Outlier-insensitive Kalman Smoothing and Marginal Message Passing

Federico Wadehn*, Lukas Bruderer*, Vijay Sahdeva*, Hang Yu†, Justin Dauwels† and Hans-Andrea Loeliger*
*Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland
†School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore

Abstract—We propose a new approach to outlier-insensitive Kalman smoothing based on Normal priors with unknown variance (NUV). In contrast to prior work, the actual computations amount essentially to iterations of a standard Kalman smoother (with few extra computations). Due to the additive noise assumption and the variational representation of the outlier noise distribution, the procedure can easily be extended to nonlinear estimation problems, e.g., by combining the outlier detection with an Extended Kalman Smoother. For the Kalman smoothing, we consider both a Modified Bryson-Frasier smoother and the recently proposed BIFM (Backward Information Filter Forward Marginal) smoother, neither of which requires matrix inversions.

I. INTRODUCTION

The performance of standard Kalman filters and smoothers is severely impaired by outliers in the measurements due to the sensitivity of the least-squares cost function to deviations from nominal noise [1], [2]. Real measurements however, are commonly contaminated with outliers, such as motion artifacts in physiological signals or glint noise in radar signals [1], [3]. Prior work on dealing with robust filtering and smoothing includes [1], [3], where Kalman smoothing is formulated as a linear regression problem and outlier detection is performed via a sparsifying $\ell_1$-penalty. Along the same lines in [4], [5], outliers are assumed to be non-Gaussian, and the mean squared error criterion is replaced by an appropriate score function in a robust M-estimation framework.

In this paper we propose a new approach to outlier-insensitive Kalman smoothing (OIKS): using an idea from sparse Bayesian learning [6], we model outliers as Gaussian impulses with unknown variances and estimate these variances by expectation maximization (EM). In contrast to the mentioned prior works, our approach is parameter-free and amounts essentially to iterated standard Kalman smoothing (with additional simple update steps for the unknown variances), i.e., we effectively stay within the linear Gaussian framework. The extension of the proposed framework to nonlinear estimation problems is straightforward and will be discussed in section IV-F.

For the Kalman smoothing step we consider two matrix inversion-free algorithms: The Modified Bryson-Frasier smoother (MBF) [7], and the recently proposed BIFM (backward information filter, forward marginal); the latter incurs the least computational costs [8]. In contrast to the commonly used Rauch-Tung-Striebel (RTS) smoother [9], neither the MBF nor the BIFM require inverting the state covariance matrix at every time step and therefore both smoothers exhibit better numerical properties, as well as a reduced computational complexity in terms of matrix multiplications.

II. SYSTEM MODEL AND BASIC IDEA

The given discrete-time signal $y = (y_1, \ldots, y_N)$ is modeled as the output of the linear state space model

$$
X_k = AX_{k-1} + BU_{k-1} + W_{k-1}
$$
$$
Y_k = CX_k + Z_k + S_k
$$

(1)

with $A \in \mathbb{R}^{d \times d}$, $X_k$, $W_k$, $B$, $C^T \in \mathbb{R}^d$, and $U_k, Y_k \in \mathbb{R}$.

The input signal $U = (U_1, \ldots, U_N)$ is a sequence of Gaussian random variables (not necessarily i.i.d., but with known means and variances).

We have two sources of observation noise: the normal distributed observation noise $Z = (Z_1, \ldots, Z_N)$ and the outlier-causing impulsive noise $S = (S_1, \ldots, S_N)$. The former is i.i.d. zero-mean Gaussian with constant variance $\sigma_Z^2$. Additionally, we have i.i.d. zero-mean Gaussian process noise $W = (W_1, \ldots, W_N)$ to account for modeling uncertainties.

The impulsive noise terms $(S_1, \ldots, S_N)$ are modeled as independent zero-mean Gaussians, each with its own variance $\gamma_k$, with $k \in \{1, \ldots, N\}$. The point of this model is that both the maximum-likelihood estimate and MAP-estimate (with
of (γ₁, …, γₙ) tend to be sparse [8], i.e. most of the estimated components γₖ will be exactly zero, which implies Sₖ = 0.

The unknown variances (γ₁, …, γₙ) will be estimated by an EM algorithm, as described in Section IV. In each E-step of the EM algorithm, the variances (γ₁, …, γₙ), are (momentarily) fixed; the resulting model is purely linear Gaussian, and each E-step amounts to one round of Kalman smoothing. We first turn to the Kalman smoothing procedure.

### III. Inversion-free Kalman Smoothing

To describe the two Kalman smoother variations, we will use the factor graph block diagram representation [10] of the state space model in (1) shown in Fig. 1. Using this representation, tabulated message passing rules as shown in Table I can be devised, which allow the composition of algorithms ranging from state estimation to output smoothing (interpolation) and input estimation [8].

Let γ = (γ₁, …, γₙ) in Fig. 1 be fixed. Inference in linear state space models, i.e. Kalman filtering and smoothing, boils down to sum-product message passing on the given factor graph [10]. There are two general approaches to compute the smoothing distribution [11]. Firstly, by a two-filter (2-Filt.) smoother [10], based on the observation that:

\[ p(x_k|y_{1:N}) \propto p(x_k|y_{1:k})p(y_{k+1:N}|x_k), \]  \hspace{1cm} (2)

which means that the smoothing distribution can be computed by a forward pass and an independent backward pass (Kalman filter that runs backwards in time). Secondly, via recursive Bayesian estimation as in (3), which requires the computation of the filtering distribution \( p(x_k|y_{1:k}) \) beforehand.

\[ p(x_k|y_{1:k}) = \int p(x_k|x_{k+1})p(x_{k+1}|y_{1:N})dx_{k+1} \]  \hspace{1cm} (3)

The RTS, as well as the MBF and the BIFM smoother, are all instances of this representation of recursive Bayesian estimation.

#### A. Notation

We will express the filtering and smoothing update steps in message passing notation (cf. [10]). The Gaussian (forward) filtering distribution \( p(x_k|y_{1:k}) \), computed with the covariance Kalman filter, is parametrized by the mean \( \tilde{m}_X_k \) and covariance \( \tilde{V}_X_k \). The backward filtering distribution \( p(x_k|y_{k:N}) \) computed with the backward information filter, will be parametrized by the precision matrix \( \tilde{W}_X_k \) (inverse of covariance matrix) and the transformed mean \( \tilde{\xi}_X_k \equiv \tilde{W}_X_k m_X_k \).

The right-pointing arrow is used for quantities in the forward filter that depend only on past observations [10], the left pointing arrow for quantities, that depend only on future and present observations, whereas message quantities without an arrow indicate parameters of the marginal distribution \( p(x_k|y_{1:N}) \).

#### B. MBF and BIFM Smoother

Table I shows MBF smoother and BIFM update rules for each node of the factor graph. Both smoothers consist of three message updates, through the “equality”-, “multiplier”- and “plus”-node. The smoothing distribution computed via the MBF, is parametrized by the dual precision \( \tilde{W}_X \equiv (\tilde{V}_X + \tilde{V}_X)^{-1} \) and the dual mean \( \tilde{\xi}_X \equiv \tilde{W}_X (\tilde{m}_X - \tilde{m}_Y) \) as described in [8]. The (forward) filtering distribution \( p(x_k|y_{k:N}) \), parametrized by \( m_X_k \) and \( V_X_k \) is computed via the standard covariance Kalman filter. Note that the invariance of the message parameters \( W_X \) and \( \xi_X \) at the “plus”-node, renders the MBF an ideal candidate for systems with multiple inputs and state noise or for input estimation as in [12].

For the BIFM in contrast, the smoothing distribution is parametrized by the mean \( m_X \) and the covariance matrix \( V_X \). To determine the filtering distribution \( p(x_k|y_{k:N}) \), parametrized by \( \xi_X_k \) and \( W_X_k \) the BIFM uses the backward information filter [10]. Note that due to the invariance of the marginal mean and the covariance at the “equality”-node, the BIFM is well suited for systems with multiple outputs.

The duality between BIFM and the MBF can be appreciated by observing that the MBF smoother is inversion-free when working with scalar observations, whereas BIFM is inversion-free when working with scalar inputs. However, in our simulations we did not find any significant difference in numerical precision between these two smoothers. Table II shows the computational complexity, the number of matrix inversion and the storage requirements of the RTS, MBF, BIFM and the 2-Filter smoother. The MBF and BIFM compare favourably with other Kalman smoothers and we advocate their usage.

In addition to these two smoothers, using the same parametrization of the Gaussian messages, we can devise
two additional smoothers, namely a forward-pass MBF and a backwards version of the BIFM, denoted Forward Information Filter, Backward Marginal (FIBM). The forward-pass MBF uses the time-reversed covariance Kalman filter and performs the forward steps with the dual mean and dual precision, whereas the FIBM computes the filtering distribution with the forward information filter, and performs the backward pass with the marginal mean and covariance. Note that for these two smoothers the A-matrix needs to be invertible, which renders them impractical for time-varying systems, e.g., like the ones obtained from linearization of the state space model as in section IV-F. Nonetheless, given invertible state transition matrices, one could envision estimation on moving windows, which employ the latter two Kalman smoother versions.

IV. OUTLIER-INSENSITIVE KALMAN SMOOTHER

Our proposed OIKS combines MBF/BIFM smoothing with the time-varying noise variance model for impulsive noise. Given a signal \( y = (y_1, \ldots, y_N) \), the MAP/ML estimates

\[
\hat{\gamma}_k \triangleq \arg\max_{\gamma_k} p(y|\gamma_1, \ldots, \gamma_k, \ldots, \gamma_N) \prod_{\ell=1}^N p(\gamma_\ell)
\]

will result in \( \gamma_k > 0 \) when an outlier is detected at time step \( k \) and 0 otherwise. To compute the ML estimate in a local manner, akin to message passing, we devise EM [13], [14] updates that iteratively compute all \( \hat{\gamma}_k \) in parallel.

A. Expectation Step

By the Markov property of the state space model, the expectation in the \( i \)-th iteration is:

\[
Q(\gamma) = \sum_{\ell=1}^{\gamma_k} Q(\gamma_i | \gamma_{i-1}, y_i, y_{i+1}) = \sum_{\ell=1}^{\gamma_k} E_{X_k | \gamma_{i-1}, y_i} \left[ \log(p(y_i, x_i, \gamma)) \right]
\]

With \( \nu_k \triangleq \sigma^2_Z + \gamma_k \) and the second moment \( \mu^\nu_{X_k} = (V_{X_k} + m_{X_k} m_{X_k}^T) \) we get:

\[
E_{X_k | \gamma_{i-1}, y_i} \left[ \log(p(y_i | x_i, \gamma)) \right] \propto \log(\nu_k) + \left( y_i^2 - 2y_i C_{X_k} m_{X_k} + C_{X_k} m_{X_k}^T \right) / \nu_k
\]

The marginal mean and covariance \( m_{X_k} \) and \( V_{X_k} \) are computed via a forward-backward sweep of the MBF or the BIFM. For the first iteration \( \gamma^0 \) is initialized as the all-zero vector.

B. Maximization Step

The maximization of (17) w.r.t. to \( \gamma_k \) yields:

\[
\frac{1}{\nu_k} - \frac{(y_k^2 - 2y_k C_{X_k} m_{X_k} + C_{X_k} m_{X_k}^T)}{\nu_k^2} = - \frac{p(\gamma_k)}{p(\gamma_k)}
\]

When using an improper prior, i.e. \( p(\gamma_k) \propto 1 \), which imposes an improper outlier distribution \( p(s_k) \propto 1/s_k \) and is heavily sparsifying, we obtain an analytic expression for the updates:

\[
\gamma_k = \max \left( y_k^2 - \sigma^2_Z - 2C_{X_k} m_{X_k} + C_{X_k} m_{X_k}^T, 0 \right)
\]

C. Termination Criterion

We iterate the EM steps until convergence, defined as:

\[
\max_{k \in \{1, \ldots, N\}} |\gamma_k - \gamma_k^{i-1}| \leq 10^{-4} \gamma_k^{i-1},
\]

or up to a maximum number of iterations (10 in our case).

D. Noise Floor Estimation:

To estimate the noise floor \( V_Z = \sigma^2_Z \), we resort again to the EM and alternate between \( \gamma \) and \( V_Z \) estimation. The key step to estimate \( V_Z \) is to only consider observations that were not detected as outliers in the previous iteration. We use an initial noise floor estimate \( \sigma^2_{Z_0} \) (which can for instance be inferred from the output signal’s energy) to mark an observation \( y_k \) as outlier if: \( \gamma_k \geq 10\sigma^2_{Z_0} \). The resulting update equation for \( V_Z \), with \( N_o \) being the number of outliers, is therefore:

\[
V_Z = \frac{1}{N - N_o} \sum_{k: \gamma_k < 10\sigma^2_{Z_0}} y_k^2 - 2y_k C_{X_k} m_{X_k} + C_{X_k} m_{X_k}^T
\]

E. Extension to Multi-Channel Observations

Given a multi-dimensional observation vector \( y_k = (y_k^{(1)}, \ldots, y_k^{(L)}) \in \mathbb{R}^L \), which is affected by outliers (cf. Fig. 3), we can apply the same procedure as in the scalar observation case if we assume that both the vector-valued noise floor \( Z[k] \) and impulsive noise source \( S[k] \) on each channel are

<table>
<thead>
<tr>
<th>matrix mult.</th>
<th>matrix inv.</th>
<th>storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTS</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2-Filter</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>MBF</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>BIFM</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Table II

Matrix multiplications, inversions and storage at every time-step for different Kalman smoothers.
are accounted for by Gaussian i.i.d. input, state and measurement noise, whereas the outliers are modeled with Laplace distributed noise. The output’s SNR when considering only the Gaussian noise floor is 7.4 dB, whereas if also impulsive noise is accounted for (here, 20% Laplace-distributed outliers) the S(I)NR drops to −6.1 dB. The estimation performance of the OIKS is significantly less deteriorated by impulsive noise (cf. Fig. 2) compared to a regular Kalman smoother implemented as MBF (KS) and even compared to an adhoc robustified Kalman smoother we will call MedKS. To remove outliers the MedKS filters the signal $y$ with a 5-th order median filter and then runs a regular Kalman smoother.

To evaluate the performance of our OIKS we ran a regular Kalman smoother, the MedKS and the so called "doubly robust smoother" from [1], (which is based on an L1 penalty on the outliers), on the state space model and noise floor level as described before, but varying the output contamination levels. In contrast to [1], where a parameter $\lambda$, setting the sparsity level, has to be determined via a grid-search, the OIKS is free of tuning parameters. In Fig. 4, we can see that both the robust smoother from [1] and the OIKS clearly outperform both the regular Kalman smoother and the ad hoc robustified version. For low contamination levels the OIKS' performance is comparable to [1], but for high contamination levels the OIKS outperforms the algorithm from [1]. Unlike [1] which estimates the mean of the outlier value (which becomes increasingly hard when contamination levels are very high), the OIKS does not follow a generative approach, but instead assigns to outliers a large variance and thus discards them from the smoothing procedure.

For low contamination levels the performance of the MedKS is comparable to the OIKS, but with an increasing number of outliers (or contiguous noise bursts) the strength of the model-based approach becomes evident. Note that when a signal is (almost) free of outliers, most $\gamma$’s are set to zero and the OIKS reverts to a regular KS.

### B. Nonlinear State Space Model: Van der Pol Oscillator

The Van der Pol oscillator describes a stable oscillation subject to nonlinear damping, which can be used to model a variety of relaxation oscillations such as for instance ECG waveforms [16]. The discrete-time model can be obtained via the Euler method resulting in:

$$
X_1[k+1] = X_1[k] + X_2[k] \Delta + W_1[k]
$$
$$
X_2[k+1] = \mu(1 - X_2^2[k])X_2[k] - X_1[k] \Delta + W_2[k]
$$
$$
Y[k] = CX[k] + Z[k] + S[k],
$$

where $Y[k]$ is the multichannel output signal and $C = [1, 0; 0, 1]$, $\mu$ the damping coefficient and $\Delta$ the step size. Assuming diagonal covariance matrices for both the noise floor $Z[k]$ and impulsive noise source $S[k]$, we can treat each channel separately as described in section IV-E.

Fig. 5 shows the simulated two-channel output of a Van der Pol oscillator with $\mu = 0.5$ and $\Delta = 0.01$ subject to impulsive noise with a contamination level of 10\%.
insensitive extended Kalman Smoother, implemented as an MBF-EKS, assigns large variances to observations affected by outliers, the remaining impulsive noise variances are set to zero, indicating the absence of outliers.

VI. CONCLUSION

We have presented an outlier-insensitive Kalman smoother denoted OIKS which can be used both for linear systems as well as nonlinear systems when combined with an Extended Kalman Smoother. Based on ideas from sparse Bayesian learning, the outliers are modeled as Gaussians with unknown variances, which are estimated by EM. The resulting posterior is an improper heavy-tailed distribution, which results in a sparse outlier detection. The actual computations boil down to iterations of standard Kalman smoothers.

We have considered two Kalman smoothers: the Modified Bryson-Fraser (MBF) and the Backward Information Filter Forward Marginal (BIFM), neither of which requires inversions of the state covariance matrix. In the simulated examples the numerical stability of both approaches is comparable, but BIFM has a lower complexity.

Regarding future work, the combination of the proposed outlier-insensitive Kalman smoother with system identification and input signal estimation as in [12] seems promising for applications such as heart beat detection in ballistocardiogram recordings [12], which are usually heavily affected by motion artifacts. Finally, the current cost function has a per sample penalty on the i.i.d. outliers, which works well for the considered scenarios, but is not tailored to burst noise, which is characterized by contiguous blocks of outliers.

REFERENCES