Outlier-insensitive Kalman Smoothing and Marginal Message Passing

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

***Abstract*—Based on an idea from sparse Bayesian learning, we propose a new approach to outlier-insensitive Kalman smoothing using Normal priors with unknown variance (NUV). In contrast to prior work, the actual computations essentially amount to iterations of a standard Kalman smoother. 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 outlier detection with an Extended Kalman smoother. We consider two matrix-inversion free Kalman smoothers, the Modified Bryson-Frasier and the recently proposed BIFM (Backward Information Filter Forward Marginal) smoother, whose reduced computational burden will be pointed out.**

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 ℓ_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 (NUV-EM OIKS): using an idea from sparse Bayesian learning [6], we model outliers as Normal distributed impulses with unknown variances (NUV) 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 as shown in Table II.

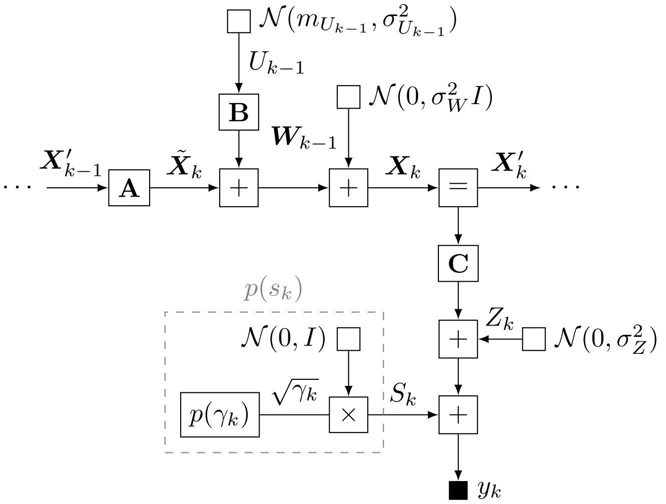


Fig. 1. Factor graph segment of the system model. (The whole factor graph consists of many such sections, one for each time step.)

II. SYSTEM MODEL AND BASIC IDEA

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

$$\begin{aligned} \boldsymbol{X}_k &= \mathbf{A}\boldsymbol{X}_{k-1} + \mathbf{B}U_{k-1} + \boldsymbol{W}_{k-1} \\ Y_k &= \mathbf{C}\boldsymbol{X}_k + Z_k + S_k \end{aligned} \tag{1}$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$, $\boldsymbol{X}_k, \boldsymbol{W}_k, \mathbf{B}, \mathbf{C}^\mathsf{T} \in \mathbb{R}^d$, and $U_k, Y_k \in \mathbb{R}$.

The input signal $\boldsymbol{U} = (U_1, \ldots, U_N)$ 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 $\boldsymbol{Z} = (Z_1, \ldots, Z_N)$ and the outlier-causing impulsive noise $\boldsymbol{S} = (S_1, \ldots, S_N)$. The former is i.i.d. zero-mean Gaussian with constant variance σ_Z^2. Additionally, we have i.i.d. zero-mean Gaussian process noise $\boldsymbol{W} = (W_1, \ldots, W_N)$ 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 γ_k, with $k \in \{1, ..., N\}$. The point of this model is that both the maximum-likelihood estimate and MAP-estimate (with suitable priors) of $(\gamma_1, \ldots, \gamma_N)$ tend to be sparse [8], i.e. most of the estimated components $\hat{\gamma}_k$ will be exactly zero, which implies $S_k = 0$.

The unknown variances $(\gamma_1, \dots, \gamma_N)$ will be estimated by an EM algorithm, as described in Section IV. In each E-step of the EM algorithm, the variances $(\gamma_1, \dots, \gamma_N)$, 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 (cf. [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 $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)$ 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:

$$p(\boldsymbol{x}_k | y_{1:N}) \propto p(\boldsymbol{x}_k | y_{1:k}) p(y_{k+1:N} | \boldsymbol{x}_k), \tag{2}$$

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 $p(\boldsymbol{x}_k | y_{1:k})$ beforehand.

$$p(\boldsymbol{x}_k | y_{1:N}) = p(\boldsymbol{x}_k | y_{1:k}) \int \frac{p(\boldsymbol{x}_{k+1} | \boldsymbol{x}_k) p(\boldsymbol{x}_{k+1} | y_{1:N})}{p(\boldsymbol{x}_{k+1} | y_{1:k})} \mathrm{d}\mathbf{x}_{k+1} \tag{3}$$

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 $p(\boldsymbol{x}_k | \boldsymbol{y}_{1:k})$, computed with the covariance Kalman filter, is parametrized by the mean $\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}_k}$ and covariance $\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$. The backward filtering distribution $p(\boldsymbol{x}_k | \boldsymbol{y}_{k:N})$ computed with the backward information filter, will be parametrized by the precision matrix $\overleftarrow{\mathbf{W}}_{\boldsymbol{X}_k}$ (inverse of covariance matrix) and the transformed mean $\overleftarrow{\boldsymbol{\xi}}_{\boldsymbol{X}_k} \triangleq \overleftarrow{\mathbf{W}}_{\boldsymbol{X}_k} \overleftarrow{\boldsymbol{m}}_{\boldsymbol{X}_k}$.

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 $p(\boldsymbol{x}_k | \boldsymbol{y}_{1:N})$.

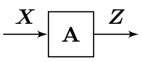
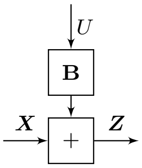
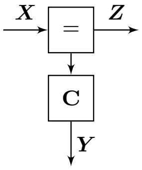
Node	**Update rule**
	$\boldsymbol{m}_{\boldsymbol{Z}} = \mathbf{A}\boldsymbol{m}_{\boldsymbol{X}}, \qquad \mathbf{V}_{\boldsymbol{Z}} = \mathbf{A}\mathbf{V}_{\boldsymbol{X}}\mathbf{A}^{\mathsf{T}} \quad (4)$ - - - - - - - - - - - - - - - - - - $\tilde{\boldsymbol{\xi}}_{\boldsymbol{X}} = \mathbf{A}^{\mathsf{T}}\tilde{\boldsymbol{\xi}}_{\boldsymbol{Z}}, \qquad \tilde{\mathbf{W}}_{\boldsymbol{X}} = \mathbf{A}^{\mathsf{T}}\tilde{\mathbf{W}}_{\boldsymbol{Z}}\mathbf{A} \quad (5)$
	$\boldsymbol{m}_{\boldsymbol{Z}} = \tilde{\mathbf{F}}^{\mathsf{T}}\boldsymbol{m}_{\boldsymbol{X}} + \mathbf{B}H\left(\mathbf{B}^{\mathsf{T}}\overleftarrow{\boldsymbol{\xi}}_Z + \overrightarrow{\boldsymbol{\xi}}_U\right) \quad (6)$ $\mathbf{V}_{\boldsymbol{Z}} = \tilde{\mathbf{F}}^{\mathsf{T}}\mathbf{V}_X\tilde{\mathbf{F}} + \mathbf{B}H\mathbf{B}^{\mathsf{T}} \quad (7)$ $\tilde{\mathbf{F}} = \mathbf{I} - \overleftarrow{\mathbf{W}}_Z\mathbf{B}H\mathbf{B}^{\mathsf{T}} \quad (8)$ $H = (\overrightarrow{W}_U + \mathbf{B}^{\mathsf{T}}\overleftarrow{\mathbf{W}}_Z\mathbf{B})^{-1} \quad (9)$ - - - - - - - - - - - - - - - - - - $\tilde{\boldsymbol{\xi}}_{\boldsymbol{X}} = \tilde{\boldsymbol{\xi}}_{\boldsymbol{Z}}, \qquad \tilde{\mathbf{W}}_{\boldsymbol{X}} = \tilde{\mathbf{W}}_{\boldsymbol{Z}} \quad (10)$
	$\boldsymbol{m}_{\boldsymbol{Z}} = \boldsymbol{m}_{\boldsymbol{X}}, \qquad \mathbf{V}_{\boldsymbol{Z}} = \mathbf{V}_{\boldsymbol{X}} \quad (11)$ - - - - - - - - - - - - - - - - - - $\tilde{\boldsymbol{\xi}}_{\boldsymbol{X}} = \mathbf{F}^{\mathsf{T}}\tilde{\boldsymbol{\xi}}_{\boldsymbol{Z}} + \mathbf{C}^{\mathsf{T}}G(\mathbf{C}\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}} - \overleftarrow{m}_Y) \quad (12)$ $\tilde{\mathbf{W}}_{\boldsymbol{X}} = \mathbf{F}^{\mathsf{T}}\tilde{\mathbf{W}}_Z\mathbf{F} + \mathbf{C}^{\mathsf{T}}G\mathbf{C} \quad (13)$ $\mathbf{F} = \mathbf{I} - \overrightarrow{\mathbf{V}}_X\mathbf{C}^{\mathsf{T}}G\mathbf{C} \quad (14)$ $G = (\overleftarrow{V}_Y + \mathbf{C}\overrightarrow{\mathbf{V}}_X\mathbf{C}^{\mathsf{T}})^{-1} \quad (15)$

TABLE I
UPDATE RULES FOR BIFM (ABOVE DASHED LINE) AND MBF (BELOW DASHED LINE).

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 $\tilde{\mathbf{W}}_{\boldsymbol{X}} \triangleq (\overrightarrow{\mathbf{V}}_{\boldsymbol{X}} + \overleftarrow{\mathbf{V}}_{\boldsymbol{X}})^{-1}$ and the dual mean $\tilde{\boldsymbol{\xi}}_{\boldsymbol{X}} \triangleq \tilde{\mathbf{W}}_{\boldsymbol{X}}(\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}} - \overleftarrow{\boldsymbol{m}}_{\boldsymbol{X}})$ as described in [8]. The (forward) filtering distribution $p(\boldsymbol{x}_k | y_{k:N})$, parametrized by $\overrightarrow{\boldsymbol{m}}_{X_k}$ and $\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$ is computed via the standard covariance Kalman filter. Note that the invariance of the message parameters $\tilde{\mathbf{W}}_{\boldsymbol{X}}$ and $\tilde{\boldsymbol{\xi}}_{\boldsymbol{X}}$ 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 $\boldsymbol{m}_{\boldsymbol{X}}$ and the covariance matrix $\mathbf{V}_{\boldsymbol{X}}$. To determine the filtering distribution $p(\boldsymbol{x}_k | y_{k:N})$, parametrized by $\overleftarrow{\boldsymbol{\xi}}_{\boldsymbol{X}_k}$ and $\overleftarrow{\mathbf{W}}_{\boldsymbol{X}_k}$ 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

	matrix mult.	matrix inv.	storage
RTS	4 (+ 0)	1	$\overrightarrow{\mathbf{V}}_{\boldsymbol{X}'_k}, \overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$
2-Filter	10 (+ 0)	0	$\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$
MBF	6 (+ 0)	0	$\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$
BIFM	4 (+ 2)	0	$\overleftarrow{\mathbf{W}}_{\boldsymbol{X}_k}$

TABLE II
MATRIX MULTIPLICATIONS (WITH ADDITIONAL COMPUTATIONS DUE TO STATE NOISE IN PARENTHESIS), INVERSIONS AND STORAGE AT EVERY TIME-STEP FOR DIFFERENT KALMAN SMOOTHERS.

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 $\mathbf{A}$-matrix needs to be invertible, which renders them unpractical 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 NUV-EM OIKS combines MBF/BIFM smoothing with the time-varying noise variance model for impulsive noise. Given a signal $\boldsymbol{y} = (y_1, \ldots, y_N)$, the MAP/ML estimates

$$\hat{\gamma}_k \triangleq \underset{\gamma_k}{\operatorname{argmax}}\, p(\boldsymbol{y}|\gamma_1, \ldots, \gamma_k, \ldots, \gamma_N) \prod_{\ell=1}^{N} p(\gamma_\ell) \qquad (16)$$

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:

$$\begin{aligned} Q(\boldsymbol{\gamma}) &= \textstyle\sum_{\ell=1}^{N} Q(\gamma_\ell) = E_{\boldsymbol{X}|\boldsymbol{\gamma}^{i-1},\boldsymbol{y}}[\log(p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\gamma}))] \\ &= \textstyle\sum_{\ell=1}^{N} E_{\boldsymbol{X}_\ell|\boldsymbol{\gamma}^{i-1},\boldsymbol{y}}[\log(p(y_\ell|\boldsymbol{x}_\ell,\gamma_\ell))] + \log(p(\gamma_\ell)) + \text{const} \end{aligned} \qquad (17)$$

With $\nu_k \triangleq \sigma_Z^2 + \gamma_k$ and the second moment $\mu^{\mathrm{II}}_{X_k} = (\mathbf{V}_{\boldsymbol{X}_k} + \boldsymbol{m}_{\boldsymbol{X}_k}\boldsymbol{m}_{\boldsymbol{X}_k}^{\mathsf{T}})$ we get:

$$\begin{aligned} &E_{\boldsymbol{X}_k|\boldsymbol{\gamma}^{i-1},\boldsymbol{y}}\left[\log(p(y_k|\boldsymbol{x}_k,\gamma_k))\right] \propto \\ &\log(\nu_k) + \left(y_k^2 - 2y_k\mathbf{C}\boldsymbol{m}_{\boldsymbol{X}_k} + \mathbf{C}\mu^{\mathrm{II}}_{X_k}\mathbf{C}^{\mathsf{T}}\right)/\nu_k \end{aligned} \qquad (18)$$

The marginal mean and covariance $\boldsymbol{m}_{\boldsymbol{X}_k}$ and $\mathbf{V}_{\boldsymbol{X}_k}$ are computed via a forward-backward sweep of the MBF or the BIFM. For the first iteration $\boldsymbol{\gamma}^0$ is initialized as the all-zero vector.

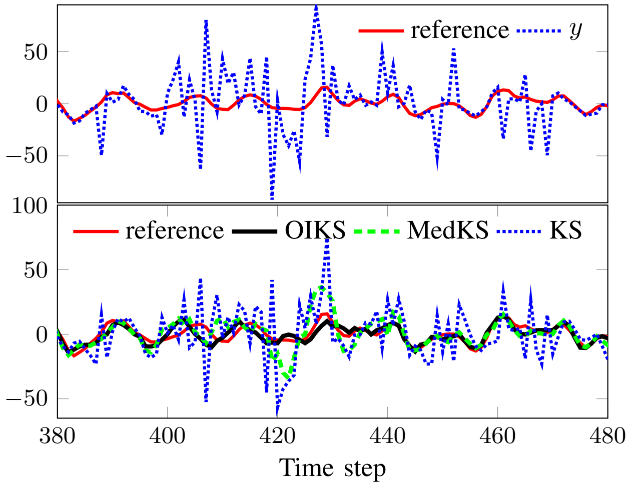


Fig. 2. Top: Ground-truth (reference) and observation $\boldsymbol{y}$ with outliers. Bottom: Kalman smoother (KS), Kalman smoother with median filter (MedKS) and the proposed outlier-insensitive Kalman smoother (OIKS).

B. Maximization Step

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

$$\frac{1}{\nu_k} - \frac{(y_k^2 - 2y_k\mathbf{C}\boldsymbol{m}_{\boldsymbol{X}_k} + \mathbf{C}\mu^{\mathrm{II}}_{X_k}\mathbf{C}^{\mathsf{T}})}{\nu_k^2} = -\frac{p'(\gamma_k)}{p(\gamma_k)} \qquad (19)$$

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:

$$\gamma_k^i = \max\left(y_k^2 - \sigma_Z^2 - 2\mathbf{C}y_k\boldsymbol{m}_{\boldsymbol{X}_k} + \mathbf{C}\mu^{\mathrm{II}}_{X_k}\mathbf{C}^{\mathsf{T}}, 0\right) \qquad (20)$$

C. Termination Criterion

We iterate the EM steps until convergence, defined as:

$$\max_{k\in\{1,\ldots,N\}} |\gamma_k^i - \gamma_k^{i-1}| \leq 10^{-4}\ \gamma_k^{i-1}, \qquad (21)$$

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

D. Noise Floor Estimation:

To estimate the noise floor $V_Z = \sigma_Z^2$, we resort again to the EM and alternate between $\boldsymbol{\gamma}$ 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 $\sigma_{Z_0}^2$ (which can for instance be inferred from the output signal's energy) to mark an observation y_k as outlier if: $\gamma_k \geq 10\sigma_{Z_0}^2$. The resulting update equation for V_Z, with N_o being the number of outliers, is therefore:

$$V_Z^i = \frac{1}{N - N_o} \sum_{k:\gamma_k^i < 10\sigma_{Z_0}^2} y_k^2 - 2y_k\mathbf{C}\boldsymbol{m}_{\boldsymbol{X}_k} + \mathbf{C}\mu^{\mathrm{II}}_{X_k}\mathbf{C}^{\mathsf{T}} \qquad (22)$$

E. Extension to Multi-Channel Observations

Given a multi-dimensional observation vector $\boldsymbol{y}_k = (y_k^{(1)}, \ldots, y_k^{(L)}) \in \mathbb{R}^L$, 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 $\boldsymbol{Z}[\boldsymbol{k}]$ and impulsive noise source $\boldsymbol{S}[\boldsymbol{k}]$ on each channel are independent. This allows us to split the $\mathbf{C}$-matrix into row

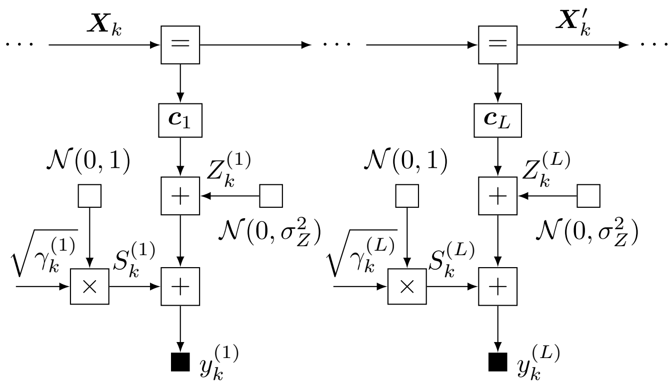


Fig. 3. The outlier insensitive smoother can be extended to multi-channel outputs by splitting the C matrix and assuming i.i.d. outliers.

vectors (cf. Fig. 3) and use the same computation rules described for the scalar output case.

F. Extension to Nonlinear State Space Models

Assuming additive input, state and measurement noise, the outlier insensitive state estimation can be extended to nonlinear dynamical systems. Given the discrete-time nonlinear dynamics, together with an additive impulsive noise source, we use an Extended Kalman smoother combined with the sparsifying outlier noise variance estimation. Here we will limit ourselves to autonomous systems with nonlinear state dynamics $\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1}$, where $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear function and $\boldsymbol{w}$ is the process noise. The measurement process is assumed to be linear, as considered before. The main step in the linearization of the nonlinear state dynamics is:

$$\boldsymbol{f}(\boldsymbol{X}_k) \approx \boldsymbol{f}(\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}'_{k-1}}) + \mathbf{A}_{k-1}(\boldsymbol{X}_k - \overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}'_{k-1}}), \tag{23}$$

where $\mathbf{A}_{k-1}$ is the Jacobi matrix of $\boldsymbol{f}(\cdot)$ evaluated at $\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}'_{k-1}}$.

The Extended Kalman smoother (EKS) can be implemented either via an extended version of the MBF or BIFM. In the MBF-EKS version, the filtering problem is solved using the standard Extended Kalman Filter [11], and the backward step is performed on the linearized system (23), using the MBF smoother. Conversely, the BIFM-EKS uses the backward extended information filter [15] and the BIFM update rules on the linearized system as described in Table I.

V. NUMERICAL SIMULATIONS

A. Linear Time-Invariant System

To evaluate the NUV-EM OIKS we will first use a linear single-input single-output system with system matrices: $\mathbf{A} = [1.12, -0.49, 0.11, -0.35; 1, 0, 0, 0; 0, 1, 0, 0; 0, 0, 1, 0]$, $\mathbf{B} = [-0.38, 0.59, 0.51, 0.3]$ and $\mathbf{C} = [1, 0, 0, 0]$. Note that this system is both stable and observable.

Fig. 2 shows the noisy output $\boldsymbol{y}$ and the corresponding noise-free output (reference) of this model subject to sinusoidal inputs. Uncertainty in the dynamics and measurements 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 $7.4\,\text{dB}$, whereas if also impulsive noise is accounted for (here, 20% Laplace-distributed outliers) the S(I)NR drops to $-6.1\,\text{dB}$. The estimation performance of the NUV-EM 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 $\boldsymbol{y}$ with a 5-th order median filter and then runs a regular Kalman smoother.

To evaluate the performance of our NUV-EM 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 λ, setting the sparsity level, has to be determined via a grid-search, the NUV-EM OIKS is free of tuning parameters.

In Fig. 4, we can see that both the robust smoother from [1] and the NUV-EM OIKS clearly outperform both the regular Kalman smoother and the ad hoc robustified version. For low contamination levels the NUV-EM OIKS' performance is comparable to [1], but for high contamination levels the NUV-EM 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 NUV-EM 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 NUV-EM 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 γ's are set to zero and the NUV-EM 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:

$$\begin{aligned}
X_1[k+1] &= X_1[k] + X_2[k]\Delta + W_1[k] \\
X_2[k+1] &= X_2[k] + [\mu(1 - X_1^2[k])X_2[k] - X_1[k]]\Delta + W_2[k] \\
\boldsymbol{Y}[k] &= \mathbf{C}\boldsymbol{X}[k] + \boldsymbol{Z}[\boldsymbol{k}] + \boldsymbol{S}[\boldsymbol{k}],
\end{aligned} \tag{24}$$

where $\boldsymbol{Y}[k]$ is the multichannel output signal and $\mathbf{C} = [1, 0; 0, 1]$, μ the damping coefficient and Δ the step size. Assuming diagonal covariance matrices for both the noise floor $\boldsymbol{Z}[\boldsymbol{k}]$ and impulsive noise source $\boldsymbol{S}[\boldsymbol{k}]$, 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%. The outlier insensitive extended Kalman Smoother, implemented as an

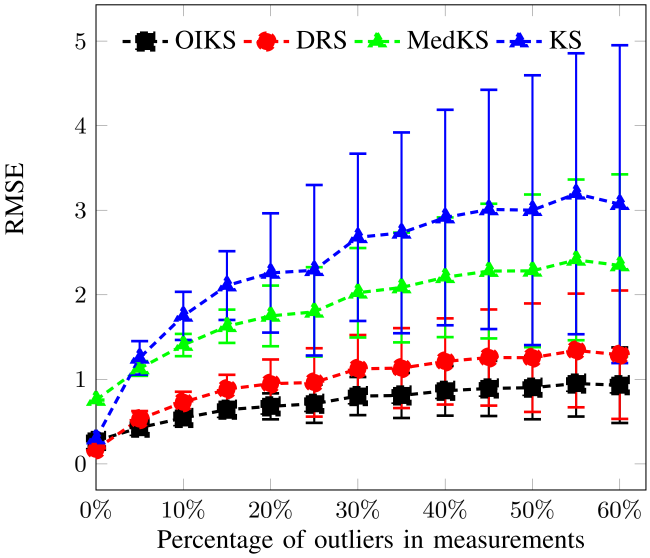


Fig. 4. Comparison of RMSE of state estimates of a regular Kalman smoother (KS), the MedKS, the DRS from [1] and the proposed OIKS against percentage of outliers in the measurements.

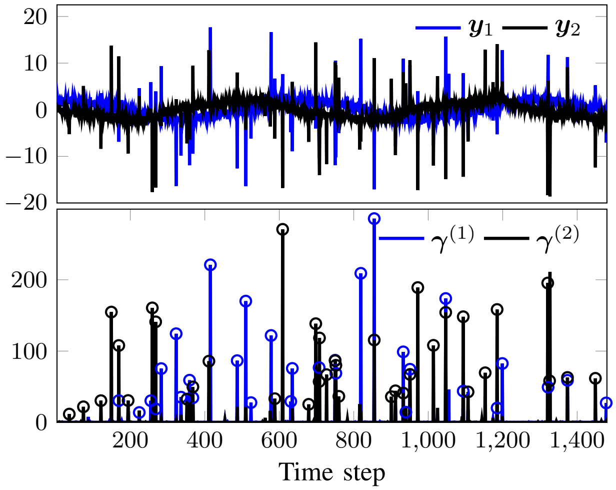


Fig. 5. The NUV-EM OIEKS detects most of the outliers on both channels and assigns them a large variance γ. Correctly detected outliers are marked with circles.

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 NUV-EM OIKS, which can be used both for linear as well as nonlinear systems when combined with an Extended Kalman smoother. Based on ideas from sparse Bayesian learning, outliers are modeled as Gaussians with unknown variances, which are estimated by the EM algorithm, resulting in a sparse outlier detection. The actual computations essentially boil down to iterations of standard Kalman smoothers.

We have considered two Kalman smoothers: the *Modified Bryson-Frasier* (MBF) and the *Backward Information Filter Forward Marginal* (BIFM) from [8], neither of which requires inversions of the state covariance matrix. We have also pointed out the advantages of the new BIFM Kalman smoother which has a lower complexity in terms of matrix multiplications and and a comparable numerical stability, which was corroborated in the simulated examples.

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

[1] S. Farahmand, G.B. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4529–4543, Oct. 2011.

[2] J. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1514–1519.

[3] A.Y. Aravkin, B.M. Bell, J.V. Burke, and G. Pillonetto, "An $\ell 1$ Laplace robust Kalman smoother," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2898–2911, Dec. 2011.

[4] Z.M. Durovic and B.D. Kovacevic, "Robust estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 44, no. 6, pp. 1292–1296, June 1999.

[5] S.C. Chan, Z.G. Zhang, and K.W. Tse, "A new robust Kalman filter algorithm under outliers and system uncertainties," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2005, pp. 4317–4320.

[6] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6236–6255, 2011.

[7] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, vol. 1, Prentice Hall Upper Saddle River, NJ, 2000.

[8] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, "On sparsity by NUV-EM, gaussian message passing, and Kalman smoothing," in *Information Theory and Applications Workshop (ITA), San Diego, CA, USA*, Feb. 2016.

[9] H. E Rauch, C.T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[10] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[11] S. Särkkä, *Bayesian Filtering and Smoothing*, Number 3. Cambridge University Press, 2013.

[12] L. Bruderer, H. Malmberg, and H.-A. Loeliger, "Deconvolution of weakly-sparse signals and dynamical-system identification by Gaussian message passing," in *IEEE International Symposium on Information Theory (ISIT)*, July 2015.

[13] P. Stoica and Y. Selén, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 112–114, 2004.

[14] S. Roweis and Z. Ghahramani, "A unifying review of linear Gaussian models," *Neural Computation*, vol. 11, no. 2, pp. 305–345, 1999.

[15] S. Thrun, Y. Liu, D. Koller, A. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *The International Journal of Robotics Research*, vol. 23, no. 7-8, pp. 693–716, 2004.

[16] BZ Kaplan, I Gabay, G Sarafian, and D Sarafian, "Biological applications of the filtered Van der Pol oscillator," *Journal of the Franklin Institute*, vol. 345, no. 3, pp. 226–232, 2008.