State Space Models with Dynamical and Sparse Variances, and Inference by EM Message Passing

Federico Wadehn*, Thilo Weber*, Hans-Andrea Loeliger
Signal and Information Processing Laboratory
ETH Zurich, Switzerland

***Abstract*—Sparse Bayesian learning (SBL) is a probabilistic approach to estimation problems based on representing sparsity-promoting priors by Normals with Unknown Variances. This representation blends well with linear Gaussian state space models (SSMs). However, in classical SBL the unknown variances are a priori independent, which is not suited for modeling group sparse signals, or signals whose variances have structure. To model signals with, e.g., exponentially decaying or piecewise-constant (in particular block-sparse) variances, we propose SSMs with dynamical and sparse variances (SSM-DSV). These are two-layer SSMs, where the bottom layer models physical signals, and the top layer models dynamical variances that are subject to abrupt changes. Inference and learning in these hierarchical models is performed with a message passing version of the expectation maximization (EM) algorithm, which is a special instance of the more general class of variational message passing algorithms. We validated the proposed model and estimation algorithm with two applications, using both simulated and real data. First, we implemented a block-outlier insensitive Kalman smoother by modeling the disturbance process with a SSM-DSV. Second, we used SSM-DSV to model the oculomotor system and employed EM-message passing for estimating neural controller signals from eye position data.**

***Index Terms*—Expectation maximization, factor graphs, hierarchical state space models, sparse Bayesian learning.**

I. Introduction

Linear Gaussian state space models (SSMs), in combination with efficient inference algorithms such as Kalman filters and smoothers [1], have long been a popular model class in signal processing, econometrics and control. The expressiveness of linear Gaussian SSMs by themselves is quite limited though. Therefore, various extensions such as switching SSMs [2] and many more [3] have been proposed. Another useful extension is obtained by combining linear SSMs with sparsity [4], [5], in particular with techniques from sparse Bayesian learning (SBL) [6]. The key idea of SBL (also known as automatic relevance determination [7]) is to represent a sparsity-promoting prior

$$p(x) \mathrel{\hat{=}} \prod_{k=1}^{N} p(x_k), \tag{1}$$

via the variational representation [8], [9]

$$p(x_k) = \sup_{s_k \geq 0} \mathcal{N}(x_k; 0, s_k^2)\rho(s_k), \tag{2}$$

with $k = 1, \ldots, N$, and with a suitably chosen non-negative function ρ. Such representations have been called Normals with Unknown Variances (NUV) [4], [9] and can be used to represent a large class of sparsity-promoting priors, such as the Laplace distribution, Student's t-distribution, etc. [8].

NUV priors are a good match for otherwise linear Gaussian models. Conditioned on the NUV parameters $s = (s_1, \ldots, s_N)$ in (2), the remaining model is Gaussian and thus amenable to inference by Gaussian message passing [4], [10], which possesses closed-form updates and is computationally very attractive. To estimate the unknown NUV parameters one can for example use the expectation maximization (EM) algorithm [4], [8]. For linear Gaussian SSMs, the E-step can be performed by Kalman smoothing and the M-step has a closed-form solution [4], [11]. To account for structure in the unknown variances, e.g., to model block sparse signals, we extend the NUV-EM framework to dynamical variances. For this, unlike in the classical SBL approach [6], the NUV parameters s are not a priori independent, but instead linked by SSMs that are driven by sparse inputs. This results in a hierarchical SSM that can conveniently be represented by a factor graph [10], [12]. In this factor graph representation, the estimation can be expressed by the message passing version of the EM algorithm [13], [14], which is a special case of variational message passing [15]. To remain in the Gaussian message passing setting, we approximate the resulting non-Gaussian EM messages via the Laplace method [16].

Modeling variances of stochastic processes is not a new idea and a large variety of models have been proposed, such as GARCH models [17] used in time series analysis, and hierarchical Gaussian filters [18]. The combination of volatility models with sparsity, however, appears to be new. Regarding inference in hierarchical SSMs, most approaches resort to some form of variational inference [19], which result in variants of the variational Kalman filter [20].

II. The System Model

In the following we present SSMs with dynamical and sparse variances (SSM-DSV) shown in Fig. 1. These are two-layer SSMs. The bottom layer (SSM$^{(0)}$) models physical signals $y^{(0)} = (y_1^{(0)}, \ldots, y_N^{(0)})$, and the top layer (SSM$^{(1)}$) models the dynamical NUV parameters $S^{(1)} = (S_1^{(1)}, \ldots, S_N^{(1)})$ of the signals $U^{(0)}$ or $Z^{(0)}$ feeding into the bottom layer. To allow for abrupt changes in the dynamical variances, the top layer is driven by independent sparse inputs $U^{(1)} = (U_1^{(1)}, \ldots, U_N^{(1)})$,

*These authors contributed equally.

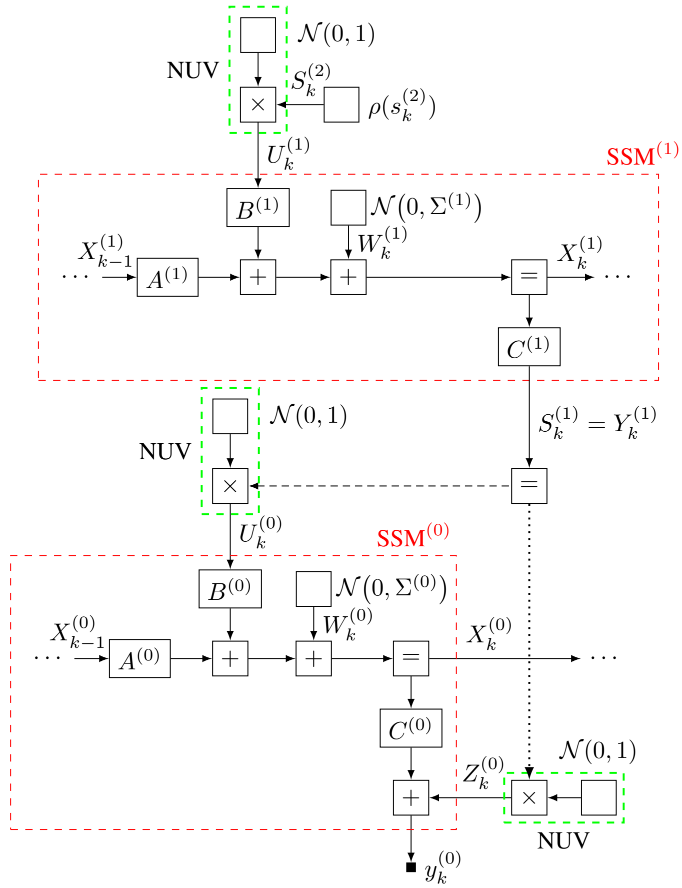

Fig. 1: Two-layer non-linear SSM with dynamical standard deviations $S^{(1)}$ of, the input process $U^{(0)}$ (dashed arrow) or output process $Z^{(0)}$ (dotted arrow), feeding into the bottom layer. The top layer is driven by sparse inputs $U^{(1)}$, modeled by NUVs.

which are modeled by NUVs. The top and bottom layer SSMs follow

$$X_k^{(\ell)} = A^{(\ell)} X_{k-1}^{(\ell)} + B^{(\ell)} U_k^{(\ell)} + W_k^{(\ell)} \tag{3}$$

$$Y_k^{(\ell)} = C^{(\ell)} X_k^{(\ell)} + Z_k^{(\ell)}, \tag{4}$$

with $k = 1, \dots, N$ and $\ell \in \{0, 1\}$. The SSM variables are $X_k^{(\ell)}, W_k^{(\ell)} \in \mathbb{R}^{d_x^{(\ell)}}$, $U_k^{(\ell)} \in \mathbb{R}^{d_u^{(\ell)}}$, and $Y_k^{(\ell)} \in \mathbb{R}^{d_y^{(\ell)}}$. The output of the top layer $Y^{(1)} = S^{(1)}$ is used as the standard deviation of the input $U^{(0)}$, or alternatively, of the output process $Z^{(0)}$

$$U_k^{(0)} \sim \mathcal{N}\left(0, (s_k^{(1)})^2\right), \text{ or } Z_k^{(0)} \sim \mathcal{N}\left(0, (s_k^{(1)})^2\right). \tag{5}$$

This interconnection results in a *non-linear* SSM. In our examples in Section IV, the top layer SSMs are first-order dynamical systems (random walks with sparse increments). Note that the expressiveness of SSM-DSV is quite broad though; by choosing appropriate system matrices in the top layer, it is possible to model signals with sparse, piecewise-constant (e.g., block-sparse), exponentially decaying, periodic variances, and more.

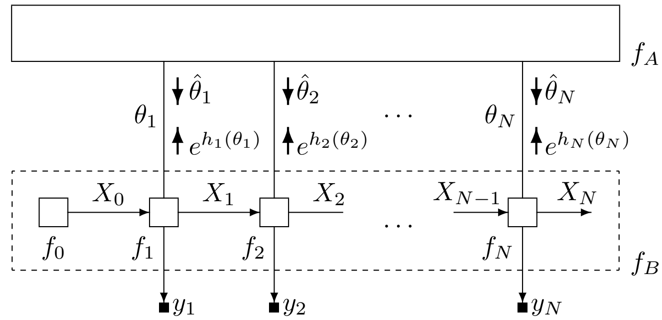

Fig. 2: Factor graph of two-layer SSM and EM messages.

III. Inference and Learning

Here, we review EM message passing and describe how it is used for learning the unknown variances in SSM-DSV.

A. Expectation Maximization as Message Passing

Graphical models, such as for example the factor graph in Fig. 2 have cycles, since the unknown parameters $\theta_1, \dots, \theta_N$ are not a priori independent. To learn parameters in such "loopy" graphs, the EM algorithm can be used. If the joint density represented by the graphical model has a "nice" factorization (e.g., due to the Markov property), the EM algorithm is conveniently expressed as a message passing algorithm [13], which can be seen as a special case of variational message passing [21].

Consider the two-layer model in Fig. 2, which represents

$$f(y, x, \theta) = f_A(\theta) f_B(y, x|\theta), \tag{6}$$

where

$$f_B(y, x|\theta) = f_0(x_0) \prod_{k=1}^{N} f_k(y_k, x_k, x_{k-1}|\theta_k). \tag{7}$$

Let $X = (X_0, \dots, X_N)$ be hidden variables, $y = (y_1, \dots, y_N)$ fixed observations, and $\theta = (\theta_1, \dots, \theta_N)$ unknown parameters (e.g., NUV terms) that we want to estimate via the EM. At each iteration we update the parameters according to

$$\hat{\theta}^{\text{new}} = \text{argmax}_\theta \, \mathbb{E}_{X|y,\hat{\theta}^{\text{old}}} \left[\log\left(f_B(y, X|\theta)\right)\right] + \log(f_A(\theta)). \tag{8}$$

Due to the factorization (7), the first term in (8) splits

$$\sum_{k=1}^{N} \mathbb{E}_{p_B} \left[\log\left(f_k(y_k, X_k, X_{k-1}|\theta_k)\right)\right] + \text{const} \triangleq h(\theta), \tag{9}$$

with p_B given in (12). By defining

$$h_k(\theta_k) \mathrel{\hat{=}} \mathbb{E}_{p_B} \left[\log\left(f_k(y_k, X_k, X_{k-1}|\theta_k)\right)\right], \tag{10}$$

$h_k(\cdot)$ can be interpreted as a log-domain *EM message*. To be consistent with the probabilistic view, the EM message in the factor graph (Fig. 2) is

$$\overleftarrow{\mu}_{\text{EM}}(\theta_k) \mathrel{\hat{=}} e^{h_k(\theta_k)}. \tag{11}$$

The expectations are computed w.r.t. the densities

$$p_B(x_k, x_{k-1}|y, \hat{\theta}_k) \propto \overrightarrow{\mu}_{X_{k-1}}(x_{k-1}) f_k(y_k, x_k, x_{k-1}|\hat{\theta}_k) \overleftarrow{\mu}_{X_k}(x_k), \tag{12}$$

where $\overrightarrow{\mu}_{X_{k-1}}(x_{k-1})$ and $\overleftarrow{\mu}_{X_k}(x_k)$ are sum-product messages [10], [12] containing summary information of the left and right part of the graph of f_B (Fig. 2). The EM message passing algorithm then alternates between computing the EM messages (11) and performing the maximization

$$(\hat{\theta}_1, \ldots, \hat{\theta}_N) = \underset{\theta_1,\ldots,\theta_N}{\operatorname{argmax}} \left(f_A(\theta) \prod_{k=1}^{N} e^{h_k(\theta_k)} \right), \tag{13}$$

which can efficiently be computed via max-product message passing [10], in particular if $f_A(\theta)$ has a nice factorization. If the objective function in (13) is Gaussian, then the optimization (solved by max-product message passing) can equivalently be performed via sum-product message passing (i.e., by marginalization) [10].

B. Learning the Unknown Variances by EM

We now particularize this approach to the model shown in Fig. 1. We restrict ourselves to structured input priors, the derivation of the EM updates for output priors is analogous and differs only in the choice of hidden variables.

1) Inference in the Bottom Layer SSM: **E-step.** To estimate the unknown NUV parameters $S^{(1)}$, we choose $U^{(0)}$ as hidden variables. For fixed observations $y^{(0)}$ and fixed $S^{(1)} = \hat{s}^{(1)}$, the bottom layer SSM is linear Gaussian. The posterior densities

$$p_B(u_k^{(0)}|y^{(0)}, \hat{s}^{(1)}) = \mathcal{N}(u_k^{(0)}; m_{U_k^{(0)}}, V_{U_k^{(0)}}), \tag{14}$$

required in the EM update of $S^{(1)}$ are computed by Gaussian message passing (see (12)). To obtain the upwards EM message (11), we compute

$$\mathbb{E}_{X,U|y,\hat{s}^{(1)}} \left[\log\Big(p(y_k^{(0)}|X_k^{(0)}) p(X_k^{(0)}|X_{k-1}^{(0)}, U_k^{(0)}) p(U_k^{(0)}|s_k^{(1)}) \Big) \right] \tag{15}$$

$$= \mathbb{E}_{p_B} \left[\log\Big(p(U_k^{(0)}|s_k^{(1)}) \Big) \right] + \text{const} \mathrel{\hat{=}} h_k(s_k^{(1)}) \tag{16}$$

from which we get the EM message:

$$e^{h_k(s_k^{(1)})} \propto \frac{1}{\sqrt{2\pi} s_k^{(1)}} \exp\left(-\frac{\mathbb{E}_{p_B}[(U_k^{(0)})^2]}{2(s_k^{(1)})^2} \right). \tag{17}$$

M-step. The NUV parameters $S^{(1)}$ are updated (see (13)) according to

$$\hat{s}_\ell^{(1)} = \underset{s_\ell^{(1)}}{\operatorname{argmax}} \max_{s_{k\neq\ell}^{(1)}} p(s^{(1)}) \prod_{k=1}^{N} e^{h_k(s_k^{(1)})}. \tag{18}$$

Note that neither the EM messages (17) nor $p(s^{(1)})$ are Gaussian, and therefore (18) cannot directly be solved by Gaussian message passing. Using the NUV representation (2), however, the prior $p(s^{(1)})$ can be represented via

$$p(s^{(1)}) = \max_{s_1^{(2)},\ldots,s_N^{(2)} \geq 0} p(s^{(1)}|s^{(2)}) \prod_{k=1}^{N} \rho(s_k^{(2)}), \tag{19}$$

where $p(s^{(1)}|s^{(2)})$ is a multivariate Gaussian. Next, we describe how the non-Gaussian EM messages are handled.

2) Inference in the Top Layer SSM: We approximate the EM messages via Laplace's method (see [16], Chapter 4.4) which corresponds to a second-order Taylor approximation in the log-domain around the mode of

$$e^{M \cdot h_k(s_k^{(1)})} \approx \mathcal{N}(s_k^{(1)}; \overleftarrow{m}_{S_k^{(1)}}, \overleftarrow{V}_{S_k^{(1)}}). \tag{20}$$

The scalar $M > 0$ is a free parameter used to influence the peakiness of the Gaussian approximation and

$$\overleftarrow{m}_{S_k^{(1)}} = \sqrt{E_{p_B}[(U_k^{(0)})^2]} = \sqrt{m^2_{U_k^{(0)}} + V_{U_k^{(0)}}} \tag{21}$$

$$\overleftarrow{V}_{S_k^{(1)}} = \frac{1}{2M} \left(m^2_{U_k^{(0)}} + V_{U_k^{(0)}} \right). \tag{22}$$

In the limit $M \to \infty$, the EM messages become point densities. The approximated EM messages can now be interpreted as Gaussian observations of the NUV parameters $S^{(1)}$ with mean (21) and variance (22). Optimizing w.r.t. the NUV parameters $S^{(2)}$ in the top layer SSM, is then the classical NUV-EM scenario [4], where the M-step is

$$\hat{s}_k^{(2)} = \sqrt{(m_{U_k^{(1)}})^2 + V_{U_k^{(1)}}}, \tag{23}$$

with $m_{U^{(1)}}$ and $V_{U^{(1)}}$ obtained by Gaussian message passing, from which we also get the posterior distribution

$$p(s_k^{(1)}|\hat{s}^{(2)}) = \mathcal{N}(s_k^{(1)}; m_{S_k^{(1)}}, V_{S_k^{(1)}}). \tag{24}$$

Finally, we set $S^{(1)}$ to the posterior mean

$$\hat{s}_k^{(1)} = m_{S_k^{(1)}} = C^{(1)} m_{X_k^{(1)}}. \tag{25}$$

C. Resulting Algorithm

The resulting algorithm iterates between Gaussian message passing in the top and bottom layer SSMs and scalar updates of the pertaining NUV parameters $S^{(1)}$ and $S^{(2)}$.

EM Message Passing in SSM-DSV

1: Initialize $S^{(1)}$ and $S^{(2)}$ to some constant values.
2: Compute posterior densities (14) by Gaussian message passing (Kalman smoothing) in SSM$^{(0)}$.
3: Compute EM messages (20) using (21) and (22).
4: Run Gaussian message passing in SSM$^{(1)}$ using (20) as observations and iteratively update $S^{(2)}$ using (23).
5: Update $S^{(1)}$ using (25).
6: Until convergence not reached, proceed with step 2.

The described algorithm runs the EM till convergence in the top layer before performing another round of Gaussian message passing in the bottom layer. Note that instead of running the EM algorithm for the estimation of $S^{(2)}$ at each update of $S^{(1)}$ till convergence, performing a few ascent steps on $S^{(2)}$ significantly improves the runtime and tends to yield better estimation results. With this modification, the runtime complexity of the proposed EM message passing algorithm is actually linear in the number of samples.

IV. Results and Applications

We present two use cases of SSM-DSV: Robust estimation in the presence of block outliers and estimation of highly correlated input signals to a physiologically-motivated SSM.

A. Kalman Smoothing with Block-Outliers

Due to their least-squares error criterion, standard Kalman smoothers are sensitive to outliers. Robust smoothers such as [22], [23] assume that outliers are sparse and their usage of per-sample penalties is suboptimal for measurement outliers that affect larger segments. By contrast, SSM-DSV are well suited for modeling dynamical systems that are affected by contiguous disturbance signals. *Simulation setup*: Let the bottom layer SSM be the third-order stochastic resonator model from [20] with

$$A^{(0)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega\Delta t) & \sin(\omega\Delta t)/\omega \\ 0 & -\omega\sin(\omega\Delta t) & \cos(\omega\Delta t) \end{bmatrix}, \quad (26)$$

$B^{(0)} = \mathbf{0} \in \mathbb{R}^3$ and $C^{(0)} = [1,1,0]$. Further, we chose $\Delta t = 0.1$, the angular velocity $\omega = 5$, the state noise covariance matrix $\Sigma^{(0)} = [0.005, 0, 0; 0, 0.1, 0; 0, 0, 0.1]$, and the measurement noise standard deviation $\sigma_Z^{(0)} = 1$. The disturbance signal is zero except on the interval $J = [250, 300]$ (see Fig. 3), where the measurements $y^{(0)}$ are replaced by the saturation model

$$y_k^{(0)} = y_{k_1}^{(0)} + (A - y_{k_1}^{(0)}) \tanh(\epsilon(k - k_1)), \quad (27)$$

for $k \in J$ and where $k_1 = 250$, $A = 15$ is the maximal saturation value, and $\epsilon = 2/|J|$ determines how fast the rise occurs as a function of the artifact duration. In [24] this model was used for simulating catheter flushes, which lead to saturations in the arterial blood pressure readings.

Estimation setup: Let the noisy signal $y^{(0)}$, the noise statistics and the bottom layer system matrices $A^{(0)}$, $B^{(0)}$, and $C^{(0)}$ be given. To capture the unknown disturbance signal, we model the measurement noise standard deviation as a random walk with sparse increments $U^{(1)}$:

$$X_k^{(1)} = X_{k-1}^{(1)} + U_k^{(1)} \quad (28)$$

$$S_k^{(1)} = Y_k^{(1)} = X_k^{(1)}. \quad (29)$$

Figure 3 shows the estimation results obtained by EM message passing and by regular Kalman smoothing. For EM-MP we used $M = 1$ (see Eq. (22)) and performed five ascent steps, i.e., Gaussian message passing, in the top layer SSM and ten iterations in the bottom layer SSM. Table I compares the normalized mean squared error (nMSE) $\|y_{\text{noise-free}} - y_{\text{est}}\|^2 / \|y_{\text{noise-free}}\|^2$ between the noise-free observation and its estimate using EM-MP and KS for different artifact durations.

Artifact duration [%]	0	2.5	5	10	15	20
nMSE(EM-MP) [%]	1.8	6.3	7.6	12.5	19	35.1
nMSE(KS) [%]	1.8	15.5	40.4	107.7	186.2	274.6

TABLE I: nMSE of estimated output signal (averaged over 50 runs) for EM-MP and KS for varying artifact durations.

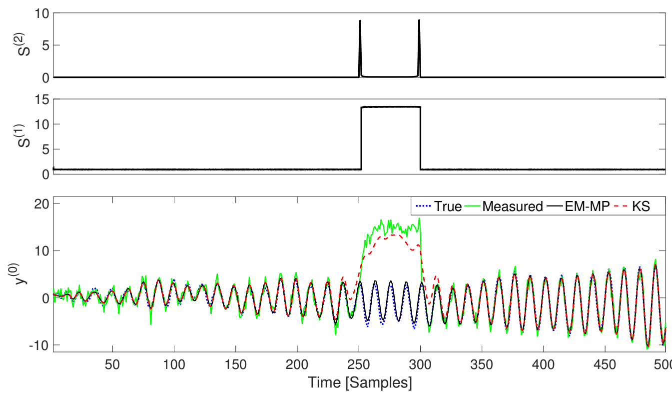

Fig. 3: Top: Estimated sparse NUV parameters $S^{(2)}$. Center: Structured NUV parameters $S^{(1)}$ representing the standard deviation of the measurement noise. Bottom: Measured (with 10% artifact contamination), ground truth, and estimated output signal obtained by EM message passing (EM-MP) and by Kalman smoothing (KS). The output nMSE for EM-MP is 6.5% and 162% for the KS.

B. Estimation of Group-sparse Inputs

Unknown Gaussian inputs to SSMs can be estimated by Gaussian message passing [10]. Similarly, sparse inputs can be estimated by placing a NUV prior on the inputs, and using the EM algorithm for estimation [4]. Here, we propose SSM-DSV and EM message passing for estimating group-sparse input signals. Consider the following application: Given eye position measurements $y^{(0)}$ (see Fig. 4), the goal is to estimate the neural controller signal that best explains the observed trajectory of the eye. Oculomotor neural controller signals, measured in spikes per second, determine the tension (resulting force) in the eye muscles, which in turn determines the eye position. The oculomotor system (eye ball and muscles) can be modeled as a linear spring-mass-damper system, which can be represented in state space form [25]. Here, we consider two types of eye movements: fixations and saccades (fast eye movements between fixations). While the neural controller signal stays approximately constant during fixations, it changes abruptly during saccades. In [26] we assumed that the neural controller signal changes at (sparse) discrete points in time, which we estimated by sparse input estimation. With this approach the estimated neural controller signal is piecewise constant (see Fig. 4). SSM-DSV provide a more realistic model for this application. For this, the bottom layer SSM represents the oculomotor system with the following SSM:

$$\begin{bmatrix} X_k^{(0),\text{Eye}} \\ X_k^{(0),\text{Neur}} \end{bmatrix} = \begin{bmatrix} A^{\text{Eye}} & B^{\text{Eye}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} X_{k-1}^{(0),\text{Eye}} \\ X_{k-1}^{(0),\text{Neur}} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} U_k^{(0)} \quad (30)$$

$$Y_k^{(0)} = (1, 0, \ldots, 0) \begin{bmatrix} X_k^{(0),\text{Eye}} \\ X_k^{(0),\text{Neur}} \end{bmatrix} + Z_k^{(0)}, \quad (31)$$

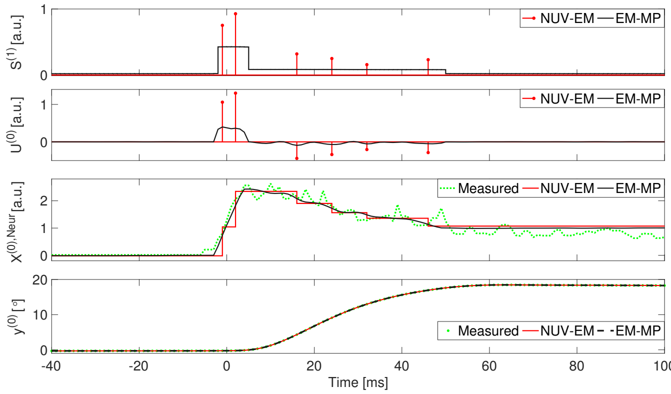

Fig. 4: Top: Estimated NUV parameter $S^{(1)}$ using sparse input estimation (NUV-EM) [26] and EM message passing (EM-MP). Second row: Estimated (group-sparse) input signal $U^{(0)}$, which is responsible for changing the neural controller signal $X^{(0),\text{Neur}}$; see (30). Third row: Estimated and measured neural controller signal from a rhesus monkey [25]. Bottom: Measured and estimated eye position $y^{(0)}$.

where $\mathbf{0} \in \mathbb{R}^4$ is the all-zeros vector and $X_k^{(0),\text{Eye}} = (\phi_k, \dot{\phi}_k, \ddot{\phi}_k, F_k) \in \mathbb{R}^4$ contains the noise-free eye position ϕ (in degrees), the velocity $\dot{\phi}$, the acceleration $\ddot{\phi}$, and the resulting force F of the eye muscles on the eye globe. The state $X^{(0),\text{Neur}}$ represents the neural controller signal and is driven by the group sparse input $U^{(0)}$. The system matrices $A^{\text{Eye}} \in \mathbb{R}^{4\times4}$ and $B^{\text{Eye}} \in \mathbb{R}^{4\times1}$ model the oculomotor system dynamics. In the top layer SSM we model the NUV parameters of $U^{(0)}$ by the zero-order hold model (28)–(29). Estimation in the resulting SSM-DSV is performed by EM message passing as described in Section III-B. We used $M = 2$ in (22) and performed twenty iterations in the bottom layer and five iterations at each run in the top layer. Figure 4 shows the eye position $y^{(0)}$ during a saccade together with the pertaining neural controller signal $X^{(0),\text{Neur}}$. In addition, the estimated group sparse input $U^{(0)}$ and the block-sparse standard deviations $S^{(1)}$ are shown.

V. Conclusion

In this paper we have proposed a flexible model class denoted by state space models with dynamical and sparse variances (SSM-DSV). For inference and learning in these two-layer models we used EM message passing. SSM-DSV can be used to model signals with dynamical variances, in particular signals with group sparse or piecewise constant variances, as shown in the neural input signal recovery and in the block-outlier insensitive Kalman smoother examples. The relative merits of EM message passing in such models, compared to free-form variational message passing [19], variational Kalman filtering based conjugate-exponential distributions [20], and particle filtering [27] remain to be investigated.

References

[1] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, New Jersey, USA: Prentice Hall, 2000.

[2] C.-J. Kim, C. R. Nelson *et al.*, *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. Cambridge, MA, USA: MIT Press, 1999.

[3] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. Oxford, UK: 2nd ed. Oxford University Press, 2012.

[4] 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 & Applications Workshop*, Feb. 2016.

[5] F. Wadehn, "State space methods with applications in biomedical signal processing," Ph.D. dissertation, ETH Zurich, 2019.

[6] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. of Machine Learning Research*, vol. 1, pp. 211–244, Jun. 2001.

[7] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination," in *Advances in Neural Inf. Process. Systems*, Dec. 2007.

[8] J. Palmer, K. Kreutz-Delgado, B. D. Rao, and D. P. Wipf, "Variational EM algorithms for non-Gaussian latent variable models," in *Advances in Neural Inf. Process. Systems*, Dec. 2006.

[9] H.-A. Loeliger, B. Ma, H. Malmberg, and F. Wadehn, "Factor graphs with NUV priors and iteratively reweighted descent for sparse least squares and more," in *Int. Symp. on Turbo Codes and Iterative Information Process.*, Dec. 2018.

[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," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, July 2007.

[11] N. Zalmai, H. Malmberg, and H.-A. Loeliger, "Blind deconvolution of sparse but filtered pulses with linear state space models," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Mar. 2016.

[12] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[13] J. Dauwels, S. Korl, and H.-A. Loeliger, "Expectation maximization as message passing," in *IEEE Int. Symp. on Inf. Theory*, Sept. 2005.

[14] J. Dauwels, A. Eckford, S. Korl, and H.-A. Loeliger, "Expectation maximization as message passing–Part i:Principles and Gaussian messages," *arXiv:0910.2832*, 2009.

[15] J. Dauwels, "On variational message passing on factor graphs," in *IEEE Int. Symp. on Inf. Theory*, Jun. 2007.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.

[17] R. S. Tsay, *Analysis of Financial Time Series*. New York, USA: John Wiley & Sons, 2005.

[18] C. D. Mathys, E. I. Lomakina, J. Daunizeau, S. Iglesias, K. H. Brodersen, K. J. Friston, and K. E. Stephan, "Uncertainty in perception and the hierarchical Gaussian filter," *Frontiers in Human Neuroscience*, vol. 8, p. 825, Nov. 2014.

[19] I. Senöz and B. de Vries, "Online variational message passing in the hierarchical Gaussian filter," in *IEEE Int. Workshop on Machine Learning for Signal Process.*, Sept. 2018.

[20] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Automat. Contr.*, vol. 54, no. 3, pp. 596–600, Feb. 2009.

[21] J. Winn and C. M. Bishop, "Variational message passing," *J. of Machine Learning Research*, vol. 6, 2005.

[22] S. Farahmand, G. B. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4529–4543, July 2011.

[23] F. Wadehn, L. Bruderer, J. Dauwels, V. Sahdeva, H. Yu, and H.-A. Loeliger, "Outlier-insensitive Kalman smoothing and marginal message passing," in *24th European Signal Process. Conf.*, Aug. 2016.

[24] Q. Li, R. G. Mark, and G. D. Clifford, "Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator," *Biomed. Eng. Online*, vol. 8, no. 1, July 2009.

[25] F. Wadehn, D. J. Mack, T. Weber, T. Heldt, and H.-A. Loeliger, "Model-based separation, detection, and classification of eye movements," *IEEE Trans. Biomed. Eng.*, May 2019.

[26] F. Wadehn, D. J. Mack, T. Weber, and H.-A. Loeliger, "Estimation of neural inputs and detection of saccades and smooth pursuit eye movements by Sparse Bayesian Learning," in *IEEE Int. Conf. on Eng. in Med. and Biol.*, July 2018.

[27] G. M. Vitetta, E. Sirignano, and F. Montorsi, "Particle smoothing for conditionally linear Gaussian models as message passing over factor graphs," *IEEE Trans. on Signal Process.*, vol. 66, no. 14, pp. 3633–3648, 2018.