BLIND DECONVOLUTION OF SPARSE BUT FILTERED PULSES WITH LINEAR STATE SPACE MODELS

Nour Zalmai, Hampus Malmberg, and Hans-Andrea Loeliger

ETH Zurich Dept. of Information Technology & Electrical Engineering {zalmai, malmberg, loeliger}@isi.ee.ethz.ch

ABSTRACT

The paper considers the problem of joint system identification and input signal estimation of an unknown linear system from noisy observations of the output signal. The input signal is assumed to be sparse, and each individual input pulse may affect the system in its own (and unknown) way. Based on ideas from sparse Bayesian learning, we derive an efficient expectation maximization (EM) algorithm for jointly estimating all unknown quantities. Unlike related prior work, the proposed algorithm does not alternate between estimating the input signal and estimating the system parameters; instead, all unknown quantities are jointly updated in each EM step. We give closed-form expressions for these EM updates, which can be efficiently computed by Gaussian message passing.

Index Terms— linear state space models, sparse Bayesian learning, expectation maximization

1. INTRODUCTION

Let $y = (y_1, \ldots, y_K) \in \mathbb{R}^K$ be a given discrete-time signal of duration $K \gg 1$. We wish to "explain" this signal as the output of a linear state space model (LSSM)

$$\begin{cases} X_k = AX_{k-1} + B_k u_k + \epsilon_k \\ y_k = CX_k + Z_k \end{cases},$$
(1)

with states $X_k \in \mathbb{R}^n$ (with $X_0 = 0$), input signal $u = (u_1, \ldots u_K) \in \mathbb{R}^K$, state-transition matrix $A \in \mathbb{R}^{n \times n}$, vectors $B_1, \ldots, B_K \in \mathbb{R}^{n \times 1}$ and $C \in \mathbb{R}^{1 \times n}$, observation noise $Z_k \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_Z^2)$, and state noise $\epsilon_k \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2 I)$. The input signal u and all model parameters $(A, B_1, \ldots, B_K, C, \sigma_Z^2)$, σ_ϵ^2) are unknown and to be estimated from y.

As an essential additional condition, we require our estimate of u to be sparse, i.e., $u_k = 0$ except for a "small" fraction of indices k. (Note also that a scale factor can be moved freely between u_k and B_k , which we will address by suitable normalizations.)

The stated problem is motivated by several applications. For example, we may want to identify an unknown mechanical system based on its response to multiple impulses of unknown timing, strength, and orientation. But the stated problem may also be viewed as parsing the signal y into individual events (where $u_k \neq 0$), each with individual features B_k , as a first step in some pattern analysis task (e.g., recovering musical scores from an audio file).

The stated problem may be viewed as a special case of the following problem: for given $y \in \mathbb{R}^K$, minimize $||y - Hu||^2$ (with $H \in \mathbb{R}^{K \times K}$) by some sparse $u \in \mathbb{R}^K$. If H is known, this is a classical compressive sensing problem [1]; if H is not known, we have a dictionary learning problem. A well-established approach for the latter (cf. [2–4]) consists in alternating between estimating u for fixed H and estimating H for fixed u. However, the methods for general H are not well suited for our problem where only one measurement vector y is available and the dictionary is strongly coherent (proper regularization on H must be introduced, cf. [5, 6]). In addition, those methods often require substantial computational power since each update has a complexity polynomial in K.

The problem of minimizing $||y - Hu||^2$ by some sparse vector u has also been addressed in a Bayesian setting. MAP estimation of u with a sparse prior was used in [7] for blind source separation with a fixed signal dictionary. More relevant for this paper are sparse Bayesian learning techniques [8], also known as automatic relevance determination (ARD) [9,10]. ARD uses compressible priors with hyper-parameters, which are determined by maximum-likelihood estimation, resulting in an appealing regularization term [11]. In [12], each element of u and H has its own hyper-parameter. The posterior densities are usually intractable, but amenable to approximative inference using variational Bayesian techniques.

In this paper, building on ideas from [6], we use the basic idea of ARD and derive an expectation maximization (EM) algorithm to simultaneously estimate all unknown parameters. Unlike related prior work (including [6]), we do not iterate between estimating the input signal u and estimating the system parameters: instead, all unknown quantities are jointly updated in every EM iteration. The computational complexity is $O(n^2K)$ per EM iteration, due to efficient computations of the required quantities by Gaussian message passing and new closed-form expressions for the EM updates.

2. LOCAL MAXIMA AT SPARSE SOLUTIONS

As stated in the introduction, we want to estimate the input vector u and the system model parameters in (1) under the constraint that u is sparse, i.e., $||u||_0 \ll K$. Sparse Bayesian technique considers U as a random vector with ARD parameters $\sigma_U^2 = (\sigma_{U_1}^2, \ldots, \sigma_{U_K}^2)$ such that

$$p(u|\sigma_U^2) = \prod_{k=1}^K \frac{1}{\sqrt{2\pi\sigma_{U_k}^2}} \exp\left(-\frac{u_k^2}{2\sigma_{U_k}^2}\right) , \qquad (2)$$

i.e., $U_k \sim \mathcal{N}(0, \sigma_{U_k}^2)$. Then, parameter estimation is performed by maximizing the type-II likelihood (cf. [13])

$$L(\theta) = p(y|\theta) = \int p(y|\theta, u)p(u|\theta) \,\mathrm{d}u \,, \tag{3}$$

with $\theta = (C, A, B_1, \dots, B_K, \sigma_Z^2, \sigma_\epsilon^2, \sigma_U^2)$ and from (1)

$$p(y|\theta, u) = \int \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{1}{2\sigma_Z^2}(y_k - Cx_k)^2} \cdot \frac{1}{(2\pi\sigma_\epsilon^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_\epsilon^2}\|x_k - (Ax_{k-1} + B_k u_k)\|^2} \, \mathrm{d}x \,.$$
(4)

Let $\hat{\theta}$ be a local maximum of $L(\theta)$ such that $||\hat{B}_k|| = 1$ for all k. Note that $L(\theta)$ is insensitive to moving a scale factor between B_k and σ_{U_k} . Using the notation of Section 5, let $\tilde{\mu}_{X_k} = \vec{m}_{X_k} - \vec{m}_{X_k}$ and $Q_k = (AF_{k-1}\vec{V}_{X_{k-1}}A^{\mathsf{T}} + \vec{V}_{X_k} + \hat{\sigma}_{\epsilon}^2 I)^{-1}$ be quantities obtained by Gaussian message passing in the factor graph in Fig. 1 of the system model (1) with $\hat{\theta}$ pluggedin (cf. [14, 15]). Then, denoting $\hat{\theta}_k$ all the parameters in $\hat{\theta}$ except \hat{B}_k and $\hat{\sigma}_{U_k}^2$, the marginal log-likelihood with respect to $(\sigma_{U_k}^2, B_k)$ can be expressed as

$$2\ln p(y|\theta_k, B_k, \sigma_{U_k}^2) = \frac{\sigma_{U_k}^2 (B_k^{\mathsf{T}} Q_k \tilde{\mu}_{X_k})^2}{1 + \sigma_{U_k}^2 B_k^{\mathsf{T}} Q_k B_k} - \ln(1 + \sigma_{U_k}^2 B_k^{\mathsf{T}} Q_k B_k) + \alpha_k ,$$
(5)

with α_k independent of $\sigma_{U_k}^2$ and B_k . Thus, at a local maximum, $p(y|\hat{\theta}_k, B_k, \sigma_{U_k}^2)$ is maximum, which leads to:

Lemma 1 If

$$\tilde{\mu}_{X_k}^{\mathsf{T}} Q_k \tilde{\mu}_{X_k} > 1 , \qquad (6)$$

then

$$\hat{\sigma}_{U_k}^2 = \left(1 - \frac{1}{\tilde{\mu}_{X_k}^{\mathsf{T}} Q_k \tilde{\mu}_{X_k}}\right) \|\tilde{\mu}_{X_k}\|^2 > 0 \tag{7}$$

and

$$\hat{B}_k = \frac{\tilde{\mu}_{X_k}}{\|\tilde{\mu}_{X_k}\|} \,. \tag{8}$$

Else, $\hat{\sigma}_{U_k}^2 = 0$ and \hat{B}_k is any vector such that $\|\hat{B}_k\| = 1$.



Fig. 1. Factor graph representation of the LSSM

An input (i.e., $\hat{\sigma}_{U_k}^2 > 0$) is introduced only to compensate a substantial error $\tilde{\mu}_{X_k}^{\mathsf{T}} Q_k \tilde{\mu}_{X_k}$ between the forward and backward state estimates. Indeed, unlike the backward state estimate \overleftarrow{m}_{X_k} that is aware of the presence of an input at time index k, the forward state estimate \overrightarrow{m}_{X_k} has no evidence of this specific input. The discrepancy $\widetilde{\mu}_{X_k} = \overrightarrow{m}_{X_k} - \overleftarrow{m}_{X_k}$ is considerable if an input actually triggers at k. Thus, we expect that the inequality in (6) holds for few k's only and thus a lot of $\hat{\sigma}_{U_k}^2$'s are zero. Consequently, at a local maximum of the likelihood $L(\theta)$, $\hat{\sigma}_U^2$ should be sparse.

In the following, we propose two versions of EM for a joint estimation of the model and ARD parameters, one with and one without state transition matrix estimation. As EM is guaranteed to converge to a local maximum, many components of $\hat{\sigma}_U^2$ are expected to converge to zero.

3. JOINT EM ALGORITHM: UNKNOWN A

Considering both $U = (U_1, \ldots, U_K)$ and $X = (X_0, \ldots, X_K)$ as hidden variables, the EM algorithm consists in iteratively computing $\mathbb{E} [\ln p(y, U, X | \theta)]$ with respect to the joint density $p(u, x | y, \hat{\theta})$ and updating the parameters according to

$$\hat{\theta} = \operatorname*{argmax}_{\theta} \mathbb{E}\left[\ln p(y, U, X|\theta)\right] \,. \tag{9}$$

We provide the update formulae for all parameters in $\theta = (C, A, B_1, \dots, B_K, \sigma_Z^2, \sigma_{\epsilon}^2, \sigma_U^2)$ even if some of them might be fixed a priori. The joint density can be expressed as

$$-2\ln p(y, u, x|\theta) = \sum_{k=1}^{K} \left(\frac{(y_k - Cx_k)^2}{\sigma_Z^2} + \ln(2\pi\sigma_Z^2) + \frac{u_k^2}{\sigma_{U_k}^2} + \ln(2\pi\sigma_{U_k}^2) + \frac{\|x_k - Ax_{k-1} - B_k u_k\|^2}{\sigma_\epsilon^2} + n\ln(2\pi\sigma_\epsilon^2) \right).$$
(10)

Looking at (10) the optimization problem in (9) splits for $(C, \sigma_Z^2), (A, B_1, \ldots, B_K, \sigma_\epsilon^2)$ and each individual $\sigma_{U_k}^2$.

For $k \in \{1, \ldots, K\}$, we have

$$\hat{\sigma}_{U_k}^2 = \operatorname*{argmin}_{\sigma_{U_k}^2} \frac{1}{\sigma_{U_k}^2} \mathbb{E}[U_k^2] + \ln\left(2\pi\sigma_{U_k}^2\right) = \mathbb{E}[U_k^2] \quad (11)$$

Denoting $\kappa = \sum_{k=1}^{K} y_k^2$, $\xi_C = \sum_{k=1}^{K} \mathbb{E}[X_k] y_k$, and $V_C = \sum_{k=1}^{K} \mathbb{E}[X_k X_k^{\mathsf{T}}]$, we have

$$\hat{C} = \underset{C}{\operatorname{argmin}} \kappa - 2C\xi_C + CV_C C^{\mathsf{T}} = \xi_C^{\mathsf{T}} V_C^{-1}$$
(12)

$$\hat{\sigma}_{Z}^{2} = \frac{1}{K} \sum_{k=1}^{K} \mathbb{E} \Big[(y_{k} - \hat{C}X_{k})^{2} \Big] = \frac{1}{K} \left(\kappa - \xi_{C}^{\mathsf{T}} V_{C}^{-1} \xi_{C} \right).$$
(13)

Then, for the remaining parameters we get

$$\hat{B}_{k} = \frac{\mathbb{E}[U_{k}X_{k}] - \hat{A}\mathbb{E}[U_{k}X_{k-1}]}{\mathbb{E}[U_{k}^{2}]} , \ k \in \{1, \dots, K\}$$
(14)

$$\hat{\sigma}_{\epsilon}^{2} = \frac{1}{nK} \sum_{k=1}^{K} \mathbb{E} \Big[\|X_{k} - \hat{A}X_{k-1} - \hat{B}_{k}U_{k}\|^{2} \Big]$$
(15)

$$\hat{A} = \underset{A}{\operatorname{argmin}} \operatorname{Tr} \left(A V_A A^{\mathsf{T}} - 2A \xi_A \right) , \qquad (16)$$

with

$$V_{A} = \sum_{k=1}^{K} \mathbb{E} \left[X_{k-1} X_{k-1}^{\mathsf{T}} \right] - \frac{\mathbb{E} \left[U_{k} X_{k-1} \right] \mathbb{E} \left[U_{k} X_{k-1} \right]^{\mathsf{T}}}{\mathbb{E} \left[U_{k}^{2} \right]}$$
(17)
$$\xi_{A} = \sum_{k=1}^{K} \mathbb{E} \left[X_{k-1} X_{k}^{\mathsf{T}} \right] - \frac{\mathbb{E} \left[U_{k} X_{k-1} \right] \mathbb{E} \left[U_{k} X_{k} \right]^{\mathsf{T}}}{\mathbb{E} \left[U_{k}^{2} \right]} .$$
(18)

The cost function (16) is a quadratic form in A and closed-form expressions can be derived for

- $A \in \mathbb{R}^{n \times n}$: $\hat{A} = \xi_A^\mathsf{T} V_A^{-1}$
- A in controllable or observable canonical form
- A in block-Jordan form

Thus, all the parameters in θ can be updated jointly with closed-form expressions. In addition, all the expectation quantities can be efficiently computed using Gaussian message passing as described in Section 5.

We conclude this section with the following remarks.

- 1. The observation noise variance σ_Z^2 controls the sparsity level and should be fixed a priori. It also sets the desired accuracy of the reconstructed signal.
- 2. The state noise variance σ_{ϵ}^2 corresponds to the LSSM mismatch and should converge to a low value. At the same time its value defines the threshold deciding whether to introduce an input or not. The initial value of σ_{ϵ}^2 must be chosen big enough to allow changes of the system model.
- 3. For numerical stability (i.e., avoiding quantities to grow to infinity), after each EM iteration, we apply the following rescaling: $B_k \leftarrow \sigma_{U_k} B_k$ and $\sigma_{U_k} \leftarrow 1$.
- During the EM algorithm, many elements of (σ_{Uk} · ||B_k||)_{k∈{1,...,K}} will converge to zero but will typically not become exact zeros. To obtain exact zeros, a

final hard update (e.g., as in Section 4 or the marginal likelihood update in (7)) can be performed.

- Restrictions of the admissible input vectors, i.e., B_k ∈ B ⊂ ℝⁿ can be incorporated in the EM update at the price of potentially giving up the closed-form expressions for both B_k and A. In case of linear constraints, closed-form expressions can still be derived.
- 6. In case of $B_1 = \cdots = B_K$, the EM update can easily be adapted.

4. JOINT EM ALGORITHM: FIXED A

In this section, we assume that the matrix A is known. Instead of adapting the algorithm of Section 3 accordingly (which is straightforward), we here consider an interesting alternative where we use EM with only X as hidden variable. Recall that the cost function is insensitive to moving a scale factor between σ_{U_k} and B_k . Here we enforce $||B_k|| = 1$ which simplifies the formulae without changing the overall estimation problem. After marginalizing over u, we have

$$-2\ln p(y, x|\theta) = \sum_{k=1}^{K} \frac{1}{\sigma_Z^2} (y_k - Cx_k)^2 + \ln(2\pi\sigma_Z^2) + \frac{1}{\sigma_\epsilon^2} \|x_k - Ax_{k-1}\|^2 + n\ln(2\pi\sigma_\epsilon^2) - \frac{\sigma_{U_k}^2 (B_k^{\mathsf{T}}(x_k - Ax_{k-1}))^2}{\sigma_\epsilon^2 (\sigma_\epsilon^2 + \sigma_{U_k}^2)} + \ln\left(\frac{\sigma_\epsilon^2 + \sigma_{U_k}^2}{\sigma_\epsilon^2}\right).$$
(19)

The EM algorithm consists in computing $\mathbb{E} \left[\ln p(y, X | \theta) \right]$ with respect to the density $p(x|y, \hat{\theta})$ and updating the parameters according to

$$\hat{\theta} = \operatorname*{argmax}_{\theta} \mathbb{E}\left[\ln p(y, X|\theta)\right] \,. \tag{20}$$

The updates for C and σ_Z^2 are similar to (12) and (13). For the input vectors, we have

$$\hat{B}_k = \operatorname*{argmax}_{\|B_k\|=1} \mathbb{E}\left[\left(B_k^\mathsf{T} (X_k - A X_{k-1}) \right)^2 \right] \,.$$
(21)

Thus, \hat{B}_k is the unit eigenvector corresponding to the maximum eigenvalue λ_k of $\mathbb{E}\left[(X_k - AX_{k-1})(X_k - AX_{k-1})^{\mathsf{T}}\right]$. Then, the updates for $\sigma_{U_k}^2$ and σ_{ϵ}^2 are

$$\hat{\sigma}_{U_k}^2 = \max\left(0, \lambda_k - \hat{\sigma}_{\epsilon}^2\right) \tag{22}$$

$$\hat{\sigma}_{\epsilon}^{2} = \operatorname*{argmin}_{\sigma_{\epsilon}} \frac{M_{A}}{\sigma_{\epsilon}^{2}} + nK \ln(\sigma_{\epsilon}^{2}) + \sum_{\lambda_{k} > \sigma_{\epsilon}^{2}} -\frac{\lambda_{k} - \sigma_{\epsilon}^{2}}{\sigma_{\epsilon}^{2}} + \ln\left(\frac{\lambda_{k}}{\sigma_{\epsilon}^{2}}\right) , \qquad (23)$$

where $M_A = \sum_{k=1}^{K} \mathbb{E} \left[\|X_k - AX_{k-1}\|^2 \right]$. This last optimization problem can be solved by considering the K + 1 sub-problems of restricting σ_{ϵ}^2 in the interval of consecutive eigenvalues ranked in decreasing order and selecting the optimum. Unlike the joint EM algorithm of Section 3, the update of $\hat{\sigma}_{U_k}^2$ can create exact zeros.

5. STABLE AND EFFICIENT MESSAGE PASSING

All required quantities for EM can be efficiently computed by Gaussian message passing in the factor graph of Fig. 1. The update rule parameterization must lead to a stable implementation while σ_{ϵ}^2 and any $\sigma_{U_k}^2$ tend to or are exactly zero. The modified Bryson-Frazier smoother is a suitable choice of update rule and in addition avoids matrix inversions. It proceeds in two sweeps (cf. [6]). The forward pass for $k = 1, \ldots, K$

$$\overrightarrow{m}_{X_k} = A(\overrightarrow{m}_{X_{k-1}} + (y_{k-1} - C\overrightarrow{m}_{X_{k-1}})g_{k-1}\overrightarrow{V}_{X_{k-1}}C^{\mathsf{T}}) (24)$$

$$V_{X_{k}} = \sigma_{U_{k}}^{2} B_{k} B_{k}^{\dagger} + \sigma_{\epsilon}^{2} I + A F_{k-1} V_{X_{k-1}} A^{\dagger}$$
(25)

$$g_k = (\sigma_Z^2 + C V_{X_k} C^{\mathsf{T}})^{-1} \tag{26}$$

$$F_k = I - g_k \overrightarrow{V}_{X_k} C^\mathsf{T} C \,, \tag{27}$$

with initialization \overrightarrow{m}_{X_0} , \overrightarrow{V}_{X_0} , and y_0 all zeros. Then, the backward pass for $k = K, \dots, 1$

$$\tilde{W}_{X_k}\tilde{\mu}_{X_k} = F_k^{\mathsf{T}}A^{\mathsf{T}}\tilde{W}_{X_{k+1}}\tilde{\mu}_{X_{k+1}} - g_k(y_k - C\overrightarrow{m}_{X_k})C^{\mathsf{T}}$$
(28)
$$\tilde{W}_{X_k} = F_k^{\mathsf{T}}A^{\mathsf{T}}\tilde{W}_{X_{k+1}}AF_k + g_kC^{\mathsf{T}}C ,$$
(29)

initialized with $\tilde{W}_{X_{K+1}} = 0$ and $\tilde{W}_{X_{K+1}}\tilde{\mu}_{X_{K+1}} = 0$. The posterior and joint posterior of X_k , U_k , and X_{k-1} are all Gaussian and thus characterized by their mean and covariance matrix. The posterior density of X_k is characterized by

$$m_{X_k} = \overrightarrow{m}_{X_k} - \overrightarrow{V}_{X_k} \widetilde{W}_{X_k} \widetilde{\mu}_{X_k}$$
(30)

$$V_{X_k} = \vec{V}_{X_k} (I - \tilde{W}_{X_k} \vec{V}_{X_k}) .$$
(31)

The joint posterior of X_{k-1} and X_k is characterized by

$$V_{X_{k-1},X_{k}^{\mathsf{T}}} = F_{k-1}\overrightarrow{V}_{X_{k-1}}A^{\mathsf{T}}(I - \widetilde{W}_{X_{k}}\overrightarrow{V}_{X_{k}}) \quad (32)$$

$$\mathbb{E}[X_{k-1}X_k^{\mathsf{T}}] = V_{X_{k-1},X_k^{\mathsf{T}}} + m_{X_{k-1}}m_{X_k}^{\mathsf{T}} .$$
(33)

The posterior density of U_k is parameterized by

$$m_{U_k} = -\sigma_{U_k}^2 B_k^\mathsf{T} \tilde{W}_{X_k} \tilde{\mu}_{X_k} \tag{34}$$

$$V_{U_k} = \sigma_{U_k}^2 - (\sigma_{U_k}^2)^2 B_k^{\mathsf{T}} \tilde{W}_{X_k} B_k$$
(35)

$$\frac{\mathbb{E}[U_k^2]}{\sigma_{U_k}^2} = 1 + \sigma_{U_k}^2 \left((B_k^\mathsf{T} \tilde{W}_{X_k} \tilde{\mu}_{X_k})^2 - B_k^\mathsf{T} \tilde{W}_{X_k} B_k \right) .$$
(36)

The expected correlations can be expressed as

$$\mathbb{E}[U_k X_{k-1}] / \sigma_{U_k}^2$$

= $-F_{k-1} \overrightarrow{V}_{X_{k-1}} A^\mathsf{T} \widetilde{W}_{X_k} B_k - B_k^\mathsf{T} \widetilde{W}_{X_k} \widetilde{\mu}_{X_k} m_{X_{k-1}}$ (37)

$$\mathbb{E}[U_k X_k] / \sigma_{U_k}^2$$

= $(I - \overrightarrow{V}_{X_k} \widetilde{W}_{X_k}) B_k - B_k^\mathsf{T} \widetilde{W}_{X_k} \widetilde{\mu}_{X_k} m_{X_k} .$ (38)

6. EXPERIMENTAL RESULTS

To demonstrate the pertinence of the algorithms, we conduct series of simulations on signals generated using the system



Fig. 2. Observed and estimated signal (after 300 iterations)

model (1) with $K = 10^4$, n = 24, $\sigma_{\epsilon}^2 = 0$, (C, A) such that each input generates a sum of exponentially damped cosines of different attenuation and frequency as illustrated in Fig. 2. The input vector is randomly generated such that the sparsity level (i.e., $||u||_0/K$) is 0.12% and each non-zero component is uniformly drawn from [0.75, 1.25]. The simulations are repeated 100 times at 3 different observation noise variances $\sigma_N^2 \in \{10^{-4}, 10^{-3}, 10^{-2}\}$. Subsequently, the desired sparsity level is controlled by tuning the parameter σ_Z^2 of the algorithm (or equivalently the ratio $\eta = \sigma_Z^2/\sigma_N^2$).

Tables 1 and 2 show the estimated sparsity level (to be compared with 0.12%) and the percentage of correctly detected inputs for the algorithms of Sections 3 and 4 respectively, after 1000 iterations (and a final hard update). We observe that a bigger η generally enforces a sparser estimate. The shaded values emphasize the regimes of interest where the algorithm outputs both a good detection ability and a sparsity level close to the actual one. Note that to achieve such accuracy the estimated system model has to be sufficiently good. A finer tuning of η would improve the estimates.

However, note that the hatched values in Table 1 emphasize cases where the algorithm must output a better data fit than expected (i.e., $\sigma_Z^2 \ll \sigma_N^2$). Thus, $\hat{A} \approx 0$ and the estimated input cannot be sparse. As a result, the final hard update fails and these results must be interpreted with caution.

$\eta_{\sigma_N^2}$	10^{-2}		10^{-1}		1		3		10	
10^{-4}	4.6	0.01	13.5	0.02	71.4	0.09	100	0.13	96.9	0.12
10^{-3}	5.2	0.01	90.6	0.24	99.0	0.18	83.6	0.11	90.9	0.12
10^{-2}	100	7.15	100	6.56	78.9	0.68	30.0	0.06	26.7	0.04

Table 1. Estimated sparsity level (blue, right) and correctly detected inputs (black), in % for A unknown, cf. Section 3

$\eta_{\sigma_N^2}$	10^{-1}		1		3		10		10^{2}	
10^{-4}	99.8	0.12	99.8	0.12	99.9	0.12	99.7	0.12	96.6	0.12
10^{-3}	99.8	2.47	99.2	1.20	98.8	0.18	96.6	0.12	88.8	0.11
10^{-2}	98.1	2.07	98.4	1.65	95.2	0.60	89.7	0.11	30.8	0.04

Table 2. Estimated sparsity level (blue, right) and correctly detected inputs (black), in % for A fixed, cf. Section 4

7. REFERENCES

- E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [2] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, 2003.
- [3] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Computer Vision–ECCV 2012*, pp. 341–355. Springer, 2012.
- [4] 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.
- [5] 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 (CVPR)*. IEEE, 2009, pp. 1964– 1971.
- [6] L. Bruderer, H. Malmberg, and H.-A. Loeliger, "Deconvolution of weakly-sparse signals and dynamicalsystem identification by Gaussian message passing," in *IEEE International Symposium on Information Theory* (*ISIT*), 2015.
- [7] 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.
- [8] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [9] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination," in Advances in Neural Information Processing Systems, 2008, pp. 1625–1632.
- [10] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [11] D. P. Wipf, B. D. Rao, and S. S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6236–6255, 2011.
- [12] V. Šmídl and O. Tichỳ, "Sparsity in Bayesian blind source separation and deconvolution," in *Machine Learning and Knowledge Discovery in Databases*, pp. 548–563. Springer, 2013.

- [13] J. O. Berger, Statistical decision theory and Bayesian analysis, Series in statistics. Springer, second edition, 1985.
- [14] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, Li 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.
- [15] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, 2004.