Deconvolution of Weakly-Sparse Signals and Dynamical-System Identification by Gaussian Message Passing

Lukas Bruderer, Hampus Malmberg and Hans-Andrea Loeliger ETH Zurich
Dept. of Information Technology and Electrical Engineering
8092 Zürich Switzerland
{bruderer, malmberg, loeliger}@isi.ethz.ch

***Abstract*—We use ideas from sparse Bayesian learning for estimating the (weakly) sparse input signal of a linear state space model. Variational representations of the sparsifying prior lead to algorithms that essentially amount to Gaussian message passing. The approach is extended to the case where the state space model is not known and must be estimated. Experimental results with a real-world application substantiate the applicability of the proposed method.**

I. Introduction

The general area of this paper is sparse Bayesian learning, which was introduced in the seminal paper [1] and has since found applications in signal processing [2] and communications [3].

The specific problems addressed in this paper is the estimation of a sparse input signal $\boldsymbol{U} = (U_1, \ldots, U_L) \in \mathbb{R}^L$ of a linear dynamical system; first, for a known dynamical system, and then, for an unknown dynamical system.

The system model is defined as

$$\boldsymbol{X}_k = \mathbf{A}\boldsymbol{X}_{k-1} + \boldsymbol{b}U_{k-1} \tag{1}$$

$$Y_k = \boldsymbol{c}\boldsymbol{X}_k + N_k \tag{2}$$

where the observations $\mathbf{Y} = (Y_1, \ldots, Y_L) \in \mathbb{R}^L$ and the model is governed by $N_k \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_N^2\right)$, $\mathbf{A} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{X}_k$, $\boldsymbol{b}$, $\boldsymbol{c}^{\mathsf{T}} \in \mathbb{R}^d$.

How to use sparsity to regularize blind estimation problems has been a research topic for many years. Related approaches include, e.g., blind source separation [4], [5] and dictionary learning [6]. The Bayesian approach to such problems provides not only an estimate of the sparse variables, but also reliability information of the posterior (e.g., posterior variance), which can be essential for blind deconvolution [7].

A key issue with any Bayesian approach is the choice of a suitable prior, which strongly influences the resulting algorithms. As observed in [8]–[10], variational representations of pdfs (probability density functions) offer interesting options in this respect. Fortunately, such variational representations include important classes of compressible priors[1] [11]. Compressible priors as in [11] are pdfs such that sorted samples exhibit a power-law decay. We model our input signals as i.i.d. processes with such priors, and such signals will be called *weakly sparse*. The main contributions of this paper are as follows:

1) We demonstrate the use of sparsifying ("compressible") priors with variational representations for input signals of state space models such that the resulting estimation algorithm amounts to Gaussian message passing (Section III).
2) We derive and demonstrate an algorithm for blind input estimation (Sections V and VI). Again, the actual computations amount to Gaussian message passing.
3) We extend the Bryson-Frazier Kalman smoother [12] to input estimation, thus obtaining a Gaussian message passing scheme for the mentioned computations that does not require a matrix inversion.

II. Examples

We begin with two examples, both in order to clarify the problem statement and to illustrate the empirical success of the proposed methods. Consider an input signal $\boldsymbol{U} \in \{-1, 0, 1\}^L$ with s non-zero components where $s \ll L$. Furthermore, observations $\boldsymbol{y}$, from the random variable $\mathbf{Y}$, are generated via a linear state space model as in (1) and (2).

Example 1 (Weakly-Sparse Input Estimation) Assume that the dynamical system, (1) and (2), is completely known. In the simulation, a (exactly) sparse signal $\boldsymbol{U}$ is passed through a strongly resonating filter of order $d = 12$, resulting in the measured signal $\boldsymbol{y}$ as depicted in Fig. 1 (top). Fig. 1 (bottom) shows the estimate of the input signal obtained by Gaussian message passing as in Sections III and IV. The corresponding LASSO estimate [13], which is much harder to compute, is also depicted for comparison. Observe that, for this example the LASSO estimate does not work well due to the strong coherence in the dictionary. □

Example 2 (Blind Deconvolution) Next consider an example where $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ in (1) and (2) are not known; all we have is the output signal $\boldsymbol{y}$ in Fig. 2 (top). Fig. 2 (bottom) shows the estimated input signal obtained by Gaussian message passing as in Section V. The assumed (and true) system order is $d = 4$.

[1] Not to be confused with the information-theoretic notion of compressibility.

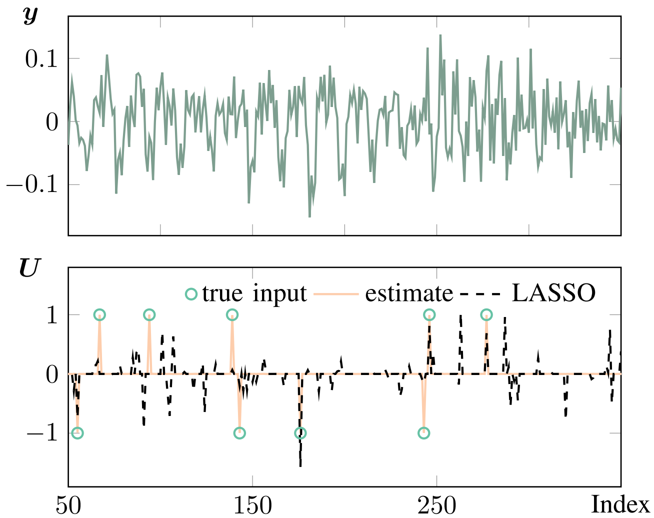

Fig. 1. Input estimation with state space model and sparsifying i.i.d. prior using simulated data with SNR of 30 dB.

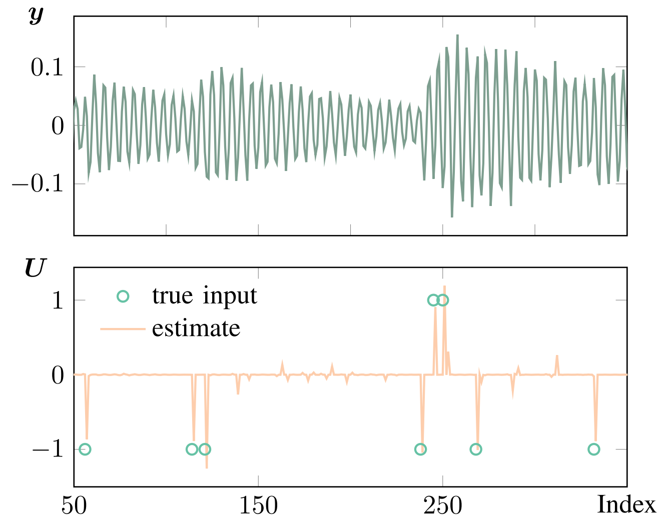

Fig. 2. Blind input estimation example using 4-th order state space model on simulated data with SNR of 30 dB.

We have not been able to make the LASSO algorithm work for this case. □

Both examples are simulated with a SNR = 30 dB, as well as i.i.d. Student's t prior over γ with $\nu = 10^{-4}$, and they are evaluated after 5 EM iterations. Additional experimental results with a real-world example will be given in Section VI.

III. Sparsity with Gaussian Algorithms

A. Basic Idea

In order to obtain a sparse estimate of the input signal $\boldsymbol{U}$, we complete the state space model in (1) and (2) with a sparsifying i.i.d. prior $p(\boldsymbol{u})$; specifically, we will choose the Student's t distribution, defined as

$$p(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{3}$$

with parameter $\nu > 0$ and where $\Gamma(\cdot)$ refers to the gamma function. The Student's t distribution with $\nu < 3$ belongs to a class of distributions that are provably compressible [11]. Clearly, such a prior makes the overall statistical model non-Gaussian.

Following the ideas of [1], [9], we will retrieve Gaussianity by the following steps. First, we will use a variational representation of $p(\boldsymbol{u})$ (as in Section III-B), which introduces new variables $\gamma_1, \ldots, \gamma_L$ such that $p(u_k|\gamma_k)$ is Gaussian and

$$p(u_k) = \sup_{\gamma_k} p(u_k|\gamma_k) f(\gamma_k) \tag{4}$$

for a function $f(\gamma_k)$. A factor-graph representation of (4) is shown in Fig. 3, where the node/factor $p(u_k)$ is expanded into a max-box, cf. [14].

In a second step, we approximate $p(\boldsymbol{u}, \boldsymbol{y})$ by the Gaussian distribution

$$\hat{p}(\boldsymbol{u}, \boldsymbol{y}) \triangleq p(\boldsymbol{y}|\boldsymbol{u}) \prod_{k=1}^{L} p(u_k|\hat{\gamma}_k) \tag{5}$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_L)$ result from the maximization discussed in Section III-C. We will see that this maximization amounts essentially to (multiple rounds of) Gaussian message passing.

Conceptually, we then estimate $\boldsymbol{U}$ by MAP estimation from the Gaussian distribution (5). However, all the required computations have already been performed in the last round of Gaussian message passing in Step 2 above.

The approximation (5) exemplifies *Type II methods* as in [15]. Sparsifying priors and related algorithms for such methods have been presented in [9], [16].

B. Variational Prior Representation

A symmetric pdf $p(x)$ is said to be *strongly super-Gaussian* [10] if $p(\sqrt{x})$ is log-convex on $(0, \infty)$. Such pdfs have heavy tails, which are key to compressibility [11]. Strongly super-Gaussian pdfs admit the following variational representation [10]: for $p(x) = e^{-g(x^2)}$, we have

$$p(x) = \sup_{\gamma > 0} p(x|\gamma)\, \phi(\gamma^{-1}), \tag{6}$$

where

$$p(x|\gamma) = (2\pi\gamma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{x^2}{\gamma}\right) \tag{7}$$

is a Gaussian distribution and

$$\phi(\alpha) = \sqrt{\frac{2\pi}{\alpha}} e^{g^{\star}(\alpha/2)} \tag{8}$$

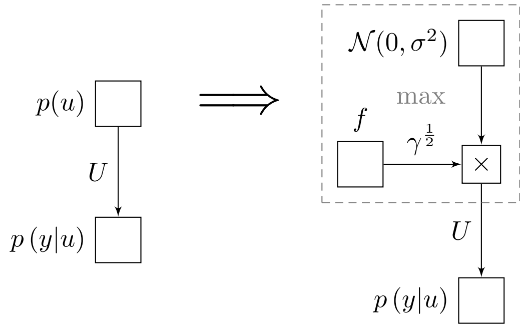

Fig. 3. Factor-graph representation of (4): the box $p(u)$ (left) is expanded into a max-box (right).

where $g^\star$ is the concave conjugate of g (see, e.g. [17]). In the sequel we define $f(\gamma) \triangleq \phi(\gamma^{-1})$.

The Student's t distribution (3) is strongly super-Gaussian, and it has the variational representation (4) with

$$f(\gamma_k) = K_\nu \, \gamma_k^{-\nu/2} e^{-\nu/2\gamma_k} \tag{9}$$

where $K_\nu \triangleq 2\pi \left(\frac{2\nu}{\nu+1}\right)^{\nu+1} e^{\nu+1}$.

C. Maximization by EM

The parameters $\hat{\boldsymbol{\gamma}}$ in (5) are defined by

$$\hat{\boldsymbol{\gamma}} \triangleq \arg\max_{\gamma_1,\ldots,\gamma_L} \int p(\boldsymbol{y}|\boldsymbol{u}) \prod_{k=1}^{L} p(u_k|\gamma_k) f(\gamma_k) \; \mathrm{d}u_k, \tag{10}$$

which may be viewed as maximizing the following lower bound on the evidence $p(\boldsymbol{y})$:

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{u}) p(\boldsymbol{u}) \; \mathrm{d}\boldsymbol{u} \tag{11}$$

$$= \int p(\boldsymbol{y}|\boldsymbol{u}) \prod_{k=1}^{L} \max_{\gamma_k} p(u_k|\gamma_k) f(\gamma_k) \; \mathrm{d}u_k \tag{12}$$

$$\geq \max_{\gamma_1,\ldots,\gamma_L} \int p(\boldsymbol{y}|\boldsymbol{u}) \prod_{k=1}^{L} p(u_k|\gamma_k)\, f(\gamma_k)\, \mathrm{d}u_k. \tag{13}$$

The maximization in (10) is naturally carried out by expectation maximization (EM) with hidden variables U_k. The EM algorithm alternates between computing new posterior probabilities by Gaussian message passing and re-estimating γ_k individually by

$$\arg\max_{\boldsymbol{\gamma}_k} \mathrm{E}_{p(u_k|\boldsymbol{y},\boldsymbol{\gamma}')}[\log p\left(u_k|\gamma_k\right)] + \log f(\gamma_k). \tag{14}$$

IV. Efficient Gaussian Message Passing

The main ingredients to (14) are the posterior densities of U_k. In linear state space models, these quantities can be computed (with linear complexity) by Kalman filtering techniques (e.g., a Rauch-Tung-Striebel smoother [12]) with additional steps, or by Gaussian message passing as in [14].

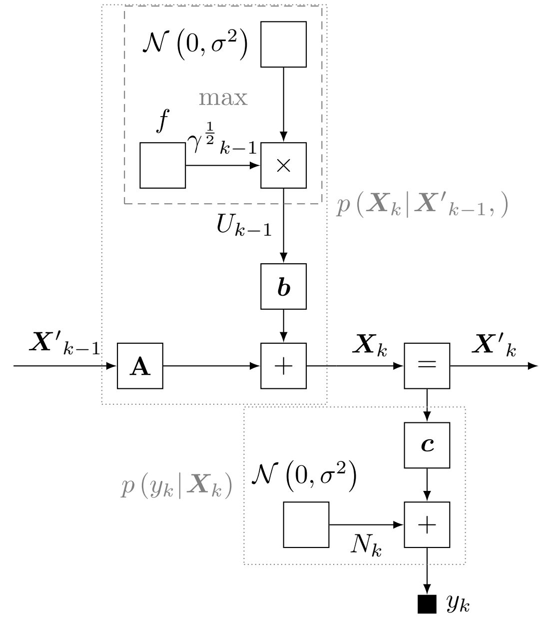

Fig. 4. A factor graph representation our weakly-sparse input state space model, where $p(\boldsymbol{y}|\boldsymbol{u})$ is decomposed into factors defined by the sate space model in (2).

We now propose an efficient Gaussian message passing scheme to compute $p(u_k|\boldsymbol{y},\boldsymbol{\gamma})$ (for fixed $\boldsymbol{y}$ and $\boldsymbol{\gamma}$) that does not require any matrix inversion. The proposed scheme turns out to be an extension of the Bryson-Frazier Kalman smoother [12] to input signal estimation. Note that, for the (infinite impulse response) models of this paper, this extension is not trivial. The following description of the algorithm generally follows [14], but focusses on the quantities

$$\tilde{\mathbf{W}}_{\boldsymbol{X}_k} = \left(\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k} + \overleftarrow{\mathbf{V}}_{\boldsymbol{X}_k}\right)^{-1} \tag{15}$$

$$\tilde{\mathbf{W}}_{\boldsymbol{X}_k} \tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_k} = \tilde{\mathbf{W}}_{\boldsymbol{X}_k} \left(\overrightarrow{m}_{\boldsymbol{X}_k} - \overleftarrow{m}_{\boldsymbol{X}_k}\right), \tag{16}$$

where $\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$ and $\overleftarrow{\mathbf{V}}_{\boldsymbol{X}_k}$ denote covariance matrices and $\overrightarrow{m}_{\boldsymbol{X}_k}$ and $\overleftarrow{m}_{\boldsymbol{X}_k}$ mean vectors being passed in the corresponding direction. The posterior distribution of U_k is then obtained as

$$\mathbf{V}_{\boldsymbol{U}_k} = \sigma^2\gamma_k - (\sigma^2\gamma_k)^2 \boldsymbol{b}^\mathsf{T} \tilde{\mathbf{W}}_{\boldsymbol{X}_{k+1}} \boldsymbol{b} \tag{17}$$

$$\mathbf{m}_{\boldsymbol{U}_k} = -\sigma^2\gamma_k \boldsymbol{b}^\mathsf{T} \tilde{\mathbf{W}}_{\boldsymbol{X}_{k+1}} \tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_{k+1}}. \tag{18}$$

The full message passing scheme to obtain (17) and (18) first performs a forward recursion over all k, equivalent to Kalman filtering, resulting in $\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}$ and $\overrightarrow{\boldsymbol{m}}_{\boldsymbol{X}_k}$. Then the quantities $\tilde{\mathbf{W}}_{\boldsymbol{X}_k}$ and $\tilde{\mathbf{W}}_{\boldsymbol{X}_k} \tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_k}$ are computed with the following two updates, starting from $k = L$: include y_k with

$$\tilde{\mathbf{W}}_{\boldsymbol{X}_k} = \mathbf{F}_k^\mathsf{T} \tilde{\mathbf{W}}_{\boldsymbol{X}'_k} \mathbf{F}_k + g_k \boldsymbol{c}^\mathsf{T} \boldsymbol{c} \tag{19}$$

$$\tilde{\mathbf{W}}_{\boldsymbol{X}_k} \tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_k} = \mathbf{F}_k^\mathsf{T} \tilde{\mathbf{W}}_{\boldsymbol{X}'_k} \tilde{\boldsymbol{\mu}}_{\boldsymbol{X}'_k} - g_k \boldsymbol{c}^\mathsf{T} \left(\boldsymbol{y}_k - \boldsymbol{c}\overrightarrow{m}_{\boldsymbol{X}_k}\right) \tag{20}$$

where

$$g_k = \left(\boldsymbol{c}\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}\boldsymbol{c}^\intercal + \sigma^2\right)^{-1} \tag{21}$$

$$\mathbf{F}_k = \mathbf{I} - g_k\boldsymbol{c}^\intercal\boldsymbol{c}\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k} \tag{22}$$

and then a time update

$$\tilde{\mathbf{W}}_{\boldsymbol{X'}_k} = \mathbf{A}^\intercal\tilde{\mathbf{W}}_{\boldsymbol{X}_{k+1}}\mathbf{A} \tag{23}$$

$$\tilde{\mathbf{W}}_{\boldsymbol{X'}_k}\tilde{\boldsymbol{\mu}}_{\boldsymbol{X'}_k} = \mathbf{A}^\intercal\tilde{\mathbf{W}}_{\boldsymbol{X}_{k+1}}\tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_{k+1}}. \tag{24}$$

Given the posterior density of U_k, the maximization problem in (14) is performed independently and the update of variational parameter γ_k is retrieved as shown in [10] via

$$\gamma_k = \left(-\frac{1}{u_k}\frac{\mathrm{d}\left(\log p\left(u_k\right)\right)}{\mathrm{d}u_k}\right)^{-1}\Bigg|_{u_k=\sqrt{\mathrm{E}_{p\left(u_k\,|\,\boldsymbol{y},\boldsymbol{\gamma}'\right)}\left[u_k^2\right]}} \tag{25}$$

V. System Identification

So far, we have assumed that the system model (1) and (2) is known. We now turn to the case where the system model is not known and must be estimated as well.

In order to derive a blind scheme, let us rewrite and extend (12) as the joint minimization problem

$$\arg\min_{\boldsymbol{\gamma},\mathcal{H}} -2\log p(\boldsymbol{y}|\boldsymbol{\gamma},\mathcal{H}) \tag{26}$$

where $\mathcal{H}$ denotes the linear operator mapping from $\boldsymbol{U}$ to the noiseless observations. From standard matrix identities, and invoking the same step as in (13), the minimization problem in (26) can be expressed as

$$\arg\min_{\boldsymbol{u},\boldsymbol{\gamma},\mathcal{H}}\log\left|\mathcal{H}\Theta\mathcal{H}^\mathsf{T} + \sigma^2\mathbf{I}\right| + \sigma^{-2}\|\boldsymbol{y}-\mathcal{H}\boldsymbol{u}\|^2 + \sum_{k=0}^{L}\frac{u_k^2}{\gamma_k} + \log f(\gamma_k). \tag{27}$$

where $\Theta = \operatorname{diag}(\boldsymbol{\gamma})$ and the columns of $\mathcal{H}$ are normalized, i.e., the state space model's impulse response $\boldsymbol{h} = h_1, \ldots, h_L$ is constrained to have energy 1 to overcome scaling ambiguity in $\mathcal{H}$ and $\boldsymbol{\gamma}$.

The objective function (27) may be computed conveniently by coordinate descent in $\mathcal{H}$ and $\boldsymbol{\gamma}$. When $\boldsymbol{\gamma}$ is fixed, we recognize a state space model identification problem where the state space is driven by non-stationary Gaussian noise. Whereas in the other case, i.e. when $\mathcal{H}$ is kept fixed, the objective is equivalent to sparse input estimation.

We propose an alternating algorithm estimates $\boldsymbol{\gamma}$ and $\mathcal{H}$ accordingly. For the input estimation, i.e., when $\mathcal{H}$ is kept fixed, we invoke the previously presented framework from Sections III and IV. In addition, when $\boldsymbol{\gamma}$ is fixed, the EM methods from [18] are used to estimate the state space model, i.e., $\mathbf{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, in (1) and (2).

To complete the EM steps from [18], the necessary posterior probabilities over $\boldsymbol{X}_k$ follow from forward-backward message-passing algorithm as in Section IV with the additional steps

$$\mathbf{V}_{\boldsymbol{X}k} = \overrightarrow{\mathbf{V}}_{\boldsymbol{X}k}\left(\mathbf{I} - \tilde{\mathbf{W}}_k\overrightarrow{\mathbf{V}}_{\boldsymbol{X}k}\right) \tag{28}$$

$$\mathbf{m}_{\boldsymbol{X}k} = \vec{m}_{\boldsymbol{X}k} - \overrightarrow{\mathbf{V}}_{\boldsymbol{X}k}\tilde{\mathbf{W}}_{\boldsymbol{X}_k}\tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_k} \tag{29}$$

and the cross covariance of $\boldsymbol{X'}_{k-1}$ and $\boldsymbol{X}^\intercal{}_k$ is computed from

$$\mathbf{V}_{\boldsymbol{X'}_{k-1},\boldsymbol{X}^\intercal{}_k} = \overrightarrow{\mathbf{V}}_{\boldsymbol{X'}_{k-1}}\mathbf{A}^\intercal\left(\mathbf{I} - \tilde{\mathbf{W}}_{\boldsymbol{X}_k}\right)\overrightarrow{\mathbf{V}}_{\boldsymbol{X}_k}. \tag{30}$$

Similarly to the maximization over $\boldsymbol{u}$ (cf. Section IV), the new messages $\tilde{\mathbf{W}}_{\boldsymbol{X}_k}$ and $\tilde{\mathbf{W}}_{\boldsymbol{X}_k}\tilde{\boldsymbol{\mu}}_{\boldsymbol{X}_k}$ result in an inversion-free algorithm, contrary to standard EM methods [18] and [19].

A. Initialization

The input estimate in the first iteration can be considered proportional to the energy in $\boldsymbol{y}$ weighted by the spectrum of the initial state space model. With no prior knowledge on model or $\boldsymbol{U}$, an instantaneous energy detector is a sensible initial choice. To this end, the initial state space model is initialized with

$$\mathbf{A} = 0, \tag{31}$$

$\boldsymbol{b}$ as all ones vector, and $\boldsymbol{c}$ is drawn randomly and scaled such that $\boldsymbol{cb} = 1$.

B. Energy-Constrained Updates

Since the EM algorithm increases the likelihood iteratively, the unspecific input estimates, during the first iterations, imply a large ambiguity on the state space model. Commonly, this results in system estimates that exhibit a gain which is far from 1.

To prevent this effect, the updates of the state space models impulse response $\boldsymbol{h} = h_1, \ldots, h_L$ are constrained such that $\|\boldsymbol{h}\|^2 = 1$. The constrained impulse response follows as

$$\|\boldsymbol{h}\|^2 = \boldsymbol{c}\left(\sum_{n=0}^{\infty}\mathbf{A}^n\boldsymbol{b}\boldsymbol{b}^\mathsf{T}\left(\mathbf{A}^\mathsf{T}\right)^n\right)\boldsymbol{c}^\mathsf{T} \tag{32}$$

$$= \boldsymbol{c}\,\mathcal{C}\left(\mathbf{A},\boldsymbol{b}\boldsymbol{b}^\mathsf{T}\right)\boldsymbol{c}^\mathsf{T}, \tag{33}$$

where we used $h_n = \boldsymbol{c}\mathbf{A}^n\boldsymbol{b}$ and $\mathcal{C}\left(\mathbf{A},\boldsymbol{b}\boldsymbol{b}^\mathsf{T}\right)$ denotes the controllability Gramian [12], which can be obtained by solving a Lyapunov equation. When $\mathbf{A}$ is (approximately) constant during an EM update of the SSM, a quadratic constraint on $\boldsymbol{c}$ can be conceived from (33) and added to the M-step of $\boldsymbol{c}$:

$$\begin{aligned}\min_{\boldsymbol{c}}\quad & \boldsymbol{c}\mathbf{W}_{\boldsymbol{c}}\boldsymbol{c}^\mathsf{T} - 2\boldsymbol{c}\mathbf{W}_{\boldsymbol{c}}\boldsymbol{m}_{\boldsymbol{c}}\\ \text{s.t.}\quad & \boldsymbol{c}\,\mathcal{C}\left(\mathbf{A},\boldsymbol{b}\boldsymbol{b}^\mathsf{T}\right)\boldsymbol{c}^\mathsf{T} = 1,\end{aligned}$$

where $\mathbf{W}_{\boldsymbol{c}}$ and $\mathbf{W}_{\boldsymbol{c}}\boldsymbol{m}_{\boldsymbol{c}}$ follows from the EM step in [18].

VI. A Real-World Example

Ballistocardiography (BCG) tries to infer heartbeats from body movement generated by the ejection of the blood at each cardiac cycle [20]. The problem is challenging because the corresponding system is often highly resonating, see Fig. 5. Moreover, the system is unknown and likely to change over time.

Some experimental results are shown in Fig. 5. The BCG signal consists of 3000 samples sampled at 20 Hz. The assumed model order for the system identification is $d = 4$. We model the input signal as a weakly sparse random signal,

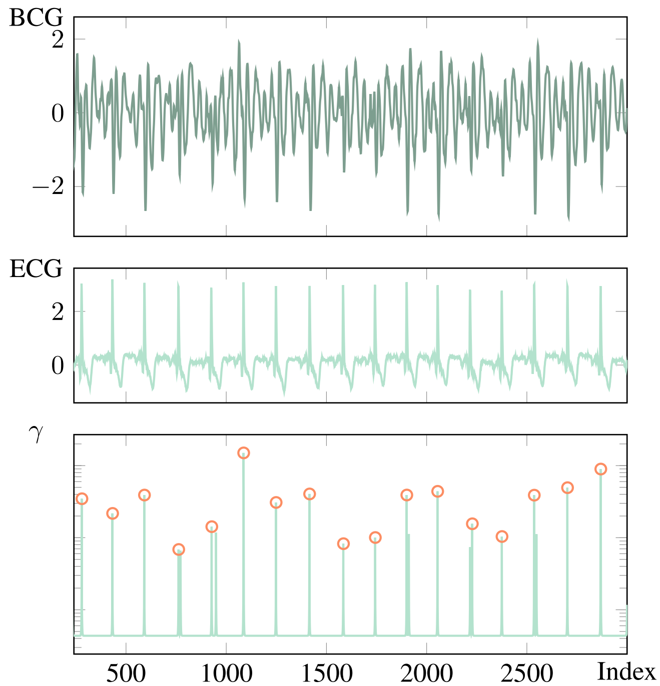

Fig. 5. Blind pulse (heart beat) detection from a ballistocardiographic (BCG) signal. The ECG signal (middle) is used for validation (i.e., to provide the ground truth). The bottom plot shows the estimated variance of each input signal sample.

specifically we select an i.i.d. Student's t prior with $\nu = 10^{-3}$. Our choice is motivated by guarantees that realizations of this prior are itself almost surely weakly sparse or compressible [11], thus encouraging soundness of our Bayesian inference scheme. Initialization according to Section V-A is followed by 20 EM iterations.

Comparing the estimated variance vector Fig. 5 (bottom) with an electrocardiographic (ECG) reference signal Fig. 5 (middle), we find that all heartbeats are detected, and there are no false alarms except that some heartbeats are split into closely adjacent beats. Such (physiologically impossible) duplications can easily be cleaned up, as illustrated by the circles in Fig. 5 (bottom).

VII. Conclusion

Variational representations of heavy-tailed priors in otherwise linear Gaussian models enable estimation by means of Gaussian message passing. This general idea was worked out for estimating the (weakly) sparse input signal of a linear state space model. The approach was then extended to the case where the linear system is unknown and must be estimated as well. The robustness and practicality of the proposed approach was demonstrated by a real-world example. Finally, we proposed a new (and very efficient) version of Gaussian message passing in linear state space models for input estimation without matrix inversion.

References

[1] M. E. Tipping, "The relevance vector machine," in *Advances in Neural Information Processing Systems*, vol. 12. MIT Press, 2000, pp. 652–658.

[2] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1150–1159, Sept 2003.

[3] D. Shutin and B. H. Fleury, "Sparse variational Bayesian SAGE algorithm with application to the estimation of multipath wireless channels," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3609–3623, 2011.

[4] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[5] C. Févotte and S. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2174–2188, Nov 2006.

[6] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[7] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1964–1971.

[8] T. S. Jaakkola and M. I. Jordan, "Bayesian parameter estimation via variational methods," *Statistics and Computing*, vol. 10, no. 1, pp. 25–37, 2000.

[9] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[10] J. A. Palmer, D. P. Wipf, K. Kreutz-delgado, and B. D. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Advances in Neural Information Processing Systems 18*. MIT Press, 2006, pp. 1059–1066.

[11] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.

[12] T. Kailath, B. Hassidi, and A. H. Sayed, *Linear Estimation*. Prentice-Hall, 2000.

[13] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. Springer, 2009, vol. 2, no. 1.

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

[15] G. Casella and R. Berger, *Statistical Inference*. Duxbury Resource Center, June 2001.

[16] M. Seeger and D. P. Wipf, "Variational Bayesian inference techniques," *IEEE Magazine on Signal Processing*, vol. 27, no. 6, pp. 81–91, 2010.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[18] 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.

[19] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.

[20] L. Giovangrandi, O. Inan, R. Wiard, M. Etemadi, and G. Kovacs, "Ballistocardiography; a method worth revisiting," in *IEEE Conference on Engineering in Medicine and Biology Society*, Aug 2011, pp. 4279–4282.