha\beta}\delta(x,z)\Big|_{\sigma=\gamma} ,$$

and the left-hand side is

$$\frac{1}{2}(\Gamma^{\mu}_{\sigma\tau}+\Gamma^{\mu}_{\tau\sigma})\delta(x,z)\Big|_{\sigma=\alpha,\tau=\beta} ,$$

as required.

Writing the particle worldline $z^{\alpha}(\lambda)$ for some parameter λ, the action (8.44) is

$$S = -m\int\big[-g_{\alpha\beta}(z)\dot{z}^{\alpha}\dot{z}^{\beta}\big]^{1/2}\mathrm{d}\lambda .$$

The computation of the energy–momentum density for the free particle is now elementary. We find

$$\begin{aligned} T^{\mu\nu} &= 2\frac{\delta S}{\delta g_{\mu\nu}} = m\int\delta^{\mu\nu}{}_{\alpha\beta}\dot{z}^{\alpha}\dot{z}^{\beta}(-\dot{z}^2)^{-1/2}\mathrm{d}\lambda \\ &\xrightarrow[\lambda\to\tau]{} m\int\delta^{\mu\nu}{}_{\alpha\beta}\dot{z}^{\alpha}\dot{z}^{\beta}\mathrm{d}\tau = \int\delta^{\mu\nu}{}_{\alpha\beta}p^{\alpha}u^{\beta}\mathrm{d}\tau . \end{aligned} \tag{8.47}$$

Look at the special form this expression takes in canonical coordinates in flat spacetime (in flat spacetime, the worldline is of course straight, but we make no use of this at this point):

$$\begin{aligned} T^{\mu\nu} &= \int p^{\mu}(\tau)u^{\nu}(\tau)\delta\big(x-z(\tau)\big)\mathrm{d}\tau \\ &= \int p^{\mu}(\tau)u^{\nu}(\tau)\delta\big(x-z(\tau)\big)\delta\big(x^0-z^0(\tau)\big)\frac{\mathrm{d}z^0(\tau)}{\dot{z}^0(\tau)} \\ &= \delta(x-z)p^{\mu}\,\frac{u^{\nu}}{u^0}\Big|_{z^0(\tau)=x^0} , \end{aligned}$$

so that

$$T^{\mu 0} = \delta(x-z)p^{\mu} \,, \qquad T^{\mu i} = \delta(x-z)p^{\mu}v^{i} \,, \qquad v^{i} = \frac{u^{i}}{u^{0}} \,.$$

The three-dimensional delta function appearing in these last equations displays the pointlike character of the particle. T^{00} is clearly the particle's energy density: all the energy p^0 is located where the particle is! T^{i0} is just as clearly the momentum density. However, if one remembers the relativistic relation $p^i = p^0 v^i$ between momentum and energy, one can alternatively regard momentum as a *rate of transport of energy*. This permits T^{i0} or T^{0i} to be interpreted also as a rate of flow of energy per unit area or *energy flux density*. In a similar vein, T^{ij} is to be regarded as a *momentum flux density*.

Returning to the curved spacetime case, we have

$$\begin{aligned} T^{\mu\nu}{}_{;\nu} &= \int \delta^{\mu\nu}{}_{\alpha\beta;\nu} p^{\alpha} u^{\beta} \mathrm{d}\tau = -\int \delta^{\mu}{}_{\alpha;\beta} p^{\alpha} \dot{z}^{\beta} \mathrm{d}\tau \\ &= -\int \dot{\delta}^{\mu}{}_{\alpha} p^{\alpha} \mathrm{d}\tau = \int \delta^{\mu}{}_{\alpha} \dot{p}^{\alpha} \mathrm{d}\tau \,, \end{aligned} \tag{8.48}$$

where we have used (8.46). Now let A_μ be an arbitrary covariant vector of compact support. Multiply both sides of this equation by A_μ and integrate over spacetime. If $T^{\mu\nu}{}_{;\nu} = 0$, one gets

$$\int A_{\alpha} \dot{p}^{\alpha} \mathrm{d}\tau = 0 \,.$$

Because A_α is arbitrary this implies $\dot{p}^{\alpha} = 0$ (the geodesic equation).

This is a sophisticated version of the result that Einstein's field equations imply that particles will follow geodesics. What exactly is the logic here? $T^{\mu\nu}$ for the free particle was derived from the action (8.44), so the assumption here is that this is the right action for a free particle. Next, if Einstein's field equations are satisfied for this system, the contracted Bianchi identity (which always holds for zero torsion) implies that $T^{\mu\nu}{}_{;\nu} = 0$, and as we have just seen, this implies that the particle follows a geodesic.

This brings us to our criticism of the claim that GR implies inertia, i.e., that this theory forces free particles to follow geodesics. We clearly have to assume that (8.44) is the action for this system, but this is tantamount to assuming that the free particle will follow a geodesic, as one finds by varying the worldline in the action. Indeed, this action is designed so that, when one varies the worldline, the geodesic equation drops out, as shown in the standard textbook demonstration at the beginning of Sect. 8.7.2.

The action S in (8.44) is invariant under coordinate changes. As mentioned earlier, there is a theorem which says that, in a torsion-free spacetime, the coordinate invariance of the matter action alone is enough to imply that the covariant divergence of an energy–momentum tensor derived from it by varying the metric will be zero if the matter satisfies the related field equations, which for the action in (8.44) means that the particle follows a geodesic. As an aside to the above discussion, let us see the more general picture painted by DeWitt in [14, Chap. 9] (for a variant see, e.g., [25, Sect. 6.5]). We assume zero torsion.

Consider an action functional S_M for some matter in a torsion-free spacetime and assume that this functional is coordinate invariant. Therefore if $\delta g_{\mu\nu}$ and $\delta\Phi^\mathrm{A}$ are the changes induced in the metric tensor and the matter dynamical variables by an infinitesimal coordinate transformation, we must have

$$0 \equiv \int \frac{\delta S_\mathrm{M}}{\delta g_{\mu\nu}} \delta g_{\mu\nu} \mathrm{d}^4 x + \frac{\delta S_\mathrm{M}}{\delta \Phi^A} \delta \Phi^A \ ,$$

with implicit summation or integration over the index A. When the matter dynamical equations are satisfied (we say that the matter is on shell), the second term is zero. Therefore, writing [14, Sect. 4.4]

$$\begin{aligned}\delta g_{\mu\nu} &= -L_{\delta\xi} g_{\mu\nu} \\ &= -g_{\mu\nu,\sigma}\delta\xi^\sigma - g_{\sigma\nu}\delta\xi^\sigma{}_{,\mu} - g_{\mu\sigma}\delta\xi^\sigma{}_{,\nu} \\ &= -\delta\xi_{\mu;\nu} - \delta\xi_{\nu;\mu} \ ,\end{aligned}$$

for some infinitesimal coordinate transformation expressed by the contravector field $\delta\xi^\mu$, assumed to have compact support, with $\delta\xi_\mu := g_{\mu\nu}\delta\xi^\nu$, and carrying out an integration by parts, we have

$$\begin{aligned}0 &= -\int \frac{\delta S_\mathrm{M}}{\delta g_{\mu\nu}} \left(\delta\xi_{\mu;\nu} + \delta\xi_{\nu;\mu}\right) \mathrm{d}^4 x \\ &= -\int T^{\mu\nu} \delta\xi_{\mu;\nu} \mathrm{d}^4 x = \int T^{\mu\nu}{}_{;\nu} \delta\xi_\mu \mathrm{d}^4 x \ .\end{aligned}$$

Because $\delta\xi_\mu$ is arbitrary, we have

$$T^{\mu\nu}{}_{;\nu} = 0 \ , \tag{8.49}$$

whenever the matter dynamical equations are satisfied.

Equation (8.49) generally holds only when the matter dynamical equations are satisfied. As DeWitt points out, in many cases, it is completely equivalent to the matter dynamical equations and can be used in place of them. This is exemplified by the case of the free particle in the above derivation from (8.48), which showed that covariant conservation implied the geodesic equation.

Although $T^{\mu\nu}$ has zero covariant divergence, this is not a true conservation law, because $T^{\mu\nu}$ accounts only for the energy and momentum of the matter. When a gravitational field is present (i.e., when spacetime is not flat), it can exchange energy and momentum with the matter. One might ask whether the energy and momentum of the gravitational field could be accounted for by treating the gravitational action functional S_G in the same way. It too is coordinate independent and hence satisfies

$$\left(\frac{\delta S_\mathrm{G}}{\delta g_{\mu\nu}}\right)_{;\nu} = 0 \ .$$

Interestingly, this relation is an identity that holds whether or not the gravitational field equations are satisfied. Its explicit form is [14, Chap. 8]

$$0 = -\frac{1}{16\pi G}\left[(-g)^{1/2}\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)\right]_{;\nu} .$$

The factor $(-g)^{1/2}$ drops out right away, since $(-g)^{1/2}$ has weight 1, whence

$$\left[(-g)^{1/2}\right]_{;\mu} := \left[(-g)^{1/2}\right]_{,\mu} - \Gamma^{\nu}_{\nu\mu}(-g)^{1/2} ,$$

and it is a standard result that

$$\Gamma^{\nu}_{\nu\mu} = (-g)^{-1/2}\left[(-g)^{1/2}\right]_{,\mu} ,$$

so

$$\left[(-g)^{1/2}\right]_{;\mu} = 0 .$$

We thus have

$$0 = \left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu} .$$

This is known as the contracted Bianchi identity, and the object in brackets is the Einstein tensor.

To show that the identity

$$0 \equiv \left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu}$$

can be obtained by contracting the Bianchi identity twice, recall that, for zero torsion, this identity is [14, Sect. 4.4]

$$R_{\tau\sigma\mu\nu;\rho} + R_{\tau\sigma\nu\rho;\mu} + R_{\tau\sigma\rho\mu;\nu} = 0 ,$$

and consider

$$\begin{aligned} 0 &= R^{\mu\nu}{}_{\mu\nu;\sigma} + R^{\mu\nu}{}_{\nu\sigma;\mu} + R^{\mu\nu}{}_{\sigma\mu;\nu} \\ &= R_{;\sigma} - R^{\mu}{}_{\sigma;\mu} - R^{\nu}{}_{\sigma;\nu} = -2g_{\sigma\mu}\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\nu} . \end{aligned}$$

The contracted Bianchi identity imposes no constraint on the gravitational field. It does, however, impose a constraint on the matter through the Einstein equations, forcing $T^{\mu\nu}$ to have zero covariant divergence. So in many cases, explicitly those where zero covariant divergence of the energy–momentum tensor implies the matter dynamical equations, the latter are superfluous, because the Einstein equations are sufficient.

But just to reiterate the point we would like to make here, the success of this theory does depend on our having the right energy–momentum tensor, or indeed the right matter action S_{M}, to describe the matter, and when we are dealing with the action (8.44) for a point particle on p. 311, we have to remember that this was chosen explictly so that the freely falling particle would follow a geodesic.

Finally, as DeWitt reminds us, it is nice to know that the equations obtained from the complete variational principle are at least consistent. They would *not* be consistent if S_{M} were not coordinate independent.

8.7.6 A Brief Conclusion

When we make suitable assumptions about the energy–momentum tensor used to describe a massive test particle, we can deduce from covariant conservation of this energy–momentum tensor that the particle will closely follow a timelike geodesic when subject only to gravitational effects, and of course, covariant conservation of this tensor is required by Einstein's equations for the gravitational field in a torsion-free region of spacetime. We do need to know what constitutes an appropriate energy–momentum tensor describing the test particle, or at least put some constraints on it. As Malament puts it [33], assumptions other than covariant conservation are required to obtain a geodesic worldline, i.e., we need to put more into our premisses in order to get the geodesic principle out.

In sophisticated arguments [33] modelling the particle as a limit of ever smaller bodies of matter, this so-called geodesic principle can be derived by making some extremely general assumption about the way energy flows within the body, viz., that its flow curves are timelike. This is referred to as an energy condition on the matter fields constituting the particle. Although this is intriguing in itself, it confirms the idea stressed throughout the above discussion that we do effectively feed in the notion of inertial motion somewhere, albeit possibly in a disguised way, when we apply variational or other methods. For example, an explicit energy–momentum tensor is assumed for the argument in Sect. 8.7.4 [see (8.34) on p. 305] and an explicit action is assumed for the argument in Sect. 8.7.5 [see (8.44) on p. 311]. In the latter case, that action was designed to deliver the geodesic principle by extremisation under variation of the worldline.

We have also noted that, when the internal makeup of the 'freely falling' particle is taken seriously, e.g., taking spin into account [14, Appendix A], or the interaction of a spatially extended charge distribution with its own fields [13], dynamical calculations show that it will be tweaked off its geodesic, clearly revealing the role of the strong equivalence principle even in cases of free fall. Finally, we note that photons follow null geodesics, not by some mysterious consequence of Einstein's equations, but because this is what is dictated by Maxwell's theory in flat spacetime and its minimal extension to curved spacetimes.

So these theorems all assume something about inertial motion in flat spacetime and carry the import of those same assumptions over to curved spacetimes via the strong equivalence principle. There is no deeper understanding of inertial motion than we had before, although it has to be said that the strong equivalence principle is a bold and sweeping hypothesis, and may of course turn out to be wrong, or need to be implemented in a non-minimal way. (This means that, in addition to replacing coordinate derivatives by covariant derivatives with respect to the metric connection, we may also need to include terms involving the curvature directly in the field equations of whatever non-gravitational phenomena we are describing.) But does the discussion in Sect. 8.7.1 really provide a deeper understanding of inertial motion?

There is an extremely elegant philosophical paper by Nerlich [42] which is worth the detour here. It claims that, if one can take it as a real entity, spacetime with accompanying metric structure does in fact explain why particles should follow geodesics when not affected by electromagnetic fields, collisions, and so on. However, the explanation is not a causal one, rather a geometrical one, since as Galileo pointed out, no cause is needed for this motion! To quote Nerlich:

> Gravity makes no sense as action across a distance by some massive things on others. It is not a force, not a cause. GR makes sense only as a local theory: it demands proximal explanation. In lots of pure gravitation situations, the *only* proximal feature available to explain anything is local spacetime structure. But surely it cannot explain matter's motion by causing it. So a style of *geometrical* explanation both local and acausal surely looks worth at least consideration.

This is strongly recommended reading.

Section 8.7.1 presents what is apparently a rather different line of thought concerning the reason why particles should resist being pushed off their geodesic, whatever agent may be the cause of that pushing. It concerns the force that a spatially extended particle must exert upon itself in that context, whenever the particle contains a continuous spatial distribution of something that is the source for any kind of force field. In GR, this self-force effect could at least explain why (by an explicit mechanism) the self-force contribution to the passive gravitational mass should be exactly equal to the self-force contribution to the inertial mass.

The argument appeals to the strong equivalence principle, and the strong equivalence principle is at least partly inspired by the idea that passive gravitational and inertial mass are exactly equal anyway, so one might just accuse this argument of circularity. The claim here is that it does nevertheless provide an insight. On the other hand, we still appeal to a version of Newton's second law which says that the total force on any object is zero [see (8.28) on p. 304]. Nerlich points out that forces are usually thought of as pulling on masses [43]. So what do the forces pull on now? The idea here must be that forces pull on each other, and in such a way that they always exactly cancel.

The reader must make his or her own decision about what is the better ontology: a real spacetime capable of geometrical explanation or dynamical effects that always act to cancel one another. But one could argue within the self-force picture that, through the metric, the geometry tells us when there is no aceleration and hence no self-force that needs to be overcome in order to maintain such motion.

Chapter 9
Dynamical Explanations for Relativistic Effects

The strong focus of modern theoretical physics on geometrical pictures and group theory sometimes makes us lose sight of the real physics that underlies phenomena like FitzGerald contraction and time dilation, the velocity dependence of particle mass, and the effect of binding energy on the inertial masses of bound state particles. Although geometry and group theory are essential tools of physics today, we should not forget their physical origins. Let us try to readjust the balance in this section, and break free from what one might describe as the Minkowski straightjacket.

9.1 Introduction

We discuss here four physical phenomena: FitzGerald contraction and time dilation (Sect. 9.2), the velocity dependence of particle mass (Sect. 9.3), and the effect of binding energy on the inertial masses of bound state particles (Sect. 9.4). Each of these effects is generally presented without physical explanation today:

- **FitzGerald Contraction and Time Dilation.** When a measuring stick is accelerated along its axis relative to some inertial frame, something happens to its length. In fact, it gets shorter. But in Minkowski's geometrical picture, a measuring stick is a 4D region of spacetime, and it is often intimated that the relativistic contraction effect is not a real effect at all, but merely some kind of illusion, depending on which spacelike hypersurface an inertially moving observer should use to define its length. Regarding time dilation, we learn that moving clocks run slow, but we should also be allowed to ask why they do so.
- **Velocity Dependence of Particle Mass.** It is common knowledge in relativity theory that the resistance of a particle to acceleration will increase as it moves faster. When at rest, this resistance will be measured by its rest mass m_0, but when moving at speed v, it will be given by a speed-dependent function

$$m(v) = m_0\gamma(v) = \frac{m_0}{\sqrt{1-v^2/c^2}}\ . \tag{9.1}$$

No explanation is usually offered up for this, except that it plays a role in the successful extension of Newton's second law to a Lorentz covariant version.

- **Inertial Consequences of Binding Energy.** When we consider a bound state particle like the proton in modern particle physics, comprising two up quarks and a down quark according to current theory, its inertial mass is taken to be made up of several components: the intrinsic inertial masses of the quarks (insofar as they can be ascertained in the difficult context of quark confinement), the kinetic energies of the quarks, and the strong, electromagnetic, or other binding energies. The usual explanation for including binding energy is simply the claim that energy and mass are equivalent!

The aim here is to show that each of these effects has a dynamical explanation. In each case, the argument makes crucial reference to acceleration, and so deserves its place in this discussion.

9.2 FitzGerald Contraction and Time Dilation

Suppose a measuring rod starts out at rest in some inertial frame, where its length is measured to be L, but later is moving at speed v along its axis relative to that same frame. Assuming it has fully recovered from the acceleration process, it will now be found to have a length $L/\gamma(v)$, where

$$\gamma(v) := \frac{1}{\sqrt{1 - v^2/c^2}} .$$

During the acceleration, it is not at all obvious what its length might be at any instant of time. That would depend on how it was accelerated, and also on its material constitution. And neither is it obvious that it would ever fully recover from the acceleration process. The worry is that it might be made of something like putty, and we might have got it moving by hitting the back of it hard with a baseball bat. So for sure it would be shorter, but not just by the generally tiny factor mentioned above.

In that sense, the measuring rod seems to be a bad example for the above thought experiment, except that by definition it is supposed to be good for measuring lengths, whatever may be happening to it. So a better idea would be to replace the measuring rod by a hydrogen atom and get it moving by accelerating the nucleus, as we discussed in Chap. 6. In this case, we can make a model for the spatial, and indeed temporal, characteristics of the atom using Maxwell's theory of electromagnetism [2]. The spatiotemporal characteristics are taken as specified by the diameter D and period P of the electron orbit.

If the atom is accelerated in such a way that it is not broken apart, and indeed if it is accelerated gently enough for the electron to get plenty of orbits in before the velocity has changed by much, one finds that the diameter of the orbit is always instantaneously approximately equal to the FitzGerald contracted length $D/\gamma(v)$ for the instantaneous speed v of the nucleus, and its period is always instantaneously approximately increased to $P\gamma(v)$ for that instantaneous speed. But this means that one could justify using the electron orbit of the atom as a clock or ruler for measuring the times and lengths in an instantaneously comoving inertial frame, as they are usually predicted by the Minkowski metric.

We still need to constrain the acceleration in this story, but in a way that is perfectly comprehensible to any practising physicist. If the acceleration broke the atom, there would be no electron orbit to talk about, and even then, if the acceleration was so great that the electron could not get round its orbit before the velocity had significantly changed, one could not justify talking about the *instantaneous* diameter or period of the orbit.

Bell describes the practicalities of all this with some care in [2], and it has in part stimulated a positive reaction, and an interesting insight into a dynamical interpretation of the metric in relativity theories [6, 8, 9, 18]. Briefly, the idea is that the metric ultimately gets its chronogeometrical interpretation from dynamical considerations. A good discussion of this very important idea can be found in [10].

Put succinctly, relativity theories provide us with a manifold and a second rank tensor field $g_{\mu\nu}$, but we have to link this mathematical structure with what is out there, and the standard practice is to do this by postulating that the times and lengths specified in the usual way by $g_{\mu\nu}$ are just what our clocks and rulers would measure. On the other hand, our clocks and rulers are complex material systems whose behaviour is assumed in the detail to be governed by the field theories of the fundamental non-gravitational forces, such as electromagnetic forces. If the metric does indeed specify what we usually measure with clocks and rulers, there must be some dynamical explanation for that. And indeed, there always is.

Bell's pre-quantum atom is a case in point. If we define proper time and length by what the atom indicates for our measurements via its period and diameter, wherever it may be and whatever it may be doing, provided that it is not being too stressed by accelerations, calculation with the appropriate field theory, viz., Maxwell's theory, shows that the Minkowski metric will indeed provide times and lengths that correspond closely to the values obtained. This idea transfers to curved spacetimes as described by general relativity, where our field theories of the fundamental non-gravitational forces are extended in the usual minimal way by application of the strong equivalence principle (see Chap. 6).

Note in passing that the fact that the atom is not understood here via quantum mechanics is irrelevant, as is the fact that we ignore the radiation disaster for such an atom, which means that we are even ignoring one of the consequences of the Maxwell theory. In fact, even the details of Maxwell's theory turn out to be irrelevant in these calculations. What counts here, in the flat spacetime case, is merely that our theories of non-gravitational forces should be Lorentz symmetric.

Note that times and lengths implied by the metric will never concur exactly with those measured by our clocks and rulers. In a nutshell, this is because no measurement in physics can be perfectly accurate. Put another way, no clock or ruler is a perfect clock or ruler. In this context, one may ask whether, or to what extent, the clock satisfies the clock hypothesis:

Clock Hypothesis. Whatever worldline it may follow in spacetime, the putative clock measures exactly the proper time along that worldline, as defined by the metric, which amounts to saying that the effect of motion on the rate of the putative clock is no more than that associated with its instantaneous velocity, while any acceleration changes nothing.

Or indeed, one may ask whether, or to what extent a measuring rod satisfies the ruler hypothesis:

Ruler Hypothesis. Whatever motion it has, the putative ruler is always instantaneously ready to measure proper length in an instantaneously comoving inertial frame.

Of course, these two requirements could never be satisfied by any real clock or measuring rod under arbitrary acceleration.

As we saw in Sect. 6.3, one way for a ruler to satisfy the ruler hypothesis would be for the material continuum of the rod to undergo rigid motion, where rigid motion is defined in a relativistic sense, viz., the proper distance between any two neighbouring particles as measured in a frame that is instantaneously comoving with either of them is constant. This can only be achieved by ensuring that each particle has exactly the right acceleration, while no two particles would actually have the same acceleration [see, for example, Sect. 2.9, and in particular (2.245) on p. 83]. And furthermore, it could never be achieved by applying a force at one end. On the other hand, one expects some objects to be close to ideal rulers under many conditions. See [31] for more discussion of these points.

One can think of the clock hypothesis as the condition that defines an ideal clock. A general discussion can be found in [1]. Of course, in the real world, a clock may be close to ideal in one situation, and not work at all in another. A lot depends on how it is accelerated, in other words, on the nature of the worldline it has to follow while doing its job.

In their classic textbook, Misner et al. claim that whether the clock is pushed beyond the point where it can still keep good time depends entirely on the construction of the clock and not at all on any universal influence of acceleration on the march of time, stressing their view that velocity produces universal time dilation, while acceleration does not [38]. This is an assumption, however. How do they know that? Of course, it could be turned into a truism, by including, if necessary, universal influences of acceleration on the putative clock among the things that push it beyond the point where it can still keep good, i.e., proper, time. This kind of issue reveals a naive trust we all seem to have in relativity theory. Relativity theory may or may not dictate the idea Misner et al. express here, but relativity theory may be wrong.

Maybe one day someone will demonstrate that there is no universal influence of acceleration on putative clocks. It does matter what happens to real putative clocks here. This seems to underlie Brown and Pooley's concerns [8, 9]. It really matters whether there are any good approximations to ideal clocks. It really matters whether there might not be some universal influences of acceleration on putative real clocks.

In fact, the Bell type of calculation gives us a handle on these questions about how good our clocks and rulers will be, at least as far as relativity theory is concerned.

Relative to some theoretical coordinate system for our spacetime, the field theories governing the dynamics of our proposed clock or ruler will tell us how they behave. In the purely theoretical context, we can then see whether the times and lengths indicated by the given systems would be expected to concur with the times and lengths predicted by the Minkowski metric (or a curved metric in general relativity).

In a sense, this would indeed appear to turn things round, taking the metric predictions as a standard (see also the discussion in Sects. 6.4 and 6.5). However, in practice, one must have a way of linking the theoretical coordinate system with actual spacetime events. If one had adopted the wrong manifold with the wrong metric, general inconsistencies would eventually show up. The theoretical metric is indeed answerable to what we put forward as clocks and rulers in the real world, despite the approximations involved.

The dynamical approach to justifying the chronogeometric properties of $g_{\mu\nu}$ can in fact be seen as eliminating geometry from physics [7], for the simple reason that the measurement of lengths and times is now viewed as a dynamical exercise involving our fundamental field theories of non-gravitational forces, while the chronogeometric significance of $g_{\mu\nu}$ is relegated to a secondary role. In a flat spacetime, where $g_{\mu\nu} = \eta_{\mu\nu}$ is the Minkowski metric, this tensor field serves only to single out coordinate systems in which our field theories take on a particularly simple form, while its chronogeometric interpretation is then a consequence of the field theories applied to putative measuring tools. In a curved spacetime, the primary role of the tensor field $g_{\mu\nu}$ is to describe the interaction between non-gravitational and gravitational effects via the strong equivalence principle used to extend our flat spacetime field theories for non-gravitational forces to the curved manifold context.

That is one way of taking the import of Bell's paper. But it has also stimulated a fierce negative reaction in some quarters [41, 47], and it is important and instructive to see why. Indeed the reversal of roles advocated by Brown and Pooley in [8, 9] is anathema to those who view relativity as a mere exercise in geometry. One reason for this is undoubtedly that the geometrical view of Minkowski spacetime provides a very good way of picturing what is happening. One does not really need a dynamical explanation. The measuring rod or electron orbit are viewed as 4D regions of spacetime, and the different spatial dimensions they appear to have for different inertially moving observers are put down to the different hypersurfaces of simultaneity those observers might choose, or would naturally choose, to use to gauge lengths.

This is indeed a very elegant and useful picture. But it does not help us to understand what happens to the measuring rod when it is accelerated, or an electron orbit when an atom is accelerated. And since something does happen, and we can describe it using field theories that live in perfect harmony with the geometrical picture of spacetime, why should we not be allowed to add this dynamical picture to our understanding?

One obvious reaction to the idea that the relativistic contraction (referred to advisedly as the FitzGerald contraction in the above) indicates a real dynamical change in the measuring stick is just this: if the observer moves into a new inertial frame with motion relative to the rod, and nothing happens to the rod, it will still be relativistically contracted. This is known in the jargon as a passive Lorentz boost, while an active Lorentz boost is nothing other than what one would normally call an acceleration in any other context. Even in this case, one has a dynamical explanation from field theories as to *why* the observer should adopt a different way of gauging lengths, or indeed times. The point is that the observer changes while accelerating to the new frame [2].

Another accusation is precisely that the dynamical picture is a one-frame picture, while much of the beauty of relativity theory comes from viewing the world as a single mathematical object, a differentiable manifold, that can be described by different coordinate choices. But from our one-frame dynamical view of length and time measurements, do we really have to build up the whole multiframe view of Minkowski's spacetime in some pseudo-axiomatic way? Can we not just say that the dynamical picture lives alongside the geometrical one, adding a touch of real physics to what has become a largely mathematical exercise in our university textbooks?

In this context, proponents of dynamical explanations for relativistic effects are sometimes written off as constructivists [44], whose aim in life is to deduce the whole of the geometrical spacetime theory starting out only from field theories of matter. As discussed earlier (see p. 260), axiomatisation is a good route to understanding and also to generalisation in the field of pure mathematics, and although it can be similarly useful in physics, it does not seem to be an obligation for our theories of nature. Surely, any piece of understanding is a good piece of understanding, so long as it is logically consistent. But Norton [44] thinks that, if the dynamical approach cannot lead in an axiomatic way to the whole theory of relativity, it is worthless, in the sense that it adds nothing to our understanding. Furthermore, he attempts to achieve such an aim, in order to show that it will fail unless it does in fact already assume the tenets of the Minkowskian dogma.

There is an obvious point that seems at first sight to be in favour of Norton's criticism, and brought up by Nerlich too [41]. In order to apply our field theories for the dynamical explanation of the FitzGerald contraction proposed above, we need a manifold with the Lorentz metric structure on it, and we need Lorentz symmetric field theories. So there is no doubt that we are starting in the middle, so to speak. But perhaps physics always has to be like that? For there remains the job of relating the theory to what we actually measure out there, and assigning abstract spacetime points to real spacetime events. And geometry alone cannot do that.

Nerlich's slant on this [41] is to point out that any interpretive problem about contracted rods also arises with the contraction of the fields of the charges making up the rod, because it is for him the same contraction. That is indeed the geometrical picture. But with Bell's calculation, we have a dynamical argument to show why they are the same. Some of us just feel that that is useful.

There is also a sense in which, with the Minkowski picture of things, nothing ever *happens*, so there need be no dynamical cause for something like a relativistic contraction. The point here is that the measuring rod is just a 4D region of spacetime, even during an acceleration. What an observer, even an accelerating one, might choose as a continuous sequence of spacelike hypersurfaces intersecting that region is neither here nor there. Indeed, the 4D region is just a 4D region, so why talk about cause? Is cause not just an obsolete notion for those who never grasped the mixing of space and time inherent in the Minkowski picture?

One might put it like that. But if we can view what is *happening* in the world by taking what seem to us to be suitable 3D spacelike sections of a 4D object, and if we can provide what would once have been called a dynamical cause for the changes in those sections, should we not be allowed to do that? Would that really add nothing to our understanding, just because it could not be taken as a new axiom for deducing the whole geometric structure of the manifold we must use to obtain a coherent multiframe description? Many who read Bell's paper [2] feel that it does add to our understanding, and another example of such heresy will be described in Sect. 9.3. After all, our non-gravitational field theories do give us dynamical equations, and that is the way we have evolved to appreciate the world around us.

To end this brief review of a vast debate, let us cast a quick glance at the thought experiment that opens Bell's paper [2]. An inertial observer sees two rockets accelerate along the same line, one in front of the other, in such a way that they always remain exactly the same distance apart for that observer. The fragile thread joining them, initially stretched to its tolerance limit, will thus break.

Petkov claims that the thread breaks because the proper distance between the two rockets, as observed by an inertial observer instantaneously comoving with either, is increasing [47]. This is indeed the geometrical way of looking at things, although perhaps slightly problematic, as he points out himself, because proper distance is a relative concept, depending on the observer, and one would prefer the breaking to depend on something that is not observer-dependent if possible.

Nerlich says that the thread breaks because the forces accelerating the rockets pull the thread apart [41]. One could say that, of course. It amounts to the following explanation: the thread breaks because something, in fact the wrong kind of thing, happens to it. This goes beyond the geometrical explanation, but it stops short of what is actually possible.

The dynamical explanation for the breakage using Maxwell's theory is real physics, of the kind we were allowed to talk about before threatened with the Minkowski straightjacket! At the beginning of the experiment, the forces within the thread are in an equilibrium that allows one to consider it as a single entity, but only just. Once it is moving, it can no longer achieve the equilibrium it once had, because the fundamental fields within it change. Put another way [2], as the rockets gather speed, the thread will become too short because of its need to FitzGerald contract. It must break when the artificial prevention of this inevitable contraction, due to being attached at each end to the rockets, imposes intolerable stress.

9.3 Velocity Dependence of Particle Mass

The notion of velocity $\mathbf{v}$ in pre-relativistic physics is replaced by the notion of four-velocity

$$V := (\gamma c, \gamma \mathbf{v})$$

in the special theory of relativity, where $\gamma(v)$ is given by (9.1). Having done this, one seeks a Lorentz covariant extension of Newton's second law $\mathbf{F} = m\mathbf{a}$ determining the acceleration $\mathbf{a}$ of a particle of mass m under the effects of a force $\mathbf{F}$, to the special relativistic context. The simplest choice by far is

$$F = \frac{\mathrm{d}P}{\mathrm{d}\tau}\,, \tag{9.2}$$

where F is a four-vector, the four-force, to be specified by looking at what effects are acting on the particle, $P := m_0 V$ is another four-vector, the four-momentum, with m_0 some constant mass, and τ is proper time along the particle worldline. P is called four-momentum because its spacelike part looks very like what one used to call momentum, provided the particle is not moving too fast. Indeed,

$$P := (\gamma m_0, \gamma m_0 \mathbf{v})\,,$$

and we call $\mathbf{p} := \gamma m_0 \mathbf{v}$ the relativistic 3-momentum. If m_0 were the inertial mass of the particle, this would be γ times the Newtonian momentum. For small v, $\gamma(v) \approx 1$, and the relativistic 3-momentum would be approximately equal to the Newtonian momentum.

Let us write

$$F = (\gamma f_0, \gamma \mathbf{f})\,, \tag{9.3}$$

We now make the crucial hypothesis that $\mathbf{f}$ is the usual 3-force measured in the laboratory. Using the fact that $\mathrm{d}t/\mathrm{d}\tau = \gamma$, the law (9.2) then requires

$$\mathbf{f} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\gamma m_0 \mathbf{v}) = \frac{\mathrm{d}}{\mathrm{d}t}(m\mathbf{v})\,, \tag{9.4}$$

where t is the coordinate time in our chosen inertial frame and we have defined $m(v) := m_0\gamma(v)$. So for example, if the particle carries a charge e and there is a magnetic field $\mathbf{B}$, we will equate the right-hand side of (9.4) with $e\mathbf{v} \times \mathbf{B}$. That still leaves the zero component of F open, at least on the face of things.

In fact, f_0 is completely determined by the proposed law of motion (9.2). The reason is that the latter can be written

$$F = m_0 A\,,$$

where

$$A := \frac{\mathrm{d}V}{\mathrm{d}\tau}$$

is called the four-acceleration. Then since $V^2 = c^2$ in special relativity, we deduce by differentiation that

$$V \cdot \frac{\mathrm{d}V}{\mathrm{d}\tau} = 0\,,$$

where the dot denotes pseudoscalar product under the Minkowski pseudometric. So $V \cdot A = 0$, and we also have $F \cdot V = 0$ for any F that satisfies our equation of motion

(9.2). But then

$$\begin{aligned} F \cdot V = 0 \quad &\Longleftrightarrow \quad \gamma f_0 \gamma c = \gamma^2 \mathbf{f} \cdot \mathbf{v} \\ &\Longleftrightarrow \quad f_0 = \frac{\mathbf{f} \cdot \mathbf{v}}{c} \,, \end{aligned}$$

so what we must choose for f_0 is completely determined by our proposed law of motion (9.2) and the statement that **f** in (9.3) is the usual 3-force measured in the laboratory.

Note that the proposal for the new law (9.2) is entirely motivated by considerations of Lorentz covariance, but along the way we have spawned a new notion of inertial mass, viz., $m := m_0 \gamma(v)$, which increases with the speed v of the particle. The hypothesis (9.4) already contains the bold suggestion that the acceleration caused by a given force will be less if the particle is moving. The m on the right-hand side of (9.4) is a very different thing to m_0, because we absorbed γ into it. So the real innovation here seems to be (9.4) along with the physical identification of **f**.

But no physical explanation is forthcoming in this account. We make these hypotheses on the basis of Lorentz symmetry considerations and then simply note that it works. However, even before Einstein's relativistic revolution, there were theoretical indications that inertial mass, or at least certain contributions to it, should increase with speed according to the function $\gamma(v)$. Here is a brief review of what has also been touched upon in Sect. 8.1, but full details can be found in [32].

We consider the momentum of the electromagnetic fields of a moving charge q_{e}. However, we shall not consider that charge to be concentrated at a point, but instead uniformly distributed over a spherical shell of radius a, at least when the system is stationary in some inertial frame. This was an early model for the electron. Naturally, there must be some binding forces in the system to stop the charge from flying apart, and we shall assume that these binding forces are due to some other fields with Lorentz symmetric dynamics. Then, when the charge shell is moving at a constant velocity **v**, we expect it to contract in the direction of motion according to the usual factor. Taking the x axis in the direction of motion, the equation for the small shell will be

$$\gamma^2 (x - vt)^2 + y^2 + z^2 = a^2 \,. \tag{9.5}$$

Note that it is the equilibrium between the EM forces and the binding forces that determines the shape the system will have when it is accelerated from one state of motion to another, and we shall return to this point shortly.

For a point P at distance r from the present position of the charge center C, such that the line CP makes an angle θ with the velocity **v**, the electric field lies radially outward from the present position of C and is in fact given by [32]

$$E_x = \frac{q_e \gamma(v)}{4\pi\varepsilon_0} \frac{x - vt}{[\gamma^2(x - vt)^2 + y^2 + z^2]^{3/2}} , \tag{9.6}$$

$$E_y = \frac{q_e \gamma(v)}{4\pi\varepsilon_0} \frac{y}{[\gamma^2(x - vt)^2 + y^2 + z^2]^{3/2}} , \tag{9.7}$$

$$E_z = \frac{q_e \gamma(v)}{4\pi\varepsilon_0} \frac{z}{[\gamma^2(x - vt)^2 + y^2 + z^2]^{3/2}} . \tag{9.8}$$

Further, the magnetic field is $\mathbf{B} = \mathbf{v} \times \mathbf{E}/c^2$ [32]. Note that, if the sign of the charge were in fact negative, both $\mathbf{E}$ and $\mathbf{B}$ would be reversed, but the momentum density given by the Poynting vector $\mathbf{g}$, viz.,

$$\mathbf{g} = \varepsilon_0 \mathbf{E} \times \mathbf{B} ,$$

would remain the same. Because the magnetic field has magnitude $vE \sin\theta/c^2$, the momentum density has magnitude

$$g = \frac{\varepsilon_0 v}{c^2} E^2 \sin\theta ,$$

and points down toward the path of the charge, making an angle θ with the vertical.

Note that it is because there is motion that there is a magnetic field, and it is because there is a magnetic field that there is a momentum carried by the electromagnetic fields. The corresponding momentum density is not a radiation field in the present case, because it drops off as quickly as the Coulomb field of a static charge with distance from the source.

If we integrate the momentum density over the whole space outside the ellipsoid in order to work out the total momentum $\mathbf{p}$ in the fields, only a component along the axis of motion will remain. This is a straightforward symmetry consideration. So we can forget about the component of $\mathbf{g}$ perpendicular to $\mathbf{v}$, and the relevant component of the momentum density is $g \sin\theta$. Integrating over all space outside the ellipsoid given by (9.5), we eventually find

$$\mathbf{p} = \frac{2}{3} \frac{q_e^2}{4\pi\varepsilon_0} \frac{\gamma(v)\mathbf{v}}{ac^2} . \tag{9.9}$$

Making the replacement $e^2 = q_e^2/4\pi\varepsilon_0$, which amounts to a change of units, this becomes

$$\mathbf{p} = \frac{2}{3} \frac{e^2}{ac^2} \gamma(v)\mathbf{v} . \tag{9.10}$$

We can thus attribute an electromagnetic mass

$$m_{\mathrm{EM}}(v) = \frac{2}{3} \frac{e^2}{ac^2} \gamma(v) \tag{9.11}$$

to an electron with such a shell structure.

Note that there is no momentum in the fields when the system is not moving. So if this spatially extended charge distribution were gently accelerated from rest to the constant velocity **v**, we would see momentum building up in the fields. A gentle acceleration is required so that a new static equilibrium can be set up between binding forces and EM forces, and since all forces are assumed to be due to Lorentz symmetric fields, we expect the final shape of the system to be roughly the FitzGerald contracted ellipsoid. Now whatever force is required to get the system into the new state of motion, some part of it is being used to get momentum into the EM fields, and this part of it is in fact overcoming an EM force that the system exerts on itself during acceleration [32].

The self-force calculation is not as easy as the momentum calculation, but it is essential to this account, because it explains the mechanism we wish to advertise here. Qualitatively, this is how it works. We imagine the centre of the charge shell moving along the x axis so that it reaches $x(t)$ at time t. We take this function to be arbitrary, so $\ddot{x}(t)$ may not be zero. (Note that t is coordinate time in our inertial frame, and dots denote derivatives with respect to this time.) We must then make some assumption about the shape of the shell, because we could not know that without detailed calculations concerning the equilibrium between the binding forces and the EM forces in the system. A common assumption is that the system is rigid in the relativistic sense discussed earlier in this book, which amounts here to saying that the shell is always exactly spherical in the inertial frame instantaneously comoving with the centre.

For each charge element on the surface we calculate the fields it produces using the exact Liénard–Wiechert solution to Maxwell's equations for a charge element with the given motion. We then calculate the forces (to be precise, the relativistic 3-forces) those fields will exert on all the other charge elements on the shell, simply adding them up, even though they act at different points of the system. And finally, we sum the result for all such forces on the system due to other charge elements on the surface. The result is a net EM force which the system exerts on itself when accelerated:

$$F_{\text{self}} = -\frac{2}{3}\frac{e^2}{ac^2}\gamma(v)^3\ddot{x} + O(a^0)\ , \tag{9.12}$$

where $v := \dot{x}$. Note here the crucial point that

$$\frac{\mathrm{d}}{\mathrm{d}t}\big[\gamma(v)v\big] = \gamma(v)^3\ddot{x}\ .$$

Comparing (9.12) with (9.11), we see that it is indeed by overcoming this self-force that we get momentum in the EM fields.

Note in passing that there are higher order terms in (9.12), going as a^0, a, a^2, and so on. The term in a^0, i.e., independent of the system size (and as it turns out, even its shape), is crucial to explaining why charged particles radiate EM energy when accelerated. It is the celebrated radiation reaction. The rate of doing work against this term in the self-force is exactly the rate of emission of EM energy given by the Larmor formula [32]. One loses this explanation for the EM radiation by an accelerating electron when the latter is treated as a point particle!

But for the moment, we are interested in the first term, which would diverge if we tried to make this into a point particle model for the electron by putting $a = 0$. This term exactly opposes the acceleration and the coefficient of $\ddot{x}$ has units of mass. In fact, the EM mass in (9.11) is part of the inertial mass of the system (see also Sect. 9.4). Let us justify this claim. The usual relativistic extension of Newton's second law can be written

$$F_{\text{self}} + F_{\text{ext}} = m_{\text{mechanical}}\gamma(v)^3\ddot{x}\,,$$

where F_{ext} is the total external force on the system, and $m_{\text{mechanical}}$ is what we might call the mechanical mass, whatever may cause that, due to other origins than the EM self-force. Inserting (9.12), dropping all terms of order a or higher, and rearranging, we find that

$$F_{\text{ext}} + \text{radiation reaction} = \left(m_{\text{mechanical}} + \frac{2}{3}\frac{e^2}{ac^2}\right)\gamma(v)^3\ddot{x}\,.$$

If one did not think about EM self-forces, and one does not usually do so except for the radiation reaction term (often left unexplained), one would say that the quantity in round brackets on the right-hand side had to be what one usually measures as the inertial mass.

How can we justify dropping the higher order terms in (9.12)? We can just say that they are negligible because a is small. Alternatively, we can set $a = 0$, thereby making a point particle model for the electron. We do not need to worry about the divergent term in the inertial mass, because we can simply point to the fact that it is

$$\boxed{m_{\text{renormalised}} := m_{\text{mechanical}} + \frac{2}{3}\frac{e^2}{ac^2}} \tag{9.13}$$

that we actually measure in practice, and this is finite. This ploy goes by the name of mass renormalisation, and the need for such a ploy plagues all point particle pictures.

It turns out that any spatially extended charge distribution will exert such a force on itself when accelerated, and the leading order term going as a^{-1}, for some length a specifying the spatial dimensions of the distribution, is always directly aligned or counteraligned with the acceleration for spherically symmetric charge distributions. (In many other cases too, but not in all, as we shall see in Chap. 10.) So this self-force effect always contributes to the inertial mass of a bound state particle made up of charged particles, for example, such as the proton or the neutron.

There is a hypothesis that one might call the self-force or bootstrap hypothesis which says in fact that all particles actually have spatial structure, and further, that all their resistance to being accelerated as represented by the letter m in Newton's second law comes from forces they exert on themselves when we try to accelerate them. The crucial thing to note then is that contributions to the inertia such as m_{EM} in (9.11) will all come with the speed-dependent factor $\gamma(v)$, provided that all the self-forces in the system, e.g., due to binding force fields, or strong force fields, or whatever fields are operating within the given particle, are indeed due to fields with Lorentz symmetric dynamics.

In this case, we do indeed have a full dynamical explanation for the fact that inertial mass increases like $\gamma(v)$ when a particle moves faster: it is simply because self-forces on the system increase in the corresponding way. And even if self-forces only lead to part of the inertial mass, we still have a dynamical explanation for why *those* contributions increase as they do with speed.

Note that it is generally considered to be a heresy to try to explain this mass increase. Perhaps the historical role of such reactions has been to protect students of physics from certain fallacies. Here is one trap that would be easy to fall into. Suppose we have a standard 1 kg mass A for measuring purposes. When we accelerate into a new inertial frame $\mathscr{I}'$, moving with respect to our original inertial frame $\mathscr{I}$, our standard mass A, carried with us, will have a higher mass as far as an observer in $\mathscr{I}$ is concerned, but since it is our standard, we still refer to it as 1 kg, by definition. Comparing with another standard 1 kg mass B which we left behind in $\mathscr{I}$, we would now want to say that the latter had a lower mass than 1 kg. On the other hand, relativity theory tells us clearly that, because B is moving at some speed relative to ourselves, it is the one that has a mass greater than 1 kg.

As usual with paradoxes in relativity theory, the solution here is to have a very clear multiframe view of things, and a good grasp of the ideas of Lorentz covariance. The point we make with our dynamical explanation above is that it is not invalidated just because someone might make a mistake somewhere down the line. The dynamical explanation lives alongside the geometric view, and is logically compatible with it, provided we exercise the usual care. It is not because it is a one-frame view that it should be declared invalid. All we are saying is that, when a bound state particle is accelerated, any self-force contributions to its inertial mass must increase as $\gamma(v)$ because the self-forces increase in the corresponding way.

9.4 Inertial Consequences of Binding Energy

In modern particle physics, many particles are considered to be bound states of others. For example, all baryons are bound states of three quarks, and mesons are quark–antiquark bound states. The inertial masses of the quarks themselves are today generally assumed to result from interaction with the Higgs field (but see the bootstrap hypothesis in Sects. 8.1 and 9.3), while the inertial masses of bound state particles are considered to comprise the inertial masses of the quarks, suitably increased depending on their motion, hence kinetic energy, within the bound state, plus another crucial term: the strong and electromagnetic binding energies of the system. The only justification for including the latter is generally to point out that mass and energy are equivalent. If pushed, we say that this in turn is because $E = mc^2$. So let us see where that last relation comes from.

We have seen that the input (9.3) in the last section together with the law (9.2) require the first component f_0 of the four-force to be given by $f_0 = \mathbf{f} \cdot \mathbf{v}/c$. But the first component of (9.2) requires

$$\gamma f_0 = \frac{\mathrm{d}(m_0 \gamma c)}{\mathrm{d}\tau} \ .$$

Then since $\mathrm{d}t/\mathrm{d}\tau = \gamma$, this means that

$$f_0 = c\frac{\mathrm{d}m}{\mathrm{d}t} \ ,$$

where $m := m_0\gamma$ is commonly called the relativistic inertial mass. So if we have $f_0 = \mathbf{f}\cdot\mathbf{v}/c$, we must have

$$\frac{\mathrm{d}(mc^2)}{\mathrm{d}t} = \mathbf{f}\cdot\mathbf{v} \ . \tag{9.14}$$

The quantity $\mathbf{f}\cdot\mathbf{v}$ is the rate at which work is done on the particle by the force $\mathbf{f}$, i.e., the rate of increase of energy of the particle, so mc^2 is identified with the total energy of the particle. This is a standard way of reaching the conclusion that $E = mc^2$ [28].

The m on the left-hand side of (9.14) is a very different thing to m_0, because we absorbed γ into it. Since $v < c$, we can expand the function $\gamma(v)$ as a series in powers of v/c to obtain

$$mc^2 = m_0c^2 + \frac{1}{2}m_0v^2 + \frac{3}{8}m_0\frac{v^4}{c^2} + \cdots \ . \tag{9.15}$$

The first term here is optimistically called rest mass energy, although it seems unlikely that the pioneers really expected to get energy out of it in the early days, and the second term is the old kinetic energy. So we have simply fed the kinetic energy into our mc^2 on the left-hand side of (9.14). There are of course the remaining terms, which are part of the boldness of the hypothesis.

Once again, all this is entirely motivated by considerations of Lorentz symmetry. We make a bold hypothesis in the form of (9.2) and the physical interpretation of (9.4), then just note that it works. But as we have seen in Sect. 8.3, there is a pre-quantum explanation for why binding energies should alter the resistance a bound state particle will show when we try to accelerate it. Here is a brief review, but full details can be found in [32].

In fact, we have already done most of the work. In the last section, we discussed the self-forces that spatially extended charge distributions exert on themselves when accelerated. These always contribute directly to the inertia of a bound state particle made up of charged particles, e.g., the proton comprising two up quarks and one down quark, because it turns out that the leading term in the EM self-force for small spherically symmetric charge distributions is always aligned or counteraligned with the acceleration. Now we can in theory calculate the EM binding energy of any spatially extended charge distribution. It is just the energy in the EM fields it produces in an inertial frame in which it is at rest, or indeed the energy required to assemble it from charges at infinity in such a frame.

What is the connection with the EM self-force? In fact, up to a physically dimensionless constant, the coefficient of the acceleration in the leading order term of the EM self-force is equal to the EM binding energy divided by c^2. For example, for the charge shell, the Coulomb energy in its EM fields is $e^2/2a$, corresponding to a mass of $e^2/2ac^2$. This is 3/4 of the contribution of the self-force to the inertial mass. The factor of 3/4 is unfortunate and needs explaining. Indeed it led to a very long debate in the literature, still going on today.

But putting that discrepancy aside until the next paragraph, our explanation for the success of simply adding binding energies (divided by c^2) to other sources of inertial mass in order to get a total inertial mass is just this: binding forces in composite particles lead to bootstrap effects, and that is why binding energy must be included in their inertial mass. Put another way, we include binding energies because they reflect the related self-forces in those bound states.

What about the discrepancy? Briefly, the difference between the energy-derived and momentum-derived (or self-force-derived) EM masses arises because we are trying to write down a four-momentum for the EM fields of our charge distribution. However, the EM energy–momentum tensor alone is not conserved, i.e., its covariant divergence is not zero, so it cannot be used to get a Lorentz covariant four-momentum for the fields by integrating over spacelike hypersurfaces.

One solution here is to make a purely ad hoc redefinition of the EM energy–momentum tensor so that one does obtain a Lorentz covariant four-momentum [32]. This approach is advocated for other purposes than considering spatially extended charge distributions. But when we are genuinely interested in the consequences of self-forces in the context of spatially extended particles, it is better to write down the total energy–momentum tensor, including EM fields but also all the other binding fields that must be present in the bound state in order to achieve an equilibrium, and then take integrals of this over spacelike hypersurfaces in order to produce a total four-momentum. The latter will be Lorentz covariant, because the total energy–momentum tensor will be conserved.

The latter ploy leads to a total self-force contribution to the inertial mass of the bound state, including terms from EM effects, but also from all the other fields operating within the bound state. In this case, if we could achieve such a calculation, the total binding energy divided by c^2 would be just the self-force-derived contribution to the inertial mass. This concludes our dynamical justification for simply adding binding energies to get a contribution to the inertial mass.

9.5 Conclusion

The big debate blowing up around dynamical explanations for the relativistic contraction effect is only part of a bigger issue of whether it is worth holding on to dynamical explanations for anything at all in what is in effect a Minkowski block universe, where the geometrical view sees everything at a glance in a 4D world. Dieks has written a more balanced view of all this, contrasting top-down and bottom-up approaches to physics [15]. The Minkowskian geometrical view, with its emphasis on unexplained principles and Lorentz symmetry, is a top-down approach, while the causal or dynamical explanations we have been resuscitating here are examples of the bottom-up approach.

In Dieks' words, the possibility of bottom-up strategies in relativity is not something new, but it is still worth stressing in view of the continuing dominance of the top-down approach both in the teaching of relativity and in philosophical accounts of the theory [15]. There is no single best way of explaining relativistic effects. It all depends what our aims are. But one thing seems clear: when we are teaching this subject, surely we need to give our students all the help we can, so that they have access to all the insights we can muster.

This is especially true when we consider the future of our theories, since if history repeats itself, we are going to need to go beyond current hypotheses. In that sense, it would be a serious mistake to become too dogmatic.

Chapter 10
Non-Renormalisability of EM Self-Force. A Classical Picture

10.1 Overview

In the last two chapters, we briefly discussed the electromagnetic forces that spatially extended charge distributions exert on themselves when accelerated, and mentioned the link with the problem of renormalisation that still plagues quantum electrodynamics [see, for example, Sect. 8.2 or (9.13) on p. 330]. It was also mentioned there that the potentially divergent leading order term in the EM self-force was *not* always aligned or counteraligned with the acceleration (see in particular the conclusion on p. 292). In such cases, this means that one cannot simply retrieve a finite version of Newton's second law in the point particle limit by renormalising the 'particle' mass. This is relevant to the themes of this book as a fundamental insight into the phenomenon of acceleration.

To prove the above claims, we consider a rigid charge dumbbell, i.e., two point charges held some distance apart by an unspecified binding force, such that, when one charge follows an arbitrary timelike worldline in Minkowski spacetime, the motion of the other is completely specified by the condition that the axis of the system is Fermi–Walker transported along that worldline. The electromagnetic force the system exerts on itself under such conditions is calculated to leading order and found as usual to go as the reciprocal of the distance between the charges. However, it is shown that this term, which diverges in the point particle limit, is not proportional to the four-acceleration of the system except in special cases, whence the relativistic extension of Newton's second law cannot be renormalised in this limit. It is shown how this problem is resolved when the charge dumbbell is replaced by a spherically symmetric charge distribution.

In Sect. 10.2, we obtain a formula for the EM fields very close to the worldline of a point charge with arbitrary motion in Minkowski spacetime, following the analysis by Dirac [16]. In this paper, Dirac showed how to renormalise the mass of a classical point electron by assuming it was a point particle and considering conservation of energy and momentum within a tube containing the worldline, an analysis which led him to the notorious Lorentz–Dirac equation for the motion of a point charge. This equation is based in part on an (infinite) adjustment of the electron mass to cater for the (infinite) energy in the EM fields close to the charge [30, Chap. 11].

In this section, the aim is to consider the electron as a spatially extended entity, making the very simple charge dumbbell model of two point charges held some distance apart by unspecified binding forces and calculating the EM force the system exerts upon itself when accelerated in some arbitrary way through Minkowski spacetime. Electromagnetic self-forces of this kind are discussed at length in [32]. The leading order term is not Coulomb, i.e., going as d^{-2}, where d is the separation of the point charges, but in fact goes as d^{-1}. In the point particle limit where one allows d to go to zero, this introduces an infinite self-force.

However, in some cases, the offending term can be absorbed into the mass times acceleration term of Newton's second law. This requires it to be proportional to the acceleration vector (four-vector or three-vector, depending on whether one considers a fully relativistic analysis or not). The mass of the charge gets adjusted by an amount that goes to infinity in the point particle limit, but this does not have physically measurable consequences if one considers the resulting m to be the physically measured mass. The fact that one has to go through such a procedure is just an inconvenient truth about the point particle approximation, and one which carries over to quantum field theory.

In [32], it was conjectured that the divergent leading order term in the EM self-force would always be aligned with the acceleration vector, so that mass renormalisation would always be possible, and that this might be the case precisely because Maxwell's theory is a gauge theory, since we know that all gauge theories are fully renormalisable in quantum field theory. In the present section it is shown that the mass of the charge dumbbell system will in fact only be renormalisable in some special cases, in particular those considered in [32], but that the situation is saved for spherically symmetric charge distributions, suggesting that the point particle limit may only be valid in such a case, viz., where the point particle is considered to be spherically symmetric.

Some assumptions have to be made about the charge dumbbell in order to build any model of its motion under acceleration (see Sect. 10.3). If the two charges are labelled A and B, we attribute an arbitrary timelike worldline $x_A(\tau_A)$ to A and specify the motion of $x_B(\tau_B)$ of B by assuming the relativistic version of rigidity widely discussed in earlier chapters (but see also [14, 31]), and further assuming that the system is as rotationless as possible, i.e., that the system axis is Fermi–Walker transported along the worldline x_A. Calculations relaxing the last assumption might also be interesting, but more difficult.

As mentioned above, it is shown that the leading order term in the EM self-force is not generally parallel to the four-acceleration $\ddot{x}_A$ of particle A (see Sect. 10.4), whence the relativistic version of Newton's dynamical equation cannot generally be renormalised. Finally, the very same formulas are used to show in Sect. 10.5 that, if the system has spherical symmetry, e.g., if one considers a rigid spherical shell of charge, the EM self-force is then proportional to the four-acceleration, and Newton's second law can be rescued by adjusting the mass of the shell, a well known result.

10.2 EM Fields Close to the Worldline of a Point Charge

Here we follow the analysis in [16]. The Liénard–Wiechert retarded potential for the EM fields produced by a point charge e following a worldline $z(s)$ in Minkowski spacetime, with metric $\operatorname{diag}(1,-1,-1,-1)$, is given as a function of the field point x by [32, Chap. 2]

$$A^\mu(x) = \left.\frac{e\dot{z}^\mu}{\dot{z}\cdot(x-z)}\right|_{s_+} ,$$

where the dot over a symbol denotes differentiation with respect to the proper time s along the worldline, and everything is evaluated at the unique retarded proper time s_+ for the given field point, i.e., the value of s such that

$$\big[x-z(s_+)\big]^2 = 0 , \qquad z^0(s_+) < x^0 . \tag{10.1}$$

Now for any two continuous functions $f(s)$ and $g(s)$ of s with $\dot{g}(s) > 0$, we have the following result from distribution theory:

$$\int f(s)\delta\big(g(s)\big)\mathrm{d}s = \int \frac{f(s)}{\dot{g}(s)}\delta\big(g(s)\big)\mathrm{d}g(s) = \frac{f(s)}{\dot{g}(s)} ,$$

evaluating the result at any value of s within the range of integration that satisfies $g(s) = 0$. By this ploy, we establish the more convenient result

$$A^\mu(x) = 2e\int \dot{z}^\mu(s)\delta\Big(\big[x-z(s)\big]^2\Big)\mathrm{d}s ,$$

integrating s from $-\infty$ to any value between the retarded and advanced proper times for the given field point x. One can relax the restriction on the range of integration by introducing a step function $\theta\big(x^0 - z^0(s)\big)$ into the integrand. But using the form given here, we have

$$\begin{aligned}
\frac{\partial A^\mu}{\partial x_\nu} &= 4e\int \dot{z}^\mu(x^\nu - z^\nu)\delta'\Big(\big[x-z(s)\big]^2\Big)\mathrm{d}s \\
&= -2e\int \frac{\dot{z}^\mu(x^\nu - z^\nu)}{\dot{z}_\lambda(x^\lambda - z^\lambda)}\frac{\mathrm{d}}{\mathrm{d}s}\delta\Big(\big[x-z(s)\big]^2\Big)\mathrm{d}s \\
&= 2e\int \frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\dot{z}^\mu(x^\nu - z^\nu)}{\dot{z}_\lambda(x^\lambda - z^\lambda)}\right]\delta\Big(\big[x-z(s)\big]^2\Big)\mathrm{d}s ,
\end{aligned}$$

whence it follows that

$$\begin{aligned}
F^{\mu\nu}(x) &:= \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} \\
&= -2e\int \frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\dot{z}^\mu(x^\nu - z^\nu) - \dot{z}^\nu(x^\mu - z^\mu)}{\dot{z}_\lambda(x^\lambda - z^\lambda)}\right]\delta\Big(\big[x-z(s)\big]^2\Big)\mathrm{d}s ,
\end{aligned}$$

and finally,

$$F^{\mu\nu}(x) = -\frac{e}{\dot{z}\cdot(x-z)}\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\dot{z}^{\mu}(x^{\nu}-z^{\nu})-\dot{z}^{\nu}(x^{\mu}-z^{\mu})}{\dot{z}\cdot(x-z)}\right]\Bigg|_{s_+}, \tag{10.2}$$

where everything in the last expression is evaluated at the retarded proper time s_+.

We now choose a field point x very close to the worldline. If it is close enough, there will be a unique proper time s_0 such that x is simultaneous with the charge at $z(s_0)$ for an observer instantaneously comoving with the charge. So s_0 is defined to be the unique proper time such that

$$\left[x^{\mu}-z^{\mu}(s)\right]\cdot v(s) = 0\,, \qquad v(s) := \dot{z}(s)\,.$$

We define the spacelike vector from $z(s_0)$ to the field point, viz.,

$$\gamma^{\mu} := x^{\mu} - z^{\mu}(s_0)\,. \tag{10.3}$$

Hence, we have

$$\gamma\cdot v(s_0) = 0\,. \tag{10.4}$$

By hypothesis, the γ^{μ} will be very small.

To apply (10.2), we need to consider the retarded time for the given field point x. We know there will be some small $\sigma\in\mathbb{R}^+$ such that the retarded time is $s_0-\sigma$. We expect σ to be of the same order of magnitude as the γ^{μ}. The aim now is to expand the right-hand side of (10.2) in powers of σ. Of course, the leading order term will go as σ^{-2}. Following Dirac, we retain terms up to $O(\sigma^0)$, even though we shall only require those up to $O(\sigma^{-1})$ for present purposes.

We begin with the Taylor expansions

$$x^{\mu}-z^{\mu}(s_0-\sigma) = \gamma^{\mu}+\sigma v^{\mu}-\frac{1}{2}\sigma^2\dot{v}^{\mu}+\frac{1}{6}\sigma^3\ddot{v}^{\mu}+O(\sigma^4)\,, \tag{10.5}$$

$$\dot{z}^{\mu}(s_0-\sigma) = v^{\mu}-\sigma\dot{v}^{\mu}+\frac{1}{2}\sigma^2\ddot{v}^{\mu}+O(\sigma^3)\,,$$

where v, $\dot{v}$, and $\ddot{v}$ on the right-hand side are all evaluated at $s=s_0$, and it should be remembered that $\gamma\sim\sigma$. Applying the trivial results

$$v^2=1\,, \qquad v\cdot\dot{v}=0\,, \qquad v\cdot\ddot{v}+\dot{v}^2=0\,,$$

and also (10.4), we obtain

$$\dot{z}\cdot(x-z) = \sigma-\sigma(\gamma\cdot\dot{v})+\frac{1}{2}\sigma^2(\gamma\cdot\ddot{v})-\frac{1}{6}\sigma^3\dot{v}^2+O(\sigma^4)\,,$$

and consequently,

$$\left[\dot{z}\cdot(x-z)\right]^{-1} = \sigma^{-1}\left[1-(\gamma\cdot\dot{v})\right]^{-1}\left[1-\frac{1}{2}\sigma(\gamma\cdot\ddot{v})+\frac{1}{6}\sigma^2\dot{v}^2+O(\sigma^3)\right]\,.$$

For the moment, we refrain from expanding out $\big[1-(\gamma\cdot\dot{v})\big]^{-1}$. We also have

$$\dot{z}^\mu(x^\nu - z^\nu) - \dot{z}^\nu(x^\mu - z^\mu)$$
$$= v^\mu\gamma^\nu - \sigma\dot{v}^\mu\gamma^\nu - \frac{1}{2}\sigma^2\dot{v}^\mu v^\nu + \frac{1}{2}\sigma^2\ddot{v}^\mu\gamma^\nu + \frac{1}{3}\sigma^3\ddot{v}^\mu v^\nu - (\mu \longleftrightarrow \nu)\,,$$

where once again the quantities v, $\dot{v}$, and $\ddot{v}$ on the right-hand side are evaluated at $s = s_0$. We now have

$$\frac{\dot{z}^\mu(x^\nu - z^\nu) - \dot{z}^\nu(x^\mu - z^\mu)}{\dot{z}\cdot(x-z)}$$
$$= \big[1-(\gamma\cdot\dot{v})\big]^{-1}\bigg[\sigma^{-1}v^\mu\gamma^\nu - \dot{v}^\mu\gamma^\nu - \frac{1}{2}\sigma\dot{v}^\mu v^\nu - \frac{1}{2}(\gamma\cdot\ddot{v})v^\mu\gamma^\nu$$
$$+\frac{1}{6}\sigma\dot{v}^2 v^\mu\gamma^\nu + \frac{1}{2}\sigma\ddot{v}^\mu\gamma^\nu + \frac{1}{3}\sigma^2\ddot{v}^\mu v^\nu - (\mu \longleftrightarrow \nu)\bigg]\,.$$

In the formula for the EM fields, this has to be differentiated with respect to s, and this can be done by differentiating with respect to σ and changing the sign, because we have not yet applied the condition that fixes σ. We thus obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\dot{z}^\mu(x^\nu - z^\nu) - \dot{z}^\nu(x^\mu - z^\mu)}{\dot{z}\cdot(x-z)}\right]$$
$$= -\big[1-(\gamma\cdot\dot{v})\big]^{-1}\bigg[-\sigma^{-2}v^\mu\gamma^\nu - \frac{1}{2}\dot{v}^\mu v^\nu + \frac{1}{6}\dot{v}^2 v^\mu\gamma^\nu$$
$$+\frac{1}{2}\ddot{v}^\mu\gamma^\nu + \frac{2}{3}\sigma\ddot{v}^\mu v^\nu - (\mu \longleftrightarrow \nu)\bigg]\,,$$

which leads us to an expansion for the EM fields in powers of σ:

$$F^{\mu\nu} = e\big[1-(\gamma\cdot\dot{v})\big]^{-2}\bigg[-\sigma^{-3}v^\mu\gamma^\nu - \frac{1}{2}\sigma^{-1}\dot{v}^\mu v^\nu + \frac{1}{2}\sigma^{-2}(\gamma\cdot\ddot{v})v^\mu\gamma^\nu$$
$$+\frac{1}{2}\sigma^{-1}\ddot{v}^\mu\gamma^\nu + \frac{2}{3}\ddot{v}^\mu v^\nu - (\mu \longleftrightarrow \nu)\bigg]\,.$$

We now determine σ itself in powers of the small quantity $\varepsilon \in \mathbb{R}^+$ defined by

$$\gamma^2 = -\varepsilon^2\,. \qquad (10.6)$$

Now σ is determined by (10.1), into which we insert the approximation (10.5) to obtain

$$\gamma^2 + \sigma^2 - \sigma^2(\gamma\cdot\dot{v}) + \frac{1}{3}\sigma^3(\gamma\cdot\ddot{v}) - \frac{1}{12}\sigma^4\dot{v}^2 = 0\,,$$

which is correct to $O(\sigma^5)$. But $\sigma = \varepsilon$ to first order, so this can be written

$$-\varepsilon^2 + \sigma^2 - \sigma^2(\gamma\cdot\dot{v}) + \frac{1}{3}\varepsilon^3(\gamma\cdot\ddot{v}) - \frac{1}{12}\varepsilon^4\dot{v}^2 = 0\,.$$

Hence

$$\sigma^2 = \big[1-(\gamma\cdot\dot{v})\big]^{-1}\left[\varepsilon^2 - \frac{1}{3}\varepsilon^3(\gamma\cdot\ddot{v}) + \frac{1}{12}\varepsilon^4\dot{v}^2\right]\,,$$

and finally,

$$\sigma = \varepsilon\big[1-(\gamma\cdot\dot{v})\big]^{-1/2}\left[1 - \frac{1}{6}\varepsilon(\gamma\cdot\ddot{v}) + \frac{1}{24}\varepsilon^2\dot{v}^2\right]\,.$$

Inserting this into the above expression for the EM fields, we have

$$\begin{aligned}F^{\mu\nu} = e\big[1-(\gamma\cdot\dot{v})\big]^{-1/2}\Bigg\{&-\varepsilon^{-3}v^\mu\gamma^\nu - \frac{1}{2}\varepsilon^{-1}\dot{v}^\mu v^\nu\big[1+(\gamma\cdot\dot{v})\big]\\ &+\frac{1}{8}\varepsilon^{-1}\dot{v}^2v^\mu\gamma^\nu + \frac{1}{2}\varepsilon^{-1}\ddot{v}^\mu\gamma^\nu + \frac{2}{3}\ddot{v}^\mu v^\nu - (\mu\longleftrightarrow\nu)\Bigg\}\,.\end{aligned}$$

This is Dirac's result [16]. It is accurate to $O(\varepsilon^0)$ and could be used to calculate EM self-forces to this order in the linear dimensions ε of a small charge distribution. Terms of $O(\varepsilon^0)$ explain EM radiation by such a system when it is accelerated [32].

In the present discussion, we shall only require terms to $O(\varepsilon^{-1})$, which diverge as the linear dimensions of our charge distribution tend to zero, i.e., $\varepsilon \to 0$. These are the terms that require renormalisation, even in the classical context, if we insist on point particles. To this order, we have

$$F^{\mu\nu} = e\big[1-(\gamma\cdot\dot{v})\big]^{-1/2}\left[-\varepsilon^{-3}v^\mu\gamma^\nu - \frac{1}{2}\varepsilon^{-1}\dot{v}^\mu v^\nu - (\mu\longleftrightarrow\nu)\right]\,.$$

We can also expand out the factor

$$\big[1-(\gamma\cdot\dot{v})\big]^{-1/2} = 1 + \frac{1}{2}(\gamma\cdot\dot{v}) + O(\varepsilon^2)\,,$$

whence finally, defining the unit spacelike vector $u := \gamma/\varepsilon$,

$$F^{\mu\nu} = e\left[\frac{u^\mu v^\nu - v^\mu u^\nu}{\varepsilon^2} + \frac{v^\mu\dot{v}^\nu - \dot{v}^\mu v^\nu + (u^\mu v^\nu - v^\mu u^\nu)(u\cdot\dot{v})}{2\varepsilon} + O(\varepsilon^0)\right]\,. \quad (10.7)$$

This is the result we shall apply below.

10.3 Charge Dumbbell and Rigidity Assumption

We wish to consider the simplest possible spatially extended charge distribution, namely two point charges e_A and e_B held some distance d apart by an unspecified binding force. When A has an arbitrary timelike worldline $x_A(\tau_A)$ in Minkowski spacetime, we know from Sect. 10.2 the EM field $F_A^{\mu\nu}(B)$ it will produce at the nearby point B, provided we also know the worldline $x_B(\tau_B)$ of B. We can then calculate the EM four-force of A on B from the standard result

$$F^\mu(A \text{ on } B) = e_B F_A^{\mu\nu}(B) v_\nu^B \,, \tag{10.8}$$

where v_B is the four-velocity of B. We can then find the four-force of B on A in a similar way and simply add the two together, even though they act at different points of the charge distribution. This gives a total EM force of the system on itself for arbitrary motion.

We call this the EM self-force F_{self}. The idea then is to expand F_{self} as a power series in the length d of the system. We expect the Coulomb terms going as d^{-2} to cancel out, but we also expect terms going as d^{-1}, d^0, and so on, to remain. In the point particle limit $d \to 0$, we need to be able to absorb the divergent term into the mass times acceleration component of the relativistic version of Newton's second law, viz.,

$$F = \frac{\mathrm{d}P}{\mathrm{d}\tau} = m\frac{\mathrm{d}a}{\mathrm{d}\tau} \,, \qquad P := mv \,,$$

where m is rest mass, v is four-velocity, F is four-force, a is four-acceleration, and τ is proper time. To do this, it must clearly be proportional to the four-acceleration of the system. If it is not, this ploy, known as renormalisation, will not work.

But how do we formulate the motions of A and B? We would like to attribute an arbitrary worldline $x_A(\tau_A)$ to A. But then this constrains the worldline of B, which is supposed to be a distance d from it. On the other hand, whenever the system has a component of its motion along its own axis, we expect some degree of relativistic contraction, and this contraction will depend on the equilibrium between the unspecified binding forces and the EM forces, whereas we wish to do the calculation without going into too much detail about the binding forces. So what would be a good rule for constraining the worldline of B?

A related problem is that we will specify the two worldlines as spacetime functions of proper time, and the proper times τ_A and τ_B of the two charges will depend on the worldlines, whereas our self-force calculation must work out the four-force of A on B and then the four-force of B on A at the same coordinate time in whatever inertial frame we have selected at the outset. A solution to these conundrums is to assume that the charge dumbbell is rigid, in the well defined relativistic sense of the word discussed earlier in the book [14, 31], and also that it is non-rotating, at least as far as this is possible. The purpose of this section is to explain briefly how this works. An in-depth discussion of rigid motion can be found in Sects. 2.3.2 and 2.4.5 (but see also [31]).

We consider an orthonormal triad $n_i^\mu(\tau_A)$ $i=1,2,3$, of spacelike vectors along x_A, orthogonal to the worldline at each value of τ_A, whence

$$n_i \cdot n_j = -\delta_{ij}\,, \qquad n_i \cdot v_A = 0\,, \qquad v_A^2 = 1\,,$$

setting $c=1$. We assume that the worldline of B can be given by

$$x_B^\mu(\tau_B) = x_A^\mu\big(\tau_A(\tau_B)\big) + \xi^i n_i^\mu\big(\tau_A(\tau_B)\big)\,, \tag{10.9}$$

for some function $\tau_A(\tau_B)$ to be determined, where ξ^i are just three fixed numbers. Naturally, $\tau_A(\tau_B)$ will depend on these three numbers. It is clear from the form of (10.9) that, given any value τ_B of the proper time of B, the function $\tau_A(\tau_B)$ delivers the unique proper time of A such that $x_B^\mu(\tau_B)$ is simultaneous with the event $x_A^\mu\big(\tau_A(\tau_B)\big)$ in the instantaneously comoving inertial frame ICIF$_A$ of A at its proper time τ_A.

For the moment we have not completely specified the choice of spacelike triad $\{n_i\}_{i=1,2,3}$. It can rotate as it moves up the worldline x_A. However, given the triad and the three numbers $\{\xi_i\}_{i=1,2,3}$, the worldline x_B of B is fully determined. Differentiating (10.9) with respect to τ_B, we have

$$v_B^\mu(\tau_B) = \dot{x}_B^\mu(\tau_B) = \Big[v_A^\mu\big(\tau_A(\tau_B)\big) + \xi^i \dot{n}_i^\mu\big(\tau_A(\tau_B)\big)\Big]\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B}\,,$$

where the dot over n_i denotes differentiation with respect to τ_A. For each value of τ_A, we can express the three four-vectors $\dot{n}_i$ in terms of the orthonormal basis $\{v_A, n_1, n_2, n_3\}$:

$$\dot{n}_i^\mu(\tau_A) = a_{0i}(\tau_A) v_A^\mu(\tau_A) + \Omega_{ij}(\tau_A) n_j^\mu(\tau_A)\,,$$

for three functions $a_{0i}(\tau_A)$ and nine functions $\Omega_{ij}(\tau_A)$. It is easy to show that only three of the latter are independent since

$$\Omega_{ij}(\tau_A) = -\Omega_{ji}(\tau_A)\,, \qquad i,j \in \{1,2,3\}\,.$$

Furthermore, the three functions $a_{0i}(\tau_A)$ are given by

$$a_{0i}(\tau_A) = -n_i(\tau_A)\cdot \dot{v}_A(\tau_A)\,, \qquad i \in \{1,2,3,\}\,,$$

whence they may be interpreted as the three components of the absolute acceleration of A at proper time τ_A, i.e., the three spatial components of the four-acceleration of A in its instantaneous rest frame ICIF$_A$ (the temporal component being zero in that frame).

We now have

$$v_B^\mu(\tau_B) = \bigg\{\Big[1+\xi^i a_{0i}\big(\tau_A(\tau_B)\big)\Big] v_A^\mu\big(\tau_A(\tau_B)\big) + \xi^i \Omega_{ij}\big(\tau_A(\tau_B)\big) n_j^\mu\big(\tau_A(\tau_B)\big)\bigg\}\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B}\,.$$

Dropping the arguments of the functions, this implies that

$$1 = v_B^2 = \left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]\left(\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B}\right)^2 ,$$

whence

$$\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} = \left[\left(1+\xi^i a_{0i}\right)^2 - \xi^i \xi^j \Omega_{ik}\Omega_{jk}\right]^{-1/2} ,$$

with the right-hand side a function of $\tau_A(\tau_B)$. The function $\tau_A(\tau_B)$ itself can then be found by integrating this along the worldline $x_B(\tau_B)$, with the boundary condition $\tau_A(0) = 0$, so that the proper times of A and B are synchronised when either is zero.

Now one case in which the system is said to be rigid in the relativistic sense is when $\Omega_{ij} = 0$, for all $i, j \in \{1,2,3\}$. We then have

$$v_B^\mu(\tau_B) = \left[1+\xi^i a_{0i}\big(\tau_A(\tau_B)\big)\right] v_A^\mu\big(\tau_A(\tau_B)\big)\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} ,$$

with

$$\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} = \left(1+\xi^i a_{0i}\right)^{-1} . \tag{10.10}$$

One very significant feature of this case is thus that

$$v_B^\mu(\tau_B) = v_A^\mu\big(\tau_A(\tau_B)\big) .$$

So for a rigid motion of our system, if we choose any event $x_B(\tau_B)$ on the worldline of B and find the unique event $x_A\big(\tau_A(\tau_B)\big)$ on the worldline of A for which an inertial observer instantaneously comoving with A considers B to be simultaneous, both charges have the same four-velocity. So an inertial observer moving instantaneously with B at the event $x_B(\tau_B)$ would also consider the event $x_A\big(\tau_A(\tau_B)\big)$ to be simultaneous.

This symmetry is going to be very useful below and is more or less the entire justification for the rather artificial rigidity assumption. Note that it is not at all obvious how such a situation might arise physically. It would have to result from the balance of forces between binding effects and EM effects within the charge dumbbell, not to mention the way the system is accelerated. Still, like many approximations in physics, it is justified by making some kind of analysis possible rather than by physical considerations!

Thinking back to the analysis in Sect. 10.2, for any choice of τ_B and considering $x_B(\tau_B)$ as a field point at which to evaluate the EM fields due to A, we have Dirac's spacelike four-vector of (10.3):

$$\gamma^\mu(\tau_B) = x_B^\mu(\tau_B) - x_A^\mu\big(\tau_A(\tau_B)\big) = \xi^i n_i^\mu\big(\tau_A(\tau_B)\big) , \tag{10.11}$$

with

$$\gamma^2 = -\xi^2 := -(\xi^1)^2 - (\xi^2)^2 - (\xi^3)^2 .$$

This is of course constant, because we assumed at the outset that (10.9) would be possible for some constant choice of the ξ^i. Naturally, this is part of the rigidity assumption. It says that whenever an inertial observer instantaneously comoving with A looks at B, or vice versa, the other charge always seems to be the same distance away. So $|\xi| =: d$ is the constant length of the charge dumbbell as judged by any inertial observer instantaneously comoving with either charge, and it will correspond to ε of (10.6) when we come to work out the EM forces of either charge on the other.

A word should be said about the assumption that $\Omega_{ij} = 0$ for all $i, j \in \{1,2,3\}$. These numbers constitute an antisymmetric matrix, and hence generate rotations. However, setting them equal to zero does not completely save our triad from rotation relative to space in some initially chosen inertial frame. In fact it amounts to saying that the triad n_i^μ is Fermi–Walker transported along the worldline of particle A, as we have seen in Sect. 2.3.3. Let us recall briefly what this means in the present context.

Recalling that $v_A(\tau_A)$ is the 4-velocity of the worldline of A, the equation for Fermi–Walker transport of a contravector M^μ along the worldline is

$$\dot{M}^\mu = -(M \cdot \dot{v}_A)v_A + (M \cdot v_A)\dot{v}_A \ . \tag{10.12}$$

This preserves inner products, i.e., if M and N are FW transported along the worldline, then $M \cdot N$ is constant along the worldline. Furthermore, the tangent vector v_A to the worldline is itself FW transported along the worldline, and if the worldline is a spacetime geodesic (a straight line in Minkowski coordinates), then FW transport is the same as parallel transport.

Now recall that the Ω_{ij} were defined by

$$\dot{n}_i^\mu = a_{0i} v_A^\mu + \Omega_{ij} n_j^\mu \ . \tag{10.13}$$

When $\Omega_{ij} = 0$, this becomes

$$\dot{n}_i^\mu = a_{0i} v_A^\mu \ . \tag{10.14}$$

This is indeed the FW transport equation for n_i^μ, found by inserting $M = n_i$ into (10.12), because we insist on $n_i \cdot v_A = 0$ and we have $a_{0i} = -n_i \cdot \dot{v}_A$.

In fact, the orientation in spacetime of the local rest frame triad n_i^μ cannot be kept constant along a worldline unless that worldline is straight (we are referring to flat spacetimes here). Under Fermi–Walker transport, however, the triad remains as constantly oriented, or as rotationless, as possible, in the following sense: at each instant of time τ_A, the triad is subjected to a pure Lorentz boost without rotation in the instantaneous hyperplane of simultaneity. (On a closed orbit, this process can still lead to spatial rotation of axes upon return to the same space coordinates, an effect known as Thomas precession.) For a general non-Fermi–Walker transported triad, the Ω_{ij} are the components of the angular velocity tensor that describes the instantaneous rate of rotation of the triad in the instantaneous hyperplane of simultaneity.

Of course, given any triad n_i^μ at one point on the worldline, it is always possible to Fermi–Walker transport it to other points by solving (10.12). We are then saying that, if the motion of B is given by

$$x_B^\mu(\tau_B) = x_A^\mu\big(\tau_A(\tau_B)\big) + \xi^i n_i^\mu\big(\tau_A(\tau_B)\big) , \tag{10.15}$$

where the ξ^i are fixed numbers, then the motion of the resulting dumbbell is rigid.

We need to check carefully that we do have all the necessary details of the above symmetry. Suppose therefore that we are trying to find the EM force of B on A for some choice τ_A. We shall need an orthogonal triad $\{n_i'(\tau_B)\}$ along the worldline of B. But we can take the triad at $x_B(\tau_B)$ to be

$$n_i'(\tau_B) := n_i\big(\tau_A(\tau_B)\big) ,$$

where $\tau_A(\tau_B)$ is the function we discussed above. This automatically satisfies

$$n_i' \cdot n_j' = -\delta_{ij} , \qquad n_i' \cdot v_B = 0 ,$$

since $v_B(\tau_B) = v_A\big(\tau_A(\tau_B)\big)$. Interestingly, $\{n_i'\}$ is also FW transported along x_B. This is shown by

$$\begin{aligned}
\dot{n}_i' &= \dot{n}_i \frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} \\
&= \Big[-(n_i \cdot \dot{v}_A)v_A + (n_i \cdot v_A)\dot{v}_A\Big]\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} \\
&= -(n_i' \cdot \dot{v}_B)v_B + (n_i' \cdot v_B)\dot{v}_B ,
\end{aligned}$$

which is the equation for FW transport of n_i' along x_B, as required.

Effectively, we now have two functions $\tau_A(\tau_B)$ and $\tau_B(\tau_A)$. They are bijective and mutual inverses (see below for the proof). If we begin with an event $x_A(\tau_A)$ on x_A, the unique event $x_B(\tau_B)$ on x_B such that $x_A(\tau_A)$ is simultaneous with it in the inertial frame instantaneously comoving with four-velocity $v_B(\tau_B)$ is $x_B\big(\tau_B(\tau_A)\big)$, where the four-velocity is $v_B\big(\tau_B(\tau_A)\big)$. Now Dirac's spacelike vector $\gamma'^\mu(\tau_A)$ from the unique $x_B\big(\tau_B(\tau_A)\big)$ to the chosen $x_A(\tau_A)$ is in this case

$$\gamma'^\mu(\tau_A) = x_A^\mu(\tau_A) - x_B^\mu\big(\tau_B(\tau_A)\big) = -\gamma^\mu\big(\tau_B(\tau_A)\big) , \tag{10.16}$$

using (10.11). We express this in terms of the triad $n_i'\big(\tau_B(\tau_A)\big) = n_i(\tau_A)$ along x_B. But we know that

$$x_B^\mu\big(\tau_B(\tau_A)\big) = x_A^\mu(\tau_A) + \xi^i n_i^\mu(\tau_A) ,$$

so

$$\gamma'^\mu(\tau_A) = -\xi^i n_i^\mu(\tau_A) = -\xi^i n_i'^\mu(\tau_B) ,$$

and

$$\gamma'^2 = -\xi^2 = -d^2 .$$

So the numbers corresponding to the ξ^i are $-\xi^i$ when we approach the problem from this way around. Furthermore, the absolute acceleration of B is

$$
\begin{aligned}
a'_{0i}(\tau_B) &= -n_i\big(\tau_A(\tau_B)\big)\cdot\dot v_B(\tau_B)\\
&= -n_i\cdot\frac{\mathrm{d}}{\mathrm{d}\tau_B}v_A\big(\tau_A(\tau_B)\big)\\
&= -n_i\cdot\frac{\mathrm{d}v_A}{\mathrm{d}\tau_A}\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B}\\
&= \frac{a_{0i}}{1+\xi^i a_{0i}}\;,
\end{aligned}
$$

using (10.10). Also by (10.10), we have

$$
\begin{aligned}
\frac{\mathrm{d}\tau_B}{\mathrm{d}\tau_A} &= \frac{1}{1-\xi^i a'_{0i}}\\
&= \frac{1}{1-\xi^i\dfrac{a_{0i}}{1+\xi^i a_{0i}}}\\
&= 1+\xi^i a_{0i} = \left(\frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B}\right)^{-1}\;,
\end{aligned}
$$

proving the above claim that the functions $\tau_A(\tau_B)$ and $\tau_B(\tau_A)$ are mutual inverses.

10.4 Leading Order Term in EM Self-Force

We now wish to use (10.7) and (10.8) to calculate the EM force of A on B for some choice of the proper time τ_B:

$$
F^\mu(A \text{ on } B) = e_B F_A^{\mu\nu}(B)v_\nu^B\;.
$$

To this we wish to add the EM force of B on A for some choice of the proper time τ_A:

$$
F^\mu(B \text{ on } A) = e_A F_B^{\mu\nu}(A)v_\nu^A\;.
$$

This is where we use the symmetry established above for rigid motion: we work out $F^\mu(A \text{ on } B)$ and $F^\mu(B \text{ on } A)$ at corresponding values of τ_A and τ_B, viz., when $\tau_A = \tau_A(\tau_B)$ with the latter function determined by integrating (10.10). The point is that the two charges then have the same four-velocity, so we can find an inertial frame in which the two events $x_A(\tau_A)$ and $x_B(\tau_B)$ are simultaneous and the two charges instantaneously at rest. If we worked out the two EM forces in this inertial frame, we could add them to get the total EM self-force as a four-vector at this instant in this frame. Since $v_A(\tau_A)$ and $v_B(\tau_B)$ are also equal for this choice of τ_A and τ_B, we could then boost back to the original inertial frame and obtain

$$
F_{\text{self}}^\mu = e_B F_A^{\mu\nu}(B)v_\nu^B + e_A F_B^{\mu\nu}(A)v_\nu^A\;.
$$

Here F_{self} can be considered as a function of either τ_A or τ_B, where τ_A has to be given by the function $\tau_A(\tau_B)$ discussed above. The rigidity assumption obviously plays a key role in this convenient picture!

Actually, the convenient picture reveals a problem when carrying out self-force calculations for any spatially extended charge distribution. We have to add up forces acting at different points of the distribution, but this then requires a choice of hyperplane of simultaneity at which to consider the sum. In a way, the above choice looks natural enough: we add up the two forces at times which an inertial observer instantaneously comoving with either point charge A or B would consider to be simultaneous, so one might call this the rest frame self-force, even though we can of course boost the result to any inertial frame, as we have done here.

We could just have added the two forces at the same time for an arbitrary inertial observer. The reader is invited to estimate the difference that could make. These considerations reveal the complexity introduced into relativistic dynamics of real extended objects by the fact that simultaneity is no longer absolute as compared with a Newtonian world.

Now by (10.7),

$$F_A^{\mu\nu}(B) = e_A\left[\frac{u^\mu v_A^\nu - v_A^\mu u^\nu}{d^2} + \frac{v_A^\mu \dot{v}_A^\nu - \dot{v}_A^\mu v_A^\nu + (u^\mu v_A^\nu - v_A^\mu u^\nu)(u\cdot\dot{v}_A)}{2d} + O(d^0)\right] ,$$

where $u = \gamma/d$ is the unit four-vector from $x_A\big(\tau_A(\tau_B)\big)$ to $x_B(\tau_B)$ and v_A and $\dot{v}_A$ are both evaluated at $\tau_A(\tau_B)$. We thus calculate

$$F^\mu(A \text{ on } B) = e_B F_A^{\mu\nu}(B) v_\nu^B(\tau_B) = e_A e_B\left[\frac{u^\mu}{d^2} + \frac{-\dot{v}_A^\mu + (u\cdot\dot{v}_A)u^\mu}{2d} + O(d^0)\right] ,$$

using the fact that $u\cdot v_B = 0$, $v_A \cdot v_B = 1$, and $\dot{v}_A \cdot v_B = 0$. The latter follows from

$$\dot{v}_B = \dot{v}_A \frac{\mathrm{d}\tau_A}{\mathrm{d}\tau_B} = \frac{\dot{v}_A}{1+\xi^i a_{0i}} ,$$

together with the fact that $\dot{v}_B \cdot v_B = 0$. Likewise,

$$F^\mu(B \text{ on } A) = e_A F_B^{\mu\nu}(A) v_\nu^A(\tau_A) = e_A e_B\left[\frac{u'^\mu}{d^2} + \frac{-\dot{v}_B^\mu + (u'\cdot\dot{v}_B)u'^\mu}{2d} + O(d^0)\right],$$

where $u' = \gamma'/d$ is the unit four-vector from $x_B\big(\tau_B(\tau_A)\big)$ to $x_A(\tau_A)$ and v_B and $\dot{v}_B$ are both evaluated at $\tau_B(\tau_A)$. By (10.16), $u' = -u$.

It is this last observation that shows why the Coulomb terms going as $O(d^{-2})$ cancel in the total self-force. There is just one more approximation to be made in determining the $O(d^{-1})$ term in the self-force:

$$\dot{v}_B = \frac{\dot{v}_A}{1+\xi^i a_{0i}} = \dot{v}_A + O(d) , \tag{10.17}$$

since $|\xi| = O(d)$. Hence, finally,

$$F_{\text{self}} = \frac{e_A e_B}{d} \Big[(u \cdot \dot{v}_A) u - \dot{v}_A \Big] + O(d^0) \; . \tag{10.18}$$

This shows that, contradicting the claim in [32], the EM self-force is not generally aligned with the four-acceleration of the system, taking the latter to be the four-acceleration of charge A here. It will not therefore be possible in general to renormalise the relativistic version of Newton's dynamical law.

There are two simple cases where the EM self-force is aligned with the four-acceleration of the system, in fact precisely the two cases considered in [32]:

- The charge dumbbell moves along a straight line perpendicular to its axis.
- The charge dumbbell moves along a straight line parallel to its axis.

10.4.1 Transverse Linear Acceleration

Taking the motion along the x axis,

$$x_A = (x_A^0, x_A^1, 0, 0) \; , \qquad v_A = (\dot{x}_A^0, \dot{x}_A^1, 0, 0) \; ,$$

where x_A^0, x_A^1 are functions of the proper time τ_A and dots over symbols denote differentiation with respect to τ_A. Likewise,

$$x_B = (x_B^0, x_B^1, d, 0) \; , \qquad v_B = (\dot{x}_B^0, \dot{x}_B^1, 0, 0) \; ,$$

where $\tau_B = \tau_A$, $x_B^0(\tau_B) = x_A^0(\tau_A)$, and $x_B^1(\tau_B) = x_A^1(\tau_A)$. We thus also have $v_B = v_A$ and $\dot{v}_B = \dot{v}_A$. Furthermore,

$$\gamma = (0, 0, d, 0) \; , \qquad u = (0, 0, 1, 0) \; .$$

Hence, $u \cdot \dot{v}_A = 0$ and we obtain simply

$$F_{\text{self}} = -\frac{e_A e_B}{d} \dot{v}_A + O(d^0) \; ,$$

agreeing with the result in [32].

10.4.2 Longitudinal Linear Acceleration

Once again, we take the motion to be along the x axis. In the local rest frame of A, $u = (0, 1, 0, 0)$, so in the fixed inertial frame relative to which A has four-velocity $v_A(\tau_A)$,

$$u = \big(\gamma(w) w, \gamma(w), 0, 0 \big) \; , \qquad v_A = \big(\gamma(w), \gamma(w) w, 0, 0 \big) \; ,$$

where w is the coordinate velocity of A in the x direction and

$$\gamma(w) := (1-w^2)^{-1/2} .$$

A little calculation shows that

$$\dot{v}_A = \gamma(w)^4 a(w,1,0,0) , \qquad a := \frac{\mathrm{d}w}{\mathrm{d}t} ,$$

where a is the coordinate acceleration of A in the x direction. We then have

$$u \cdot \dot{v}_A = -\gamma(w)^3 a ,$$

so that, using (10.18) and a little manipulation,

$$F_{\mathrm{self}} = -\frac{2e_A e_B}{d}\dot{v}_A + O(d^0) .$$

Once again this agrees with the result in [32].

10.5 Spherical Symmetry

In the last section, we identified two special cases in which the general result (10.18) does in fact lead to a renormalisable mass in the limit as the system size tends to zero. But the general result itself contains a term proportional to the separation four-vector u between the two charges which means that the leading order term in the expansion of F_{self}, going as d^{-1}, is not proportional to the four-acceleration vector of the system, whence renormalisation will not generally be possible.

One way round this problem is to hypothesise that any real particle comprising a spatially extended charge distribution is spherically symmetric. It is in fact a well known result, for example, that the leading order term in the EM self-force of a spherical charge shell will be aligned with the four-acceleration, and indeed we can prove this from (10.18). This result is not wholly surprising. What vectors are left for F_{self} to pick out when the system geometry itself does not specify any particular vector? There is of course the four-velocity, but it turns out that F_{self} picks out the four-acceleration and mass renormalisation is then possible.

We thus consider a rigid spherical charge shell of radius R whose center follows an arbitrary timelike worldline in such a way that any line segment between diametrically opposite charge elements on the surface is Fermi–Walker transported along that worldline, i.e., the sphere undergoes as little rotation as possible in the sense described in Sect. 10.3. Rigidity means here that there is always an inertial frame in which the whole charge shell is instantaneously at rest, and that the shell is always spherical in that frame [31].

We consider the charge shell in its instantaneous (inertial) rest frame. The surface charge density is $\rho = e/4\pi R^2$, assuming a total charge of e on the surface. We take surface charge elements in pairs $\mathrm{d}\sigma_1 = e\mathrm{d}\Omega_1/4\pi$ and $\mathrm{d}\sigma_2 = e\mathrm{d}\Omega_2/4\pi$, where $\mathrm{d}\Omega_1$

and $\mathrm{d}\Omega_2$ are the corresponding solid angle elements subtended at the center. For each pair of charge elements there is a unit separation four-vector $u = (0, \mathbf{n})$.

There is also an acceleration vector $\dot{v} = (0, \mathbf{a})$ which can be the four-acceleration of either of the charge elements, or better still of the sphere center, since we know that these four-acceleration vectors differ by at most a term of $O(R)$ [see (10.17), for example]. Of course, the unit three-vector $\mathbf{n}$ will vary as we change the pair of surface charge elements, but the three-vector $\mathbf{a}$ will not, because we carry out our calculation for some particular instantaneous snapshot of the charge shell.

By (10.18), the EM self-force on this pair of charge elements is

$$\delta F_{\mathrm{self}}(\Omega_1, \Omega_2) = -\frac{e^2 \mathrm{d}\Omega_1 \mathrm{d}\Omega_2}{d(\Omega_1, \Omega_2)(4\pi)^2} \Big[(\mathbf{n} \cdot \mathbf{a})\mathbf{n} + \mathbf{a} \Big] ,$$

where $d(\Omega_1, \Omega_2)$ is a distance function, specifying the separation of the two charge elements in this frame, and $\mathbf{n}$ is of course also a function of the pair (Ω_1, Ω_2). The total EM self-force is thus

$$F_{\mathrm{self}} = -\frac{e^2}{2(4\pi)^2} \iint \frac{(\mathbf{n} \cdot \mathbf{a})\mathbf{n} + \mathbf{a}}{d(\Omega_1, \Omega_2)} \mathrm{d}\Omega_1 \mathrm{d}\Omega_2 ,$$

with an extra factor of 1/2 to account for the fact that we count each pair of charge elements twice in the integral.

We now write $\mathbf{n} = \mathbf{n}_\parallel + \mathbf{n}_\perp$, where $\mathbf{n}_\parallel$ is the component of $\mathbf{n}$ parallel to $\mathbf{a}$ and $\mathbf{n}_\perp$ is its component perpendicular to $\mathbf{a}$. Then

$$(\mathbf{n} \cdot \mathbf{a})\mathbf{n} = (\mathbf{n}_\parallel \cdot \mathbf{a})\mathbf{n}_\parallel + (\mathbf{n}_\parallel \cdot \mathbf{a})\mathbf{n}_\perp .$$

However, the integral of the second term here is zero:

$$\iint \frac{(\mathbf{n}_\parallel \cdot \mathbf{a})\mathbf{n}_\perp}{d(\Omega_1, \Omega_2)} \mathrm{d}\Omega_1 \mathrm{d}\Omega_2 = 0 .$$

This is because, for every pair of solid angle elements $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$, there is another pair $\mathrm{d}\Omega_1'$ and $\mathrm{d}\Omega_2'$ obtained by rotation through 180° about $\mathbf{a}$, for which $\mathbf{n}_\parallel = \mathbf{n}_\parallel'$ but $\mathbf{n}_\perp = -\mathbf{n}_\perp'$. Contributions therefore cancel.

Hence,

$$F_{\mathrm{self}} = -\frac{e^2}{2(4\pi)^2} \iint \frac{(\mathbf{n}_\parallel \cdot \mathbf{a})\mathbf{n}_\parallel + \mathbf{a}}{d(\Omega_1, \Omega_2)} \mathrm{d}\Omega_1 \mathrm{d}\Omega_2 .$$

Now $\mathbf{n}_\parallel \cdot \mathbf{a} = n_\parallel a$, the product of the lengths of the two vectors, and

$$(a n_\parallel)\mathbf{n}_\parallel = \mathbf{a}(n_\parallel)^2 = \mathbf{a}(\mathbf{e} \cdot \mathbf{n})^2 ,$$

where $\mathbf{e}$ is a unit vector parallel to $\mathbf{a}$. We now have

$$F_{\mathrm{self}} = -\frac{e^2 \mathbf{a}}{2(4\pi)^2} \iint \frac{1 + (\mathbf{e} \cdot \mathbf{n})^2}{d(\Omega_1, \Omega_2)} \mathrm{d}\Omega_1 \mathrm{d}\Omega_2 ,$$

Note that by spherical symmetry the integral here has to be independent of **e**. If **e** points into the solid angle Ω, then

$$\iint \frac{(\mathbf{e}\cdot\mathbf{n})^2}{d(\Omega_1,\Omega_2)}\mathrm{d}\Omega_1\mathrm{d}\Omega_2 = \frac{1}{4\pi}\int \mathrm{d}\Omega \iint \frac{(\mathbf{e}\cdot\mathbf{n})^2}{d(\Omega_1,\Omega_2)}\mathrm{d}\Omega_1\mathrm{d}\Omega_2$$
$$= \frac{1}{4\pi}\iint \frac{\mathrm{d}\Omega_1\mathrm{d}\Omega_2}{d(\Omega_1,\Omega_2)}\int(\mathbf{e}\cdot\mathbf{n})^2\mathrm{d}\Omega\ ,$$

and

$$\frac{1}{4\pi}\int(\mathbf{e}\cdot\mathbf{n})^2\mathrm{d}\Omega = \frac{1}{4\pi}\int_0^{\pi}\cos^2\theta\sin\theta\mathrm{d}\theta\int_0^{2\pi}\mathrm{d}\phi$$
$$= \frac{1}{2}\int_0^{\pi}\cos^2\theta\sin\theta\mathrm{d}\theta = \frac{1}{3}\ .$$

So finally,

$$F_{\mathrm{self}} = -\frac{2}{3}\frac{e^2}{(4\pi)^2}\mathbf{a}\iint\frac{\mathrm{d}\Omega_1\mathrm{d}\Omega_2}{d} = -\frac{4}{3}U\mathbf{a}\ , \tag{10.19}$$

where

$$U := \frac{e^2}{(4\pi)^2}\frac{1}{2}\iint\frac{\mathrm{d}\Omega_1\mathrm{d}\Omega_2}{d}$$

is the EM energy of the charge shell in this frame.

Equation (10.19) is a standard result [32]. It shows that the EM self-force in this case is proportional to the four-acceleration of the spatially extended charge distribution. The mass of such a particle can thus be renormalised in the limit as the size of the charge distribution goes to zero.

10.6 Non-Relativistic Limit

Non-renormalisability remains in general in the non-relativistic limit. In this case, the hyperplanes of simultaneity of either A or B are approximately hyperplanes of simultaneity for the global inertial frame chosen at the outset, i.e., hyperplanes of constant inertial time t. Furthermore, the proper times of both A and B are both approximately equal to inertial time, i.e., $\tau_A(t) \approx t$ and $\tau_B(t) \approx t$. So

$$x_B(\tau_B) - x_A(\tau_A) \approx (0, \text{space vector from } A \text{ to } B \text{ at time } t)\ .$$

This means that $u^0 \approx 0$, and $\mathbf{u}$ is approximately a unit space vector in the direction from A to B lying in the hyperplane of constant t at each time t. The four-velocities of A and B are

$$v_A \approx (c, \mathrm{d}\mathbf{x}_A/\mathrm{d}t)\ , \qquad v_B \approx (c, \mathrm{d}\mathbf{x}_B/\mathrm{d}t)\ ,$$

and the four-accelerations

$$\dot{v}_A \approx (0, \mathrm{d}^2\mathbf{x}_A/\mathrm{d}t^2) , \qquad \dot{v}_B \approx (0, \mathrm{d}^2\mathbf{x}_B/\mathrm{d}t^2) .$$

We thus find

$$u \cdot \dot{v}_A \approx -\mathbf{u} \cdot \frac{\mathrm{d}^2\mathbf{x}_A}{\mathrm{d}t^2} ,$$

which is minus the scalar product of the unit three-vector joining A to B in the constant t hyperplane and the instantaneous three-acceleration of A at that inertial time t. When this is not zero, as is generally the case, the term $(u \cdot \dot{v}_A)u$ in the self-force (10.18) will not be zero, and nor will it generally be aligned with the acceleration of A, since it lies along the unit three-vector joining A to B in the constant t hyperplane.

10.7 Conclusion

The conjecture in [32] to the effect that the leading order term in the EM self-force of a spatially extended charge distribution is always aligned with the four-acceleration of the distribution was not correct. This is sometimes true, however, and in particular in the case of a spherically symmetric charge distribution. It was further conjectured in [32] that this situation might arise because Maxwell's theory is a gauge theory, since it is known that gauge theories are renormalisable in quantum field theory. There may still be a connection, of course, in the cases where the mass of the classical object is actually renormalisable.

It may also be that the renormalisability in quantum field theory contains a hidden spherical symmetry assumption which would be difficult to discern, given the obscure ontological characteristics of quantum field theory.

The rigidity assumption so tightly linked with the constraint of FW transport for the system axis is another place to look for physical problems. In this context, it is interesting to note that work has recently been done on EM self-forces without the need to make such assumptions [26].

One final possibility concerns the unmentioned force that prevents the dumbbell from collapsing (when e_A and e_B have opposite signs) or falling apart (when they have the same sign). If this force arises due to some field for which the two entities A and B are sources, a self-force effect can be expected due to this field. It is just possible that the component of this force that is not aligned or counteraligned with the acceleration might for some reason always cancel the offending component of the EM self-force. That seems unlikely, but it is important to remember that the model of the spatially extended particle is not physically complete without considering all the relevant forces.

Above all, it should be remembered that a problem only occurs here if we insist on taking a point particle limit. If all particles, no matter how fundamental, are actually spatially extended sources of whatever force fields they may generate, then we can live with the leading order term in the associated self-forces not being aligned with the acceleration. On the other hand, if this occurred for very small scale structures within such sources, large self-force terms would be expected and ought to be observed.

Chapter 11
Electromagnetic Radiation and the Coming of Age of the Equivalence Principle

The discussion in this section will range in a recycling and sometimes redundant way over the following:

- Uniformly accelerating charged particles.
- Supported charges in static homogeneous gravitational fields.
- Electromagnetic radiation from such charges.
- Equivalence principles.

The aim is to give an updated overview of [30], and to raise a few questions about this long-standing and sometimes heated controversy.

11.1 The Scenario

Consider first a charged particle coming down the x axis in a flat spacetime, slowing to a halt somewhere, then accelerating back up the x axis in such a way that its four-acceleration has constant relativistic length $a^\mu a_\mu$. This is eternal translational uniform acceleration, illustrated by the worldline in Fig. 11.1 (compare also with Fig. 2.4). Translational uniform acceleration means that the worldline is a hyperbola, as we have seen in Sect. 2.9, while eternal means that it goes on forever and has been going on forever.

Now accelerating charged particles usually radiate electromagnetic energy, so what about this point charge with hyperbolic motion? In order to find out, we have to solve Maxwell's equations for the fields, and luckily we always have the Liénard–Wiechert retarded solutions. In this case we find that fields are produced in the region $x+t \geq 0$, and we notice something interesting at the instant of time $t=0$, when the charge is instantaneously at rest in this particular inertial frame, namely that the magnetic fields are instantaneously equal to zero everywhere, but only for this instant of time. The same can therefore be said of the Poynting vector.

Of course, a Lorentz transformation can reduce any point on the worldline to rest, and as Pauli pointed out, the hyperbolic worldline looks exactly the same in

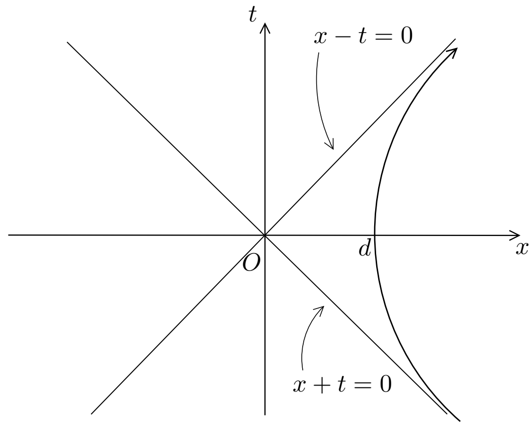

Fig. 11.1 A charged particle arrives from large positive x (*bottom right*), slows down to a halt at $x = d$ (in this frame), then accelerates back up the x axis. The worldline is asymptotic to the null cones at the origin, i.e., it is asymptotic to $x+t=0$ for large negative times, and $x-t=0$ for large positive times. Naturally, it never actually reaches the speed of light

any inertial frame, so the magnetic field and hence the Poynting vector are always zero everywhere, provided we keep changing inertial frame, so that we are always using the inertial frame instantaneously comoving with the charge. For this reason, Pauli suggested that the charge might not radiate [46].

But in any given inertial frame, the Poynting vector is going to change from zero as time goes by, and as far as we know, there is no relevance in what one would observe by continually changing inertial frame. However, in a well known non-inertial frame adapted to the motion of the charge, the magnetic field components are identically zero everywhere and at all times. As we know, for any timelike worldline in Minkowski spacetime, there are coordinates with special properties, said to be adapted to the worldline, and which we have called semi-Euclidean (SE) coordinates.

In summary, the idea is that, at each event on the given worldline, the accelerating observer borrows the hyperplane of simultaneity of an instantaneously comoving inertial observer and attributes her own proper time to all points on it. With this ploy and a few other simple tricks, we can arrange for the observer to sit permanently at the space origin of the new coordinate system, whence her worldline is just the time coordinate axis, with the time coordinate being the observer's own proper time. We can also arrange for the metric to have Minkowski form right along the time axis, but not off it. This means that the coordinate frame we are constructing is a tetrad frame along the worldline, but not off it (see also Sect. 13.1.3). And finally, by the ploy mentioned above, the geometry will be Euclidean on the constant time hypersurfaces, hence the name semi-Euclidean coordinates. But recall that the connection is *not* zero along the worldline, since it must encode the acceleration.

We can make another interesting observation here, which shows just how special this kind of motion is. It is only for an observer with eternal translational uniform acceleration that the Minkowski metric can be made to have a static form when expressed relative to SE coordinates (by choosing a Fermi–Walker transported space triad along the worldline). For any other kind of acceleration, if we carry out this kind of construction, the components of the Minkowski metric will depend on the new time coordinate or there will be nonzero components g_{0i}, $i = 1,2,3$. For a proof, see Sect. 2.3.8, and in particular (2.66) on p. 33.

11.2 The Fields

We now have the non-inertial coordinate system adapted to the charge motion, but we still need to be able to talk about magnetic fields in a situation where we are not using inertial coordinates. Physically, this is not so obvious, but mathematically it is very easy, because we have the electromagnetic field tensor $F_{\mu\nu}$ and we can express its components relative to any frame. We also know that the matrix of components of this tensor is antisymmetric in any frame so it can always be written in the form

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix} , \tag{11.1}$$

and then we can just read off **E** and **B**. What these mean physically for a general coordinate frame is another matter, to which we shall return.

So if we take the charge with hyperbolic motion, find its Liénard–Wiechert retarded fields relative to some inertial coordinate system, then transform them to the SE coordinate system adapted to the charge motion, we find that the SE magnetic field is permanently zero. In addition, the SE electric field is static, in the sense that, at any given SE space coordinate, the SE electric field does not change as SE coordinate time goes by. We may make another observation here that shows once again just how special this kind of motion is: it is *precisely and only* for the case of eternal translational uniform acceleration that this construction yields such an elegant and simple picture.

So maybe Pauli was right after all. Maybe there is no radiation of electromagnetic energy in this case. There is another reason for thinking that this may be so. If a charge radiates, we expect there to be a reaction force on it, but it turns out that the radiation reaction force is zero for translational uniform acceleration. This can be shown either from considerations of energy and momentum conservation [16], or by calculating the electromagnetic force an extended charge distribution exerts on itself when accelerated [32].

11.3 The Problem

Despite these arguments, it is generally agreed that the uniformly accelerating charge does in fact radiate, and we can of course calculate a radiation rate. That was the conclusion of Bondi and Gold in a paper they published over 50 years ago [4], but that raised another problem for them which is best introduced by a quote:

> The principle of equivalence states that it is impossible to distinguish between the action on a particle of matter of a constant acceleration or of static support in a gravitational field. This might be thought to raise a paradox when a charged particle, statically supported in a gravitational field, is considered, for it might be thought that a radiation field is required to assure that no distinction can be made between the cases of gravitation and acceleration.

So now we are talking about a gravitational field and an equivalence principle, and we are concerned about whether a static charge in a gravitational field should be able to radiate. This can be spelt out in the following way.

A static homogeneous gravitational field (SHGF) is usually modelled in general relativity (GR) by a metric interval of the form

$$\mathrm{d}s^2 = \left(1 + \frac{gy^1}{c^2}\right)^2 (\mathrm{d}y^0)^2 - (\mathrm{d}y^1)^2 - (\mathrm{d}y^2)^2 - (\mathrm{d}y^3)^2 \,, \tag{11.2}$$

where c is the speed of light and g a constant with units of acceleration (see Sect. 6.3, and also Chap. 12 for further discussion). The metric $g_{\mu\nu}$ is almost in the standard Minkowski form, except for the component g_{00}, which is a function of one of the space coordinates y^1.

Now it turns out that this is precisely the SE line element for an eternally uniformly accelerating observer with absolute acceleration g, i.e., $a^\mu a_\mu = -g^2$. So we can show that the curvature is zero and there are therefore no tidal effects, hence the name homogeneous for this spacetime. This is thus a flat spacetime, despite the fact that we are taking it to model a gravitational field, and we can show that there exist coordinates such that the metric assumes the Minkowski form everywhere and everywhen. Those would then be interpreted as the coordinates that would naturally be adopted by a freely falling observer.

So what was Bondi and Gold's problem? They do not believe that a charge sitting at fixed space coordinates in the SHGF should radiate. But the trouble is that general relativity, with a little help, really does predict that it should radiate. And the little bit of help is an equivalence principle. So if we do not believe that a charge sitting at fixed space coordinates in the SHGF should radiate, perhaps it is the equivalence principle that is wrong, or somehow inapplicable to charged particles.

11.4 Equivalence Principles

This is therefore a good point to recall the two equivalence principles that form part of any introductory course on general relativity. The first is usually called the *weak* *equivalence principle* (WEP), although there is nothing weak at all about it. In fact it forms part of the standard formulation of GR for any curved spacetime. We impose the metric condition, which says that the covariant derivative of the metric should be zero, and this fully determines the connection in the torsion-free case. It then turns out that the first coordinate derivatives of the metric components are linear combinations of the connection coefficients, so if we can arrange for the latter to be zero at some event P by a clever choice of coordinates, we will find that the metric components are slowly changing functions of the coordinates at that point.

We then have the following standard argument. For any event P in spacetime, there is always a choice of coordinates in some neighbourhood of that event for which the connection coefficients are zero at P and the metric takes the Minkowski form at P. By continuity and the above observation, this will then be approximately so in some small neighbourhood of P. Basically, WEP thereby guarantees the mathematical existence of local inertial coordinates at any spacetime event in the manifold and decrees that these correspond to the coordinates one would naturally set up in a freely falling, non-rotating laboratory (see also Sect. 6.5).

But we still need to be able to talk about electromagnetism in the framework of a general curved spacetime, and for this we need the *strong equivalence principle* (SEP). This states that, in the locally inertial frames whose existence is guaranteed by WEP, all physics looks roughly as it does in the context of special relativity. This is a rather vague statement and would be difficult to use. In practice, we take the special relativistic formulation of whatever non-gravitational physics it is we are trying to do, e.g., Maxwell's equations if we are doing electromagnetism, and replace all coordinate derivatives by covariant derivatives. At least, this is the simplest or minimal way to implement the strong equivalence principle. There are more sophisticated ways which will not concern us here. This then leads to the minimal extension of Maxwell's equations (MEME) to a general curved spacetime.

Now imagine a charge held at fixed SE space coordinates in an SHGF. It turns out that it is accelerating uniformly, and because of that, SEP tells us that it produces exactly the same fields in the global inertial frame that happens to be available in this case as a uniformly accelerating charge in a gravity-free spacetime. So if there is radiation in the latter case, there will also be radiation for the static charge in an SHGF. This is an application *par excellence* of the strong equivalence principle in the sense that there is no approximation here due to local effects, since the local inertial frame is globally inertial.

So if we think a static charge in a static spacetime cannot radiate EM energy, then here is another argument against the uniformly accelerating charge in flat spacetime without gravity being able to radiate EM energy. However, as mentioned earlier, the consensus says that it can. Alternatively, 'the' equivalence principle may be wrong. But WEP is built into standard GR, and GR would be virtually unusable without SEP, in the sense that we would require some other way of shipping our non-gravitational theories of physics into the curved spacetime framework. And another alternative is that a static charge in a static spacetime may after all be able to radiate EM energy.

We have now sketched the whole issue here. This is a tangle of at least three problems, probably more:

- Do eternally uniformly accelerating charges radiate EM energy?
- Does this debunk some form of equivalence principle?
- Can a stationary charge in a static spacetime radiate EM energy?

Let us begin by addressing the last question in more detail.

11.5 Stationary Charge in a Static Spacetime

A static spacetime is one in which there exist coordinates in which the metric interval assumes the form

$$\mathrm{d}s^2 = g_{00}(x^1,x^2,x^3)(\mathrm{d}x^0)^2 + \sum_{i,j=1,2,3} g_{ij}(x^1,x^2,x^3)\mathrm{d}x^i\mathrm{d}x^j \,, \tag{11.3}$$

so that the metric components do not depend on the time coordinate, and in addition the matrix of metric components is in block diagonal form with g_{01}, g_{02}, and g_{03} equal to zero (see Sect. 4.3.17). A static or stationary charge is one that sits at fixed space coordinates in a spacetime with one time coordinate and three space coordinates. This kind of staticity is clearly a coordinate dependent notion. To give an example, a static charge in a semi-Euclidean coordinate system is accelerating.

Now Bondi and Gold say that a static charge in a static spacetime cannot radiate EM energy. However, GR with the help of SEP and MEME predict that a freely falling observer will observe EM radiation from a static charge in an SHGF, if uniformly accelerating charges in gravity-free spacetimes do radiate. And since Bondi and Gold found that uniformly accelerating charges do radiate in the latter case, they had to come up with some other solution. And here it is: there is no such thing as an infinite static homogeneous gravitational field.

The point is that, if we do away with the SHGF, we may try to argue like this. Radiation from a charge can only be established, according to Bondi and Gold, by surveying space out to large distances. At any distance over which one could affirm the observation of EM radiation, the presence of a gravitational field would be revealed by its inhomogeneity. The EM effects do not then have to be the same as in an SHGF, and this is supposed to save the charged particle from having to radiate.

The weak point here is presumably the first claim, that one must be a long way from the charge in order to ascertain whether or not it is radiating. Of course, we must agree that the SHGF is unphysical, but this is all a matter of approximation, and there is no quantitative link between approximations to gravitational effects, which have one kind of source, viz., matter and energy, and approximations to EM effects, which have a quite different kind of source, viz., charges and the motions of charges.

Many people commented on this over the following 25 years, in particular Rohrlich, and we shall return to his views on these matters later. But in 1980, Boulware came up with a complete mathematical analysis of the EM fields due to a charge with eternal translational uniform acceleration [5], and he concluded that there would be EM radiation. So let us examine his arguments for reconciling these issues.

Boulware is interesting because there is a subtle change of tack here. He does not claim that a charge that is stationary relative to coordinates in which the metric is static will not radiate, only that an observer that is stationary relative to these coordinates will not be able to *measure* any radiation there is. So for a charge supported in an SHGF, there is radiation for the freely falling observer, but not for a co-accelerating observer sitting with the charge. Something is therefore telling him that the co-accelerating observer must not be able to see any radiation.

This seems to raise several questions. First of all, why should anyone want to show that? Here is a conjecture. Suppose I am holding a charged particle and moving inertially. Then I will not be able to tell what velocity I am moving at by looking at the EM fields of the charge. This is because Maxwell's equations are Lorentz symmetric. So perhaps the idea here is that I will not even be able to tell whether I am accelerating or not. The point is that I could be sitting still in an inertial frame and holding the charge and not see any radiation, or I could be accelerating uniformly and holding the charge and not see any radiation.

The problem is of course that this fails in the details, because the fields look different in the accelerating case, for any choice of coordinates the observer might make to express those fields. However, that does seem to raise another question: how do we know what accelerating observers will see?

Before returning to this question, let us consider the two arguments Boulware gives to try to support his claim (see the original paper [5] or the detailed discussion in [30, Chap. 15]). Figure 11.2 shows spacetime again. The cross represents the light cones at the spacetime origin. Now for reasons of causality, the charge can only produce fields in regions I and II, which is $x+t \geq 0$. But also for reasons of causality, if we travel with the charge, we can only get news from regions I and IV. We can never get news of the fields in region II. So we cannot send a friend into region II and he phones later to say that he is witness to some nice radiating fields produced by the charge we are travelling with.

In fact, we are stuck looking at the fields in region I, and Boulware tries to convince us that, if we do that, those fields will look more Coulomb than radiating, that is to say, they will look more $1/r^2$ than $1/r$ for a suitable choice of distance r. However, we should perhaps be asking whether the observer could not accurately predict the fields even into region II from sufficiently accurate measurement of the fields in the close neighbourhood of the charge worldline. And we may also wonder why we should care about what one particular observer can or cannot measure.

Boulware's second argument concerns the generalisation of the Poynting vector to the SE coordinate system, which is identically zero everywhere in region I, as pointed out earlier. This is often cited as definitive proof that the co-accelerating observer could not detect any radiation. But what is the physical meaning of this generalisation of the Poynting vector to coordinates other than inertial coordinates? According to Parrott, one of the main post-Boulware commentators on these issues,

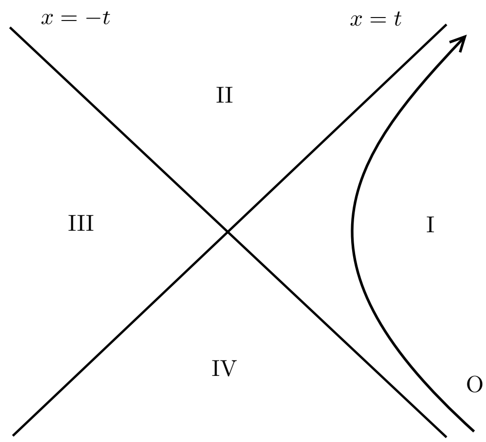

Fig. 11.2 Boulware's four regions of spacetime. The line $t = x$ is an event horizon, because the observer O can never receive any signal from regions II and III on the other side of it. Pictorially this is because the forward light cone of any potential signalling event in region II or III is entirely contained within those regions. For a similar reason, O can never signal to any event in regions III or IV, because the forward light cone of any point on the worldline of O is entirely contained within regions I and II

it does not give an energy flow at all when integrated over a spacelike hypersurface, but another radiated quantity, associated with a Lorentz boost Killing vector field [45].

11.6 Killing Vector Fields Revisited

The Killing vector fields play an important role in these discussions, so it is worth recalling the basics. As we saw in Sect. 4.3.17, a Killing vector field (KVF) is a vector field K such that the Lie derivative of the metric along the flow of K is zero. This is something very convenient from a mathematical point of view owing to the elegant formulation

$$K_{\mu;\nu} + K_{\nu;\mu} = 0 \,. \tag{11.4}$$

The flow of K is related to a symmetry of the metric, i.e., an isometry, so in a general curved spacetime, there are no Killing vector fields. However, in a static spacetime, there is always at least one Killing vector field, namely the time coordinate vector field for coordinates in which the metric assumes its static form (11.3).

But what can we do with a Killing vector field? In fact, if we also have a zero-divergence symmetric tensor $T^{\mu\nu}$, i.e., having the properties

$$T^{\mu\nu}{}_{;\nu} = 0 \,, \qquad T^{\mu\nu} = T^{\nu\mu} \,, \tag{11.5}$$

then we can construct a vector field

$$v^{\mu} := T^{\mu\nu} K_{\nu} \,, \tag{11.6}$$

and it is straightforward to show that this new vector field will have zero covariant divergence, i.e.,

$$v^{\mu}{}_{;\mu} = 0 \,. \tag{11.7}$$

It thus represents a conserved quantity, and we can use Gauss' theorem, and so on.

But, of course, the energy–momentum tensor of the EM field is symmetric and divergence-free in the right circumstances, so we can get a divergence-free vector field for every Killing vector field of the metric just by contracting with this energy–momentum tensor. This can be used to define the energy of a field in an inertial frame. The inertial time coordinate vector field in Minkowski spacetime is a timelike, normalised KVF, and it gives the density of field energy–momentum by contracting with the energy–momentum tensor.

Note, however, that not every divergence-free vector field constructed by contracting a KVF with an energy–momentum tensor can be interpreted as a density of field energy–momentum. There is a certain minimal requirement that the Killing vector field must be timelike and normalised at a given event for that to work. This is just the usual intepretation of the energy–momentum tensor in a general curved spacetime. If an observer has four-velocity u, then the contraction of u with the energy–momentum tensor is supposed to give the density of energy–momentum that this observer would measure using standard techniques. And of course, u is a unit timelike vector.

Now Minkowski spacetime is maximally symmetric, so it is absolutely full of Killing vector fields. In fact, it is absolutely full of Lorentz boost KVFs, since there is one in every space direction. Here is the one in the x direction:

$$K(x,t) := x\frac{\partial}{\partial t} + t\frac{\partial}{\partial x} \,, \tag{11.8}$$

expressed relative to an inertial coordinate system. Written like this, it may not look much, until we realise that every curve in the flow of this vector field is a uniformly accelerating worldline. Better still, when it is expressed relative to the SE coordinate system for an eternally uniformly accelerating observer in the x direction, it takes on the very simple form

$$K = \partial_{\tau} \,, \tag{11.9}$$

up to a multiplicative constant, where τ is the SE coordinate time. So it is basically the SE time coordinate vector field.

This shows that there is a close relationship between the Lorentz boost Killing vector fields and the SE coordinate systems of observers with eternal translational uniform acceleration. This is indeed what makes the latter kind of motion so very special in many respects. If we were to consider an observer with arbitrary timelike worldline, we could always construct a SE coordinate system, but the metric would not generally assume the static form (11.3), and this is because that worldline would not generally be the flow curve of any KVF.

So what about Boulware's second argument, concerning the generalisation of the Poynting vector to the SE coordinate system? According to Parrott, if the SE observer comoving with the charge uses the SE Poynting vector, she will not be calculating a flow of energy at all, but a flow of a kind of pseudo-energy constructed from the energy–momentum tensor and the Lorentz boost KVF by contracting the two. From a Minkowski standpoint, in terms of the flow of energy as it is usually defined in an inertial frame, the calculation of the SE observer looks very strange indeed [30, Chap. 16].

This may be so, but then we know that energy is a frame-dependent concept. We know how to transform the energy of a thing from one inertial frame to another by carrying out a Lorentz transformation of an energy–momentum four-vector, and then we get a different energy for the thing in each inertial frame. But here we are talking about non-inertial frames and this seems to raise several questions:

- How should an accelerating observer define energy?
- What energy would be measured by an accelerating observer using standard techniques?
- How should an accelerating observer define radiation?

Of course, as a vector, an energy–momentum four-vector can be represented relative to any coordinate or other frame, but here we are suggesting a *different definition* which favours the idea that the relevant quantity should be a conserved quantity.

11.7 Equivalence Principles Revisited

Before considering these questions from a different angle, let us just do a small detour and examine Boulware's equivalence principle, since we have unfinished business there. A striking thing about many of the papers that purport to be discussing the equivalence principle in this context is that they often give no usable statement of the equivalence principle. We usually have something like this: a uniformly accelerated frame must be indistinguishable from a gravitational field. But this kind of statement is clearly open to the kind of subjective interpretation we get from Boulware. What does *indistinguishable* mean? We are saying here that there might be a radiation field for one observer, but another one must not be able to see it. We must ask whether such an idea is really necessary.

Perhaps we are we just trying to save Bondi and Gold's opinion that a static charge in a static spacetime cannot appear to radiate? But this in turn seems to assume something about what constitutes energy and radiation in non-inertial frames, which brings us back to our earlier question. And such statements of 'the' equivalence principle are to be contrasted with WEP and SEP, which have fully objective definitions and fully mathematical implementations.

Let us return to the questions raised above by a slightly circuitous route. Parrott introduces an interesting idea of accelerating the charged particle by means of a tiny rocket with a tiny fuel tank and a fuel gauge for reading off how much fuel has been used. In gravity-free Minkowski spacetime, since the uniformly accelerating charge is radiating energy which can be detected and used, according to Parrott, conservation of energy suggests that the radiated energy must be provided by the rocket. We might then expect to burn more fuel to produce a given accelerating worldline than we would to produce the same worldline for a neutral particle of the same mass.

So according to Parrott, we have an experimental test to determine *locally* whether EM energy is being radiated or not. The key point here is indeed that such a test would be local. This is to be contrasted with Bondi and Gold's idea that we must be far away from the charge in order to find out whether or not it is radiating. In this case, we simply observe the rocket's fuel consumption.

But now consider a rocket holding the charge stationary relative to SE coordinates in the SHGF. If we burn more fuel to carry the charged particle (than to carry a neutral particle) when accelerating in the gravity-free Minkowski spacetime, we shall burn more fuel to support the charged particle in an SHGF, by an application *par excellence* of SEP. That is what the theory says if we accept the strong equivalence principle: the mathematics is strictly identical in the two cases.

But Parrott says that the equivalence principle does not apply to charged particles [45]. It is not absolutely clear what he means by the equivalence principle, because there is no clear statement of it in his paper. However, he does claim that local experiments will distinguish a stationary charged particle in an SHGF from an accelerated particle in a gravity-free Minkowski spacetime. And of course he may be right. One day we may be able to do this experiment, and we may find that he was right. In which case, we will know that the strong equivalence principle was not applicable here.

On the other hand, if SEP were not always applicable, we would be in some kind of trouble. How would we use GR? We would need some alternative way to ship our non-gravitational theories of physics into the curved spacetime context. And to save SEP, we need to admit that a stationary charge in an SHGF can radiate EM energy, at least as viewed from a freely falling frame.

But it should be said that, apart from Parrott, nobody seems to disagree with that. For example, Boulware and Rohrlich do not disagree with that. They just do not want the stationary charge in the static spacetime to *appear* to radiate to a comoving observer, for some reason. Recall the conjecture made earlier. Perhaps they consider this to be some form of EP, or an extension of a relativity principle to an accelerating situation. However, neither of these ideas are necessary to the system based on GR with WEP and SEP, and both fail in the details, since the fields in the accelerating case look different in the details for any choice of frame the accelerating observer may choose to represent them.

11.8 Which Frame?

The question remains: what will non-inertial observers actually observe? In the gravity-free Minkowski spacetime, if we use the SE Poynting vector, are we calculating radiated energy for some observer? Does that give the energy that would be measured by an accelerating observer using standard techniques, whatever that means? Or is it just a good definition? But if it is good, what is it good for? What are we trying to achieve? And does the value we obtain by calculating with the SE Poynting vector convert properly to the extra fuel that would be needed to accelerate a charged particle?

What should concern us here is that there is no obvious reason why an accelerating observer should adopt SE coordinates. After all, they are just coordinates, despite certain convenient features. They are also artificial in some ways. For example, the accelerating observer would have to use a rigid ruler, i.e., one satisfying the so-called ruler hypothesis, in order to actually measure the space coordinates in the direction of acceleration.

And why not use a tetrad frame [34]? Recall that the SE coordinate frame *is* a tetrad frame along the worldline, but not off it, and there are many ways to extend it to a tetrad frame off the worldline. But which one should we use to represent the EM fields off the worldline? More about this in Chap. 13.

11.9 Stationary Charge in a Static Spacetime Revisited

Let us now look more generally at the idea of a static charge in a static spacetime, but this time consider a general static spacetime, which may or may not be curved, as specified by the metric interval (11.3), asking once again whether it is true that a charged particle that is stationary with respect to the space coordinates in a static spacetime generates a pure electric field in that frame. Parrott gives a neat mathematical demonstration [45] that, if we have a stationary charge relative to coordinates for which the metric is static, there will be retarded field solutions for such a charge with zero magnetic field. Unfortunately, he has to assume that the electric field is static in order to derive the result, which weakens the argument somewhat, but let us gloss over that for these purposes. Then if the retarded field solution is unique, this means that the retarded field solution for such a charge always has zero magnetic field.

What is interesting about this argument is that it is entirely dependent upon the use of SEP and MEME! Of course, how else could we say anything at all about electromagnetism in a curved spacetime context?

Now the time coordinate in a static spacetime provides a timelike Killing vector field for the static metric, and it is not difficult to show that we may assume the KVF to be normalised along any given curve in its flow [see (12.3) on p. 376]. But then this KVF gives a conserved quantity in conjunction with the EM energy–momentum tensor by contracting the two together. So perhaps what a stationary charged particle in this static spacetime does not radiate is the pseudo-energy defined as the conserved quantity corresponding to translation by the formal time coordinate in this spacetime. And perhaps we should indeed define this as the energy for an observer sitting at fixed space coordinates in this spacetime. But if that is a good definition, let us not forget to say what it is good for. What are we trying to achieve? Where is the physics?

Going back to the flat spacetime context that we have been discussing here, Parrott agrees with Boulware that there is no radiation of the conserved quantity corresponding to the Lorentz boost Killing vector field. But he says that this is irrelevant to questions concerning physically observed radiation. And indeed, this is the case, until someone fills the physical gaps in these arguments.

Parrott also considers a stationary charge relative to the usual coordinates for Schwarzschild spacetime, which is a static spacetime, and asks whether it will radiate EM energy? Parrott claims that it would not, which is interesting, because this is exactly what one would say in a naive special relativistic (SR) version of gravity in which gravity is just a force. A stationary particle is inertial in SR, so Maxwell's theory says there will be no radiation.

In fact, it is interesting to contrast what GR and SR say about radiation from supported and freely falling charges, because they make diametrically opposite predictions about this. And this is because they make diametrically opposite claims about which of the two cases is actually accelerating. In GR, the supported charge is accelerating and the freely falling charge is not, while in SR, it is the freely falling charge that is accelerating.

But in order to understand the physical implications of a scenario in GR, we must first look at what is happening in the locally inertial frame guaranteed by WEP, and then deduce things about EM fields by applying SEP, since this is the only procedure we have. And when we look at the static charge in Schwarzschild spacetime as it would be described in a locally inertial frame, we find that it is accelerating, so there is then nothing obvious at all about the conclusion that this charge will not radiate. On the contrary, MEME says it will, at least to the freely falling observer.

Then if energy radiates out, and if it is true that this can be detected locally, it is tempting to consider that this must be supplied by whatever is holding the charge up against the gravitational effects. However, the zero radiation reaction in the case of eternal uniform acceleration confuses this issue. Boulware shows that there is a flow of energy in towards the charge in the case of eternal uniform acceleration, suggesting that this originates from the horizon $x+t=0$ (see the original paper [5] or the detailed discussion in [30, Chap. 15]). However, there is only a field on this horizon (in fact a distributional field) if the charge has been accelerating like this *forever*, whereas self-force calculations suggest that the radiation reaction force on the charge would be zero at any instant of time when it has uniform acceleration.

For an arbitrary, i.e., not necessarily uniform acceleration, there will be a radiation reaction, and we might then be able to argue that the radiated energy is somehow supplied by whatever is pushing the charged particle off its geodesic. It would be interesting to see concise discussions of this point.

Returning to the way energy is redefined for observers following the flow curve of a KVF, the conserved quantity corresponding to the time coordinate KVF in a static spacetime may be the only natural mathematical candidate for a conserved quantity. But is it what we would normally call energy physically? Or is it just a good definition, and if so, with what aim in mind? We ought to remember that mathematical convenience is not sufficient to be sure that we are doing physics, i.e., that we are getting a useful relationship with what is out there.

Furthermore, if a charge is stationary relative to some coordinates we happen to have chosen, we must remember that these are only coordinates. It will not generally be stationary relative to the kind of coordinates we are supposed to use to understand the theory physically, viz., inertial or locally inertial coordinates. And we should remember that GR is very different from SR as regards gravitational effects, since GR builds in an interaction of sorts between gravity and other fields via SEP. This is indeed how light is affected by gravitational effects in GR.

11.10 Interpreting Physical Quantities in Non-Inertial Frames

It is interesting to end this discussion by considering what Rohrlich has to say about these matters in his classic book [49], now in its third edition. Here we focus in particular on the way he suggests that we should interpret quantities expressed relative to non-inertial coordinate systems. A detailed discussion of all this can be found in [30]. Let us begin with a quote:

> [An SHGF] is a field whose lines of force are equidistant parallels, such as the gravitational field in the laboratory. It is known that this type of gravitational field can be simulated by uniform acceleration of a neutral particle in Newtonian mechanics and in special relativity. Is this also true for the motion of a charged particle?

So here we have a rather typical statement of an equivalence principle. Let us see how we get on with that.

He begins by presenting a tempting fallacy, and what is interesting here is to try to determine precisely what it is that he considers to be fallacious in the following statement:

> A neutral and a charged particle cannot fall equally fast in an SHGF, because the charged particle will radiate, being accelerated, and thereby lose energy, hence fall more slowly than the neutral particle.

But if the freely falling charge is accelerating, and if this is not the fallacy here, then it looks as though we are doing special relativity. Anyway, his statement of intent is now to prove that a charged and a neutral particle in an SHGF *will* in fact fall equally fast, despite the fact that, according to him, the charged one loses energy by radiation.

Before examining his argument, let us just note that the GR picture is exceedingly simple in this particular case, because the freely falling neutral and charged particles are stationary in the global Minkowski frame. The charged particle will not radiate, so this is a solution of the free particle equation of motion, i.e., we do not need to consider the kind of sophisticated arguments expounded in the classic paper by DeWitt and Brehme [13], which show that curvature can intervene directly in the equation of motion and change the whole notion of free fall for charged particles. Note also that this charge would only radiate in the SR picture!

Returning to Rohrlich's discussion, he thus sets out to prove that a charged and a neutral particle in an SHGF will in fact fall equally fast, i.e., have the same worldline or the same shape of worldline, despite the fact that the charged one, according to him, loses energy by radiation. His argument is basically GR+SEP leading to MEME, but note that he still claims that there is radiation. However, we then discover that the freely falling observer will not see any radiation, and this because the charge is just sitting still in an inertial frame.

So this is precisely the GR picture, and we begin to wonder what we must do in order to see this radiation. In fact, it turns out that we have to be stationary relative to the SE coordinates for the SHGF to see it. However, the field of the charged particle in the freely falling frame is Coulomb, so what we are claiming here is that a Coulomb field will *look like* a radiating one to an accelerating observer, whatever that means. But even if it did, is that how the accelerating observer should understand what is happening, by looking at the electric and magnetic fields relative to some coordinates that happen to be adapted to her worldline?

After all, these are not the only possible coordinates that such a person could use. There are other adapted coordinate systems, and there are tetrad frames that could be used to express the fields off the worldline. But which picture should the accelerating observer use?

We may consider another example of what seems unholy in this account of things. The geodesic equation in the SHGF says precisely that the four-acceleration of a thing is zero, and then we say that the thing is freely falling. Fiddling around with the coordinates will not make free fall in this flat spacetime, or indeed in any other spacetime, become a uniform acceleration, because it is zero acceleration. But what Rohrlich suggests here is that, if the supported observer using SE coordinates should somehow be duped into thinking that her coordinate system has any real significance, the freely falling particle may *appear* to have this or that acceleration. So in this view of things, coordinates can be taken by their resident observer, if there is one, to have some real physical significance.

It is interesting that Rohrlich should seek coordinates for the supported observer relative to which free fall looks like uniform acceleration, because free fall in an SHGF *is* uniform acceleration in the naive special relativistic model of gravity in which gravity is just a force [30, Chap. 3], another striking result concerning uniform acceleration.

Anyway, transforming the Coulomb field in the freely falling frame, which we know to solve MEME in the GR version of the SHGF, Rohrlich claims that we obtain a radiating field in the SE coordinates. So in his view the supported observer will 'see' this charge as radiating. He goes further, giving the standard formula for the radiation rate in ordinary Minkowski coordinates in SR, specifying how it is found algebraically from the components of the EM field tensor relative to such a frame and noting that this rate is Lorentz invariant. But what he then suggests is that, when we transform to arbitrary coordinates for this spacetime (no longer a Lorentz transformation), we can use the same algebraic combination of the new components of the EM field tensor to deliver a rate of energy radiation.

This is just to show how we can get a nonzero rate for the supported observer when the radiation rate is resolutely zero for the freely falling observer. But there are two difficulties here. First of all, the prescription is potentially ambiguous, given the various possible ways of expressing the fields off the worldline. But note that if the radiation rate depends only on a specific point on the worldline, i.e., a specific proper time of the charge, it may be possible to circumvent this difficulty, using the Lorentz invariance of the rate. On the other hand, this leaves us with the problem of interpreting the radiation rate calculated in this way. Is it supposed to be what the accelerating observer would measure using standard techniques, whatever that means, or is it just a good definition? And if so, what is it good for? What are we trying to achieve by it?

Here is an exercise for the reader. Think up an equivalence principle that would make you want the Coulomb field to look like a radiating field to an accelerating observer. One answer is a Newtonian, naive special relativistic, pre-GR kind of equivalence principle which we do not need, and which fails in the details, because this field will never look exactly like the radiating field of an accelerating charge, for any choice of coordinate system the accelerating observer may choose to express the Coulomb field.

However, this is not the end of the mysteries. In 1964, Mould invented an entirely theoretical radiation detector that would bear out such predictions [30, 39]. In other words, when it is moving inertially and there is no radiation, it does not record any radiation, and when it is moving inertially and there is radiation, it records radiation, but when it moves in uniform acceleration past a Coulomb field, it also excites. This is a striking result, if the theory in his paper is correct. There is a close parallel with the Unruh–DeWitt detector in quantum field theory (QFT) (see Chap. 14).

Rohrlich also asks how we know that a charged particle at rest relative to the supported frame will not radiate. That would be the prediction of an SR version of gravity. In GR, it is perhaps better to say that this charged particle will radiate and that the supported observer can spot this if she wants to! Rohrlich agrees that the freely falling observer will see the supported charge radiating at the well known constant rate. He then transforms the fields to the SE coordinate system and deduces that there is no radiation because the magnetic field is zero in the SE system, an argument we have already discussed twice here.

Such a claim is perhaps best answered by a question: should we treat the SE magnetic field as the kind of magnetic field we know and love from our school days? And let us note once again that, apart from having zero magnetic field and being static, the SE version of these fields does not look anything like the Coulomb field.

11.11 Conclusion

This discussion has raised many questions, but two in particular. First of all, how should we formulate our equivalence principles? A good rule might be to stick to WEP and SEP and forget any statements that talk about whether something can be distinguished from something else. The latter are likely to be Newtonian, naive SR, and pre-GR principles that are no longer needed in the GR framework, and liable to fail there.

Another question concerns the way we interpret quantities expressed relative to non-inertial coordinate systems. Here we should perhaps be clearer about whether we are interested in what accelerating observers actually measure, or whether we are just trying to make good definitions for them. But what will accelerating observers observe? What will they consider to be good definitions? And if they are good, what are they good for? What exactly are we trying to achieve? What will accelerating observers measure using accelerating detectors? Indeed, does it help to know what accelerating detectors will detect?

We should remember that there is a major theoretical difference between inertial motion and accelerating motion, both for observers and for detectors. When an observer is moving inertially, we know what are the best coordinates for such a person to use: they are inertial or locally inertial coordinates. This is because all our field theories of matter are Lorentz symmetric or locally Lorentz symmetric, and these are the coordinate systems in which they assume their simplest forms.

Regarding detectors, imagine designing two different detectors to measure the same physical quantity. Whenever they are moving inertially in the same physical context, we expect them to deliver the same value for whatever quantity it is they are supposed to measure. This is once again because all our field theories of matter, which govern both the internal constitution of the detectors and the environment of the detectors, are Lorentz symmetric or locally Lorentz symmetric. But what can we say when they are accelerating? Will they always deliver the same result for the given physical quantity? After all, there is no corresponding acceleration symmetry in our field theories of matter.

Note on the Exercise for the Reader

Consider a situation in the naive SR version of gravity, in which gravity is just a force, and we have Maxwell's theory in flat spacetime. Imagine an observer sees a set of particles of various masses all accelerating away with the same acceleration. She may construe this as a situation in which there is a gravitational field and the particles are in free fall while she is held up against fall (case 1), or one in which there is no gravitational field, the particles are moving inertially, and she is accelerating away from the particles (case 2).

A Newtonian EP based on the equality of passive gravitational mass and inertial mass claims that these situations will be indistinguishable to that observer. But what happens if the particles are charged? When they are freely falling, they will generate

EM radiation in the inertial frame of the observer, because they are accelerating in this SR picture. But if they are stationary in an inertial frame, they will produce Coulomb fields, and in order for the two situations to be ‘indistinguishable’, the accelerating observer in this second scenario must ‘see’ these Coulomb fields as radiating fields.

Unfortunately, whatever coordinates or other frame the accelerating observer uses in the second case, the Coulomb fields of the charges will not look exactly like the radiating fields of accelerating charges to an inertially moving observer, so that would nevertheless fail in the details.

What would be a GR version of this thought experiment? Actually, the two situations can be construed in many ways, but the one to watch is this. If we consider case 1 to be free fall of the particles and the observer supported against free fall, and case 2 to be inertial motion of the particles and acceleration of the observer, then the two scenarios are identical in GR and there is no more to be said about distinguishing them.

Chapter 12
KVFs, SHGFs, and Uniform Gravitational Fields

Static homogeneous gravitational fields (SHGF) were first discussed in Sect. 6.3, but mentioned again in Sect. 11.3. There is some doubt in the literature as to whether they really are homogeneous. For example, in his classic paper, analysed at length in [30], Boulware tries to derive the metric interval

$$\mathrm{d}s^2 = \left(1 + \frac{gy^1}{c^2}\right)^2 (\mathrm{d}y^0)^2 - (\mathrm{d}y^1)^2 - (\mathrm{d}y^2)^2 - (\mathrm{d}y^3)^2 \ , \tag{12.1}$$

given on p. 356, usually taken to model an SHGF in general relativity, from some simple assumptions [5]: invariance under time translations for some time coordinate, invariance under the Euclidean group E_2 of translations and rotations in a 2D spatial coordinate plane perpendicular to the other spatial coordinate axis, and flat (zero curvature). Interestingly, he never refers to this spacetime as homogeneous, only static. This is presumably because later on he claims that an extremely strong gravitational field produces the future event horizon $z = t$. However, this has to be a misunderstanding, because there are no tidal effects, an assumption he makes explicitly, of course, by assuming zero curvature.

One can see why he might make claims about an extremely strong gravitational field. He is imagining observers supported at certain coordinate values. However, that has nothing to do with the strength of any gravitational field. Coordinates are just coordinates. What makes this spacetime homogeneous, and thus justifies the appelation SHGF, is the assumption that there are no tidal effects. Someone can set up funny coordinates, hold a set of observers there, and conclude that they are held up against a gravitational field that could be varying in quite a wild way, even in a flat spacetime, if that were how we defined the gravitational field. This was the point about 'stationary' charges in gravitational fields discussed in Sect. 11.5. That is a coordinate-dependent notion.

Of course, if one can see the source of a gravitational field, as happens in the case of the Schwarzschild spacetime when it is taken to describe the gravitational field around a massive spherical body, one could perhaps take proper distance from the source, relative to some 'natural' choice of spacelike hypersurfaces, to define stationarity. That was certainly the pre-relativistic way of viewing things. This issue does therefore raise a question of what it means, if anything, to be at rest in relativity theories.

There is another spacetime which is sometimes taken to model a *uniform* gravitational field and which may also be a good candidate for homogeneity [29]. Actually, Boulware makes the claim that there do not exist static coordinate systems in which bodies at rest at different points undergo the same proper acceleration [5], neglecting to say that he assumes flatness! As a matter of fact, the uniform gravitational field can be derived from the general class of spacetimes he considers at the beginning of his paper [30]. This is the class with metric interval of the form

$$\mathrm{d}s^2 = \mathrm{e}^{2\phi(X)}\mathrm{d}T^2 - \mathrm{d}X^2 \,, \tag{12.2}$$

which includes (12.1) when the smooth function ϕ has the form

$$\phi(X) := \ln(1+gX) \,,$$

for some constant g. For any smooth ϕ, it is straightforward to show that [30]:

- $\partial/\partial T$ is a Killing vector field.
- The proper acceleration of an observer at coordinate rest is $\phi'(X)$.
- The curvature scalar of the metric is $\phi'' + \phi'^2$.

The uniform gravitational field is the one with $\phi(X) := gX$, on the grounds that all observers sitting at fixed coordinates have the same proper acceleration g. The curvature in this case is also constant, taking the value g^2 throughout spacetime, so this is not a flat spacetime unless $g = 0$. In that sense, this gravitational field, whether considered homogeneous or not, is not an artefact of the choice of coordinate system, as one might accuse the SHGF with metric interval (12.1).

Certainly, one could describe this spacetime as coordinate-homogeneous for the coordinates (T, X). But do these coordinates have any real physical significance, e.g., because of other properties of the metric with respect to these coordinates? Of course, we may perfectly well define that to be a homogeneous spacetime if we want, but the question is always whether one definition is better than another, and that depends on what we want to achieve.

In GR, coordinate independence of an object is taken as a sign that it may correspond to some real physical entity. Such objects are tensors. (A given tensor has different representations depending on the choice of coordinates, but the tensor remains independent of the choice of coordinates.) However, we do set up coordinate systems in the real world in order to be able to describe it, and almost everything we actually measure is coordinate dependent. But the 'solution' proposed here for attributing physical meaning in all these coordinate-dependent situations is to do so only when the coordinates are inertial (in flat spacetime) or locally inertial.

So there is an issue here, for example, about what it means to be at rest. As one would expect, the notion of being at coordinate rest could hardly be physically significant in itself. But there does exist a more geometric notion of being at rest which could possibly be attributed physical meaning, and indeed often is. If $\mathscr{M}$

is a spacetime with a timelike Killing vector field (KVF), say T, then the integral curves of T are the orbits of a one-parameter family of isometries, so in a sense these worldlines might be said to represent an observer who does not see the spacetime geometry changing.

Would such a person be able to justify saying she was at rest, and thus be able to interpret her four-acceleration in terms of a gravitational force? Is this not precisely what happens in a Schwarzschild spacetime, noting that ∂_t is a timelike KVF for the usual coordinates in such a spacetime, with integral curves

$$\left\{(t,r,\theta,\phi) : t \in \mathbb{R},\ r,\theta,\phi \text{ constant}\right\}$$

for each choice of r,θ,ϕ, whereupon we can establish a relationship between GR and the Newtonian approximation? Here is an idea that does not depend on any notion of keeping some proper distance fixed, being firmly based on the very geometry of the spacetime.

But then we have the problem of Minkowski spacetime with its multitude of KVFs, and in particular, its multitude of Lorentz boost KVFs, all associated with different four-accelerations, whence a different choice of KVF may lead us to decree that a different force is required to support us 'at rest' in the gravitational field. In answer to this, one might say that it was analogous to the Newtonian situation, where we can have a 'real' gravitational force and a 'fictitious' inertial force due to the coordinates adapted to one's motion. One would just have to bear in mind that the 'gravitational force' attributed in this way is a quantity associated with a particular choice of timelike KVF.

The lengthy wording in the last paragraph reminds us that there is an important switch of terminology in GR because a freely falling object has zero four-acceleration, so one prefers to say that there is no force on it. In other words, gravity itself is not a force in this picture. There is a force of some kind when the object is pushed off its geodesic, e.g., when it is supported in a gravitational field. An example would be an object held at fixed space coordinates in the usual coordinate system for Schwarzschild spacetime, or at fixed SE space coordinates in an SHGF with metric interval (12.1). When we take an object on our outstretched palm, we prevent it from falling, hence push it off its geodesic, by exerting an upward force on it.

In the SHGF, the problem is that one could deduce any value for the gravitational field, by supporting objects at different fixed SE space coordinates for some TUA observer. This is just the remains of the old Newtonian equivalence principle mentioned in the note at the end of Chap. 11 (see p. 369). Such effects do not have the absolute status of tidal effects as instantiated by nonzero curvature and associated geodesic deviation, which are usually taken as the sign of a somehow more genuine gravitational effect.

But does this mean that we really can talk about the gravitational field of the Rindler wedge (region I in Fig. 11.2) and say that it becomes stronger and stronger as one approaches the surface $z = t$? The trouble is that this requires us to impute physical meaning to coordinates, because it requires us to impute physical meaning to the way we have chosen to support different objects at different locations in the spacetime.

Regarding homogeneity, it can be viewed as genuinely geometric, rather than just a coordinate based idea [29]. A spacetime is homogeneous if, given any two points p,q, there is an isometry ϕ such that $\phi(p)=q$. Minkowski spacetime clearly has this property, and it is not alone in this. A spacetime is spatially homogeneous if it can be written as a product $\Sigma \times \mathbb{R}$, where we denote $\Sigma \times \{t\}$ as Σ_t, such that for any t, and any $p,q \in \Sigma_t$, there is an isometry ϕ of the spacetime such that $\phi(p)=q$. This is satisfied by many cosmological solutions, and indeed it is built into the Robertson–Walker cosmologies at the outset.

Note then that the spacetime with metric interval (12.2) and $\phi(X) := gX$, referred to above as modelling a uniform gravitational field, is not homogeneous in this sense: X translations are not isometries, because $\partial/\partial X$ is not a KVF. However, both X and T translations preserve the integral curves of $\partial/\partial T$ and their proper accelerations, so at a pinch one might claim that there is a sense in which the gravitational field, if not the metric geometry, is homogeneous. The coordinates seem to have some kind of geometric significance, but how could one claim any physical significance for them? The issue here is not really over how one defines homogeneity, but whether associated concepts have any physical interest. And it is striking that all observers at coordinate rest for these coordinates have the same proper acceleration of magnitude g. The gravitational field, if it is established by these observers, is homogeneous, even if the metric geometry is not. But note that the Einstein tensor is not zero anywhere, so this is not an empty spacetime. What could give such a matter–energy distribution? What kind of energy condition would it satisfy? Are we doing real physics with such a field?

Naturally, one expects being at rest to be a problem in relativity theories! It is thus surprising how much store is laid by it, apparently, in discussions of EM radiation. The explanation must certainly be related to the idea that geometry is unchanging when one follows the flow of a KVF. We shall see this claim again in the context of the Unruh effect (see Chap. 14). One can solve the Klein–Gordon scalar field equation in the semi-Euclidean coordinate system of an observer with eternal translational uniform acceleration, obtaining positive and negative frequency solutions for the proper time of that observer (provided one distributes that proper time appropriately, in the SE coordinate way, over events off the observer worldline). One can obtain a complete set of such Rindler modes and expand the quantum field in the usual way in terms of those functions, defining creation operators as the coefficients of the negative frequency modes and annihilation operators as the coefficients of the positive frequency modes, and thereby making a whole new construction of the quantum field theory (related to the usual one in a certain way that can be established). We then see claims like this [11]:

> The quantisation of the scalar field in Rindler modes is just the natural quantisation a Rindler observer would perform. To him spacetime always looks the same, so he ‘feels’ at rest just like an object sitting on the surface of an infinite flat Earth.

We should baulk at this glib way of putting things. The only sure way of constructing anything is to use WEP and SEP, because they are indeed necessary and sufficient, and observer independent, ways of getting the whole of physics in the curved spacetime context. And the procedure these principles suggest is always to interpret things physically in locally inertial frames.

This is true despite the fact that SEP can be implemented in different ways. This principle requires curved space laws to reduce to flat space laws when the metric is flat, and there are many ways to implement such a principle. The simplest or minimal way is to replace coordinate derivatives by covariant derivatives in the flat space laws, and then we can add all kinds of functions $f(\kappa)$ (κ being curvature) provided that $f(0)=0$, or multiply by others $g(\kappa)$ such that $g(0)=1$, and so on. Naturally, we begin with the simplest hypothesis, until such times as predictions no longer accord with measurement, and then tweak that in some intelligent way if that helps. The fact that there are many possibilities may look bad, but it is not. When the time comes, we will see intuitively what kind of extra curvature dependent term is needed, if that is the solution, because we will know more about the problem.

The gravitational aspect is crucial to the KVF version of rest put forward here. Otherwise we can have uniformly accelerating observers in a gravity free spacetime and say they are at rest, which sounds strange. What could stop that seeming strange would be the presence of a gravitational field which those people consider they are being supported against, although here again, the idea of being supported against a gravitational field cannot be made precise unless there is some notion of proper distance from the source. One should ask how a KVF picture is superior to one using the kind of quasi-stationary quasi-canonical coordinates generally employed to obtain Newtonian gravitational theory as an approximation from GR [14, 30].

In a context where one sees the source, proper distance from it does seem to be an important factor, and this is something completely lacking in the idealistic flat spacetime case generally referred to as an SHGF. Is there a source for that? It is easy to find one in Newtonian gravity theory, but what about in GR [30, Chap. 2]? Everything seems to work fine in the Schwarzschild case, for example. With the usual coordinates, when we sit at fixed r,θ,ϕ, we are on the flow curve of a KVF, we are coordinate stationary, and our proper distance from the source is constant for proper distance as suggested by this coordinate system. Here are three reasons one might invoke for claiming to be at rest! And yet one must be accelerated in order to achieve this state! This is surely a counterintuitive aspect of the GR picture.

But it remains problematic to claim that we really can talk about the gravitational field of the Rindler wedge as becoming stronger and stronger as one approaches the line $z=t$, given that there are no tidal effects here. It reveals the limitations of this way of viewing things, at least in this highly idealistic situation. Indeed, it is probably a reason for not bothering with observers and what they think. If anything, it shows that we need to use the SHGF only as an approximation in regions where curvature can be treated as roughly constant.

Those who do talk about observers are looking for some kind of explanatory picture. Indeed the KVF picture purports to be an explanatory picture for an observer. But what exactly are we aiming to do with such pictures, e.g., for an observer moving with a detector? What is the observer apart from the detector? This was exactly the problem raised in Chap. 11. It is confounded by the fact that detectors are likely to be unreliable when accelerating (either in SR or in GR) since there is no acceleration symmetry in our fundamental field theories of matter to match the velocity symmetries (Lorentz symmetries) they have. Two different detectors designed to measure the same physical quantity will always deliver the same value of that physical quantity when in free fall under the same physical conditions, but they are unlikely to do so when accelerating in any way under the same physical conditions.

Another weak point about KVF pictures is that they only work for people following the flow curve of a KVF, i.e., not even for someone with arbitrary motion in a flat spacetime. And we should also be concerned about the fact that they cannot be normalised everywhere? Supposing there is an observer following one of the flow curves, we can always arrange for the KVF to be normalised to unity along that curve so that it is always equal to the four-velocity of our observer, since the pseudolength $\sqrt{K^2}$ of a KVF K is always constant along a flow curve:

$$K^\mu \nabla_\mu (K^2) = 2K^\mu K^\nu \nabla_\mu K_\nu = 0 \,, \quad \text{because } \nabla_\mu K_\nu + \nabla_\nu K_\mu = 0 \,. \tag{12.3}$$

But then the KVF will not normally be normalised along any other flow curves, so somehow, those other curves do not quite represent observers. This happens precisely because the flow curves are all accelerating worldlines, so it is a critical feature for Killing observers anywhere, even in flat spacetimes. To see this, note that, if we could normalise K everywhere, so that $K^2 = 1$, we would have $K^\mu K_{\mu;\nu} = 0$, hence $K^\mu K_{\nu;\mu} = 0$ when K is a KVF, whence the flow curves of K would all be geodesics, with no acceleration.

We see this for the Lorentz boost KVFs in Minkowski spacetime. The main observer at the space origin of the SE coordinate frame can arrange for the time coordinate to be her proper time, but then this time coordinate, distributed in the usual way by the SE coordinate system to other SE space points, will never be the proper time of any other observers following other flow curves of this same KVF. This is worth mentioning because the so-called Rindler observer perspective about which so much is made in the context of the Unruh effect (see Chap. 14) is built upon this particular KVF-motivated way of distributing the proper time of the main observer to other SE space points. This perspective, if it is one, is highly coordinate dependent in that sense, whether there be an underlying KVF structure to the coordinates or not. And as already mentioned several times, the whole approach leaves us in the lurch for observers with other timelike worldlines.

So normalising the KVF is important if it is a natural picture we are after. In general, only one observer can arrange for his parameter to be his proper time. The fact that all the others going with the KVF flow are forced to use some other parameter than their own proper time means that no one can pretend that this is somehow an inertial coordinate frame. And this happens precisely because there is acceleration.

Of course, each flow curve of a KVF can be used as the spatial origin of a coordinate system, with proper time for time. So an observer on one flow curve can use his proper time, and an observer on another can use hers, and although their time coordinates do not agree, they can still communicate because the relationship between the time coordinates is perfectly well-behaved. This is not a problem. There is a quite general sense in which coordinate systems in GR serve largely only to communicate. But there is an interesting point here if we think about inertial observers rather than accelerating ones: observers moving with an inertial observer in her own frame would be inertial, with synchronisable times.

We said above that the gravitational aspect is crucial to the KVF notion of rest, because otherwise we can have uniformly accelerating observers in a gravity free spacetime and say they are at rest, which sounds strange. One reply to this is that we should be willing to accept such strangeness. At this level of understanding of the world, why should it not look strange? Perhaps our brains never evolved to understand such things. Such a problem occurs in some quantum theories of microphysics. However, in the latter case, rather than saying that we understand a strange theory, we might be better admitting that we have a theory that works but we do not understand it, or maybe could never understand it.

On the other hand, such a stance does not seem appropriate in the present case, because there is nothing particularly mysterious about spacetime. Is this really a case where we need to re-educate or refine our intuition? What we have here is just a strange definition of being at rest, and it is strange because it does not accord with our usual understanding. Being at rest has something to do with distances to other things, and in gravitational fields, those other things can only be the sources of the fields. If saying that observer A is at rest is just another way of saying that observer A's worldline is the integral curve of a timelike KVF, on the grounds that we have nothing better to replace it by, then that is fine, of course, and there would be no point just arguing about words.

Now an observer at rest some distance from a spherically symmetric mass distribution has a four-acceleration in GR which to a high degree of approximation matches the force required to hold an analogous observer at real, Newtonian rest relative to the analogous mass distribution in Newtonian gravity. In this context, how is a KVF picture superior to one using quasi-stationary quasi-canonical coordinates (see [14, Chap. 6] or [30, Chap. 2])? One could argue that it is conceptually cleaner, in that it really is a bit of intrinsic geometry, and not coordinate dependent. And are we not supposed to find intrinsic geometric ideas more fundamental than coordinate-based ones?

This is a well worn groove in the textbooks. But surely what is fundamental is the way we relate coordinates to measurements in the real world, and that passes in principle (WEP and SEP) by locally inertial frames. The philosophical idea behind an intrinsic geometric idea is presumably that it corresponds to a thing in itself, a real thing out there that we can only know through the shadows it casts, or something like that. The manifold itself, for example, corresponds to the Universe as it is, not just some description of it. That is certainly a fine idea, but we still need to relate our descriptions of the manifold to descriptions of what is out there. Geometric elegance, it should be remembered, is mathematics. Our problem is to relate that to the real world.

As mentioned above, a common derivation of the link between Einstein's equation and Newtonian gravity goes through the assumption that there are quasi-canonical quasi-stationary coordinates, which leads to a very general argument (for any spacetime), but quasi-stationarity is founded on the understanding that the coordinate system is fixed in some intuitive sense with respect to gravitational sources (see [14, Chap. 6] or [30, Chap. 2]). It is not clear that we could really make the link with the Newtonian approximation without at least intuitively referring to proper distance to the source. If one were not somehow fixing the proper distance to the source, the coordinates one came up with here might be free fall coordinates, in which case there would appear to be no gravitational effect at all.

What do we actually gain by the Killing vector idea? Is it just a picture? If the starting point is always SEP, i.e., the selection principle for which theories are counted as plausible if they look reasonable when expressed in a free-falling inertial frame, why make that coexist with another principle based on KVFs, which is very likely incompatible with that, or can at best merely agree?

Non-Euclidean geometry is not necessarily the problem here for our intuitions. The issues arise identically in flat spacetime. For example, why pretend that a uniformly accelerating rigid frame is an inertial frame? We do not need to believe that. And anyway, accelerating frames are not inertial, they are accelerating, as attested by the non-disappearance of the connection coefficients along the worldline of the observer who uses them. There is a real mathematical difference with inertial frames, even in flat spacetime.

Of course, we can see what the theory GR+SEP says about the picture a Killing observer would obtain. However, a Killing observer is not an inertial observer unless the flow curve of the KVF happens to be a geodesic. The problem is perhaps not counterintuitive results, but hoping for something to be true which is not in general. After all, the equivalence principle (WEP+SEP) provides a map everywhere, and everyone uses it (although of course it may ultimately be wrong), and that map says that the picture for a Killing observer will not look like the picture for an inertial observer in general, because the Killing observer is usually accelerating.

The main advantages of the WEP and SEP solution are:

- There are no observers.
- The only coordinates mentioned are the canonical local inertial coordinates provided by WEP, and they do not actually need to be mentioned to implement that programme.
- We do not need anything else.

As an example, consider the geodesic 'principle' which says that a point test particle subject to no external non-gravitational effects will follow a geodesic. This and all variants for spinning particles or charged particles or spatially extended particles or whatever, follow from WEP+SEP together with the special relativistic versions of Newton's first and second laws and the appropriate field theories of matter (see Sect. 8.7). Note that SEP is needed for all test particles, massive or massless, the physical input being Newton's laws and the appropriate field theories of matter for massive particles and Maxwell's theory, for example, for photons. And as discussed in Chap. 6, without SEP, GR would be useless. Even the usual interpretation of $g_{\mu\nu}$ as a metric field would just have to be posited, and it usually is, whereas SEP allows us to model specific clocks and rulers to show that they will measure the metric field to this or that level of accuracy, thereby justifying the usual interpretation of the metric field.

The KVF idea is just that the metric looks the same in some sense as one moves along a flow curve of the KVF. This is also true in another sense (not the Lie derivative sense) if one moves along a geodesic and adopts the right coordinates, although something intrinsic like the spacetime curvature may vary in that case. But flat spacetime seems to illustrate a problem with our trust in KVFs. Which KVF should we use? A free fall KVF whose flow curves are geodesics or a Lorentz boost KVF? Would most people not opt for the former? In the latter case, we will not be able to arrange for comoving coordinates relative to which the connection coefficients are zero on the worldline. The metric looking the same thus seems a long way from saying one is somehow at rest. After all, one might always be able to claim that one is at rest, despite the fact that one presumably feels accelerations.

The upshot of all the problems discussed here is just that acceleration is not inertial motion. Are we sure that we are taking accelerating motion seriously? What good are these accelerating frame pictures? Do we need them?

Chapter 13
Using Tetrad Fields as Accelerating Frames

The aim here is to discuss a nice idea for improving our concept of accelerating frame, exposed in a paper by Maluf and Ulhoa entitled *Electrodynamics in accelerated frames revisited* [34]. They suggest expressing the Faraday tensor and electric and magnetic fields relative to tetrad fields, in order to decide:

- whether the Coulomb field of a stationary charge in an inertial frame can appear to radiate to an accelerating observer,
- whether the EM fields of an accelerating charge can appear Coulomb, or at least appear not to be radiating to a coaccelerating observer.

They claim to be able to resolve what they refer to as a paradox: the idea that the fields of the stationary charge in the inertial frame might look to an accelerating observer the same as the fields of an accelerating charge would look to an inertially moving observer.

Is this a paradox? Why should those two things look the same? One reason, discussed at length in [30] and reviewed in Chap. 11, is that it would save what is actually an unfounded intuition about EM radiation by charged particles in static spacetimes. The point was that the metric components for a static homogeneous gravitational field (SHGF) look, in the usual coordinates for such a spacetime, like the Minkowski metric components expressed relative to semi-Euclidean (Rindler) coordinates adapted to a uniformly accelerating worldline. Now it is widely held, although the assumption is unfounded, that a stationary charge in a static spacetime cannot radiate. If this were true, and if we take stationary to mean not moving relative to the usual coordinates for the SHGF spacetime, then an application *par excellence* of the weak and strong equivalence principles tells us that a uniformly accelerating charge in a gravity-free Minkowski spacetime cannot radiate in the comoving SE coordinate frame.

Note also that anyone who really wants to save the unfounded intuition just mentioned is compelled either to question the transfer from the SHGF spacetime to the gravity-free spacetime, i.e., the relevant equivalence principles, or perhaps to show that the radiating accelerating charge in the gravity-free spacetime does not appear to be radiating when we coaccelerate with it. But would the latter really satisfy someone who feels intuitively that a stationary charge in a static spacetime cannot radiate. They would have to admit that it does radiate for a freely falling observer, whatever happens, if the weak and strong equivalence principles are valid.

In any case, at this point there enter subjective versions of the equivalence principle which talk about what can or cannot be distinguished, and as we saw in Chap. 11, some commentators try to convince themselves that it is enough for some observer coaccelerating with a uniformly accelerating charged particle in gravity-free Minkowski spacetime to be physically unable to measure radiation which is in fact there.

That brings us back to the paradox for Maluf and Ulhoa. But whatever one feels about the above debate, these authors point out that it may not be the best policy, when trying to understand what the coaccelerating, or any accelerating observer would observe, to refer to coordinate frames. They advocate orthonormal frames, or tetrads. The aim here is to examine the relationship between semi-Euclidean coordinates in Minkowski spacetime and various possible tetrad fields, then look at the choice made by these authors and ask whether one really does solve the problem of what an accelerating observer would observe.

In Sect. 13.1, we examine the orthonormal frame associated with the SE coordinate frame adapted to an observer with translational uniform acceleration, and in Sect. 13.2, an orthonormal frame associated with a field of uniformly accelerating observers all of whom have the same uniform acceleration. Section 13.3 discusses the general problem of extending a tetrad defined along a worldline to some neighbourhood of that worldline and comments in particular on the choice made by Maluf and Ulhoa. Section 13.4 then obtains the components of the Faraday tensor relative to the tetrad derived from the SE coordinate frame and shows that the magnetic field is zero for that tetrad.

13.1 Tetrad for Semi-Euclidean Coordinate Frame

Here we consider the TUA observer of Sect. 2.9 and examine the associated SE coordinate frame in order to obtain a natural tetrad from it. In fact, we shall see that the coordinate frame is orthogonal, and we only need to normalise one of the vector fields, viz., the temporal coordinate vector field, in order to obtain a tetrad.

According to (2.230) on p. 78, the transformation from SE coordinates (y^0, y^1) to Minkowski coordinates (t, x) is

$$\begin{cases} t = \dfrac{c}{g}\left(1 + \dfrac{gy^1}{c^2}\right)\sinh\dfrac{gy^0}{c^2}\ , \\ x = \dfrac{c^2}{g}\left(1 + \dfrac{gy^1}{c^2}\right)\cosh\dfrac{gy^0}{c^2} - \dfrac{c^2}{g}\ , \end{cases} \tag{13.1}$$

where c is the speed of light and g a constant with units of acceleration.

13.1.1 Uniformly Accelerating Worldline at $y^1 = 0$

The curves with constant y^1 are uniformly accelerating worldlines. For example, when $y^1 = 0$, we get the worldline

$$t(y^0) = \frac{c}{g}\sinh\frac{gy^0}{c^2}\,, \qquad x(y^0) = \frac{c^2}{g}\left(\cosh\frac{gy^0}{c^2} - 1\right)\,, \tag{13.2}$$

parametrised by y^0. We can eliminate y^0 to get x as a function of t along the world-line:

$$x(t) = \frac{c^2}{g}\left[\left(1+\frac{g^2t^2}{c^2}\right)^{1/2} - 1\right]\,. \tag{13.3}$$

This implies

$$v(t) := \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{gt}{(1+g^2t^2/c^2)^{1/2}}\,, \tag{13.4}$$

from which we can find the proper time $\tau(t)$ along the curve, setting it to zero when $t = 0$. We have

$$c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 = \mathrm{d}t^2\left[c^2 - v(t)^2\right]\,,$$

and, after a short calculation,

$$\left(c\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2 = c^2 - v(t)^2 = c^2 - \frac{g^2t^2}{1+g^2t^2/c^2}\,,$$

implying finally that

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{1}{(1+g^2t^2/c^2)^{1/2}}\,. \tag{13.5}$$

Hence,

$$\tau(t) = \int_0^t \frac{1}{(1+g^2t^2/c^2)^{1/2}}\mathrm{d}t\,. \tag{13.6}$$

Making the substitution

$$t = \frac{c}{g}\sinh\alpha\,, \qquad \frac{\mathrm{d}t}{\mathrm{d}\alpha} = \frac{c}{g}\cosh\alpha\,,$$

we then find that

$$\tau(t) = \frac{c}{g}\sinh^{-1}\frac{gt}{c}\,, \tag{13.7}$$

or

$$t(\tau) = \frac{c}{g}\sinh\frac{g\tau}{c}\,. \tag{13.8}$$

Looking back at the first relation of (13.2), we see that

$$\tau(y^0) = \frac{1}{c}y^0\,, \tag{13.9}$$

along this worldline, i.e., the SE time coordinate is, up to a factor of c, the proper time of an observer following this worldline through the origin of the (t,x) coordinates. Furthermore, when

$$t(\tau)=\frac{c}{g}\sinh\frac{g\tau}{c}\ ,$$

the equation (13.3) for the worldline can be parametrised by τ as

$$x(\tau)=\frac{c^2}{g}\left(\cosh\frac{g\tau}{c}-1\right)\ . \tag{13.10}$$

This allows us to calculate the four-velocity $v(\tau)$ of this worldline:

$$v(\tau):=\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix}t(\tau)\\x(\tau)\end{pmatrix}=\begin{pmatrix}\cosh(g\tau/c)\\c\sinh(g\tau/c)\end{pmatrix}\ . \tag{13.11}$$

This has pseudolength c.

13.1.2 Uniformly Accelerating Worldline at $y^1=\kappa$

We now examine the constant y^1 curve obtained for some value $y^1>0$, following exactly the same pattern as in the last section. The path of a point at fixed $y^1=\kappa>0$ is

$$\begin{cases}t(y^0)=\dfrac{c}{g}\left(1+\dfrac{g\kappa}{c^2}\right)\sinh\dfrac{gy^0}{c^2}\ ,\\[2ex] x(y^0)=\dfrac{c^2}{g}\left(1+\dfrac{g\kappa}{c^2}\right)\cosh\dfrac{gy^0}{c^2}-\dfrac{c^2}{g}\ .\end{cases} \tag{13.12}$$

Eliminating the parameter y^0, we can parametrise this worldline by the Minkowski time t as

$$x(t)=\frac{c^2}{g}\left[\sqrt{\left(1+\frac{g\kappa}{c^2}\right)^2+\frac{g^2t^2}{c^2}}-1\right]\ . \tag{13.13}$$

This implies

$$v(t):=\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{gt}{\sqrt{\left(1+\dfrac{g\kappa}{c^2}\right)^2+\dfrac{g^2t^2}{c^2}}}\ , \tag{13.14}$$

and, after a short calculation,

$$\left(c\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2=c^2-v(t)^2=c^2-\frac{g^2t^2}{\left(1+\dfrac{g\kappa}{c^2}\right)^2+\dfrac{g^2t^2}{c^2}}\ ,$$

implying finally that

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{1+g\kappa/c^2}{\sqrt{\left(1+\dfrac{g\kappa}{c^2}\right)^2 + \dfrac{g^2t^2}{c^2}}} . \tag{13.15}$$

Comparing with (13.5), this is the same but with g replaced by another constant, viz.,

$$g \longrightarrow \frac{g}{1+g\kappa/c^2} ,$$

so the solution this time is

$$t(\tau) = \frac{c(1+g\kappa/c^2)}{g} \sinh \frac{g\tau}{c(1+g\kappa/c^2)} , \tag{13.16}$$

replacing g by the new constant in (13.8).

It must be remembered, of course, that this is not the same τ as before. It is now the proper time for the worldline specified by κ. According to the first relation of (13.12), along this worldline,

$$\sinh \frac{gy^0}{c^2} = \frac{gt}{c(1+g\kappa/c^2)} ,$$

and comparing with (13.16),

$$\sinh \frac{g\tau}{c(1+g\kappa/c^2)} = \sinh \frac{gy^0}{c^2} ,$$

whence

$$\tau(y^0) = \frac{1}{c}\left(1+\frac{g\kappa}{c^2}\right) y^0 . \tag{13.17}$$

For the worldline through the spacetime origin, we retrieve (13.9).

The worldline (13.12) with $\kappa \neq 0$ can be parametrised by its proper time using (13.17):

$$\begin{cases} t(\tau) = \dfrac{c(1+g\kappa/c^2)}{g} \sinh \dfrac{g\tau}{c(1+g\kappa/c^2)} , \\ x(\tau) = \dfrac{c^2(1+g\kappa/c^2)}{g} \cosh \dfrac{g\tau}{c(1+g\kappa/c^2)} - \dfrac{c^2}{g} . \end{cases} \tag{13.18}$$

This allows us to calculate the four-velocity $v(\tau)$ of this worldline:

$$v(\tau) := \frac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix} t(\tau) \\ x(\tau) \end{pmatrix} = \begin{pmatrix} \cosh \dfrac{g\tau}{c(1+g\kappa/c^2)} \\ c\sinh \dfrac{g\tau}{c(1+g\kappa/c^2)} \end{pmatrix} . \tag{13.19}$$

This has pseudolength c. It can also be expressed as a function of the SE coordinates. Since τ depends only on y^0 along the curve, the same goes for the four-velocity. Since

$$\frac{g\tau}{c(1+g\kappa/c^2)} = \frac{gy^0}{c^2} \;,$$

we have

$$v(y^0) = \begin{pmatrix} \cosh(gy^0/c^2) \\ c\sinh(gy^0/c^2) \end{pmatrix} \;. \tag{13.20}$$

It is a remarkable thing that this is independent of κ, a point already established in Sect. 2.9, which leads to the phenomenon of HOS sharing discussed on pp. 28 and 60. This is illustrated for the present case in Fig. 2.7. Another notation for writing the last relation is

$$v(y^0) = \cosh\frac{gy^0}{c^2}\,\partial_t + c\sinh\frac{gy^0}{c^2}\,\partial_x \;. \tag{13.21}$$

13.1.3 SE Coordinate Frame

We now return to (13.1) in order to express the SE coordinate frame $\{\partial_{y^0},\partial_{y^1}\}$ in terms of the Minkowski coordinate frame $\{\partial_t,\partial_x\}$. We have

$$\frac{\partial}{\partial y^0} = \frac{\partial t}{\partial y^0}\frac{\partial}{\partial t} + \frac{\partial x}{\partial y^0}\frac{\partial}{\partial x} \;, \qquad \frac{\partial}{\partial y^1} = \frac{\partial t}{\partial y^1}\frac{\partial}{\partial t} + \frac{\partial x}{\partial y^1}\frac{\partial}{\partial x} \;,$$

whence

$$\partial_{y^0} = \frac{1}{c}\left(1+\frac{gy^1}{c^2}\right)\cosh\frac{gy^0}{c^2}\,\partial_t + \left(1+\frac{gy^1}{c^2}\right)\sinh\frac{gy^0}{c^2}\,\partial_x \tag{13.22}$$

and

$$\partial_{y^1} = \frac{1}{c}\sinh\frac{gy^0}{c^2}\,\partial_t + \cosh\frac{gy^0}{c^2}\,\partial_x \;. \tag{13.23}$$

These are obviously pseudoorthogonal, since

$$\begin{aligned}\partial_{y^0}\cdot\partial_{y^1} &= c^2\frac{1}{c}\left(1+\frac{gy^1}{c^2}\right)\cosh\frac{gy^0}{c^2}\,\frac{1}{c}\sinh\frac{gy^0}{c^2} - \left(1+\frac{gy^1}{c^2}\right)\sinh\frac{gy^0}{c^2}\cosh\frac{gy^0}{c^2}\\ &= 0\;,\end{aligned}$$

using

$$\partial_t\cdot\partial_t = c^2\;, \quad \partial_t\cdot\partial_x = 0\;, \quad \partial_x\cdot\partial_x = -1\;.$$

By similar calculations,

$$\begin{aligned}\partial_{y^0}\cdot\partial_{y^0} &= c^2\frac{1}{c^2}\left(1+\frac{gy^1}{c^2}\right)^2\cosh^2\frac{gy^0}{c^2} - \left(1+\frac{gy^1}{c^2}\right)^2\sinh^2\frac{gy^0}{c^2}\\ &= \left(1+\frac{gy^1}{c^2}\right)^2\;,\end{aligned}$$

and

$$\partial_{y^1} \cdot \partial_{y^1} = c^2 \frac{1}{c^2} \sinh^2 \frac{gy^0}{c^2} - \cosh^2 \frac{gy^0}{c^2} = -1 \ .$$

In summary,

$$\partial_{y^0} \cdot \partial_{y^1} = 0 \ , \quad \partial_{y^0} \cdot \partial_{y^0} = \left(1 + \frac{gy^1}{c^2}\right)^2 \ , \quad \partial_{y^1} \cdot \partial_{y^1} = -1 \ . \tag{13.24}$$

We observe that this tetrad is normalised right along the worldline $y^1 = 0$, and nowhere else. Note that it is the main observer who follows this worldline, i.e., the one who would set up these SE coordinates. In a sense, all the other observers in this rigid observer field, sitting at fixed SE space coordinates for the the main observer, are stuck with the proper time of the main observer, rather than their own.

Note also that rigidity is defined here, in the usual way, to mean that, in the SE coordinate reckoning of the main observer, each of the other observers sits at fixed values of y^1. This in turn means that, at any proper time of the main observer, any of the $y^1 = \kappa$ observers is always the same distance away as gauged in the inertial frame instantaneously comoving with the main observer.

These are some of the wonderful properties of the SE coordinate system, but as already mentioned in Sect. 2.9, they require other observers $y^1 = \kappa$ to have different uniform accelerations to the main observer, as is easily checked by differentiating $v(\tau)$ in (13.19) with respect to τ:

$$a(\tau) := \frac{\mathrm{d}v(\tau)}{\mathrm{d}\tau} = \frac{g}{c(1+g\kappa/c^2)} \begin{pmatrix} \sinh \dfrac{g\tau}{c(1+g\kappa/c^2)} \\ c\cosh \dfrac{g\tau}{c(1+g\kappa/c^2)} \end{pmatrix} . \tag{13.25}$$

This is a uniform acceleration, because

$$a \cdot a = -\frac{g^2}{(1+g\kappa/c^2)^2} \ , \tag{13.26}$$

which is constant (but different for each value of κ). So the value of the uniform acceleration of the observer at $y^1 = \kappa$ is $g/(1+g\kappa/c^2)$, which is equal to g when $\kappa = 0$, i.e., for the main observer, but smaller for everyone at positive values of κ.

It is easy to obtain a tetrad from $\{\partial_{y^0}, \partial_{y^1}\}$ by normalising ∂_{y^0}. A tetrad adapted to the rigid field of UA observers would be

$$\begin{cases} e_0 = \dfrac{1}{c} \cosh \dfrac{gy^0}{c^2} \, \partial_t + \sinh \dfrac{gy^0}{c^2} \, \partial_x \ , \\ e_1 = \dfrac{1}{c} \sinh \dfrac{gy^0}{c^2} \, \partial_t + \cosh \dfrac{gy^0}{c^2} \, \partial_x \ , \end{cases} \tag{13.27}$$

obtained from (13.22) and (13.23). Like ∂_{y^0}, the vector field e_0 is tangent everywhere to all the worldlines of our rigid observer field. There is nothing more to check in claiming this, because those worldlines are the constant y^1 curves in the spacetime, i.e., they are coordinate curves, and they have tangent ∂_{y^0}.

The above expression for the tetrad is somewhat hybrid, because we express $e_0(t,x)$ and $e_1(t,x)$ in terms of the SE time coordinate y^0. It is better if everything on the right-hand side is given as a function of the inertial coordinates (t,x). We need the inverse of the coordinate transformation (13.1) back on p. 382, which is easily found to be

$$\begin{cases} y^0 = \dfrac{c^2}{g}\tanh^{-1}\dfrac{ct}{x+c^2/g} \,, \\ y^1 = \left[\left(x+\dfrac{c^2}{g}\right)^2 - c^2t^2\right]^{1/2} - \dfrac{c^2}{g} \,. \end{cases} \tag{13.28}$$

Now in (13.27), we have to replace $\cosh(gy^0/c^2)$ and $\sinh(gy^0/c^2)$, so we only need the first relation of (13.28). This means working out the cosh and sinh of an inverse tanh. With the obvious notation, we have

$$T := \frac{S}{C}\,, \qquad C^2 - S^2 = 1\,,$$

and hence,

$$C = \frac{1}{\sqrt{1-T^2}}\,, \qquad S = \frac{T}{\sqrt{1-T^2}}\,.$$

Hence,

$$\cosh\frac{gy^0}{c^2} = \cosh\tanh^{-1}\frac{ct}{x+c^2/g} = \frac{x+c^2/g}{\sqrt{(x+c^2/g)^2 - c^2t^2}}$$

and

$$\sinh\frac{gy^0}{c^2} = \sinh\tanh^{-1}\frac{ct}{x+c^2/g} = \frac{ct}{\sqrt{(x+c^2/g)^2 - c^2t^2}}\,,$$

so finally,

$$\begin{cases} e_0(t,x) = \dfrac{1}{c}\dfrac{x+c^2/g}{\sqrt{(x+c^2/g)^2 - c^2t^2}}\,\partial_t + \dfrac{ct}{\sqrt{(x+c^2/g)^2 - c^2t^2}}\,\partial_x \,, \\ e_1(t,x) = \dfrac{1}{c}\dfrac{ct}{\sqrt{(x+c^2/g)^2 - c^2t^2}}\,\partial_t + \dfrac{x+c^2/g}{\sqrt{(x+c^2/g)^2 - c^2t^2}}\,\partial_x \,. \end{cases} \tag{13.29}$$

13.2 Tetrad for Another Field of Uniformly Accelerating Observers

We now consider a field of uniformly accelerating observers, filling spacetime, which looks rigid to some inertial observer. To do this we consider an observer who, for that inertial observer, passes through the spacetime origin along the worldline (13.3) on p. 383, viz.,

$$x(t) = \frac{c^2}{g}\left[\left(1+\frac{g^2t^2}{c^2}\right)^{1/2} - 1\right] , \tag{13.30}$$

also given by (13.10) on p. 384, viz.,

$$x(\tau) = \frac{c^2}{g}\left(\cosh\frac{g\tau}{c} - 1\right) , \tag{13.31}$$

when parametrised by its own proper time. The other observer worldlines are then obtained by translation along the x axis. The one going through $x = X$ at time $t = 0$ is given by

$$x_X(\tau) = X + \frac{c^2}{g}\left(\cosh\frac{g\tau}{c} - 1\right) . \tag{13.32}$$

Note that the parameter τ here will also be the proper time along this worldline, taken as being zero when $t = 0$. Since this is also obviously given by

$$x_X(t) = X + \frac{c^2}{g}\left[\left(1+\frac{g^2t^2}{c^2}\right)^{1/2} - 1\right] , \tag{13.33}$$

this observer field looks rigid to the inertial observer with coordinates (t,x), in the sense that, at any given time t, the x coordinate distance between any two such observers is always the same. But it is not rigid in the usually accepted sense of Sect. 13.1.

We now specify the tetrad field $\{f_0, f_1\}$ as follows. At any $(t,x) \in \mathbb{M}$, we require $f_0(t,x)$ to be tangent to the unique worldline x_X passing through there, and $f_1(t,x)$ to be the parallel transport along a curve of constant t of the vector $f_1(t,x-X)$ we get by the SE construction for the worldline passing through $(t,x) = (0,0) \in \mathbb{M}$. In other words, since parallel transport in $\mathbb{M}$ does not change vector components, we can think of this as just taking the tetrad field along the main worldline (through the origin of $\mathbb{M}$ for the given inertial coordinates) and sliding it parallel to the x axis to all the different X values.

Looking at (13.27), we thus decree

$$\begin{cases} f_0(t,x) = \dfrac{1}{c}\cosh\dfrac{g\tau}{c}\,\partial_t + \sinh\dfrac{g\tau}{c}\,\partial_x , \\ f_1(t,x) = \dfrac{1}{c}\sinh\dfrac{g\tau}{c}\,\partial_t + \cosh\dfrac{g\tau}{c}\,\partial_x , \end{cases} \tag{13.34}$$

where τ is the proper time along the unique curve x_X through (t,x). This is not the best representation, because we would like our tetrad to be a function of the inertial coordinates (t,x). But we know from (13.8) on p. 383 that

$$t(\tau) = \frac{c}{g}\sinh\frac{g\tau}{c} .$$

Hence also,

$$\cosh\frac{g\tau}{c} = \sqrt{1+\frac{g^2t^2}{c^2}} .$$

This means that

$$\begin{cases} f_0(t,x) = \dfrac{1}{c}\sqrt{1+\dfrac{g^2t^2}{c^2}}\,\partial_t + \dfrac{gt}{c}\,\partial_x , \\ f_1(t,x) = \dfrac{gt}{c^2}\,\partial_t + \sqrt{1+\dfrac{g^2t^2}{c^2}}\,\partial_x . \end{cases} \tag{13.35}$$

This should be compared with the first tetrad (13.29), which is a function of x as well as t.

13.3 Tetrads Adapted to a Given Worldline

Given a timelike worldline, we can always find a tetrad field along it whose zeroth member is the unit tangent to the worldline (basically, the velocity four-vector, up to a factor of c). Then there are many ways to choose the other three members of the tetrad at any point on the worldline. They only need to be tangent to a spacelike hypersurface cutting the worldline at that point in such a way that the velocity four-vector is perpendicular to it there. There is also plenty freedom if we wish to extend the tetrad on the worldline to a tetrad over some neighbourhood of the worldline.

Now Maluf and Ulhoa are considering charges with accelerating worldlines and observers moving with them, or accelerating observers and inertially moving charges. They want their observer to use a tetrad field along her worldline to decide whether the EM fields due to the charge are radiating or not. In fact it is easy to formulate electromagnetic theory in terms of tetrad components. If $\{e_a\}_{a\in\{0,1,2,3\}}$ is a tetrad and $F_{\mu\nu}$ are the components of the Faraday tensor relative to some coordinates, then the tetrad components of this tensor are

$$F_{ab} := e_a{}^{\mu} e_b{}^{\nu} F_{\mu\nu} . \tag{13.36}$$

Of course, we can write down Maxwell's equations for the tetrad components, but there is no need to. If we can solve them for some set of coordinates (and we can, for inertial coordinates, when the spacetime is flat), then we can transform using (13.36) to get this picture of the EM field.

This is indeed what M/U do for the fields due to the point charge, whatever it is doing, transforming the Liénard–Wiechert solution of Maxwell's equations.

On the face of things, it looks a very good idea to express the fields relative to a tetrad field along the observer worldline, rather than just express them relative to SE coordinates, since the coordinate frame for SE coordinates is not even normalised. As we have seen, however, the SE coordinate frame is in fact normalised precisely along the main worldline of the SE system, presumably a point in its favour.

But there is another point here. In order to assess the fields, however we intend to do it, we must look at the values of the field components F_{ab} off the worldline. But it is obvious from (13.36) that this will depend, not only on the way the components $F_{\mu\nu}$ of the Faraday tensor vary off the worldline, but also on the choice of tetrad, since $e_a{}^\mu$ are functions of spacetime. Now once again, at first glance, one might think that it would be better to refer the field to tetrad components, since the SE coordinate frame is not normalised as soon as we leave the worldline. In some ways, that does indeed seem to be a better solution than the usual arguments by Boulware or Rohrlich, who suggest that we should pretend the SE coordinates are inertial, and judge the SE components of the EM field as though they were inertial components (see Chap. 11). That is precisely the kind of view deplored in [30]. Here we definitely seem to have a better idea.

But which tetrad field shall we take when we leave the worldline? We have already found two such fields, viz., $\{e_0, e_1\}$ in (13.29) and $\{f_0, f_1\}$ in (13.35), identical on the main worldline, i.e., the uniformly accelerating worldline through the origin of the inertial coordinate system. What is more e_0 and f_0 are both equal to the observer four-velocity along that worldline. One tetrad is adapted to a rigid observer field in the standard relativistic sense of the word, and the other is adapted to what would appear to be a rigid observer field to an inertial observer.

As mentioned at the beginning of this section, there is freedom in the choice of spacelike vectors in the tetrad along the worldline. Of course, it is thoroughly reasonable to adopt an FW transport of some particular choice at some event on the observer worldline, say the origin of the inertial coordinate system. Then the tetrad is at least fully determined along the worldline, and alternative choices of starting orientation for the spatial triad only lead to a physically irrelevant fixed rotation.

But there is plenty of freedom in extending off the worldline, and that is the problem here. The tetrad field is not unique, so the results we get by referring to this tetrad will also not be unique. They will depend on the neighbouring observers we must choose for our initial observer. The example of the rigid and non-rigid sets of observers in Minkowski spacetime illustrates this in a simple situation.

In the most general context, i.e., for a general timelike worldline in a curved spacetime, one could do the following, for example. First obtain a tetrad field along the worldline by FW transport of some choice at some preselected event, so that the zero vector is the four-velocity everywhere along the worldline. At each event on the worldline consider all spacelike geodesics through that event that are orthogonal to the four-velocity there. There will be some neighbourhood of the chosen event such that those geodesics intersect nowhere else within the neighbourhood. Furthermore, there will be some neighbourhood of the worldline such that no two of these spacelike geodesics from *different* events along the worldline actually intersect within that neighbourhood. We could then extend the tetrad by parallel transport along the spacelike geodesics.

This can be looked at in a slightly different way. We expect there to be some neighbourhood of the worldline such that, for any event within that neighbourhood, there is a unique event on the worldline such that there is a unique spacelike geodesic joining the two events with the property that it is orthogonal to the worldline where it intersects it. The tetrad just suggested is found at the selected event by parallel transport along that unique spacelike geodesic from the unique event on the worldline.

That is one way to extend the tetrad. One can imagine others. If we refer back to the tetrad $\{e_0, e_1\}$ given by (13.27) on p. 387, we started with the uniformly accelerating worldline through the origin of the inertial coordinates and chose the obvious vector fields along that worldline. We then in fact extended that tetrad by parallel transport along those spacelike geodesics through the worldline that cut it orthogonally to its four-vector at each event. That remains to be proved, but is rather obvious with some knowledge of the rigid observer field and SE coordinates. The fact is best seen by examining (13.27). The vector fields vary only with y^0, but are 'constant' on each hyperplane of simultaneity of the main observer, i.e., they do not vary with y^1. In any case, we understand that we obtained this tetrad by precisely the very general idea of the last two paragraphs.

Note also, as shown in Fig. 2.5 on p. 79, that the spacelike geodesics in question do intersect eventually, all at the same spacetime event, viz., $y^0 = 0$, $y^1 = -c^2/g$, which is an SE coordinate singularity, i.e., the g_{00} component of the SE metric goes to zero there. This is a long way down the x axis at $x = -c^2/g$, with $t = 0$.

The other tetrad field $\{f_0, f_1\}$, given by (13.35) on p. 390, is not obtained by the above general method from the tetrad along the main observer worldline. It is still obtained by parallel transport along geodesics, but they are not orthogonal to the worldline. In fact, the geodesics here are $t =$ constant, so it is easy to picture in the (t,x) diagram. The claim here is best understood by looking at (13.35). The vector fields vary only with t, but are 'constant' on each hyperplane $t =$ constant, i.e., they do not vary with x. This illustrates just how much freedom there is in extending the tetrad off the worldline.

Now one might well criticise the choice $\{f_0, f_1\}$ because it depends on a choice of inertial frame, whereas the other extension seems much more closely adapted to the SE coordinate system, parallel translating along hyperplanes of simultaneity (HOS) of the main observer. Note in passing that this observer borrows the HOS of the instantaneously comoving inertial observer at any event on her worldline, attributing a time $y^0 = \tau$ (her proper time) to all points on it. On the other hand, this process in itself has something artificial about it. Why spread time like this over the HOS? And why take this to be the HOS?

Let us now ask what method is used in [34] to extend the tetrad off the main uniformly accelerating worldline, and what justification is given for selecting that extension. Note that their conventions for the metric and positions of indices are different from those used above, but we can see enough to understand what they are doing, in particular, which tetrad field they choose. They consider first tetrad fields adapted to observers at rest in Minkowski spacetime, so

$$e^a{}_\mu(ct,x,y,z) = \delta^a_\mu \ .$$

They then consider a time-dependent boost in the x direction to get the tetrad field

$$e^a{}_\mu(ct,x,y,z) = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \qquad (13.37)$$

where

$$\gamma = \frac{1}{(1-\beta^2)^{1/2}} , \qquad \beta = v/c , \qquad v = v(t) \ .$$

They say that this frame is adapted to observers with four-velocity

$$u^\mu = (\gamma, \beta\gamma, 0, 0) \ ,$$

so we already get an idea of which tetrad we have! Basically, the four-velocity field of these observers is constant over hyperplanes $t =$ constant. The boost is time dependent, so the four-velocity field of the observers varies with t, but not with x.

So let us close the gap between (13.37) and (13.35). Considering the formulas given by M/U, they say that we have uniform acceleration for each observer in their observer field when

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{v}{\sqrt{1-v^2/c^2}} = a \text{ (constant)} \ .$$

Assuming $v(0) = 0$, this implies that

$$\frac{v}{\sqrt{1-v^2/c^2}} = at \ ,$$

and it is easy to check that a solution is

$$v(t) = \frac{at}{(1+a^2t^2/c^2)^{1/2}} \ .$$

Then

$$\gamma(t) = \frac{1}{\sqrt{1-v^2/c^2}} = (1+a^2t^2/c^2)^{1/2} \ ,$$

so

$$\gamma\beta = \gamma v/c = at/c \ ,$$

and the tetrad field (13.37) becomes

$$e^a{}_\mu(ct,x) = \begin{pmatrix} (1+a^2t^2/c^2)^{1/2} & -at/c \\ -at/c & (1+a^2t^2/c^2)^{1/2} \end{pmatrix} . \qquad (13.38)$$

This should be compared with (13.35), which specifies our tetrad field $\{f_0, f_1\}$.

13.4 Faraday Tensor Components Relative to SE Tetrad

The aim here is to express the fields of the uniformly accelerating charge (passing through the origin of our inertial coordinates) relative to the SE tetrad (13.27) on p. 387, viz.,

$$\begin{cases} e_0 = \dfrac{1}{c}\cosh\dfrac{gy^0}{c^2}\,\partial_t + \sinh\dfrac{gy^0}{c^2}\,\partial_x\,, \\ e_1 = \dfrac{1}{c}\sinh\dfrac{gy^0}{c^2}\,\partial_t + \cosh\dfrac{gy^0}{c^2}\,\partial_x\,. \end{cases} \tag{13.39}$$

We do this by transforming the Faraday tensor $F_{\mu\nu}$ expressed relative to SE coordinates, for which the frame field is given by (13.22) and (13.23) on p. 386, viz.,

$$\begin{cases} \partial_{y^0} = \left(1+\dfrac{gy^1}{c^2}\right)\left(\dfrac{1}{c}\cosh\dfrac{gy^0}{c^2}\,\partial_t + \sinh\dfrac{gy^0}{c^2}\,\partial_x\right)\,, \\ \partial_{y^1} = \dfrac{1}{c}\sinh\dfrac{gy^0}{c^2}\,\partial_t + \cosh\dfrac{gy^0}{c^2}\,\partial_x\,. \end{cases} \tag{13.40}$$

Basically, we have

$$\partial_{y^0} = \left(1+\frac{gy^1}{c^2}\right)e_0\,,\quad \partial_{y^1} = e_1\,,\quad \partial_{y^2} = e_2\,,\quad \partial_{y^3} = e_3\,. \tag{13.41}$$

It is shown in [30, p. 231] that the Faraday tensor components relative to the SE coordinate frame are

$$F_{\mu\nu}^{\mathrm{SE_{coord}}} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & 0 & 0 \\ -E_2 & 0 & 0 & 0 \\ -E_3 & 0 & 0 & 0 \end{pmatrix}\,, \tag{13.42}$$

where

$$\mathbf{E} = -\begin{pmatrix} \partial A_\tau/\partial z \\ \partial A_\tau/\partial x \\ \partial A_\tau/\partial y \end{pmatrix}\,, \tag{13.43}$$

with

$$A_\tau := -\frac{eg}{4\pi}\frac{g^{-2}+\rho^2+X^2}{\left[\left(g^{-2}+\rho^2+X^2\right)^2 - 4X^2g^{-2}\right]^{1/2}} \tag{13.44}$$

and

$$X^2 := x^2 - t^2\,, \qquad \rho^2 = y^2 + z^2\,. \tag{13.45}$$

So we have the well known fact that the SE magnetic field is zero, so often glibly interpreted as meaning that there is no radiation. The details of the electric field are given to show that it looks *nothing* like the Coulomb field to the coaccelerating observer, even under the assumption that such an observer would use this picture of things. It is a highly complex function. Note, however, that it is static, i.e., independent of the SE coordinate time (not immediately obvious).

Let us convert the above version of the Faraday tensor from an expression relative to the SE coordinate frame to an expression relative to the SE coordinate tetrad. The following is somewhat laborious, especially as it is clear that we are going to get zero magnetic field again for the transformation implied by (13.41). We have

$$F_{ab}^{\mathrm{SE_{tetrad}}} = \overline{\theta}_a{}^\mu \overline{\theta}_b{}^\nu F_{\mu\nu}^{\mathrm{SE_{coord}}} ,$$

where $\overline{\theta}_a{}^\mu$ converts covectors expressed in components with respect to SE coordinates to covectors expressed in components with respect to the SE tetrad. For any vector V, we have

$$V = V^\mu \partial_{y^\mu} , \qquad V = \overline{V}^a e_a .$$

How is $\overline{V}^a$ related to V^μ? Looking at (13.41), it is clear that

$$\overline{V}^1 = V^1 , \quad \overline{V}^2 = V^2 , \quad \overline{V}^3 = V^3 ,$$

and

$$V^0 \partial_{y^0} = V^0 \left(1 + \frac{gy^1}{c^2}\right) e_0 ,$$

so

$$\overline{V}^0 = \left(1 + \frac{gy^1}{c^2}\right) V^0 . \tag{13.46}$$

We thus have the transformation matrix

$$\theta = (\theta^a{}_\mu) = \begin{pmatrix} 1 + gy^1/c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

with μ indexing columns and a indexing rows. Then for any vector V, we have

$$\overline{V}^a = \theta^a{}_\mu V^\mu .$$

What about a covector ω now? We have

$$\overline{\omega}_a = \overline{\theta}_a{}^\mu \omega_\mu ,$$

and

$$\overline{\omega}_a \overline{V}^a = \omega_\mu V^\mu ,$$

a scalar quantity for any vector V. Hence,

$$\overline{\theta}_a{}^\mu \omega_\mu \theta^a{}_\nu V^\nu = \omega_\mu V^\mu \ ,$$

which implies

$$\overline{\theta}_a{}^\mu \theta^a{}_\nu = \delta^\mu_\nu \ . \tag{13.47}$$

Now if we arrange for each matrix

$$\theta = (\theta^a{}_\mu) \ , \qquad \overline{\theta} = (\overline{\theta}_a{}^\mu) \ ,$$

to be such that a labels rows and μ labels columns, then (13.47) says that

$$\theta^{\mathrm{T}}\overline{\theta} = \mathrm{Id} \ , \qquad \overline{\theta} = (\theta^{\mathrm{T}})^{-1} \ .$$

Finally, then

$$\overline{\theta} = \begin{pmatrix} (1+gy^1/c^2)^{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ ,$$

and we have

$$F^{\mathrm{SE}_{\mathrm{tetrad}}} = \overline{\theta} F^{\mathrm{SE}_{\mathrm{coord}}} \overline{\theta}^{\mathrm{T}} = \frac{1}{1+gy^1/c^2} \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & 0 & 0 \\ -E_2 & 0 & 0 & 0 \\ -E_3 & 0 & 0 & 0 \end{pmatrix} \ . \tag{13.48}$$

We have the obvious fact that the magnetic field is zero everywhere. The electric field is static and clearly will not look like a Coulomb field relative to SE coordinates, barring miracles, due to the complexity of the given functions (combined with the fact that there is absolutely no reason why it should look Coulomb).

13.5 A Brief Conclusion

Maluf and Ulhoa fill their space with identically uniformly accelerating observers. When they express the Faraday tensor components near the worldline in order to picture what the field is doing (be it the field of a coaccelerating charge or an inertially moving charge), the picture they get will depend to some extent on this choice.

Indeed, they find that the magnetic field due to the uniformly accelerating charge passing through the origin of the inertial coordinates will not be zero relative to their tetrad. On the other hand, relative to the SE tetrad, where space is filled with a rigid field of uniformly accelerating observers, the magnetic field due to this uniformly accelerating charge will in fact be zero. But which choice of tetrad field should we make? Which choice would an accelerating observer make? The fact is that we do not know.

Even if the suggestion in [34] does turn out to be an improvement on what is advocated by Boulware and Rohrlich (see Chap. 11), there is still an ambiguity. In such a situation, would it not be preferable to picture the fields in the inertial frame? And in the general context of a curved spacetime, would it not be preferable to picture fields in locally inertial frames, as authorised by the weak and strong equivalence principles which we must invoke in any case, whatever frame we are choosing?

It seems odd to see how electromagnetic fields expressed relative to non-inertial coordinate frames are often interpreted at face value. For example, a uniformly accelerating observer using semi-Euclidean (Rindler) coordinates adapted to her worldline is supposed to be duped somehow into thinking that her coordinates are in fact inertial coordinates, and that the EM field components relative to that coordinate frame can be treated as though they were the electric and magnetic fields we know and love from our school days.

There is a sense in which a tetrad approach is an improvement. It is true that the semi-Euclidean (SE) coordinate frame is orthonormal along the accelerating worldline, but only there. Even in its immediate vicinity, the timelike coordinate vector field is no longer normalised. That alone seems to be an argument against trying to interpret the SE components of the Faraday tensor as though the tensor had been expressed relative to an inertial coordinate frame.

On the other hand, one could just normalise the timelike SE coordinate vector everywhere, since the SE coordinate frame is in fact orthogonal, and the three spacelike SE coordinate vectors are normalised everywhere. This tetrad field is obtained from the one along the accelerating worldline by parallel transport along hypersurfaces of constant SE time coordinate. One could then use this tetrad field to express the components of the Faraday tensor.

But for some reason, Maluf and Ulhoa do not do that. Their tetrad field is obtained from the one along the accelerating worldline by parallel transport along hypersurfaces of constant inertial time coordinate, for some arbitrarily chosen inertial coordinate system. We should wonder then why that would be a better choice. It does make a difference to the final interpretation of the field which we hope to attribute to the uniformly accelerating observer, because we have

$$F_{ab} := e_a{}^{\mu} e_b{}^{\nu} F_{\mu\nu} \ ,$$

and the tetrad field components $e_a{}^{\mu}$ are functions of spacetime. In fact, relative to the tetrad derived by normalising the already orthogonal SE coordinate frame, the magnetic field will be zero for the uniformly accelerating charge.

However, as discussed in Chap. 11, we should not take that to mean that there is no radiation of EM energy by the charge, no more than we should take the vanishing of the magnetic field of the uniformly accelerating charge in the SE coordinate frame adapted to the charge to mean that there is no EM radiation. The present view is that we should not take it to mean anything at all.

While it may well make more sense to say that the accelerating observer will interpret the field by expressing its components relative to a tetrad field than to say that she would just consider its SE coordinate frame components, we do not really escape from a much more fundamental problem, viz., we do not actually know what accelerating observers would do. And it seems there are fundamentally different choices, if the claim above is correct.

On the other hand, we do know what observers with inertial motion would do because we know (at least we think so, it is a hypothesis) what their measuring instruments would do under inertial motion, and we know (if the hypothesis is right) that there will be especially nice coordinates for such a person to use. Then in the more general case of a curved spacetime, our only recourse is surely to refer to a locally inertial frame, such as always exists (weak equivalence principle), where we make the explicit further hypothesis (an add-on to all the other assumptions of general relativity, often called the strong equivalence principle) that all non-gravitational physics, e.g., electrodynamics, will look roughly as it would in the absence of gravity, where we have Maxwell's theory. So perhaps it is a fundamental mistake here to try to say anything at all about what accelerating observers would do.

Chapter 14
Unruh Effect

Here is another situation where some quantities expressed relative to a semi-Euclidean coordinate system are interpreted physically as being relevant to uniformly accelerating observers. To illustrate this, we may begin with a quote from an epic review of everything that has been done in this area over the past 40 years [11]:

> [The Unruh effect] has played a crucial role in our understanding that the particle content of a field theory is observer dependent.

This is no longer classical electrodynamics. Here we are talking about quantum field theory, where the concept of particle is subject to a certain ontological fuzziness, and the field is no longer an electromagnetic field, because the Unruh effect is usually introduced by discussing the Klein–Gordon scalar field. Any reference to field equations here can be taken to mean the Klein–Gordon (KG) equation.

14.1 Observer with Translational Uniform Acceleration

If we want to set up a quantum field theory, the key question is: what constitutes a positive frequency solution to the field equations? Once we have that, we can try to make the usual expansion of the field in terms of creation and annihilation operators. But the answer to this question depends on what time coordinate we use. Now in Minkowski spacetime, we normally choose an inertial time coordinate for the QFT construction. But it turns out that we can find solutions to the field equations that are positive frequency with respect to the SE time coordinate for an eternally uniformly accelerating observer. The QFT construction then delivers a different vacuum and different particles.

In inertial coordinates in Minkowski spacetime, the natural positive frequency solutions to the field equations lead to what we normally call particles. Let us call them Minkowski particles for the present purposes. In SE coordinates for an eternally uniformly accelerating observer, associated as we have seen with a Lorentz boost Killing vector field, the natural solutions to the field equations lead to different particles. Let us call these Rindler particles here. How are we to interpret this new Rindler vacuum and these new particles?

The first thing we note is that the usual QFT vacuum, the Minkowski vacuum, is full of Rindler particles. Of course, this is not really so surprising since we have always known that the QFT vacuum is *not* nothing. But we also note that formally, for the right formal choice of 'Hamiltonian', the density operator for the Rindler particles is precisely the density operator for a system of particles in equilibrium at a certain nonzero temperature. This is the Unruh temperature. It is linearly proportional to the absolute acceleration of the eternally uniformly accelerating observer who sets up this alternative view of the quantum field.

But how do we know that the formal 'Hamiltonian', constructed by the usual Lagrangian field theoretical techniques but in the framework of a semi-Euclidean coordinate system, is what an accelerating observer would call energy? Are we saying that this is what such a person would naturally measure, or is it just a good definition? And if it is a good definition, what is it good for? Note that the naturalness of the 'temperature' attributed to the thermal bath of Rindler particles depends on the naturalness of this definition of the Hamiltonian.

Certain features of the new Hamiltonian are not so natural. It comes from a classical field 'energy' for the eternally uniformly accelerating observer, and this classical field energy is defined to be

$$E := \int_{\Sigma} K^{\mu} T_{\mu\nu} \mathrm{d}\Sigma^{\nu} \,, \tag{14.1}$$

where K^{μ} is the appropriate Lorentz boost Killing vector field, $T_{\mu\nu}$ is the formal energy–momentum tensor for the KG scalar quantum field, and Σ can be any spacelike hypersurface cutting all timelike curves in the spacetime, since we know that $K^{\mu}T_{\mu\nu}$ is covariantly conserved, in the sense that its covariant divergence is zero (see Sect. 11.6). A simple choice for Σ is the hyperplane of zero SE time for the chosen observer.

This quantity E is supposed to be the energy of the field as gauged by an observer following a flow line of the KVF. Note that K can be assumed normalised along the observer worldline. However, the expression (14.1) for E integrates $K^{\mu}T_{\mu\nu}$ over places where K^{μ} is not normalised, whence we cannot claim that this quantity is the density of energy–momentum (a four-vector with components equal to the energy density and the rates of flow of energy per unit area in three space directions) that would be measured by the observer following the flow line of the KVF at that particular point.

Worse, all observers following flow curves of the Lorentz boost KVF are accelerating, and we shall soon see that they themselves will not measure the same things with their detectors as instantaneously comoving inertial observers, according to the results of the discussion about detectors below. So this really does look like a case of making the best definition of energy we can, constrained by the requirement that the thing we integrate must be conserved.

Anyway, the usual QFT vacuum is thus described as a thermal bath of Rindler particles at the Unruh temperature. It is usual to joke at this point that, if we accelerate our lunch at high enough acceleration, it will cook, remembering of course not to try to keep up with it! But can we demonstrate that an SE observer will interact with those Rindler particles, just by the fact that she is accelerating? Can we show that such an observer will end up in 'thermal equilibrium' with them? Surprisingly, there are indeed arguments to support such claims.

Things become a lot clearer when we stop talking about observers and start talking about detectors. Consider the Unruh–DeWitt (UD) detector, which is a point-like detector with a linear interaction with the quantum field and two energy levels (ground and excited). Clearly this was not chosen on the grounds of physical realism, but more sophisticated models have been investigated, and we shall assume that the following discussion is also borne out in more realistic cases.

We can consider this detector in four different situations:

1. Stationary in an inertial frame in the Minkowski vacuum, it does not excite.
2. Accelerating uniformly in an inertial frame in the Minkowski vacuum, it does excite. This case is also described by saying that the detector is stationary, i.e., sitting at fixed SE space coordinates, in the Rindler thermal bath. So we can understand its excitation through absorption of the ontologically fuzzy Rindler particles.
3. Stationary in an inertial frame in a Minkowski thermal bath of the usual (but nevertheless ontologically fuzzy) Minkowski particles at the Unruh temperature corresponding to some eternal uniform acceleration, the detector will excite, but not at the same rate as in case 2.
4. Accelerating uniformly in an inertial frame in the Rindler vacuum, a state of the field in which there are no Rindler particles but which is full of Minkowski particles, the detector will not excite, provided that it has the right absolute acceleration as determined by specifying the motion of the observer who set up this particular Rindler vacuum. (There are different Rindler vacuums, depending on the worldline of the observer who sets them up.) This case is also described by saying that the detector is stationary in the Rindler vacuum, i.e., sitting at fixed SE space coordinates for the observer who set up this Rindler vacuum construction. The striking thing about this case is that the detector does not 'see' the Minkowski particles in the field state, provided it has the right motion.

A large part of the discussion of the Unruh effect concerns interpretation of the eternally uniformly accelerating cases in terms of Rindler particles, or their absence, using the alternative expansion of the quantum field in terms of positive and negative frequency solutions to the KG equation. Note, however, that these expansions are *not* necessary in order to calculate the excitation of the UD detector for arbitrary motion through the Minkowski vacuum. This can always be done using the standard expansion of the quantum field in terms of creation and annihilation operators associated with negative and positive frequency operators for an inertial time coordinate, and in this sense, these results are just standard results about the QFT vacuum.

Cases 1 and 3 above suggest that this detector does function in some sense as a Minkowski particle detector, while cases 2 and 4 suggest that it functions in some sense as a Rindler particle detector, provided that it is doing the right thing in each case, i.e., provided that it has the right motion. Of course, such claims are hampered as always by the ontological fuzziness of the particle notion in quantum field theory.

What about temperature? We can consider not only excitation but also deexcitation of the detector, and we find that the associated rates satisfy a detailed balance relation. If we imagine a ‘gas’ of these pointlike detectors held at some fixed SE space coordinate, i.e., uniformly accelerating, in the Minkowski vacuum (case 2), and if we consider their excitation and deexcitation rates, we find that their energies will distribute over the two available energies (ground and excited) in precisely the way we would expect for a gas at the Unruh temperature. Likewise for a similar detector ‘gas’ held stationary in an inertial frame in a Minkowski particle thermal bath at the corresponding temperature, despite the fact that the excitation rates differ in the two cases. So there is some value in this temperature interpretation and it does suggest that one’s lunch might indeed cook if accelerated sufficiently, provided that one’s lunch does behave like a gas of UD detectors. A proviso nevertheless.

To end, let us just compare the UD and Mould detectors (see Sect. 11.10):

- The Mould detector detects nothing when stationary in a Coulomb field, just as the UD detector detects nothing when stationary in the Minkowski (usual QFT) vacuum.
- The Mould detector detects radiation when accelerated through the Coulomb field, just as the UD detector detects something when accelerated through the Minkowski vacuum.
- The Mould detector detects nothing when accelerated with the charge source, just as the UD detector detects nothing when ‘stationary’, i.e., uniformly accelerating, in the appropriate Rindler vacuum.
- The Mould detector detects something when the charge is accelerated and the detector is moving inertially, just as the UD detector will detect something when moving inertially in a Minkowski thermal bath.

The parallel continues slightly:

- The excitation rate of the Mould detector when accelerated through a Coulomb field is not the same as when it moves inertially through the field of an accelerating charge (and in any case, the field of an accelerating charge does not look exactly like a Coulomb field for any choice of coordinates adapted to the motion of the accelerating charge).
- The excitation rate of the Unruh–DeWitt detector when uniformly accelerated through the Minkowski vacuum with some absolute acceleration a, construed as being stationary in a Rindler thermal bath, is not the same as its excitation rate when stationary in a Minkowski thermal bath at a temperature equal to the Unruh temperature corresponding to the acceleration a.

Those who study the Unruh effect have referred to the whole subject of the discussion in Chap. 11 about uniformly accelerating charges and the equivalence principle as being merely a semantic issue [24]! But what is physics if not the semantics of our mathematical models, i.e., an attempt to extract meaning from such models in the context of physical measurement in the real world?

In any case, the same could be said about accounts of the Unruh effect. And a problem remains with these accounts: they claim to give the perspective of an accelerating observer, while it is quite clear that this approach only works (only exists) when the accelerating worldline is a flow curve of a Killing vector field. Without this, there is apparently no elegant alternative construction of the quantum field theory. So what would be the perspective of an arbitrarily accelerating observer?

As already suggested, not every timelike worldline in Minkowski spacetime is the flow curve of some KVF. This can be understood by noting that, in flat spacetime, KVFs are related to transformations in the Poincaré group, with 3 space translations, 1 time translation, 3 Lorentz boosts, and 3 space rotations, making the total of 10, whence there can be no others, apart from combinations, e.g., a regular spiral in spacetime, corresponding to a particle in circular motion with constant angular speed. Basically, the Lie algebra of KVFs under Lie vector product is ten-dimensional, and we can account for all the 10 dimensions without needing all kinds of curve.

More rigorously, we can use the theorem that any affinely parametrised geodesic cuts a KVF at the same angle everywhere [29]. If λ is the affine parameter and $T := \partial/\partial\lambda$ is the tangent vector, then

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(T \cdot K) = T \cdot \nabla(T \cdot K) = T^{\mu}(T_{\nu;\mu}K^{\nu} + T^{\nu}K_{\nu;\mu}) ,$$

and

$$T^{\mu}T_{\nu;\mu} = 0$$

is just the geodesic equation, while

$$T^{\mu}T^{\nu}K_{\nu;\mu} = \frac{1}{2}T^{\mu}T^{\nu}(K_{\nu;\mu} + K_{\mu;\nu}) = 0 ,$$

by the equation that says that K is a KVF.

With this in mind, it is easy to construct a timelike worldline that is not a flow curve of a KVF. The geodesics in Minkowski spacetime are coordinate straight lines, like the time axis, for example. So take a curve that follows the time axis, veers off it slightly, comes back and crosses it, then rejoins it. If this were the flow curve of a KVF, it would have to make the same angle with the time axis everywhere, which it does not.

14.2 Observer with Circular Motion

Another case considered by Unruh theorists is circular motion at constant angular speed [11]. They analyse this using corotating cyclindrical coordinates, rather than the rigid semi-Euclidean coordinate systems discussed in Sect. 2.11. Let us look briefly at the relevant mathematics.

14.2.1 Cylindrical Coordinates and Killing Vector Field

In cylindrical coordinates (t,r,θ,z), the observer worldline is specified by r and z being constants, viz., R and 0, respectively, while $\theta = \nu t$. We shall now show that this worldline is the flow curve of a Killing vector field. We need to check a few things here:

- Find the metric components in cylindrical coordinates.
- Find the connection coefficients in cylindrical coordinates.
- Show that $K := \partial_t + \nu\partial_\theta$ is a KVF.
- Show that K is tangent to the spiralling worldline.

14.2.2 Metric in Cylindrical Coordinates

Naturally, we have

$$\begin{cases} t = t\ , \\ x = r\cos\theta\ , \\ y = r\sin\theta\ , \\ z = z\ . \end{cases}$$

The quick route to the metric is to note that the metric interval is

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2\mathrm{d}\theta^2 - \mathrm{d}z^2\ , \tag{14.2}$$

whence

$$g_{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\ . \tag{14.3}$$

14.2.3 Connection in Cylindrical Coordinates

We use the formula

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2}g^{\alpha\delta}(g_{\delta\beta,\gamma} + g_{\gamma\delta,\beta} - g_{\beta\gamma,\delta})\ . \tag{14.4}$$

Only derivatives with respect to r could be nonzero, and then only g_{22} is actually a function of r. In short, only $g_{22,1} = -2r$ is actually nonzero. We find

$$\Gamma^2_{21} = \frac{1}{2}g^{2\delta}(g_{\delta 2,1}) = \frac{1}{2}g^{22}g_{22,1}\ ,$$

noting that

$$g^{\alpha\beta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1/r^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} , \tag{14.5}$$

whence

$$\Gamma^2_{21} = \Gamma^2_{12} = \frac{1}{r} .$$

The only other nonzero component is

$$\Gamma^1_{22} = \frac{1}{2} g^{11} (g_{12,2} + g_{21,2} - g_{22,1}) = -\frac{1}{2}(-2r) = r .$$

So finally, the only nonzero connection coefficients are

$$\boxed{\Gamma^2_{21} = \frac{1}{r} = \Gamma^2_{12} , \qquad \Gamma^1_{22} = r} . \tag{14.6}$$

14.2.4 Killing Vector Field

We now show that

$$K := \partial_t + v \partial_\theta \tag{14.7}$$

is a Killing vector field. In component form,

$$K^0 = 1 , \qquad K^1 = 0 , \qquad K^2 = v , \qquad K^3 = 0 , \tag{14.8}$$

and

$$K_0 = 1 , \qquad K_1 = 0 , \qquad K_2 = -r^2 v , \qquad K_3 = 0 . \tag{14.9}$$

The Killing equation is

$$K_{\alpha;\beta} + K_{\beta;\alpha} = 0 , \tag{14.10}$$

or

$$K_{\alpha,\beta} + K_{\beta,\alpha} = 2\Gamma^\gamma_{\alpha\beta} K_\gamma . \tag{14.11}$$

In this case,

$$\begin{aligned} 2\Gamma^\gamma_{\alpha\beta} K_\gamma &= 2\big(\Gamma^0_{\alpha\beta} K_0 + \Gamma^1_{\alpha\beta} K_1 + \Gamma^2_{\alpha\beta} K_2 + \Gamma^3_{\alpha\beta} K_3\big) \\ &= 2\big(\Gamma^0_{\alpha\beta} - \Gamma^2_{\alpha\beta} r^2 v\big) \\ &= -2\Gamma^2_{\alpha\beta} r^2 v , \end{aligned}$$

noting in the last step that $\Gamma^0_{\alpha\beta} = 0$ for all α, β. From the connection coefficients in (14.6), we see that we must have $(\alpha, \beta) = (1,2)$ or $(2,1)$ to get anything from this term. But this is also true on the other side of the Killing equation (14.11), because only K_2 has a nonzero coordinate derivative and that has to be a derivative with respect to r. Then we just note that

$$K_{2,1} + K_{1,2} = -2r\nu$$

and

$$-2\Gamma^2_{12} r^2 \nu = -2r\nu \ .$$

We conclude that the Killing equation is satisfied.

Let us just check that our spiralling worldline is a flow curve of K, i.e., that K is tangent to that worldline. Let the spiral be λ. Then for any function $f : \mathbb{M} \to \mathbb{R}$, possibly only defined on a small neighbourhood of part of λ, the tangent vector T to λ has the effect

$$\begin{aligned} Tf &= \left.\frac{\partial f}{\partial \tau}\right|_\lambda \quad \text{(derivative along } \lambda) \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau} f\big(\gamma(\tau)\big) \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau} f\big(\gamma\tau, R\cos(\nu\gamma\tau), R\sin(\nu\gamma\tau), 0\big) \quad \text{(in Cartesian coordinates)} \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau} f(\gamma\tau, R, \nu\gamma\tau, 0) \quad \text{(in cylindrical coordinates)} \\ &= \gamma\partial_\tau f + \nu\gamma\partial_\theta f \\ &= \gamma K \ . \end{aligned}$$

So γK is the tangent vector when the worldline is parametrised by proper time. This vector has unit length all along the worldline, of course. In fact, K was not normalised:

$$K^2 = g_{\alpha\beta} K^\alpha K^\beta = 1 - r^2\nu^2 \ , \tag{14.12}$$

and along the worldline $r = R$, so

$$K^2(\text{on worldline}) = 1 - R^2\nu^2 = \gamma^{-2} \ . \tag{14.13}$$

This bears out the general fact that a Killing vector field always has constant length along any of its flow curves, proven by observing that

$$K^\mu \nabla_\mu (K^2) = 2K^\mu K^\nu \nabla_\mu K_\nu = K^\mu K^\nu (\nabla_\mu K_\nu + \nabla_\nu K_\mu) = 0 \ ,$$

and hence can be normalised along any preselected flow curve (although not generally everywhere in spacetime).

Note also from (14.12) that the KVF is timelike within the light cylinder $r < 1/\nu$, and spacelike without. This will be important in a moment.

14.2.5 Rotating Cylindrical Coordinates

The observer on the spiral at R and with angular speed ν does not just move up the time axis in the cylindrical coordinates of the last section. We thus change to what might be called rotating cylindrical coordinates (t, r, θ', z), where

$$\theta' := \theta - \nu t \ . \qquad (14.14)$$

The observer worldline is now

$$x(\tau) = \big(t(\tau) = \gamma\tau, R, 0, 0\big) \ . \qquad (14.15)$$

The observer is moving up a worldline 'parallel' to the time axis, i.e., with fixed space coordinates, but with $R \neq 0$. Contrast with the rigid SE coordinate system we attached to an observer with this kind of motion in Sect. 2.11.

We now carry out the same investigation as above:

- Find the metric components in rotating cylindrical coordinates.
- Find the connection coefficients in rotating cylindrical coordinates.
- Rewrite the Killing vector field $K := \partial_t + \nu\partial_\theta$ in the new coordinates and check the Killing equation, just to make sure that there have been no errors.

14.2.6 Metric in Rotating Cylindrical Coordinates

A short cut to the metric is to write heuristically

$$\mathrm{d}\theta = \mathrm{d}\theta' + \nu\mathrm{d}t \ ,$$

then substitute this into the expression for the metric interval:

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}t^2 - \mathrm{d}r^2 - r^2(\mathrm{d}\theta' + \nu\mathrm{d}t)^2 - \mathrm{d}z^2 \\ &= \mathrm{d}t^2 - \mathrm{d}r^2 - r^2\mathrm{d}\theta'^2 - r^2\nu^2\mathrm{d}t^2 - 2r^2\nu\mathrm{d}\theta'\mathrm{d}t - \mathrm{d}z^2 \ ,\end{aligned}$$

so

$$\mathrm{d}s^2 = (1 - r^2\nu^2)\mathrm{d}t^2 - \mathrm{d}r^2 - r^2\mathrm{d}\theta'^2 - 2r^2\nu\mathrm{d}\theta'\mathrm{d}t - \mathrm{d}z^2 \ , \qquad (14.16)$$

and

$$\boxed{g^{\mathrm{rot}}_{\alpha\beta} = \begin{pmatrix} 1/\gamma(r)^2 & 0 & -r^2\nu & 0 \\ 0 & -1 & 0 & 0 \\ -r^2\nu & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \ , \qquad 1/\gamma(r)^2 = 1 - r^2\nu^2} \ . \qquad (14.17)$$

The metric is in stationary, but not static form.

The raised form is

$$\boxed{g^{\alpha\beta}_{\text{rot}} = \begin{pmatrix} 1 & 0 & -\nu & 0 \\ 0 & -1 & 0 & 0 \\ -\nu & 0 & -1/r^2\gamma(r)^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}} . \tag{14.18}$$

14.2.7 Connection in Rotating Cylindrical Coordinates

We use the formula

$$^{\text{rot}}\Gamma^{\alpha}_{\beta\gamma} = \frac{1}{2} g^{\alpha\delta}_{\text{rot}} (g^{\text{rot}}_{\delta\beta,\gamma} + g^{\text{rot}}_{\gamma\delta,\beta} - g^{\text{rot}}_{\beta\gamma,\delta}) . \tag{14.19}$$

In the calculations, we shall drop the suffix indicating that this is relative to *rotating* cylindrical coordinates. As before, only derivatives with respect to r, which is coordinate 1, actually give something here. This gives four nonzero coordinate derivatives of $g^{\text{rot}}_{\alpha\beta}$, viz.,

$$g^{\text{rot}}_{00,1} = -2r\nu^2 , \quad g^{\text{rot}}_{02,1} = -2r\nu , \quad g^{\text{rot}}_{20,1} = -2r\nu , \quad g^{\text{rot}}_{22,1} = -2r .$$

Now, regarding the prefactor $g^{\alpha\delta}$,

$$\begin{cases} \alpha = 0 & \Longrightarrow \quad \delta = 0 \text{ or } 2 , \\ \alpha = 1 & \Longrightarrow \quad \delta = 1 , \\ \alpha = 2 & \Longrightarrow \quad \delta = 0 \text{ or } 2 , \\ \alpha = 3 & \Longrightarrow \quad \delta = 3 . \end{cases}$$

Consider first

$$\Gamma^0_{\beta\gamma} = \frac{1}{2} g^{00} (g_{0\beta,\gamma} + g_{\gamma 0,\beta} - g_{\beta\gamma,0}) + \frac{1}{2} g^{02} (g_{2\beta,\gamma} + g_{\gamma 2,\beta} - g_{\beta\gamma,2}) .$$

We know that $g_{\beta\gamma,0}$ and $g_{\beta\gamma,2}$ are zero for any β,γ, and we obtain something nonzero from the rest in the cases $(\beta,\gamma) = (0,1)$, $(1,0)$, $(1,2)$, and $(2,1)$. However, the calculation leads to

$$\boxed{^{\text{rot}}\Gamma^0_{01} = {}^{\text{rot}}\Gamma^0_{10} = 0 = {}^{\text{rot}}\Gamma^0_{12} = {}^{\text{rot}}\Gamma^0_{21}} . \tag{14.20}$$

Now consider

$$\Gamma^1_{\beta\gamma} = \frac{1}{2} g^{11} (g_{1\beta,\gamma} + g_{\gamma 1,\beta} - g_{\beta\gamma,1}) = \frac{1}{2} g_{\beta\gamma,1} .$$

We thus obtain

$$\boxed{^{\text{rot}}\Gamma^1_{00} = -r\nu^2 , \quad {}^{\text{rot}}\Gamma^1_{02} = -r\nu = {}^{\text{rot}}\Gamma^1_{20} , \quad {}^{\text{rot}}\Gamma^1_{22} = -r} . \tag{14.21}$$

Next we have

$$\Gamma^2_{\beta\gamma} = \frac{1}{2}g^{20}(g_{0\beta,\gamma} + g_{\gamma 0,\beta} - g_{\beta\gamma,0}) + \frac{1}{2}g^{22}(g_{2\beta,\gamma} + g_{\gamma 2,\beta} - g_{\beta\gamma,2}) ,$$

where $g_{\beta\gamma,0}$ and $g_{\beta\gamma,2}$ are zero, so

$$\Gamma^2_{\beta\gamma} = -\frac{\nu}{2}(g_{0\beta,\gamma} + g_{\gamma 0,\beta}) + \frac{1}{2}\left(\nu^2 - \frac{1}{r^2}\right)(g_{2\beta,\gamma} + g_{\gamma 2,\beta}) .$$

This is only nonzero for $(\beta,\gamma) = (0,1)$, $(1,0)$, $(1,2)$, or $(2,1)$. We soon obtain

$$\boxed{{}^{\mathrm{rot}}\Gamma^2_{01} = \frac{\nu}{r} = {}^{\mathrm{rot}}\Gamma^2_{10} , \qquad {}^{\mathrm{rot}}\Gamma^2_{12} = \frac{1}{r} = {}^{\mathrm{rot}}\Gamma^2_{21}} . \tag{14.22}$$

Finally,

$$\Gamma^3_{\beta\gamma} = \frac{1}{2}g^{33}(g_{3\beta,\gamma} + g_{\gamma 3,\beta} - g_{\beta\gamma,3}) = 0 ,$$

whence

$$\boxed{{}^{\mathrm{rot}}\Gamma^3_{\beta\gamma} = 0 , \quad \forall\, \beta,\gamma} . \tag{14.23}$$

14.2.8 Killing Vector Field Revisited

Let us reexpress the KVF $K = \partial_t + \nu\partial_\theta$ relative to the rotating cylindrical coordinates and check Killing's equations as a way to ensure that there are no errors in the above calculations. The components in cylindrical coordinates are

$$K^0 = 1 , \qquad K^1 = 0 , \qquad K^2 = \nu , \qquad K^3 = 0 , \tag{14.24}$$

and in rotating cylindrical coordinates they are given by

$$K^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\nu} K^\nu , \tag{14.25}$$

where

$$\left(\frac{\partial x^{\mu'}}{\partial x^\nu}\right) = \begin{pmatrix} \partial x^{0'}/\partial x^0 & \partial x^{0'}/\partial x^1 & \partial x^{0'}/\partial x^2 & \partial x^{0'}/\partial x^3 \\ \partial x^{1'}/\partial x^0 & \partial x^{1'}/\partial x^1 & \partial x^{1'}/\partial x^2 & \partial x^{1'}/\partial x^3 \\ \partial x^{2'}/\partial x^0 & \partial x^{2'}/\partial x^1 & \partial x^{2'}/\partial x^2 & \partial x^{2'}/\partial x^3 \\ \partial x^{3'}/\partial x^0 & \partial x^{3'}/\partial x^1 & \partial x^{3'}/\partial x^2 & \partial x^{3'}/\partial x^3 \end{pmatrix} ,$$

whence

$$\left(\frac{\partial x^{\mu\prime}}{\partial x^{\nu}}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\nu & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{14.26}$$

We now observe that

$$\begin{pmatrix} K^{0\prime} \\ K^{1\prime} \\ K^{2\prime} \\ K^{3\prime} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\nu & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \nu \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} ,$$

whence

$$K = \partial_t , \tag{14.27}$$

in the rotating cylindrical coordinate system. This is obviously tangent to the world-line

$$x(\tau) = \big(t(\tau) = \gamma(R)\tau, R, 0, 0\big) .$$

Equation (14.27) gives the contravariant vector field, with components

$$K^0 = 1 , \qquad K^1 = 0 , \qquad K^2 = 0 , \qquad K^3 = 0 , \tag{14.28}$$

dropping the primes.

The covariant vector has components found by multiplying by the matrix of $g^{\mathrm{rot}}_{\alpha\beta}$ in (14.17) on p. 407, which gives

$$K_0 = 1 - \nu^2 r^2 , \qquad K_1 = 0 , \qquad K_2 = -\nu r^2 , \qquad K_3 = 0 . \tag{14.29}$$

Let us check that this does satisfy the Killing equation

$$K_{\alpha,\beta} + K_{\beta,\alpha} = 2\Gamma^{\gamma}_{\alpha\beta} K_\gamma .$$

This requires that

$$K_{\alpha,\beta} + K_{\beta,\alpha} = 2\Gamma^{0}_{\alpha\beta} K_0 + 2\Gamma^{2}_{\alpha\beta} K_2 ,$$

which is equivalent to

$$K_{\alpha,\beta} + K_{\beta,\alpha} = -2\Gamma^{2}_{\alpha\beta} r^2 \nu , \quad \forall\, \alpha, \beta ,$$

since we know that $\Gamma^0_{\alpha\beta} = 0$ for all α, β, by (14.20). Now referring to (14.22), we note that we must have $(\alpha,\beta) = (0,1)$, $(1,0)$, $(1,2)$, or $(2,1)$ to get something on the right-hand side, but this is also true on the left-hand side, since the covariant components of K are only functions of r, and then only the 0 and 2 components of K are actually functions of r. It is a simple matter to check that the equation is satisfied for these cases.

14.2.9 Detectors

It turns out that, in Minkowski spacetime, a two-level detector interacting linearly with a KG quantum field, i.e., the Unruh–DeWitt detector, will excite with nonzero probability when moving in a circular orbit with constant angular speed in the usual Minkowski vacuum of that field. Not surprisingly, this is true whatever the angular speed, even changing, but discussions focus on the case of constant angular speed, and not only because it is of course simpler. In fact it highlights a problem with the kind of discussion Unruh theorists engage in. The point is that, when the angular speed is constant, the worldline of the detector is a flow curve of a Killing vector field, as we have just seen, but when it varies, it is not. So what particle interpretation could we give in the latter case?

However, there is an interesting twist in the case of constant angular speed Ω. If we restrict the region of the spacetime available to the quantum field to some cylindrical region of radius R about the axis of rotation, where $\Omega R < c$, so that the region occupied by the field lies entirely within the light cylinder, but where the radius r of the worldline satisfies $r < R$, it turns out that the detector does not excite.

This is not the place to go into the details of the Unruh–DeWitt detector, although it is surprisingly straightforward. Many accounts can be found by looking at [11] and references therein. The QFT construction itself is also straightforward. We solve the KG equations with some boundary conditions on the cylinder of radius R, e.g., Dirichlet boundary conditions, which stipulate that the field is zero on and beyond the cylinder. When we do the expansion of the quantum field in terms of annihilation and creation operators, the resulting operator-valued distribution has support within the cylinder of radius R.

Incidentally, this shows how very different a QFT vacuum is from what we used to call a vacuum. The latter was some region of spacetime where there was nothing. The former is a state of a field in which there is not nothing in general, although there may be nothing in some regions, e.g., outside the cylinder of radius R in the case just described.

There is a strong sense in which the *physical* interest of the Unruh theory ends here, with these albeit interesting insights into the nature of the quantum vacuum. Although it may be amusing to try to establish a picture or perspective for someone moving with the detector, it is almost never possible using the Killing vector field idea, and it seems hardly necessary anyway. Surely what matters is just the interaction and resulting behaviour of our detectors, something we can always predict, or at least estimate, for any detector motion? But let us nevertheless review the twist in the story that is supposed to be exemplified by uniform circular motion.

Returning to the unrestricted Minkowski spacetime, we were not surprised to find the detector exciting, even if the field is in its vacuum state. This is because the detector itself is not in an inertial state, but accelerating, which requires some input of energy. So that seems comprehensible enough. But now we wonder how this could be understood by someone moving with the detector, and using rotating cylindrical coordinates, for example. The point about these coordinates is that our observer is stationary relative to such coordinates, i.e., there is no acceleration for such a person, at least in the naive coordinate sense.

We are trying now to make this look paradoxical. Here is someone who is not accelerating (in that naive, coordinate-based sense). If the field state were the vacuum for this person, then that is supposed to seem odd, because of the non-accelerating state of the detector in this 'perspective'. So for this 'perspective' to make sense, we would prefer to say that the usual Minkowski vacuum is not the vacuum for such an observer, but a state of the field containing some kind of particles for this observer. Then it does not matter if the detector is not accelerating in this perspective, because even an 'inertially moving' detector should be able to detect particles if there are any.

However, this is actually a very silly idea. In fact, all our theories of physics are generally covariant in the sense that they can of course be formulated relative to any coordinate system, precisely because a choice of coordinates by an observer should be irrelevant to what happens physically. Not moving relative to some coordinates is a meaningless fact according to the basic principles of general relativity, which also underpin what happens in flat spacetimes. Coordinates are just coordinates. (The idea is pushed to an extreme in the general theory of relativity, because in this theory there is *only* a generally covariant formulation.)

The presence of a Killing vector field slightly improves the situation for people who think it worth giving this kind of particle interpretation. The observer moving on the circular orbit with constant angular speed, whose worldline is therefore a regular spiral in flat spacetime, is following the flow curve of a KVF, so there is a sense in which the metric always 'looks the same' to this person, since the Lie derivative of the metric is zero along the worldline. This is supposed to give a kind of geometrical version of being 'at rest'.

On the other hand, this person is accelerating. We have the same situation with the eternally uniformly accelerating observer in flat spacetime, who is also following the flow curve of a KVF. This person can likewise claim to be 'at rest' in the above geometrical sense. It is certainly interesting in a mathematical sense that we can achieve alternative perspectives, e.g., for the case of eternally uniformly accelerating detectors and observers, with the associated notion of temperature. But it is not necessary or useful and could not possibly have any deep physical meaning. If it did, we would be able to do it for *any* motion of detector and observer, even motions that are not flow curves of Killing vector fields.

We have already encountered a similar very silly idea in classical electromagnetism with respect to the eternally uniformly accelerating observer (see Chap. 11). If this person carries an electron, it will radiate electromagnetic energy, at least according to the usual notion of EM radiation for any inertially moving detector (or observer). But what about a detector (or observer) moving with the electron?

For the detector, we need to know something about its design and interaction with the EM field, and there is every chance that what it does when it accelerates will depend on both those features, and not be the same for two detectors with different designs and interactions, even if they always register the same results when moving inertially in the same EM fields. The reason is that, although our field theories are all Lorentz symmetric, there is no corresponding acceleration symmetry as far as we know.

For the observer, then, there is a problem: how to define radiation when one is accelerating? Now for various reasons, people who have discussed this problem over the past 100 years (yes, that long!) have wanted the accelerating observer not to see any radiation, or not to detect any, or to believe that there isn't any. A common idea has been to adopt a semi-Euclidean coordinate system in which this observer remains at the space origin and then point out that the generalisation of the Poynting vector for these particular EM fields is zero. But why use those coordinates? And what is the physical meaning of this generalisation of the Poynting vector?

Indeed, why should we not want the observer moving with the charge to believe that there is EM radiation? Just to keep our feet on the ground, physically, the only important thing here is to be able to predict what this or that design of detector should do when moving with the electron in this motion. But psychologically, perhaps what we are trying to do is to extend the principle of relativity (which refers to relativity of velocities) to a relativity of accelerations: however I am moving, if I carry an electron with me, I can never deduce anything about my motion by looking at its fields.

This is true if we restrict the motions to inertial motions, due to the Lorentz symmetry of Maxwell's theory, but it is not actually true if we allow accelerating motions, because even though the generalisation of the Poynting vector to the comoving semi-Euclidean coordinate system is zero for eternal uniform acceleration, the EM fields of the electron nevertheless look quite different to Coulomb fields when expressed relative to the comoving semi-Euclidean coordinate system. The observer could tell whether she was accelerating by looking at the EM fields of the electron.

And why should we want to extend the principle of relativity to accelerations? It looks as though we have the same kind of mistaken intention with the Unruh theory in the case of eternal uniform acceleration. There is in fact a perfect parallel with the classical example. Let us sketch that.

What the observer wants to say is that, when she is accelerating, the Minkowski vacuum of the quantum field is not the vacuum. She likes to put it that way because a detector she carries is registering something and she does not think it should register something when they are both 'sitting still' in a vacuum. So she prefers to say that this state of the field contains particles as far as she is concerned. Then she does not need to be surprised to find that her detector registers something.

Actually, we do not need to be surprised anyway. Acceleration is acceleration. It does not go away just because we can find coordinates relative to which we do not move. It is better to understand the situation like this, because for most accelerating motions there is absolutely no particle picture anyway, so there could be no deep physical significance to these notions.

What is the classical parallel? Some authors (e.g., Rohrlich, Boulware) claim that an observer accelerating through the Coulomb fields of an inertially moving electron will think it is radiating. Some people (Mould) have even invented detectors to prove this! But what coordinate or other frame should the accelerating observer use to see the Coulomb fields as radiating? Each choice will give a different picture, and not one of those pictures can make the fields look in the details exactly like the fields due to an electron with the same acceleration as this observer but when viewed from an inertial frame. There is no acceleration symmetry in Maxwell's theory.

Of course, one can make definitions for accelerating observers. But why bother? What matters is just what this or that detector will do, and we can predict that.

Now let us return to the tricky (but physically irrelevant) issue of the observers in circular orbit at constant angular speed in flat spacetime. We said before that we were trying to make the exciting detector look paradoxical to someone moving with it, because this person considers herself stationary (and she is, in the rotating cylindrical coordinate system). If the field state (the usual Minkowski vacuum) were the vacuum for this person, then that is supposed to seem odd, because of the non-accelerating state of the detector in this 'perspective'. So for this 'perspective' to make sense, we would prefer to say that the usual Minkowski vacuum is not the vacuum for such an observer, but a state of the field containing some kind of particles for this observer. Then it does not matter if the detector is not accelerating in this perspective, because even an 'inertially moving' detector should be able to detect particles if there are any.

But the trouble is that, if we think we can apply the 'axioms' for building the QFT in the rotating cylindrical coordinate perspective, the usual Minkowski vacuum turns out to be the vacuum for that new construction too. So this idea of changing the definition of 'particle' for someone stationary relative to the rotating cylindrical coordinate perspective would not appear to save us here. And, as the story goes, we know why that is: it is because the Killing vector field associated with this worldline goes spacelike beyond some distance from the centre of rotation [see (14.12) on p. 406]. A proper QFT construction requires a *global* timelike Killing vector field. So one of the 'axioms' authorizing this alternative construction has not been fulfilled.

We can 'prove' that this is the correct interpretation by looking at the truncated quantum field that goes to zero outside a cylinder that contains the worldline but lies within the light cylinder. As mentioned above, the detector on the regular spiralling worldline does not excite in the usual Minkowski vacuum of such a field configuration, even though it is accelerating. In addition, the vacuum constructed using the 'axioms' for building the QFT in the rotating cylindrical coordinate perspective is exactly the same as the usual Minkowski vacuum.

Apart from the fact that we may consider this to 'solve' a non-problem for the reasons sketched above (why should we care if an observer stationary in some coordinate system should see a detector excite, because being stationary in some coordinate system is physically irrelevant?), should we not now be concerned that a detector accelerating through a QFT vacuum does not register? After all, it is really accelerating, not just doing something relative to some coordinates!

Here are two other questions intended to highlight the above points:

- Why does the observer with circular motion use rotating cylindrical coordinates for the QFT construction? The only convenient feature about them is apparently that she sits at the space origin of these coordinates throughout the motion. Why not use a Fermi–Walker transported tetrad along the worldline to construct a truly rigid coordinate system (in the sense discussed earlier in this book)? Or some other coordinate system in which the observer remains at the space origin throughout the motion? Which choice should be decreed as giving the 'perspective' of the rotating observer?

- In the classical example of Chap. 11 and discussed again above, some commentators have the idea that the EM fields of an eternally uniformly accelerating charge should not look like radiating fields when expressed relative to the comoving semi-Euclidean (and rigid) coordinate system, even that they should look like Coulomb fields. But they do not, although they share some qualitative features with Coulomb fields (zero magnetic field, time constant electric field). These fields are perhaps analogous therefore to the Rindler vacuum in the QFT construction for an eternally uniformly accelerating observer. But does the Rindler vacuum really look (to the eternally uniformly accelerating observer) in all the details like a Minkowski vacuum (to the inertially moving observer)?

The last question extends to something more general. Does the whole Rindler vacuum and particle construction look, to the eternally uniformly accelerating observer, in the details like the Minkowski vacuum and particle construction to the inertially moving observer? If it did, there ought to be enough symmetry between the two 'perspectives' to be able to show that the Rindler vacuum is actually a thermal state of Minkowski particles. Is that the case?

14.3 A Conclusion of Sorts

In this section, we have baulked at the claim that the quantisation of the scalar field in Rindler modes is just the natural quantisation a Rindler observer would perform. The defence for such a claim is that spacetime always looks the same to this observer, so he feels at rest just like an object sitting on the surface of an infinite flat Earth. The arguments against have been spelt out in enough detail by now. But surely we need a way of thinking about what we mean by QFT and particles in a general spacetime?

If we want to define particles, we have to be able to use the positive and negative frequency solutions of the field equations to define creation and annihilation operators. But this construction depends on the spacetime having a natural time coordinate. What could we do if it did not? Surely we can still do quantum field theory? The answer is clearly that we should carry out the construction, albeit approximately, in locally inertial systems, applying the weak and strong equivalence principles. This will always work, and it is precisely in such locally inertial frames that we can claim to understand things physically. But this already rules out the Rindler picture.

So the aim here is not to make an exception of the Rindler wedge alone, but of all so-called pictures by accelerating observers, including any that is supposed to come from a nice KVF-based solution of some field equations with associated construction of a quantum field. Of course, these constructions are there and they are interesting, but as physicists, or natural philosophers, we should ask for more justification of the physical interpretation. Geometric elegance, it should be remembered, is mathematics. Our problem is to relate that to the real world.

In an absolutely general spacetime with nothing to help us, what would we do? We would use a locally inertial frame and construct there. That is justified on principle (the strong equivalence principle). The only way to understand any other construction, e.g., relative to a KVF if there is one, is to refer to these locally inertial systems. But there seems to be no attempt to do that. Instead, we are encouraged to pretend that some non-inertial coordinates are actually inertial (compare with the account of freely falling charges criticised in Sect. 11.10).

And of course we can always analyse the behaviour of detectors with arbitrary worldlines in arbitrary spacetimes without any reference to the particle interpretations that might be imputed to comoving observers. For instance, regarding the pointlike Unruh–DeWitt detectors in QFT, one can analyse the behaviour of a particle detector with an arbitrary motion in Minkowski space. Coordinates make no difference, fortunately, to what the theory says it will register. So we do not actually need the alternative Rindler wedge construction of the quantum field. That serves only to attempt to give a particle picture.

References

1. R.T.W. Arthur: Minkowski's proper time and the status of the clock hypothesis. In: V. Petkov (ed.), *Space, Time, and Spacetime. Physical and Philosophical Implications of Minkowski's Unification of Space and Time*, Springer, Berlin, Heidelberg, New York (2010)
2. J.S. Bell: *Speakable and Unspeakable in Quantum Mechanics*, 2nd edn., Cambridge University Press, Cambridge (2004) Chap. 9
3. M. Blau: *Lecture Notes on General Relativity*, available on the Web
4. H. Bondi, T. Gold: The field of a uniformly accelerated charge, with special reference to the problem of gravitational acceleration, Proc. Roy. Soc. A **229**, 416–424 (1955)
5. D. Boulware: Radiation from a uniformly accelerated charge, Annals of Physics **124**, 169–188 (1980)
6. H.R. Brown: *Physical Relativity: Space–Time Structure from a Dynamical Perspective*, Clarendon Press, Oxford (2005)
7. H.R. Brown: The behaviour of rods and clocks in general relativity, and the meaning of the metric field. In: *Beyond Einstein: Essays on Geometry, Gravitation, and Cosmology*, ed. by D.E. Rowe, Einstein Studies, Vol. 12, Birkhäuser, Boston (2009)
8. H.R. Brown, O. Pooley: Minkowski spacetime. A glorious non-entity. In: *The Ontology of Spacetime*, ed. by D. Dieks, Elsevier (2006). Also available at `philsci-archive.pitt.edu/archive/00001661`
9. H.R. Brown, O. Pooley: The origin of the spacetime metric. Bell's 'Lorentzian pedagogy' and its significance in general relativity. In: *Physics Meets Philosophy at the Planck Scale*, ed. by C. Callender, N. Huggett, Cambridge University Press (2001). Also available at `gr-qc/9908048`
10. J. Butterfield: Reconsidering relativistic causality, Intl. Studies Phil. Sci. **21**, issue 3 (2007), Sect. 4
11. L.C.B. Crispino, A. Higuchi, G.E.A. Matsas: The Unruh effect and its applications, `arXiv:0710.5373v1 [gr-qc]` (29 Oct 2007). This reference gives a complete overview with bibliography. A nice introduction to the Unruh effect can also be found in F. Hoffmann: *On aspects of particle interpretation of quantum field theory, as exemplified by the Unruh effect*, PhD thesis, Ludwig-Maximilians-Universität, Munich (2010)
12. F. de Felice, C.J.S. Clarke: *Relativity on Curved Manifolds*, Cambridge University Press, Cambridge (1990)
13. B.S. DeWitt, R.W. Brehme: Radiation damping in a gravitational field, Annals of Physics **9**, 220–259 (1960)
14. B.S. DeWitt: *Bryce DeWitt's Lectures on Gravitation*, ed. by S.M. Christensen, Springer, Berlin, Heidelberg (2011)
15. D. Dieks: Understanding in physics: Bottom-up versus top-down. In: *Scientific Understanding: Philosophical Perspectives*, ed. by H. de Regt, S. Leonelli, K. Eigner, University of Pittsburgh Press, Pittsburgh (2009) Chap. 12

16. P.A.M. Dirac: Classical theory of radiating electrons. Proc. Roy. Soc. A **167**, 148 (1938)
17. C. Doran, A. Lasenby: *Geometric Algebra for Physicists*, Cambridge University Press, Cambridge (2003)
18. M. Dorato: Relativity theory between structural and dynamical explanations, International Studies in the Philosophy of Science **21** (1), 95–102 (2007)
19. S. Dürr et al.: Ab initio determination of light hadron masses, `arXiv:0906.3599v1 [hep-lat]` (19 Jun 2009)
20. A. Eisele: On the behaviour of an accelerated clock, Helvetica Physica Acta **60**, 1024–1037 (1987)
21. R.P. Feynman, R.B. Leighton, M. Sands: *The Feynman Lectures on Physics*, Vol. II. Addison-Wesley, Reading MA (1964)
22. M. Friedman: *Foundations of Space–Time Theories*, Princeton University Press, NJ (1983)
23. Y. Friedman, T. Scarr: *Covariant Uniform Acceleration*, `arXiv:1105.0492v2 [phys.gen-ph]` (3 Oct 2011)
24. S.A. Fulling: Review of some recent work on acceleration radiation, J. Mod. Opt. **52**, 2207–2213 (2005)
25. M. Göckeler, T. Schücker: *Differential Geometry, Gauge Theories, and Gravity*, Cambridge University Press, Cambridge (1987)
26. S.E. Gralla, A.I. Harte, R.M. Wald: A rigorous derivation of electromagnetic self-force, `arXiv:0905.2391v2 [gr-qc]` (21 July 2009)
27. S.W. Hawking, G.F.R. Ellis: *The Large Scale Structure of Space–Time*, Cambridge University Press, Cambridge (1973)
28. M. Longair: *Theoretical Concepts in Physics*, Cambridge University Press, Cambridge (1984)
29. R.J. Low: Personal communication. Many thanks for bringing the uniform gravitational fields to my attention in Sect. 12 and suggesting the proof in Sect. 14.1.
30. S.N. Lyle: *Uniformly Accelerating Charged Particles. A Threat to the Equivalence Principle*, Springer, Berlin, Heidelberg (2008)
31. S.N. Lyle: Rigidity and the ruler hypothesis. In: V. Petkov (ed.), *Space, Time, and Spacetime. Physical and Philosophical Implications of Minkowski's Unification of Space and Time*, Springer, Berlin, Heidelberg, New York (2010); and almost unchanged in [32]
32. S.N. Lyle: *Self-Force and Inertia. Old Light on New Ideas*, Lecture Notes in Physics 796, Springer, Berlin, Heidelberg (2010). Chapter 8 is a synopsis of this book
33. D. Malament: A remark about the 'geodesic principle' in general relativity, to appear in a festschrift for W. Demopoulos but available as a preprint on the Web. See also J.O. Weatherall: On the status of the geodesic principle in Newtonian and relativistic physics, Studies in the History and Philosophy of Modern Physics, **42** (4) 276–281 (2011)
34. J.W. Maluf, S.C. Ulhoa: Electrodynamics in accelerated frames revisited, available from the archive as `arXiv:1009.3968v1 [physics.class-ph]` (20 Sep 2010)
35. B. Mashhoon: *Acceleration-Induced Nonlocality*, `arXiv:gr-qc/0301065v1` (17 Jan 2003)
36. B. Mashhoon: *The Hypothesis of Locality and Its Limitations*, `arXiv:gr-qc/0303029v1` (7 March 2003)
37. B. Mashoon: *Nonlocal Special Relativity*, `arXiv:0805.2926v1 [gr-qc]` (19 May 2008)
38. C.W. Misner, K.S. Thorne, J.A. Wheeler: *Gravitation*, Freeman, San Francisco (1973)
39. R.A. Mould: Internal absorption properties of accelerating detectors, Annals of Physics **27**, 1–12 (1964)
40. J.V. Narlikar: *Introduction to Cosmology*, Cambridge University Press, Cambridge (1993)
41. G. Nerlich: Bell's 'Lorentzian pedagogy': A bad education. This can be obtained from the archives as `philsci-archive.pitt.edu/5454/1/Bell.pdf` (2010)
42. G. Nerlich: Why spacetime is not a hidden cause: A realist story. In: *Space, Time, and Spacetime. Physical and Philosophical Implications of Minkowski's Unification of Space and Time*, Springer, Berlin, Heidelberg, New York (2010)
43. G. Nerlich: Private communication (2010)

44. J.D. Norton: Why constructive relativity fails. In: *Symposium on Time and Relativity*, Minneapolis, MN, (25–27 October 2007). Also available at `philsci-archive.pitt.edu/archive/00003655`
45. S. Parrott: Radiation from a uniformly accelerated charge and the equivalence principle, available from `arXiv:gr-qc/9303025 v8` (2001)
46. W. Pauli: Encycl. Math. Wiss. **19**, 648, 653 (1920)
47. V. Petkov: *Relativity and the Nature of Spacetime*, 2nd edn., Frontiers Series, Springer-Verlag, Berlin, Heidelberg (2009)
48. W. Rindler: *Introduction to Special Relativity*, Oxford University Press, New York (1982)
49. F. Rohrlich: *Classical Charged Particles*, 3rd edn., World Scientific Publishing, Singapore (2007)
50. L. Ryder: *Introduction to General Relativity*, Cambridge University Press, Cambridge (2009)
51. M.R. Spiegel, *Advanced Calculus*, Schaum's Outline Series, McGraw-Hill, (1974), p. 331
52. C.M. Will: The confrontation between general relativity and experiment, available at `www.livingreviews.org/lrr-2001-4` (2001)

Index

accelerating clock, 226
accelerating detector, 226, 368, 369, 375, 401–402, 411–415
acceleration due to gravity, 249
acceleration length, 265, 266
acceleration symmetry, 1, 2, 5, 226, 267, 376, 413, 414
acceleration tensor, 37, 38, 40
  antisymmetry of, 37, 48, 75
  constant, 75, 102
  for non-uniform circular motion, 91–93
  for uniform circular motion, 84
  Lorentz boost of, 49, 50
  non-constant, 43, 47, 98–102
  transformation of, 49
  translational form, 39, 49, 50, 75, 76
acceleration time, 265
action
  coordinate invariance, 308, 314, 315, 317
  design of, 307, 309, 310, 317
  diffeomorphism invariance, 307
  for charged particles in curved spacetime, 301
  for free particle in curved spacetime, 300, 309
  for gravity in GR, 307
  for point particle, 311
action principle, 300
adapted coordinate frame, 4
  in GR, 6, 146
  in SR, 12
atom as clock, 226, 231, 253, 257, 258, 321
  in de Sitter spacetime, 234, 236
  in Schwarzschild spacetime, 243, 244
  in SHGF, 252
atom as ruler, 231, 253, 257, 258, 320, 321
  in de Sitter spacetime, 234, 236
  in Schwarzschild spacetime, 243, 244
  in SHGF, 248, 252
axiomatisation, 260, 268, 324

bare mass, 287
baryon, 293, 294, 331
Bell, J.S., 7, 227, 229, 306, 321, 325
Bell-type argument, 227, 253, 266, 320, 322, 324
  in de Sitter spacetime, 233–237
  in GR, 229–256
  in Schwarzschild spacetime, 237–245
  in SHGF, 246–252
  negative reaction to, 323–325
Bianchi identity, 314, 316
block universe, 324
Bondi, H., 356, 358, 362
Born rigidity, 24, 286
Boulware, D., 359, 362, 363, 365, 371, 372, 391, 397, 413
bound state particles, 282, 292, 293, 296, 300, 330
  inertial mass, 283, 331
Brehme, R.W., 367
Brown, H.R., 219, 245, 255, 259, 266, 267, 306, 322
Butterfield, J., 308, 310

charge dipole, 299
charge dumbbell, 287
  arbitrary rigid motion, 292, 335–352
  binding force, 352
  Fermi–Walker transport, 341–346
  longitudinal linear motion, 290, 348
  longitudinal rotational motion, 291
  mass renormalisation, 290
  pion model, 296

rigidity assumption, 341–346
self-force, 346–349
transverse linear motion, 288, 348
transverse rotational motion, 291
charged rocket, 362
Christoffel symbols, 160
circular motion, 5, 9
cylindrical coordinates, 404–406
connection, 404
metric, 404
length calculations, 270–279
length measurement, 268–270
quantum field theory, 403–415
rotating cylindrical coordinates, 407–410
connection, 408
metric, 407
Unruh–DeWitt detector, 411, 414
with constant angular speed, 83–89
as flow curve of a KVF, 404–406
FS metric, 88
HOS intersection, 114–121
light cylinder, 112
range of validity of SE coordinates, 121
with varying angular speed, 89–109
acceleration tensor, 91
coordinates for non-precessing tetrad, 102–104
FW transported tetrad, 94–98
metric for FW transported tetrad, 108
metric for non-precessing tetrad, 104–108
non-precessing tetrad, 90
range of validity of SE coordinates, 111–112
surface $g_{00} = 0$, 135–139
clock hypothesis, 3, 7, 220, 245, 322
applied to atom, 227
evaluation of, 232, 254–258, 267
for muon decay, 266
clock synchronisation, 17, 28, 65, 161, 176
congruence of timelike worldlines, 7
connection, 7
for cylindrical coordinates in flat spacetime, 404
for de Sitter spacetime, 234
for normal frame, 151–152
for rotating cylindrical coordinates in flat spacetime, 408
for Schwarzschild spacetime, 239
for SE coordinates with FW transported axes, 30
for SE coordinates with rotating axes, 34, 74
in tetrad frame, 169
metric condition, 34, 141, 146, 151, 160
torsion free, 141, 146, 160, 170
transformation of, 159, 192
vanishing at a point, 142–146
consistency, 257
constructivist approach to relativity theories, 254, 324
continuous medium, 7, 193–218
dynamics, 206–214
energy density, 200–202
energy–momentum–stress density, 215–216, 218
law of motion, 208, 214
proper energy density, 194, 195
proper metric, 194, 197
stress tensor, 202–206
tetrad frame, 194, 195
three reference frames, 195–197
coordinate independence, 372, 377
Coriolis force, 19
cosmic time, 233
cosmological constant, 233
Coulomb field as EM radiation, 367, 368, 370, 381, 402, 413
covariance of all physical theory, 412
covariant derivative, 159
curvature scalar, 305

de Sitter spacetime, 233–237
connection, 234
metric, 233
normal coordinates, 234
definitions in inertial frames, 398
definitions in non-inertial frames, 8–10, 362, 364–369, 375, 397, 398, 400, 401, 411, 413, 414
DeWitt, B.S., 315, 317, 367
Dieks, D., 333
Dirac, P.A.M., 8, 281, 287, 335, 340
dogma in relativity theory, 324, 334
dust, 305, 306
energy–momentum tensor, 305
dynamical approach to relativity theories, 219, 245, 259–261, 306, 322
dynamical explanations for relativistic effects, 8, 229, 282, 290, 292, 297, 319–334

Einstein tensor, 305, 307
covariant divergence of, 307
Einstein's equations, 304, 377
as explanation for inertia, 304–317
from an action, 307
Einstein–Hilbert action, 307
Eisele, A., 266, 267
electromagnetic mass, 281–286

discrepancy between energy- and momentum-derived values, 284
energy-derived, 281
in quantum theory, 295, 296
momentum-derived, 282
self-force-derived, 285, 290
electromagnetic radiation, 353–370
ascertaining the existence of, 358, 359, 363, 382
detector, 226, 368, 402
explained by self-force, 286, 290
local test for, 363
viewed from accelerating frame, 5, 359, 362, 364, 381, 397, 412, 413
electron
as point particle, 283
as spatially extended particle, 281
classical radius, 284
inertial mass, 283
toy model, 287
elementary particles, 282, 293
energy condition, 317
energy–momentum conservation, 309
energy–momentum tensor, 193–218, 305, 361
for charge shell, 284
for dust, 305
for point particle, 305, 309, 311, 313, 317
equivalence principle, 356–358
application to charged particles, 363
Newtonian, 368, 369, 373
strong, *see* strong equivalence principle
unusable version, 362–363, 366, 382
useful formulations, 369
weak, *see* weak equivalence principle
Euler–Lagrange equations, 300
exponential map, 142

Faraday tensor, 355, 381
tetrad components, 390, 391, 394–397
Fermi rotation coefficients, 168, 171
transformation of, 192–193
Fermi–Walker transport, 4, 20, 25, 48–52, 74, 152, 335, 336, 344–345, 355
for non-uniform circular motion, 94–98
Feynman, R.P., 284, 296
FitzGerald contraction, 264, 282, 286, 319, 329
dynamical explanation for, 229, 320–325
flavour, 294, 296
form tetrad, 164
four-acceleration, 17, 326
four-force, 326
four-momentum, 326
four-velocity, 17, 325
free fall, 7, 247, 300, 304, 373
as uniform acceleration, 367
of charged particle, 317, 365–367, 378
of spatially extended particle, 317, 378
of spinning particle, 309, 378
viewed from normal frame, 153
viewed from SE frame, 19, 72
Friedman, M., 12, 303
Friedman, Y., 37, 219, 256
Friedman–Scarr transport, 41, 75, 102
and FW transport, 48–52
for uniform circular motion, 85
Frobenius theorem, 166, 175

Galileo, G., 318
gauge theory, 352
generalised uniform acceleration, 36–62, 75
acceleration transformations, 67–73
circular motion, 83–89
definition, 37
determinant of FS metric, 56
extent of generalisation, 53–55
metric for FS coordinate frame, 55–57, 75, 88
motion of fixed space points, 57–62, 72
nature of generalisation, 38–40
rigidity of coordinate frame, 44–48
SE coordinate frame, 40–44
geodesic, 7, 20, 142
null, 309, 310, 317
geodesic completeness, 142
geodesic coordinates, 102, 149
geodesic equation, 305, 311, 314, 367
geodesic principle, 8, 297, 300–302, 317
consequence of WEP and SEP, 378
for photons, 309
from Einstein's equations, 304–317
physical motivation, 301
geometric elegance, 377, 416
Gold, T., 356, 358, 362
gravity in SR, 3, 39, 365–367, 369

Hamiltonian for accelerating observer, 400
heresy in relativity theory, 325, 331
Higgs field, 283, 331
holonomic frame, 155
homogeneous spacetime, 374
horizon, 81, 359, 365
in SHGF, 371
hyperbolic motion, 77, 353
hyperplanes of simultaneity, 13, 161, 194, 220, 227
construction of, 16
intersection of, 6, 14, 21, 79, 80, 114–121

sharing, 27, 60, 76, 82, 386
hypersurfaces of simultaneity, 147–148, 161, 175
in de Sitter spacetime, 233

indistiguishability, 362
inertia, 8, 281–318
geometrical explanation, 318
self-force explanation, 318
inertial frame, 11
inertial mass, 281
ab initio calculations for light hadrons, 283
and energy, 292, 320, 331–332
and passive gravitational mass, 39, 297–300, 302–304
bootstrap hypothesis, 283, 293, 296, 330, 331
hybrid model, 283, 331
in quantum theory, 295
inclusion of binding energy, 292, 297, 320, 331–333
negative, 299
of bound state particles, 292, 331
of electron, 283
of elementary particles, 283
of neutron, 295
of proton, 283, 295
of quarks, 296
speed dependence, 282, 290, 297, 319, 325–331
strong force contribution, 295
isospin, 294, 295
symmetry breaking, 296

Jacobi identity, 157

Killing observer, 56, 57, 76, 365, 366, 375, 376, 378, 399
in flat spacetime, 376
limitations of, 378, 403, 411, 412
Killing vector field, 8, 9, 56, 187, 360–362
and conserved quantities, 360
and notion of rest, 9, 373–375, 377, 379
associated with uniform circular motion, 405–406, 409–410
in flat spacetime, 403
in Schwarzschild spacetime, 240
in static spacetime, 360, 364
in uniform gravitational field, 372
Lorentz boost, 360, 361, 373, 376, 399
normalising, 361, 376, 400
Klein–Gordon equation, 374, 399

label coordinates, 21, 194, 195

Lagrangian
for free particle in curved spacetime, 300
for spinning particle, 309
Larmor formula, 286, 329
lepton, 282, 293
Levi-Civita connection, 309
Lie bracket, 7, 156
Lie derivative, 56, 157
of metric, 187
Liénard–Wiechert potential, 229, 288, 289, 329, 337, 353, 355, 390
light deflection, 238, 366
light signals, 7, 221–224, 269
little group, 49
locality hypothesis, 5, 7, 256, 263–280
applied to lengths, 269
as approximation, 265
evaluation of, 258
practical interpretation, 268
weak, 7, 219–228, 256
locally inertial frame, 7, 141, 146, 154, 162, 257, 258, 266, 397, 398
for constructing QFT, 415
in SHGF, 246, 247
Lorentz force law, 40
in curved spacetime, 301
Lorentz symmetry, 1, 5, 226, 232, 254, 259, 263, 264, 267, 324, 369, 376, 413
and Bell argument, 227, 321
and inertia, 293, 331
local, 2, 369
of Maxwell's theory, 231, 290, 359, 413
Lorentz–Dirac equation, 335

Mach principle, 293
main observer, 21
Malament, D., 317
Maluf, J.W., 381, 382, 390, 392, 396, 397
Mashhoon, B., 256, 263, 268, 269, 279
mass renormalisation, 8, 281, 286–287, 330, 335
for charge dumbbell, 290
impossibility in non-relativistic limit, 351
impossibility of, 292, 335–352
in QED, 283, 284, 335
mass splitting, 294–296
Maxwell's equations, 229
and Newton's first law, 289
for tetrad components, 390
minimal extension to curved spacetime, 234, 257, 301, 309, 317, 357, 364, 378
meson, 293, 294, 331
metric

dynamical interpretation, 259, 308, 321, 323, 379
for cylindrical coordinates in flat spacetime, 404
for de Sitter spacetime, 233
for FS coordinate frame, 55–57, 75, 88
for normal frame, 149, 150
for observer with circular motion, 104–109
for rotating cylindrical coordinates in flat spacetime, 407
for Schwarzschild spacetime, 237
for SE coordinates with FW transported axes, 28, 74
for SE coordinates with rotating axes, 33, 74
for SHGF, 246
for TUA observer, 79
geometrical significance, 259, 378
in normal coordinates, 146
in static spacetime, 188
in stationary spacetime, 188
in tetrad frame, 155
Minkowski form, 11
standard interpretation, 231, 233, 234, 253, 254, 258
Mould detector, 226, 368, 402, 413
Mould, R.A., 226, 368, 413
multiplet, 293, 295, 296
muon decay, 266
as clock, 267

naturalness of operational procedures, 3–5, 7, 90, 220, 224, 227, 228, 400, 415
Nerlich, G., 318, 324, 325
neutron, 294
inertial mass, 295
quark flavour model, 294
Newton's first law, 285, 289, 293, 304, 378
Newton's second law, 281, 287, 302, 378
impossibility of renormalisation, 348
Lorentz covariant version, 320, 326
new version, 293, 297, 299, 304, 318
renormalised, 287, 289, 291, 330, 335, 336
Newton's third law, 285, 298
Newtonian approximation, 378
Newtonian gravity, 302, 377
geometrisation, 303
Newtonian world, 1, 3, 5, 12, 14, 17, 83, 264
non-holonomic frame, 7, 155
normal coordinates, 142–146, 149
cheap version, 230
in de Sitter spacetime, 234
in Schwarzschild spacetime, 240–242
in SHGF, 249
physical interpretation, 232, 253
normal frame, 146–153
connection, 151–152
construction, 146–149
non-rotating, 152–153
properties, 149–150
Norton, J.D., 324

ontology of spacetime, 318
operational definitions, 228, 231, 280
for lengths, 279, 280
for proper distance, 221
for times, 280

paradox in relativity theory, 331
Parrott, S., 359, 362–365
particle physics, 282, 293–296, 331
passive gravitational mass, 8, 39, 297–300, 303, 304
in GR, 299
self-force contributions, 299, 304
Pauli, W., 353
Petkov, V., 325
photon, 317, 378
geodesic principle, 309
in curved spacetime, 310
physics as semantic issue, 402
pion, 294
as charge dumbbell, 296
length, 296
quark model, 294
Poincaré group, 403
Poincaré, H., 284
point particle, 281, 283, 302, 308, 310, 330, 335
action, 311
as limit, 285, 317
energy–momentum tensor, 305, 309, 311, 313, 317
renormalisation problem, 330
Pooley, O., 322
Poynting vector, 353
in SE frame, 359, 362, 364, 413
projection tensor, 163, 165, 177, 194
proper distance, 224, 233, 253
established by Bell atom, 236, 243, 244, 252
established by light signals, 221–224, 227, 269
in de Sitter spacetime, 233
in Schwarzschild spacetime, 238, 244
in SHGF, 252
proper metric, 23–25
in curved spacetime, 177, 178
of continuous medium, 194, 197
proper time, 225, 227, 253, 267, 269

established by Bell atom, 236, 243, 244, 252
in Schwarzschild spacetime, 239, 243
in SHGF, 252
of Killing observers, 376
proton, 283, 294, 296
inertial mass, 283, 295
internal structure, 283
quark flavour model, 294
pseudo-energy, 8, 9, 362, 364–366, 400

quantum chromodynamics, 283
quantum electrodynamics, 283, 284, 335
quantum field theory, 9, 226, 283
axioms, 414
canonical construction, 399
for observer with circular motion, 403–415
for TUA observer, 374, 399–403
in general spacetime, 415
Minkowski thermal bath, 401, 402
ontological fuzziness, 399, 401, 402
renormalisability, 292, 352
Rindler particles, 374, 400, 401
truncated, 411
quark, 282, 293–296, 331
inertial mass, 283, 296
quasi-stationary quasi-canonical coordinates, 377, 378

radiation problem, 5, 8, 353–370, 381–398, 413
radiation reaction, 290, 329, 365
for charge with TUA motion, 355
rate of strain tensor, 172, 179–187
in curved spacetime, 182
in flat spacetime, 181
nonrelativistic, 179
relativity principle for accelerations, 359, 363, 413
rest, 17, 372, 374
as proper distance from source, 371, 375, 377, 378
coordinate dependence of, 358, 366, 371
different interpretations, 8
geometric notion, 9, 372–375, 377, 379, 412, 415
in accelerating frame, 57–62, 66, 72, 75, 81
in gravitational field, 356, 373
in Minkowski thermal bath, 402
in Rindler thermal bath, 402
in Schwarzschild spacetime, 365, 371, 373
in SHGF, 356–358, 363, 371
in static spacetime, 358–360, 364–366, 381
Newtonian, 377
Ricci rotation coefficients, 169–171
Ricci tensor, 305
rigid motion, 23–25, 75, 88, 248, 322
as approximation, 264, 286
in curved spacetime, 177–179, 187
nonrelativistic, 180
rigidity, 4, 8, 23, 24, 74, 248, 309, 329, 343, 344, 387
and HOS sharing, 60
for GUA coordinate frame, 44–48, 75
for inertial observer, 389
Rindler coordinates, 3, 248
Rindler observer, 376, 399
Rindler wedge, 79, 81, 373, 375
Robertson–Walker cosmologies, 374
rocket thought experiment, 325
Rohrlich, F., 363, 366–368, 391, 397, 413
ruler hypothesis, 3, 7, 25, 220, 245, 248, 267, 322, 364
evaluation of, 254–258
Ryder, L.H., 221, 238

Scarr, T., 37, 219, 256
Schwarzschild spacetime, 229, 237–245
connection, 239
distance from source, 371, 375
interpretation of, 238–239
metric, 237
normal coordinates, 240–242
stationary charge in, 365
timelike KVF, 373
self-force, 8, 281–297
absence for constant velocities, 285, 289, 293, 300, 304
and equality of inertial mass and PGM, 297–300
antigravity effect, 299
as breakdown of Newton's third law, 285
calculations, 287–292
calculations not assuming rigidity, 286
contributions to PGM, 299, 303
electromagnetic, 285–292, 329–333, 335–352
explanation of EM radiation, 286, 290, 297, 329
semi-Euclidean frame, 5, 73, 354
and locality hypothesis, 264
connection, 17, 30, 34, 74, 354
for circular motion, 83–109
for GUA motion, 43
for TUA motion, 76–83, 110, 376, 382–388
horizon, 81, 359
mathematical construction, 20, 27
metric, 28, 33, 74, 109, 355
metric determinant, 34, 109

motion of fixed space points, 66, 75, 81
normalised, 382–388
qualitative construction, 12
region of validity, 80, 109–139
surface $g_{00} = 0$, 122–139
time coordinate, 66
velocity transformations, 62–66
silly ideas, 412
simultaneity, 6
spatially extended charge distribution, 264, 267, 281, 327–331
binding forces, 282, 284, 329
bootstrap hypothesis, 330
discrepancy between energy- and momentum-derived masses, 284
energy-derived mass, 281
equality of momentum- and self-force-derived masses, 286
momentum-derived mass, 282
overall electrically neutral, 296
self-force on, 285, 329, 330
self-force-derived mass, 285, 290
spherically symmetric, 336, 349–351
spatially extended particles, 281
bootstrap hypothesis, 283, 293, 296
free fall, 317
spatially homogeneous spacetime, 374
spin currents, 305, 309
static homogeneous gravitational field, 5, 9, 230, 246–252, 298, 356, 366, 381
artefact of choice of coordinates, 372
doubt about homogeneity, 371, 373, 375
horizon, 371
locally inertial frame, 246, 247
metric, 246
non-existence of, 358
normal coordinates, 249
practical interpretation, 249
source, 249, 375
static spacetime, 8, 56, 188, 357–360, 364–366, 381
Killing vector field, 360, 364
stationary spacetime, 56, 187
strangeness in physics, 377
strong equivalence principle, 230–232, 234, 236, 244, 245, 256, 258, 266, 298, 301, 308–310, 321, 323, 357, 364, 369, 374, 378, 397, 398, 415
application *par excellence*, 357, 363, 381
as bold hypothesis, 2, 317
different implementations, 375
role in Bell-type arguments, 245, 253, 254
role in free fall, 317, 378
strong force, 294
structure coefficients, 7, 157, 166
in holonomic frame, 158
transformation of, 158, 190–192
superhelical motion, 25, 37, 46, 48, 75, 88
supporting against gravity, 242, 300, 304, 363, 371, 373, 375

tensor, 372
tetrad formalism, 7, 161–193
tetrad frame, 9, 155, 161
adapted to a worldline, 390–394
connection in, 169
for continuous medium, 194, 195
for expressing EM fields, 364, 381–398
from SE frame for TUA observer, 382–388
of Maluf and Ulhoa, 392–394, 396
propagation, 166–167
tensor components in, 165
things in themselves, 377
Thomas precession, 26, 94, 344
tidal effects, 371, 373, 375
time dilation, 65, 219, 319
dynamical explanation for, 229, 320–325
in Schwarzschild spacetime, 239
timelike congruence, 161–193
expansion, 171, 189
vorticity, 173, 189
vorticity free, 175–177
torsion, 7, 146, 159, 170, 305–307, 309
in holonomic frame, 160
source for, 160
translational uniform acceleration, 34, 39, 49–51, 75–83, 110, 247, 248
and Lorentz boost KVF, 361
metric for FS coordinate frame, 79
motion of fixed space points, 81
non-covariance of, 39
of charged particle, 353–355
quantum field theory, 399–403
tetrad frame, 382–390

Ulhoa, S.C., 381, 382, 390, 392, 396, 397
uniform acceleration, 4, 9, 29
generalised, *see* generalised uniform acceleration
translational, *see* translational uniform acceleration
uniform gravitational field, 372
non-homogeneity of, 374
universal acceleration effects, 221, 227, 228, 267, 322
Unruh effect, 5, 9, 374, 376, 399–416
Minkowski particles, 399
Rindler particles, 374, 400, 401

Rindler thermal bath, 400–402
Unruh temperature, 400
physical meaning of, 402
Unruh–DeWitt detector, 9, 226, 368, 401–402, 411
with arbitrary motion, 416
with circular motion, 411–415
with TUA motion, 401–403

vacuum, 9, 226, 401
as Rindler thermal bath, 400–402
Minkowski, 400, 401
nature of, 400, 411
Rindler, 400, 401
viewed from rotating frame, 403–415
viewed from TUA frame, 399–403
velocity symmetry, *see* Lorentz symmetry

weak equivalence principle, 2, 6, 141, 162, 230, 256, 298, 357, 369, 374, 378, 397, 415
weight in GR, 299
Weyl postulate, 233
world of happenings, 325