SMLR-Type Blind Deconvolution of Sparse Pulse Sequences Under a Minimum Temporal Distance Constraint

Georg Kail, *Member, IEEE*, Franz Hlawatsch, *Fellow, IEEE*, and Clemens Novak

***Abstract*—We consider Bayesian blind deconvolution (BD) of an unknown sparse sequence convolved with an unknown pulse. Our goal is to detect the positions where the sparse input sequence is nonzero and to estimate the corresponding amplitudes as well as the pulse shape. For this task, we propose a novel evolution of the *single most likely replacement* (SMLR) algorithm. Our method uses a modified Bernoulli-Gaussian prior that incorporates a minimum temporal distance constraint. This prior simultaneously induces sparsity and enforces a prescribed minimum distance between the pulse centers. The minimum distance constraint provides an effective way to avoid overfitting (i.e., spurious detected pulses) and improve resolution. The proposed BD method overcomes certain weaknesses of the traditional SMLR-based BD method, which is verified experimentally to result in improved detection/estimation performance and reduced computational complexity. Our simulation results also demonstrate performance and complexity advantages relative to the iterated window maximization (IWM) algorithm and a recently proposed partially collapsed Gibbs sampler method.**

***Index Terms*—Bayesian blind deconvolution, Bernoulli-Gaussian prior, iterated window maximization (IWM) algorithm, single most likely replacement (SMLR) algorithm, sparse deconvolution.**

I. Introduction

THE problem of blind deconvolution (BD) arises in various fields including digital communications [1]–[5], seismology [6]–[9], biomedical signal processing [10]–[13], and astronomy [14], [15]. Because the result of BD is inherently nonunique, additional assumptions or constraints—such as monotonicity [16], positivity [17]–[19], and sparsity [20]–[23]—are typically used. In this paper, we study BD under a *combined sparsity and minimum distance constraint* as introduced recently in [24]. In the setting considered, a sparse random sequence is convolved with a pulse of unknown shape. The (few) nonzero entries of the sparse sequence thus mark the positions and weights (amplitudes) of potentially overlapping replicas of the pulse in the observed sequence. The temporal distances between these positions are constrained to be not smaller than some prescribed minimum distance. This minimum distance may be smaller than the (effective) pulse length, thus allowing for overlap between successive pulses. Such a minimum distance constraint is physically relevant and appropriate in many applications, including layer detection [12], medical imaging [10], [13], seismology [8], and multipath parameter estimation [25], [26]. For example, various biomedical signals, such as in electrocardiography and electromyography, contain pulses whose centers cannot have arbitrarily small time separation. In scenarios where the pulses correspond to reflections from layer boundaries, such as in seismology or optical coherence tomography, there may be a lower limit on the thickness of layers. Typically, signals that satisfy a minimum distance constraint are also sparse. A minimum distance constraint was considered in our previous work on BD [24]. For nonblind deconvolution, theoretical implications of a minimum distance were studied in [27].

Our goal is to detect the positions where the sparse input sequence is nonzero and to estimate the corresponding amplitudes as well as the pulse shape. For this task, we adopt a Bayesian setting because of the resulting ease of performing calculations (both formally and algorithmically) and the possibility of introducing prior information and constraints via the choice of a prior distribution [28]. More specifically, we use a modified Bernoulli-Gaussian prior incorporating the minimum distance constraint. The basic Bernoulli-Gaussian model is well established as a convenient means to model sparsity [7], [8], [20], [23], [29]–[31]. It yields an intuitive relation between the sparsity of a sequence and its prior probability and at the same time often helps make computations tractable. As previously observed in [24], augmenting the Bernoulli-Gaussian model by a minimum distance constraint provides an effective way to avoid overfitting and improve resolution. Whereas a partially collapsed Gibbs sampler [32] based method was used in [24] for detection/estimation, here we propose a different method that is inspired by the *single most likely replacement* (SMLR) algorithm [29]. As we will demonstrate, the proposed method can outperform the method of [24] with respect to both performance and computational efficiency.

The SMLR algorithm is an iterative technique that is computationally efficient but may converge to a local maximum of

Manuscript received July 15, 2014; revised December 11, 2014 and April 13, 2015; accepted April 18, 2015. Date of publication June 09, 2015; date of current version August 13, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Adel Belouchrani. This work was supported by the Austrian Science Fund (FWF) under grants S10603, J3495, and P27370. This work was done while C. Novak was with the Institute of Telecommunications, Vienna University of Technology.

G. Kail and F. Hlawatsch are with Institute of Telecommunications, Vienna University of Technology, A-1040 Vienna, Austria (e-mail: gkail@nt.tuwien.ac.at; fhlawats@nt.tuwien.ac.at).

C. Novak is with Frequentis AG, A-1100 Vienna, Austria (e-mail: clemens.novak@frequentis.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2442951

the posterior distribution rather than the global one. In its original form [29], it is suited for *nonblind* Bayesian deconvolution, i.e., the pulse shape is supposed known. Various modifications, extensions, and applications of the SMLR algorithm have been presented in [31], [33]–[40]; however, none of them considers a minimum distance constraint. In particular, the SMLR algorithm was used for BD in [33] (for seismic signals) and in [34] (for discrete-valued sparse signals). A powerful generalization of the SMLR algorithm is given by the *iterated window maximization* (IWM) algorithm [36]. The performance and complexity of the IWM algorithm strongly depend on the choice of hypothesis sets and of a tuning parameter that influences the sparsity of the deconvolution result. In contrast to this tuning parameter, the hyperparameters of the method proposed here have an intuitive interpretation, which may make it easier to choose them appropriately.

The proposed method is different from existing methods in that it is tailored to the minimum distance constraint. In contrast to classical SMLR, which often fails in the presence of such a constraint, it exploits the constraint to achieve good performance, computational efficiency, and sparse results. Computational efficiency is further enhanced by choosing the order of the algorithmic steps differently from existing SMLR-type methods. Finally, our method allows for prior assumptions about the time and frequency localization of the unknown pulse shape, which leads to improved performance. To make such assumptions possible, the method includes the explicit determination of an optimal amplitude scale and time shift.

This paper is organized as follows. In Section II, we present the signal model, our choice of prior distributions, and the resulting posterior distribution. In Section III, we propose a modified SMLR algorithm that incorporates the minimum distance constraint. Section IV describes an efficient implementation of this algorithm following [30], [31]. Section V presents a BD method that uses the modified SMLR algorithm. Simulation results assessing the performance and complexity of the proposed BD method are presented in Section VI.

II. Signal Model and Parameter Priors

The signal model and prior distributions are essentially as in [24]; we briefly describe them here for the sake of a self-contained exposition.

A. Signal Model

The observed sequence is given by

$$x_k = \sum_{l=1}^{K} f_{k-l} a_l + n_k, \quad k = 1, \ldots, K, \tag{1}$$

where a_k, $k \in \{1, \ldots, K\}$ is an unknown sparse complex sequence of length K, f_k is an unknown complex pulse (defined to be zero outside $\{-Q, \ldots, Q\}$, where typically $Q \ll K$), and n_k is independent and identically distributed (iid) circularly symmetric complex Gaussian noise with unknown variance σ_n^2. Using $\mathbf{x} \triangleq (x_1 \; x_2 \cdots x_K)^{\mathrm{T}}$, $\mathbf{a} \triangleq (a_1 \; a_2 \cdots a_K)^{\mathrm{T}}$, $\mathbf{f} \triangleq (f_{-Q} \; f_{-Q+1} \cdots f_Q)^{\mathrm{T}}$, and $\mathbf{n} \triangleq (n_1 \; n_2 \cdots n_K)^{\mathrm{T}}$, the signal model (1) can be written as[1]

$$\mathbf{x} = \mathbf{F}\mathbf{a} + \mathbf{n}, \tag{2}$$

where $\mathbf{F} = \mathrm{toep}(\mathbf{f})$ denotes the $K \times K$ Toeplitz matrix that has $(f_0 \; f_1 \; \cdots f_Q \; 0 \; \cdots \; 0)^{\mathrm{T}}$ as its first column and $(f_0 \; f_{-1} \cdots f_{-Q} \; 0 \; \cdots \; 0)$ as its first row.

Following [29]–[31], [41], we use a binary indicator sequence $b_k \in \{0, 1\}$ that is 0 if a_k is zero and 1 if a_k is nonzero. We define the indicator vector $\mathbf{b} \triangleq (b_1 \; b_2 \cdots b_K)^{\mathrm{T}} \in \{0, 1\}^K$, and denote by L the number of nonzero b_k (equivalently, nonzero a_k); note that $L = \sum_{k=1}^{K} b_k = \sum_{k=1}^{K} b_k^2 = \|\mathbf{b}\|^2$. We can write (2) as

$$\mathbf{x} = \mathbf{F}_b \mathbf{a}_b + \mathbf{n}, \tag{3}$$

where $\mathbf{F}_b \in \mathbb{C}^{K \times L}$ is obtained from $\mathbf{F}$ by removing all columns k such that $b_k = 0$ and $\mathbf{a}_b \in \mathbb{C}^L$ contains the corresponding a_k, i.e., the nonzero entries of $\mathbf{a}$. The *minimum distance constraint* postulates that $\mathbf{b} \in \mathcal{C}$, where $\mathcal{C} \subseteq \{0, 1\}^K$ is the set of all length-K binary sequences $\mathbf{b}$ such that the temporal distance of any two nonzero entries $b_k = 1$ and $b_{k'} = 1$ (equivalently, of any $a_k, a_{k'} \neq 0$) satisfies $|k - k'| \geq d_{\min}$, for a given $d_{\min}$. Thereby, the set of admissible sequences $\mathbf{b}$ is significantly reduced ($\mathbf{b} \in \mathcal{C}$ rather than $\mathbf{b} \in \{0, 1\}^K$).

As a parsimonious parametric model for the pulse shape vector $\mathbf{f} \in \mathbb{C}^{2Q+1}$, we use the basis expansion

$$\mathbf{f} = \sum_{m=1}^{M} \gamma_m \mathbf{h}_m = \mathbf{H}\boldsymbol{\gamma}. \tag{4}$$

Here, the basis vectors $\mathbf{h}_m = (h_{m,-Q} \quad h_{m,-Q+1} \cdots h_{m,Q})^{\mathrm{T}}$ are suitably time-scaled versions of the first M Hermite functions [12], [42], [43], which are sampled, truncated, and centered at the $(Q+1)$th entry; $\boldsymbol{\gamma} \triangleq (\gamma_1 \cdots \gamma_M)^T$ is a random coefficient vector; and $\mathbf{H} = (\mathbf{h}_1 \cdots \mathbf{h}_M)$. Typically, $M \ll 2Q+1$. Using (4), the signal model in (2) can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{n} = \mathbf{A}\mathbf{H}\boldsymbol{\gamma} + \mathbf{n}, \tag{5}$$

where $\mathbf{A}$ is the $K \times (2Q+1)$ Toeplitz matrix that has $(a_{Q+1} \; a_{Q+2} \cdots a_K \; 0 \cdots 0)^{\mathrm{T}}$ as its first column and $(a_{Q+1} \; a_Q \cdots \; a_1 \; 0 \cdots 0)$ as its first row.

B. Parameter Priors

Within the Bayesian framework, we have to specify prior probability density functions (pdfs) for the unknown quantities $\mathbf{a}$, $\boldsymbol{\gamma}$, and σ_n^2 to be estimated. For $\mathbf{a}$, we adopt a modified Bernoulli-Gaussian prior that incorporates the minimum distance constraint [24]. Following [30], [31], [41], we specify the prior pdf $p(\mathbf{a})$ by specifying $p(\mathbf{a}|\mathbf{b})$ and $p(\mathbf{b})$. We set

$$p(\mathbf{a}|\mathbf{b}) = \prod_{k=1}^{K} p(a_k|b_k), \tag{6}$$

with (note that $b_k = 0$ implies $a_k = 0$)

$$p(a_k|b_k) = \begin{cases} \delta(a_k), & b_k = 0 \\ \mathcal{CN}\left(a_k; 0, \sigma_a^2\right), & b_k = 1. \end{cases} \tag{7}$$

[1]Our choice to define $\mathbf{x}$ and $\mathbf{a}$ such that they have the same length K is arbitrary and not critical for our method. Depending on the application, it may be more suitable to define $\mathbf{a}$ such that its length is, e.g., either smaller or larger than that of $\mathbf{x}$ by $2Q$.

Here, $\delta(\cdot)$ denotes the Dirac delta function, σ_a^2 is a fixed hyperparameter, and $\mathcal{CN}(\cdot;\mu,\sigma^2)$ denotes the circularly symmetric complex Gaussian pdf with mean μ and variance σ^2. From (6) and (7), we obtain the following conditional prior of $\mathbf{a}_b$:

$$p(\mathbf{a}_b|\mathbf{b}) = \mathcal{CN}\big(\mathbf{a}_b;\mathbf{0},\sigma_a^2\mathbf{I}\big)\,, \tag{8}$$

where $\mathcal{CN}(\cdot;\boldsymbol{\mu},\mathbf{C})$ denotes the multivariate circularly symmetric complex Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Note that $\mathbf{a}_b$ depends on $\mathbf{b}$ (and, in (8), is conditioned on $\mathbf{b}$) due to the fact that its dimension is $L = \|\mathbf{b}\|^2$. Regarding the choice of $p(\mathbf{b})$, the classical Bernoulli-Gaussian model [7], [8], [20], [23], [29]–[31] would be obtained by using $p(\mathbf{b}) = \mathcal{B}(\mathbf{b};\pi_1)$, where $\mathcal{B}(\cdot;\pi_1)$ denotes an iid Bernoulli probability mass function (pmf) with "1-probability" π_1. To incorporate the minimum distance constraint, we replace this unconstrained Bernoulli prior by

$$p(\mathbf{b}) \propto I_{\mathcal{C}}(\mathbf{b})\,\mathcal{B}(\mathbf{b};\pi_1) = I_{\mathcal{C}}(\mathbf{b})\,\pi_1^{\|\mathbf{b}\|^2}(1-\pi_1)^{K-\|\mathbf{b}\|^2}, \tag{9}$$

where $\propto$ means "proportional to," $I_{\mathcal{C}}(\cdot)$ denotes the indicator function of the minimum distance constraint set $\mathcal{C}$ (i.e., $I_{\mathcal{C}}(\mathbf{b})$ is 1 if $\mathbf{b}\in\mathcal{C}$ and 0 otherwise), and π_1 is considered as a fixed hyperparameter. Together, π_1 and $d_{\min}$ determine[2] $\mathrm{E}\{L/K\}$, i.e., the *a priori* mean rate of 1's in $\mathbf{b}$. For $d_{\min} = 1$ (i.e., no minimum distance constraint), the priors (6)–(9) simplify to the classical Bernoulli-Gaussian model.

The prior of $\boldsymbol{\gamma}$ is chosen (differently from [24]) as independent circularly symmetric complex Gaussian, i.e.,

$$p(\boldsymbol{\gamma}) = \mathcal{CN}(\boldsymbol{\gamma};\mathbf{0},\boldsymbol{\Lambda}), \tag{10}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1,\ldots,\lambda_M\}$ with fixed variances λ_m. Here, λ_m decreases with increasing m; because of the time-frequency localization of the Hermite functions [44], [45, p. 26], this implies that the pulse shape $\mathbf{f}$ is likely to be strongly concentrated around the time and frequency origins. This is more plausible and realistic than the pulse shapes obtained for a prior with iid γ_m (as in [24]). Fig. 1(a) illustrates this time-frequency concentration by comparing two examples of the expected squared magnitude of the pulse shape, $\mathrm{E}\{|f_k|^2\} = \sum_{m=1}^{M}\lambda_m|h_{m,k}|^2$, for decreasing and for constant coefficient variances λ_m. Some realizations of f_k are shown in Fig. 1(b), demonstrating that the proposed prior using decreasing λ_m is general enough to allow a wide variety of pulse shapes.

The noise variance σ_n^2 is treated as a random hyperparameter whose prior is chosen as an inverse gamma pdf, i.e.,

$$p\big(\sigma_n^2\big) = \mathcal{IG}\big(\sigma_n^2;\xi,\eta\big) = \frac{\eta^{\xi}}{\Gamma(\xi)}\,\frac{e^{-\eta/\sigma_n^2}}{\sigma_n^{2(\xi+1)}}\,I_{\mathbb{R}^+}\big(\sigma_n^2\big)\,, \tag{11}$$

where $\Gamma(\xi)$ is the gamma function, $I_{\mathbb{R}^+}(\cdot)$ is the unit step function, and ξ and η are fixed hyperparameters. This is a conjugate prior [28, p. 152] of the Gaussian likelihood function to be presented in (12). The same is true for the Gaussian priors $p(\mathbf{a}_b|\mathbf{b})$ in (8) and $p(\boldsymbol{\gamma})$ in (10).

[2]In fact, it can be shown that $\mathrm{E}\{L/K\} \approx \bar{\pi}_1 \triangleq 1/(\varphi^{-K} + d_{\min} - 1/2)$, where φ is the positive real solution of the equation $\varphi^{4d_{\min}} + (4d_{\min}^2/K)\,\varphi^{4d_{\min}+K} = \left(\frac{\pi_1}{1-\pi_1}\right)^{4d_{\min}/K}$.

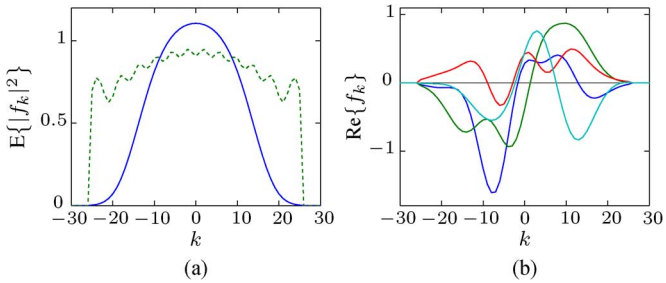

Fig. 1. Construction of the pulse shape f_k: (a) $\mathrm{E}\{|f_k|^2\}$ for decreasing λ_m (solid line) and for constant λ_m (dashed line). (b) Four realizations of f_k for the λ_m corresponding to the solid line in (a) (the real part is shown).

C. Posterior Distribution

Bayesian detection/estimation relies on the posterior distribution of the quantities of interest, i.e., of $\mathbf{a}$ (or, equivalently, $\mathbf{b}$ and $\mathbf{a}_b$), $\boldsymbol{\gamma}$, and σ_n^2. The posterior distribution depends on the likelihood function and the priors [28, p. 9]. The likelihood function of our model is, according to (3), (5), and the iid Gaussian prior for the n_k,

$$p\big(\mathbf{x}|\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2\big) = \mathcal{CN}\big(\mathbf{x};\mathbf{F}_b\mathbf{a}_b,\sigma_n^2\mathbf{I}\big) = \mathcal{CN}\big(\mathbf{x};\mathbf{A}\mathbf{H}\boldsymbol{\gamma},\sigma_n^2\mathbf{I}\big)\,. \tag{12}$$

Assuming that $\mathbf{a}$, $\boldsymbol{\gamma}$, and σ_n^2 are *a priori* independent, the joint posterior distribution of $\mathbf{b}$, $\mathbf{a}_b$, $\boldsymbol{\gamma}$, and σ_n^2 is obtained from the likelihood function and priors as [24]

$$p\big(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big) \propto p\big(\mathbf{x}|\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2\big)\;p(\mathbf{a}_b|\mathbf{b})p(\mathbf{b})p(\boldsymbol{\gamma})\;p\big(\sigma_n^2\big), \tag{13}$$

where the factors on the right hand side are given in (8)–(12). This joint posterior will be used in Section V.

In Section III, we will use the joint posterior with the amplitudes $\mathbf{a}_b$ marginalized out, i.e.,

$$\begin{aligned} p(\mathbf{b},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}) &= \int p\big(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big)\mathrm{d}\mathbf{a}_b \\ &\propto \left[\int p\big(\mathbf{x}|\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2\big)\;p(\mathbf{a}_b|\mathbf{b})\mathrm{d}\mathbf{a}_b\right] p(\mathbf{b})p(\boldsymbol{\gamma})\;p\big(\sigma_n^2\big). \end{aligned}$$

Inserting (12) and (8) and dropping factors that are constant with respect to $\mathbf{b}$, $\boldsymbol{\gamma}$, and σ_n^2, we obtain

$$\begin{aligned} &p\big(\mathbf{b},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big) \\ &\propto \left[\int \frac{1}{\sigma_n^{2K}\sigma_a^{2L}}\exp\left(-\frac{\|\mathbf{x}-\mathbf{F}_b\mathbf{a}_b\|^2}{\sigma_n^2}\right)\exp\left(-\frac{\|\mathbf{a}_b\|^2}{\sigma_a^2}\right)\mathrm{d}\mathbf{a}_b\right] \\ &\qquad\times p(\mathbf{b})\;p(\boldsymbol{\gamma})\;p\big(\sigma_n^2\big) \\ &= \left[\int \exp\left(-\mathbf{a}_b^{\mathrm{H}}\left(\frac{\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b}{\sigma_n^2}+\frac{\mathbf{I}}{\sigma_a^2}\right)\mathbf{a}_b + 2\mathrm{Re}\left\{\mathbf{a}_b^{\mathrm{H}}\frac{\mathbf{F}_b^{\mathrm{H}}\mathbf{x}}{\sigma_n^2}\right\}\right)\mathrm{d}\mathbf{a}_b\right] \\ &\qquad\times \exp\left(-\frac{\|\mathbf{x}\|^2}{\sigma_n^2}\right)\frac{p(\mathbf{b})}{\sigma_a^{2L}}\;p(\boldsymbol{\gamma})\;\frac{p\big(\sigma_n^2\big)}{\sigma_n^{2K}}, \end{aligned}$$

where the superscript $^{\mathrm{H}}$ denotes conjugate transposition. Letting

$$\boldsymbol{\Sigma}_b \triangleq \left(\frac{\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b}{\sigma_n^2}+\frac{\mathbf{I}}{\sigma_a^2}\right)^{-1}, \quad \boldsymbol{\mu}_b \triangleq \frac{\boldsymbol{\Sigma}_b\mathbf{F}_b^{\mathrm{H}}}{\sigma_n^2}\mathbf{x}, \tag{14}$$

we obtain further

$$\begin{aligned} p\big(\mathbf{b},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big) &\propto \left[\int \exp\Big(-\mathbf{a}_b^{\mathrm{H}}\boldsymbol{\Sigma}_b^{-1}\mathbf{a}_b + 2\,\mathrm{Re}\left\{\mathbf{a}_b^{\mathrm{H}}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b\right\}\Big)\mathrm{d}\mathbf{a}_b\right] \\ &\qquad \times \exp\left(-\frac{\|\mathbf{x}\|^2}{\sigma_n^2}\right)\frac{p(\mathbf{b})}{\sigma_a^{2L}}\,p(\boldsymbol{\gamma})\,\frac{p\big(\sigma_n^2\big)}{\sigma_n^{2K}} \\ &= \left[\int \exp\Big(-(\mathbf{a}_b-\boldsymbol{\mu}_b)^{\mathrm{H}}\boldsymbol{\Sigma}_b^{-1}(\mathbf{a}_b-\boldsymbol{\mu}_b)\Big)\mathrm{d}\mathbf{a}_b\right] \\ &\qquad \times \exp\left(\boldsymbol{\mu}_b^{\mathrm{H}}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b - \frac{\|\mathbf{x}\|^2}{\sigma_n^2}\right)\frac{p(\mathbf{b})}{\sigma_a^{2L}}\,p(\boldsymbol{\gamma})\,\frac{p\big(\sigma_n^2\big)}{\sigma_n^{2K}} \\ &\propto |\boldsymbol{\Sigma}_b|\,\exp\left(\boldsymbol{\mu}_b^{\mathrm{H}}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b - \frac{\|\mathbf{x}\|^2}{\sigma_n^2}\right)\frac{p(\mathbf{b})}{\sigma_a^{2L}}\,p(\boldsymbol{\gamma})\,\frac{p\big(\sigma_n^2\big)}{\sigma_n^{2K}}, \quad (15) \end{aligned}$$

where $|\boldsymbol{\Sigma}_b|$ denotes the determinant of $\boldsymbol{\Sigma}_b$.

III. Modified SMLR Algorithm

In this section, we present the core of the proposed detection/estimation method, which is a modification of the SMLR algorithm introduced in [29]. For now, our goal is to detect the binary indicator sequence $\mathbf{b}$ and estimate the amplitudes $\mathbf{a}_b$ while all other parameters are considered fixed and known, i.e., the deconvolution is nonblind. In Section V, we will present an iterative BD method that uses the proposed modified SMLR algorithm.

A straightforward approach to finding $\mathbf{b}$ and $\mathbf{a}_b$ would be joint maximum a posteriori (MAP) detection and estimation, i.e., maximizing the posterior pmf $p(\mathbf{b},\mathbf{a}_b,\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x})$ (see (13)), where $\hat{\boldsymbol{\gamma}}$ and $\widehat{\sigma_n^2}$ denote some fixed estimates. However, [29] discourages joint optimization of $\mathbf{b}$ and $\mathbf{a}_b$ because it often leads to the detection of spurious nonzero indicators b_k with very small amplitudes a_k. As explained in [29], this problem is avoided by first detecting $\mathbf{b}$ based on (15), which is (13) with $\mathbf{a}_b$ marginalized out, and then estimating $\mathbf{a}_b$ by maximizing expression (13) in which the detected $\mathbf{b}$ is inserted. The actual computation is very similar to that of joint MAP optimization of $\mathbf{b}$ and $\mathbf{a}_b$, which is reobtained by simply removing the factor $|\boldsymbol{\Sigma}_b|$ in (15). In the remainder of this section, we will discuss only the detection of $\mathbf{b}$, while the estimation of $\mathbf{a}_b$ is considered in Section V.

We first provide a brief review of the original SMLR algorithm [29]. Our presentation is slightly different from that in [29] because in our signal model (as presented in Section II), $\mathbf{b}$ satisfies the minimum distance constraint, $\mathbf{f}$ has finite length and is represented by a basis expansion, and $\mathbf{a}$, $\mathbf{f}$, and $\mathbf{n}$ are complex. We note that using a finite-length pulse (as proposed in [35]) rather than the ARMA pulse model used in [29] has been common in recent SMLR-type methods because it leads to simpler algorithms and lower complexity.

A. Review of the SMLR Algorithm

SMLR [29] is an efficient algorithm for detecting binary sequences $\mathbf{b}\in\{0,1\}^K$. The guiding principle is MAP detection, which finds the $\mathbf{b}$ maximizing the posterior pmf $p(\mathbf{b},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x})$ (see (15)). The MAP sequence detector

$$\hat{\mathbf{b}}_{\mathrm{MAP}}(\mathbf{x}) \triangleq \underset{\mathbf{b}'\in\{0,1\}^K}{\arg\max}\; p\Big(\mathbf{b}',\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x}\Big)$$

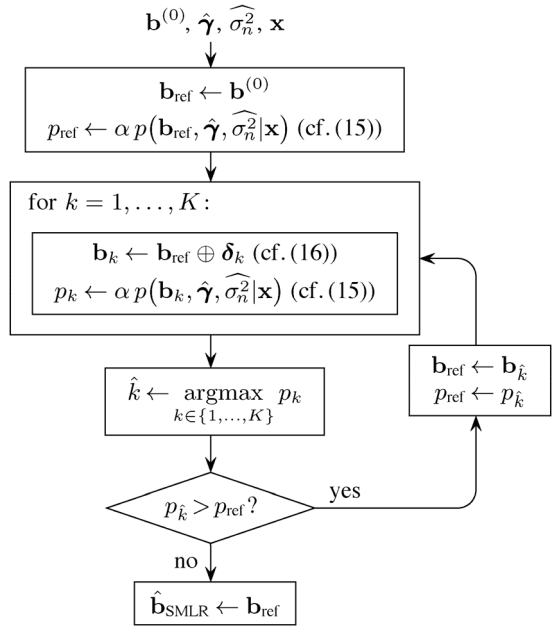

Fig. 2. Flow chart of the classical SMLR detector [29].

is optimal in that it minimizes the probability of a sequence error, $\mathrm{P}\{\hat{\mathbf{b}}\neq\mathbf{b}\}$ [46]. However, it generally amounts to an exhaustive search, requiring the evaluation of $p(\mathbf{b},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x})$ for all the 2^K possible hypotheses $\mathbf{b}\in\{0,1\}^K$ [29]. As an efficient suboptimal alternative, the SMLR detector iteratively computes a $\mathbf{b}$ corresponding to a local maximum of the posterior pmf $p(\mathbf{b},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x})$.

The algorithm is stated in Fig. 2. The initialization (choice of $\mathbf{b}^{(0)}$) will be discussed in the context of BD in Section V. One iteration corresponds to one execution of the loop in Fig. 2. Within each iteration, a reference sequence $\mathbf{b}_{\mathrm{ref}}\in\{0,1\}^K$ obtained from the previous iteration is used to generate K hypotheses

$$\mathbf{b}_k = \mathbf{b}_{\mathrm{ref}} \oplus \boldsymbol{\delta}_k, \quad k = 1,\ldots,K, \qquad (16)$$

where $\oplus$ denotes elementwise modulo-2 addition and $\boldsymbol{\delta}_k$ is the binary sequence of length K that is 1 at position k and zero otherwise. Thus, $\mathbf{b}_k$ differs from $\mathbf{b}_{\mathrm{ref}}$ exactly in the kth entry. For each of the K hypotheses $\mathbf{b}_k$, the posterior pmf is calculated (up to some normalization factor α):

$$p_k \triangleq \alpha\, p\Big(\mathbf{b}_k,\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x}\Big). \qquad (17)$$

Let $\hat{k}\triangleq\arg\max_{k\in\{1,\ldots,K\}}\; p_k$, and consider the corresponding hypothesis, $\mathbf{b}_{\hat{k}}$, and its posterior pmf, $p_{\hat{k}}$. The latter is compared to

$$p_{\mathrm{ref}} \triangleq \alpha\, p\Big(\mathbf{b}_{\mathrm{ref}},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x}\Big). \qquad (18)$$

If $p_{\hat{k}} > p_{\mathrm{ref}}$, the SMLR algorithm proceeds to the next iteration, now using $\mathbf{b}_{\hat{k}}$ as the reference sequence $\mathbf{b}_{\mathrm{ref}}$. Otherwise, $\mathbf{b}_{\mathrm{ref}}$ is a local maximum of $p(\mathbf{b},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2}|\mathbf{x})$, and therefore it is returned as the detected sequence $\hat{\mathbf{b}}_{\mathrm{SMLR}}$. We will also denote the result of the SMLR algorithm as $\hat{\mathbf{b}}_{\mathrm{SMLR}}(\mathbf{x},\mathbf{b}^{(0)},\hat{\boldsymbol{\gamma}},\widehat{\sigma_n^2})$.

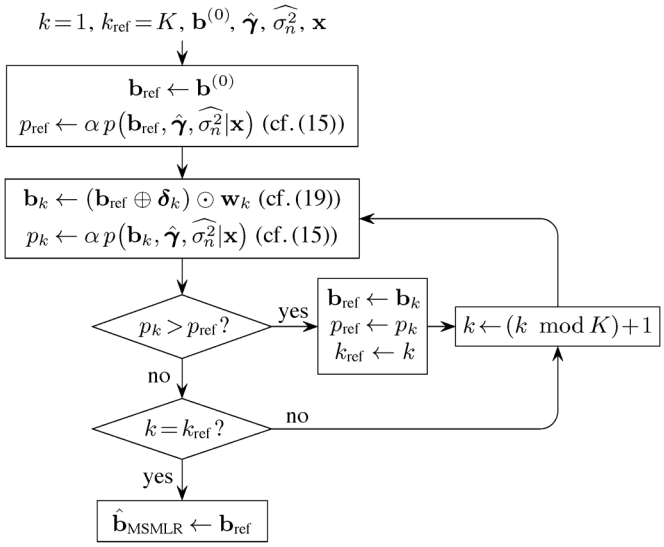

Fig. 3. Flow chart of the proposed MSMLR detector.

B. The Proposed Modified SMLR Algorithm

The proposed modified SMLR (MSMLR) algorithm is presented in Fig. 3. It differs from the original SMLR algorithm in two aspects, which concern the generation of the hypotheses $\mathbf{b}_k$ and the scheduling of different steps within the algorithm. In what follows, we will discuss these two modifications of the SMLR algorithm.

1) First Modification: The first modification is motivated by the minimum distance constraint $\mathbf{b} \in \mathcal{C}$. Differently from [29], our posterior probability $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ enforces the minimum distance constraint through the factor $p(\mathbf{b})$ in (15). Thus, $p_k = 0$ for any $\mathbf{b}_k \notin \mathcal{C}$. If the hypotheses were generated according to (16), we would obtain many hypotheses $\mathbf{b}_k \notin \mathcal{C}$—namely, all $\mathbf{b}_k$ where a new 1 is added within a distance of less than $d_{\min}$ from some 1 in $\mathbf{b}_{\text{ref}}$ (note that $\mathbf{b}_{\text{ref}} \in \mathcal{C}$). More specifically, with $\{k_m\}_{m=1}^{L}$ denoting the positions of the 1-entries in $\mathbf{b}_{\text{ref}}$, we would have $\mathbf{b}_k \notin \mathcal{C}$ for all k such that $1 \leq |k - k_m| \leq d_{\min} - 1$, for any $m = 1, \ldots, L$. Rather than generating such useless hypotheses, we propose to replace each of them by a hypothesis that satisfies the minimum distance constraint. This is achieved by modifying (16) such that every new 1 has at least $d_{\min}-1$ zeros on each side, i.e., by generating the kth hypothesis $\mathbf{b}_k$ using the following rule (cf. the second box in Fig. 3) instead of (16):

$$\mathbf{b}_k = (\mathbf{b}_{\text{ref}} \oplus \boldsymbol{\delta}_k) \odot \mathbf{w}_k, \quad k = 1, \ldots, K. \tag{19}$$

Here, $\odot$ denotes the elementwise vector product and $\mathbf{w}_k$ is a length-K binary "mask vector" with $d_{\min} - 1$ zeros to the left of k and $d_{\min} - 1$ zeros to the right of k, i.e.,

$$(\mathbf{w}_k)_{k'} = \begin{cases} 0, & 1 \leq |k - k'| \leq d_{\min} - 1 \\ 1, & \text{else,} \end{cases}$$

for $k' = 1, \ldots, K$. An example comparing the hypotheses generated according to (19) (MSMLR) and (16) (SMLR) is shown in Fig. 4.

For a closer analysis of the new hypothesis generation rule (19), let $\mathcal{K}_k$ denote the set of all positions within a two-sided interval of length $2d_{\min} - 1$ around k where $\mathbf{b}_{\text{ref}}$ is one, i.e., $\mathcal{K}_k \triangleq \{k' \in \{k - d_{\min} + 1, \ldots, k + d_{\min} - 1\} : (\mathbf{b}_{\text{ref}})_{k'} = 1\}$. Due to our minimum distance constraint, $\mathcal{K}_k$ can only have zero, one, or two elements, and if it has two elements, one is strictly smaller than k and the other is strictly larger than k. Depending on $\mathcal{K}_k$, we can distinguish four complementary cases, in which (19) simplifies in a specific manner. These four cases and corresponding simplifications of (19) are as follows:

$$\begin{aligned}
&1)\ \mathbf{b}_k = \mathbf{b}_{\text{ref}} - \boldsymbol{\delta}_k && \text{if } \mathcal{K}_k = \{k\} \\
&2)\ \mathbf{b}_k = \mathbf{b}_{\text{ref}} - \boldsymbol{\delta}_{\tilde{k}} + \boldsymbol{\delta}_k && \text{if } \mathcal{K}_k = \{\tilde{k}\} \text{ such that } \tilde{k} \neq k \\
&3)\ \mathbf{b}_k = \mathbf{b}_{\text{ref}} - \boldsymbol{\delta}_{\tilde{k}} - \boldsymbol{\delta}_{\tilde{k}'} + \boldsymbol{\delta}_k && \text{if } \mathcal{K}_k = \{\tilde{k}, \tilde{k}'\} \\
&4)\ \mathbf{b}_k = \mathbf{b}_{\text{ref}} + \boldsymbol{\delta}_k && \text{if } \mathcal{K}_k = \{\}.
\end{aligned}$$

Here, in Cases 2 and 3, $\tilde{k}$ and $\tilde{k}' \neq \tilde{k}$ are any elements of $\{k - d_{\min}+1, \ldots, k+d_{\min}-1\}$ except k. In Fig. 4(b), for example, $\mathbf{b}_3$, $\mathbf{b}_6$, and $\mathbf{b}_8$ correspond to Case 1; $\mathbf{b}_2$, $\mathbf{b}_4$, and $\mathbf{b}_5$ correspond to Case 2; $\mathbf{b}_7$ corresponds to Case 3; and $\mathbf{b}_1$ corresponds to Case 4. We note that the results of (19) and (16) differ only in Cases 2 and 3. In these cases, simply adding a 1 at position k according to (16) violates the minimum distance constraint. The MSMLR rule (19), on the other hand, not only adds a 1 at position k but also removes the 1 at position $\tilde{k}$ (Case 2) or the 1's at positions $\tilde{k}$ and $\tilde{k}'$ (Case 3). This can be interpreted as *shifting* the 1('s) from position(s) $\tilde{k}$ (or $\tilde{k}$ and $\tilde{k}'$) to k. In fact, inspection of the set $\{\mathbf{b}_k\}_{k=1}^{K}$ for a given $\mathbf{b}_{\text{ref}} \in \mathcal{C}$ shows that this set contains all possible hypotheses that can be obtained by shifting the 1 at one position k_m to anywhere within $\{k_m - d_{\min} + 1, \ldots, k_m + d_{\min} - 1\}$. In Fig. 4(b), for example, the 1 at $k_m = 3$ is shifted left in $\mathbf{b}_2$ and right in $\mathbf{b}_4$, the 1 at $k_m = 6$ is shifted left in $\mathbf{b}_5$ and right in $\mathbf{b}_7$, and the 1 at $k_m = 8$ is shifted left in $\mathbf{b}_7$. This resolves a known weakness of the classical SMLR algorithm, namely the absence of hypotheses with shifted 1's (cf. [36], [37]).

The importance of hypotheses with shifted 1's can be illustrated as follows. Suppose that $\mathbf{b}_{\text{ref}}$ contains a 1 at position k', but the posterior $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ would be larger if $\mathbf{b}$ instead contained a 1 at position $k \neq k'$, with $|k - k'| \leq d_{\min} - 1$. (This scenario often occurs in practice.) Now recall that, using (16), each iteration in classical SMLR may change only one entry of $\mathbf{b}_{\text{ref}}$. Therefore, the "desirable" hypothesis with the larger posterior can only be reached in at least two steps, by first adding a 1 at k and then removing the 1 at k' or by first removing the 1 at k' and then adding a 1 at k. In either case, there is at least one intermediate hypothesis with 1's at both k and k' or neither k nor k'. Such intermediate hypotheses often have a lower posterior $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ than the original $\mathbf{b}_{\text{ref}}$, which prevents the algorithm from ever reaching the "desirable" hypothesis. With the minimum distance constraint, this problem is aggravated considerably, since the intermediate hypothesis with 1's at both k and k' has zero probability. In [36] and [37], shifting is included through additional hypotheses or additional iterations. In our MSMLR algorithm, the hypotheses with shifted 1's replace those that are useless due to the minimum distance constraint.

To summarize, the MSMLR hypothesis generation rule (19) preserves all hypotheses $\mathbf{b}_k$ generated according to (16) that comply with the minimum distance constraint, and replaces all others by hypotheses with suitably shifted 1's. If $\mathbf{b}_{\text{ref}}$ satisfies the minimum distance constraint, then the constraint is also

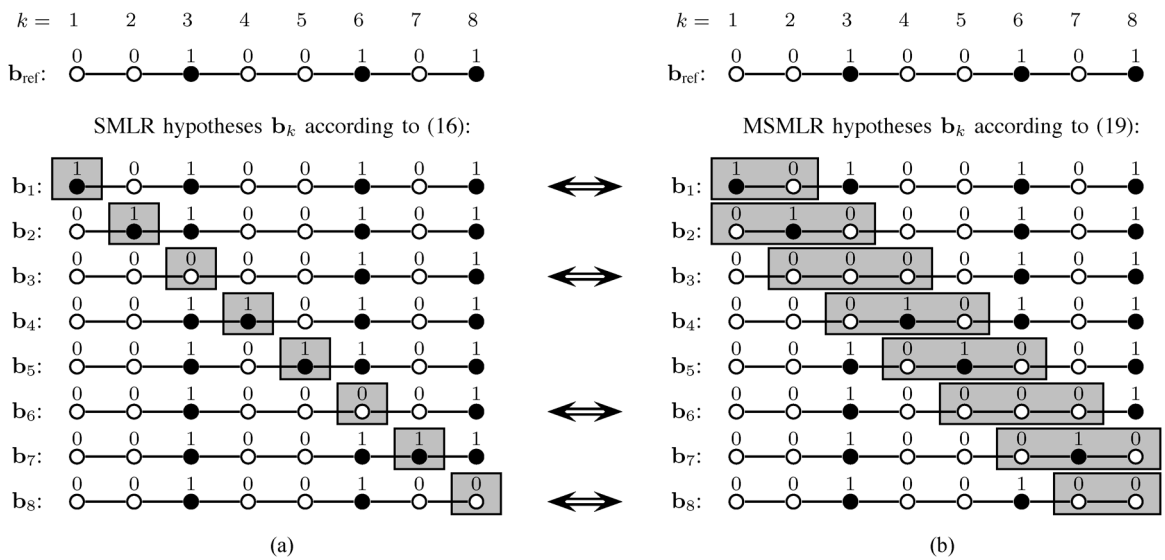

Fig. 4. Hypotheses generated from the same $\mathbf{b}_{\mathrm{ref}}$ according to (a) the SMLR hypothesis generation rule (16) and (b) the MSMLR hypothesis generation rule (19), for $K = 8$ and $d_{\min} = 2$ (i.e., there must be at least one 0 between any two 1's). White (black) nodes depict zero (nonzero) indicators. In each SMLR hypothesis $\mathbf{b}_k$, a gray box highlights the kth entry, which is the only entry where $\mathbf{b}_k$ differs from $\mathbf{b}_{\mathrm{ref}}$. In each MSMLR hypothesis $\mathbf{b}_k$, a gray box highlights the entries $(\mathbf{b}_k)_{k'}$ such that $|k - k'| \leq d_{\min} - 1$, i.e., $(\mathbf{b}_k)_k$ as well as the surrounding entries that are zero regardless of the corresponding entries in $\mathbf{b}_{\mathrm{ref}}$. All MSMLR hypotheses satisfy the minimum distance constraint. Among the SMLR hypotheses, only $\mathbf{b}_1$, $\mathbf{b}_3$, $\mathbf{b}_6$, and $\mathbf{b}_8$ satisfy the minimum distance constraint. These hypotheses appear also among the MSMLR hypotheses, as indicated by the arrows.

satisfied by all $\mathbf{b}_k$, and, hence, by the next $\mathbf{b}_{\mathrm{ref}}$. Therefore, choosing an initial vector $\mathbf{b}^{(0)} \in \mathcal{C}$ ensures that the final result of the MSMLR algorithm, $\hat{\mathbf{b}}_{\mathrm{MSMLR}}$, is in $\mathcal{C}$.

2) Second Modification: Our second modification, abbreviated M2 in the following, concerns the scheduling of the different calculation steps and yields large computational savings. As shown in Fig. 3, each time a hypothesis $\mathbf{b}_k$ is constructed according to (19), we calculate its posterior probability p_k and immediately compare it to p_{ref}. If $p_k > p_{\mathrm{ref}}$, the hypothesis $\mathbf{b}_k$ instantly replaces $\mathbf{b}_{\mathrm{ref}}$, i.e., *before* the next hypothesis $\mathbf{b}_{k+1}$ is constructed. Thus, whenever $p_k > p_{\mathrm{ref}}$, generating $\mathbf{b}_{k+1}$ according to (19) amounts to calculating $\mathbf{b}_{k+1} = (\mathbf{b}_k \oplus \boldsymbol{\delta}_{k+1}) \odot \mathbf{w}_{k+1}$. We repeat the two steps of generating a hypothesis $\mathbf{b}_k$ and calculating p_k (with k cycling through $1, \ldots, K$) until a local maximum of the posterior probability $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ is found. When $\mathbf{b}_{\mathrm{ref}}$ has reached a local maximum of $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$, K successively generated hypotheses fail to increase the posterior probability, i.e., these hypotheses do not replace $\mathbf{b}_{\mathrm{ref}}$. This marks the end of the algorithm, because applying (19) to one fixed $\mathbf{b}_{\mathrm{ref}}$ cannot produce more than K different new hypotheses. Let k_{ref} denote the position where $\mathbf{b}_{\mathrm{ref}}$ was changed most recently, i.e., the previous $\mathbf{b}_{\mathrm{ref}}$ was replaced by $\mathbf{b}_{k_{\mathrm{ref}}}$. If the next K hypotheses $\mathbf{b}_k$ with $k = k_{\mathrm{ref}} + 1, k_{\mathrm{ref}} + 2, \ldots, K, 1, 2, \ldots, k_{\mathrm{ref}}$ fail to replace $\mathbf{b}_{\mathrm{ref}}$, i.e., $p_k \leq p_{\mathrm{ref}}$ (cf. the first case distinction in Fig. 3), then $\mathbf{b}_{\mathrm{ref}}$ corresponds to a local maximum of the posterior, and it is thus our detected sequence $\hat{\mathbf{b}}_{\mathrm{MSMLR}}$. The MSMLR algorithm checks whether the K most recently generated hypotheses failed to replace $\mathbf{b}_{\mathrm{ref}}$ through the second case distinction in Fig. 3, i.e., by checking whether k equals k_{ref}.

The effects of M2 on the estimation accuracy and on the computational complexity are not immediately obvious. Regarding estimation accuracy, we note that the detected sequence $\hat{\mathbf{b}}_{\mathrm{MSMLR}}$ obtained with M2 may evidently be different from that obtained without M2. Without M2, each update of $\mathbf{b}_{\mathrm{ref}}$ is obtained by maximizing p_k over K hypotheses. Starting from the same $\mathbf{b}_{\mathrm{ref}}$, the update with M2 is obtained by maximizing p_k only over a subset of these K hypotheses. Let $p_{\mathrm{ref,updated}}$ and $p_{\mathrm{ref,updated}}^{(\mathrm{M2})}$ denote the respective probabilities after the update. It follows from the above that $p_{\mathrm{ref,updated}}^{(\mathrm{M2})} \leq p_{\mathrm{ref,updated}}$, i.e., the update step with M2 increases p_{ref} not more—but typically less—than the update step without M2. It is important to note that, even with these smaller update steps, the algorithm may still converge to a higher local maximum than with the larger steps. However, the algorithm without M2 at least avoids all local maxima that are smaller than $p_{\mathrm{ref,updated}}$, which is an advantage over the algorithm with M2 because $p_{\mathrm{ref,updated}}^{(\mathrm{M2})} \leq p_{\mathrm{ref,updated}}$. Therefore, the problem of the multimodality of $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ becomes more serious when M2 is used.

The solution to this problem is provided by the proposed modified hypothesis generation rule (19), i.e., by a new definition of "neighbor" sequences. Indeed, hypotheses with shifted 1's are not "neighbors" of $\mathbf{b}_{\mathrm{ref}}$ in the SMLR sense, because they differ from $\mathbf{b}_{\mathrm{ref}}$ in more than one entry. Since the MSMLR hypothesis generation rule (19) contains such hypotheses instead of the hypotheses with zero probability, each $\mathbf{b}$ has more neighbors in the MSMLR sense with nonzero probability than it has neighbors in the SMLR sense with nonzero probability. This increase in the number of neighbors with nonzero probability significantly reduces the number of local maxima of $p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$ compared to SMLR. The relaxation of the multimodality problem through (19) is, in many cases, an effective means to overcome the theoretical disadvantage of M2 mentioned above, i.e., the smaller update steps. Extensive simulations (cf. Section VI-C) show that the performance of MSMLR with and without M2 is effectively the same. The only—trivial—exception is the case $d_{\min} = 1$

(i.e., no constraint), since the MSMLR hypotheses here do not contain shifted 1's. Furthermore, simulations show that combining SMLR with M2 leads to very bad performance, again due to the lack of hypotheses with shifted 1's.

Regarding computational complexity, we recall that, without M2, MSMLR generates K hypotheses before it makes one update (which, according to Section III-B1, may change up to three entries within an interval of length $2d_{\min}$). With M2, on the other hand, generating K hypotheses typically comes with several updates. The number of these updates is between 1 and K (assuming that $\mathbf{b}_{\mathrm{ref}}$ is not a local maximum itself). Some of the updates may be ineffective because they are overwritten by the subsequent update (e.g., if an update places a 1 at position k and the subsequent update shifts this 1 to position $k+1$, leading to a configuration that could also have been reached without the first update). Some of the updates may even be counterproductive in the sense that they lead to a bad configuration which can only be changed after several subsequent updates. Clearly, these effects increase the complexity of the algorithm. On the other hand, some of the updates typically anticipate updates which would be chosen by the algorithm without M2 only in later iterations, i.e., after generating more sets of K hypotheses. This effect greatly reduces the complexity. Simulations confirm that the latter effect clearly outweighs the former two on average, making MSMLR with M2 even more efficient than other SMLR-based methods (cf. Sections VI-B and VI-C).

IV. Efficient Implementation

Next, we present an efficient implementation of the MSMLR algorithm. This implementation combines results of [30] with the Cholesky factor approach of [31], using some adaptations to accommodate the MSMLR hypothesis generation rule (19) and our signal model.

The central step of each iteration is to check if $p_k > p_{\mathrm{ref}}$ (cf. Fig. 3). According to (17) and (18), this amounts to deciding if $p(\mathbf{b}_k, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x}) > p(\mathbf{b}_{\mathrm{ref}}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$, or, equivalently, if

$$p\Big(\mathbf{b}_k|\hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}, \mathbf{x}\Big) > p\Big(\mathbf{b}_{\mathrm{ref}}|\hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}, \mathbf{x}\Big). \tag{20}$$

Here, $p(\mathbf{b}|\boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ can be obtained, up to a normalization, from (15) by dropping all factors that do not depend on $\mathbf{b}$, i.e.,

$$p\big(\mathbf{b}|\boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\big) \propto |\boldsymbol{\Sigma}_b| \exp\Big(\boldsymbol{\mu}_b^{\mathrm{H}} \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\mu}_b\Big) \frac{p(\mathbf{b})}{\sigma_a^{2L}}. \tag{21}$$

As shown in [31], equation (20) can be reformulated such that we can avoid calculating $p(\mathbf{b}_k|\hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}, \mathbf{x})$ directly and instead exploit the relation between $\mathbf{b}_k$ and $\mathbf{b}_{\mathrm{ref}}$ for improved efficiency. To this end, it is helpful to define (cf. (14))

$$\mathbf{e}_b \triangleq \sigma_n^2 \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\mu}_b = \mathbf{F}_b^{\mathrm{H}} \mathbf{x} \tag{22}$$

$$\boldsymbol{\Psi}_b \triangleq \sigma_a^2 \boldsymbol{\Sigma}_b^{-1} = \frac{\sigma_a^2}{\sigma_n^2} \mathbf{F}_b^{\mathrm{H}} \mathbf{F}_b + \mathbf{I} \tag{23}$$

and

$$\boldsymbol{\Phi}_b \triangleq \mathrm{chol}\{\boldsymbol{\Psi}_b\},$$

where $\mathrm{chol}\{\cdot\}$ denotes the lower triangular Cholesky factor [47, p. 500], i.e., $\boldsymbol{\Phi}_b \boldsymbol{\Phi}_b^{\mathrm{H}} = \boldsymbol{\Psi}_b$. We can then write (21) as

$$\begin{aligned} p\big(\mathbf{b}|\boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\big) &\propto \frac{\sigma_a^{2L}}{|\boldsymbol{\Psi}_b|} \exp\left(\frac{\sigma_a^2}{\sigma_n^4} \mathbf{e}_b^{\mathrm{H}} \boldsymbol{\Psi}_b^{-1} \mathbf{e}_b\right) \frac{p(\mathbf{b})}{\sigma_a^{2L}} \\ &= \frac{1}{|\boldsymbol{\Phi}_b|^2} \exp\left(\frac{\sigma_a^2}{\sigma_n^4} \left\|\boldsymbol{\Phi}_b^{-1} \mathbf{e}_b\right\|^2\right) p(\mathbf{b}). \end{aligned}$$

Using the Cholesky factor in this context was also proposed in [40] and [30]. Inserting (9) and dropping some constant factors yields

$$\begin{aligned} p\big(\mathbf{b}|\boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\big) \propto \frac{1}{|\boldsymbol{\Phi}_b|^2} \exp\left(\frac{\sigma_a^2}{\sigma_n^4} \left\|\boldsymbol{\Phi}_b^{-1} \mathbf{e}_b\right\|^2\right) I_{\mathcal{C}}(\mathbf{b}) \\ \times \left(\frac{\pi_1}{1-\pi_1}\right)^{\|\mathbf{b}\|^2}. \end{aligned}$$

We can also drop the factor $I_{\mathcal{C}}(\mathbf{b})$ because $\mathbf{b} \in \mathcal{C}$ for all hypotheses used in our algorithm. Furthermore, it will be convenient to subtract the constant $\|\mathbf{b}_{\mathrm{ref}}\|^2$ from the exponent in the last factor. We thus obtain

$$\begin{aligned} p\big(\mathbf{b}|\boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\big) \propto \tilde{q}(\mathbf{b}) \triangleq \frac{1}{|\boldsymbol{\Phi}_b|^2} \exp\left(\frac{\sigma_a^2}{\sigma_n^4} \left\|\boldsymbol{\Phi}_b^{-1} \mathbf{e}_b\right\|^2\right) \\ \times \left(\frac{\pi_1}{1-\pi_1}\right)^{\|\mathbf{b}\|^2 - \|\mathbf{b}_{\mathrm{ref}}\|^2}. \end{aligned}$$

Following [29], we perform the comparison step (20) in the logarithmic domain, which helps avoid numerical problems with very large or small (unnormalized) probabilities. Thus, (20) is reformulated as

$$q(\mathbf{b}_k) > q(\mathbf{b}_{\mathrm{ref}}), \tag{24}$$

where

$$\begin{aligned} q(\mathbf{b}) &\triangleq \log \tilde{q}(\mathbf{b}) \\ &= -2 \log |\boldsymbol{\Phi}_b| + \frac{\sigma_a^2}{\sigma_n^4} \left\|\boldsymbol{\Phi}_b^{-1} \mathbf{e}_b\right\|^2 + \big(\|\mathbf{b}\|^2 - \|\mathbf{b}_{\mathrm{ref}}\|^2\big) \tilde{\pi}_1, \end{aligned} \tag{25}$$

with $\tilde{\pi}_1 \triangleq \log\left(\frac{\pi_1}{1-\pi_1}\right)$. Note that in (24), typically, only $q(\mathbf{b}_k)$ needs to be calculated anew in each iteration, whereas $\mathbf{b}_{\mathrm{ref}}$ (and, thus, $q(\mathbf{b}_{\mathrm{ref}})$) usually stays the same for many iterations.

Let $\mathbf{F}_{b,\mathrm{ref}}$, $\mathbf{e}_{b,\mathrm{ref}}$, $\boldsymbol{\Psi}_{b,\mathrm{ref}}$, and $\boldsymbol{\Phi}_{b,\mathrm{ref}}$ denote the respective quantities $\mathbf{F}_b$ etc. evaluated for $\mathbf{b} = \mathbf{b}_{\mathrm{ref}}$. Furthermore, we denote by $\mathbf{F}_{b,k}$, $\mathbf{e}_{b,k}$, $\boldsymbol{\Psi}_{b,k}$, and $\boldsymbol{\Phi}_{b,k}$ the respective quantities evaluated for $\mathbf{b} = \mathbf{b}_k$ (possibly up to a permutation of the columns, entries, or columns/rows, as indicated below). Then, as was shown in [30] and [31], we can calculate $q(\mathbf{b}_k)$ efficiently by using the relations existing between $\boldsymbol{\Psi}_{b,\mathrm{ref}}$ and $\boldsymbol{\Psi}_{b,k}$ and between $\mathbf{e}_{b,\mathrm{ref}}$ and $\mathbf{e}_{b,k}$. We obtain $\boldsymbol{\Psi}_{b,k}$ and $\mathbf{e}_{b,k}$ from $\boldsymbol{\Psi}_{b,\mathrm{ref}}$ and $\mathbf{e}_{b,\mathrm{ref}}$ in different ways corresponding to the four complementary cases of how $\mathbf{b}_k$ may be related to $\mathbf{b}_{\mathrm{ref}}$. (We recall that these cases were listed in Section III-B1.) Subsequently, we calculate $q(\mathbf{b}_k)$ by inserting $\mathbf{b}_k$, $\mathbf{e}_{b,k}$, and $\boldsymbol{\Phi}_{b,k} \triangleq \mathrm{chol}\{\boldsymbol{\Psi}_{b,k}\}$ for $\mathbf{b}$, $\mathbf{e}_b$, and $\boldsymbol{\Phi}_b$ in (25). The four cases are discussed in the following.

Case 1: $\mathbf{b}_k = \mathbf{b}_{\mathrm{ref}} - \boldsymbol{\delta}_k$. Here, $\mathbf{F}_{b,k}$ is $\mathbf{F}_{b,\mathrm{ref}}$ with the column corresponding to k removed. Similarly, $\mathbf{e}_{b,k}$ (cf. (22)) is obtained from $\mathbf{e}_{b,\mathrm{ref}}$ by removing the entry corresponding to k,

and $\mathbf{\Psi}_{b,k}$ (cf. (23)) is obtained from $\mathbf{\Psi}_{b,\mathrm{ref}}$ by removing the corresponding column and row.

Case 2: $\mathbf{b}_k = \mathbf{b}_{\mathrm{ref}} - \boldsymbol{\delta}_{\tilde{k}} + \boldsymbol{\delta}_k$. In analogy to Case 1, we form $\tilde{\mathbf{F}}_{b,k}$, $\tilde{\mathbf{e}}_{b,k}$, and $\tilde{\mathbf{\Psi}}_{b,k}$ by removing the column of $\mathbf{F}_{b,\mathrm{ref}}$, entry of $\mathbf{e}_{b,\mathrm{ref}}$, and column and row of $\mathbf{\Psi}_{b,\mathrm{ref}}$, respectively, that correspond to $\tilde{k}$. Then, $\mathbf{e}_{b,k}$ and $\mathbf{\Psi}_{b,k}$ are defined as

$$\mathbf{e}_{b,k} = \begin{pmatrix} \tilde{\mathbf{e}}_{b,k} \\ e_k \end{pmatrix}, \quad \mathbf{\Psi}_{b,k} = \begin{pmatrix} \tilde{\mathbf{\Psi}}_{b,k} & \frac{\sigma_a^2}{\sigma_n^2}\mathbf{g}_k \\ \frac{\sigma_a^2}{\sigma_n^2}\mathbf{g}_k^{\mathrm{H}} & \frac{\sigma_a^2}{\sigma_n^2}\|\mathbf{f}_k\|^2 + 1 \end{pmatrix}, \quad (26)$$

where $\mathbf{f}_k$ denotes the kth column of $\mathbf{F}$, $e_k \triangleq \mathbf{f}_k^{\mathrm{H}}\mathbf{x}$, and $\mathbf{g}_k \triangleq \tilde{\mathbf{F}}_{b,k}^{\mathrm{H}}\mathbf{f}_k$. Note that $\mathbf{e}_{b,k}$ and $\mathbf{\Psi}_{b,k}$ are permuted versions of, respectively, (22) and (23) evaluated for $\mathbf{b} = \mathbf{b}_k$. Here, for the sake of a simpler notation, we exploit the fact that consistently permuting the entries of $\mathbf{e}_{b,k}$ and the columns/rows of $\mathbf{\Psi}_{b,k}$ has no effect on $q(\mathbf{b}_k)$.

Case 3: $\mathbf{b}_k = \mathbf{b}_{\mathrm{ref}} - \boldsymbol{\delta}_{\tilde{k}} - \boldsymbol{\delta}_{\tilde{k}'} + \boldsymbol{\delta}_k$. Similarly to Case 2, we first remove the entries of $\mathbf{e}_{b,\mathrm{ref}}$ and the columns/rows of $\mathbf{\Psi}_{b,\mathrm{ref}}$ that correspond to $\tilde{k}$ and $\tilde{k}'$ and then add an entry and a column/row calculated as in (26).

Case 4: $\mathbf{b}_k = \mathbf{b}_{\mathrm{ref}} + \boldsymbol{\delta}_k$. Here, as an alternative to calculating $\mathbf{e}_{b,k}$, $\mathbf{\Psi}_{b,k}$, and $\mathbf{\Phi}_{b,k}$ as in the other three cases, we can pursue a more efficient approach. Namely, as shown in [30] and [31], one can use the relations (presented without proof)

$$\left\|\mathbf{\Phi}_{b,k}^{-1}\mathbf{e}_{b,k}\right\|^2 = \left\|\mathbf{\Phi}_{b,\mathrm{ref}}^{-1}\mathbf{e}_{b,\mathrm{ref}}\right\|^2 + \frac{\left|\mathbf{f}_k^{\mathrm{H}}\mathbf{F}_{b,\mathrm{ref}}\,\mathbf{z} - e_k\right|^2}{\tau}$$
$$|\mathbf{\Phi}_{b,k}|^2 = \tau|\mathbf{\Phi}_{b,\mathrm{ref}}|^2,$$

with

$$\mathbf{z} \triangleq \frac{\sigma_a^2}{\sigma_n^2}\mathbf{\Psi}_{b,\mathrm{ref}}^{-1}\mathbf{e}_{b,\mathrm{ref}}$$
$$\tau \triangleq 1 + \frac{\sigma_a^2}{\sigma_n^2}\|\mathbf{f}_k\|^2 - \frac{\sigma_a^4}{\sigma_n^4}\left\|\mathbf{\Phi}_{b,\mathrm{ref}}^{-1}\mathbf{F}_{b,\mathrm{ref}}^{\mathrm{H}}\mathbf{f}_k\right\|^2,$$

to obtain from (25)

$$q(\mathbf{b}_k) = q(\mathbf{b}_{\mathrm{ref}}) - \log\tau + \frac{\sigma_a^2}{\sigma_n^4}\frac{\left|\mathbf{f}_k^{\mathrm{H}}\mathbf{F}_{b,\mathrm{ref}}\mathbf{z} - e_k\right|^2}{\tau} + \tilde{\pi}_1.$$

As proposed in [36], the calculation of (26)—i.e., of the quantities $\mathbf{g}_k$, $\|\mathbf{f}_k\|^2$, and e_k—can be significantly simplified as follows. Let k_m denote the position corresponding to the mth column of $\tilde{\mathbf{F}}_{b,k}$. Then $(\mathbf{g}_k)_m = (\tilde{\mathbf{F}}_{b,k}^{\mathrm{H}}\mathbf{f}_k)_m = \mathbf{f}_{k_m}^{\mathrm{H}}\mathbf{f}_k$, which simplifies to

$$(\mathbf{g}_k)_m = \rho_{k_m-k}^{(f)} \quad \text{with } \rho_k^{(f)} \triangleq \sum_{l=-Q}^{Q} f_{l-k}^* f_l$$

(where the superscript * denotes complex conjugation), unless k and k_m are both smaller than $Q+1$ or both larger than $K-Q$. To verify this, recall that the length-K vector $\mathbf{f}_k$ contains $\mathbf{f}$ centered around the kth element, padded with zeros and sometimes truncated, namely if $k \le Q$ or $k \ge K-Q+1$. Analogous considerations apply to $\mathbf{f}_{k_m}$. The inner product of two shifted copies of $\mathbf{f}$ equals the autocorrelation function $\rho_k^{(f)}$ of f_k. If one of the two copies has its entire support within $\{1,\ldots,K\}$ (i.e., it is not truncated), then it is irrelevant whether the other copy is truncated. However, if both copies are truncated, the simplification does not apply and $\mathbf{f}_{k_m}^{\mathrm{H}}\mathbf{f}_k$ must be calculated directly.

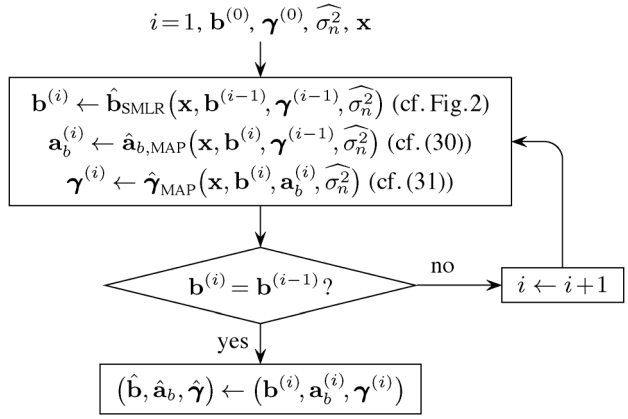

Fig. 5. Flow chart of classical SMLR-based BD [33].

Similar considerations show that $\|\mathbf{f}_k\|^2$ in the additional diagonal entry of $\mathbf{\Psi}_{b,k}$ (see (26)) is given by

$$\|\mathbf{f}_k\|^2 = \begin{cases} \sum_{l=1-k}^{Q}|f_l|^2, & 1 \le k \le Q \\ \|\mathbf{f}\|^2 = \rho_0^{(f)}, & Q+1 \le k \le K-Q \\ \sum_{l=-Q}^{K-k}|f_l|^2, & K-Q+1 \le k \le K. \end{cases}$$

It follows from $Q \ll K$ that $(\mathbf{g}_k)_m = \rho_{k_m-k}^{(f)}$ and $\|\mathbf{f}_k\|^2 = \rho_0^{(f)}$ for most k and m. Finally, calculating $e_k = \mathbf{f}_k^{\mathrm{H}}\mathbf{x}$ amounts to convolving x_k with f_{-k}^*, i.e., $e_k = \sum_{l=1}^{K} f_{l-k}^* x_l$ for $k \in \{1,\ldots,K\}$.

Note that $\rho_k^{(f)}$, $\|\mathbf{f}_k\|^2$, and e_k have to be calculated only at the start of the algorithm (for $k = -2Q,\ldots,2Q$, for $k = 1,\ldots,K$, and for $k = 1,\ldots,K$, respectively), because $\mathbf{f}$ is fixed throughout the nonblind MSMLR algorithm. Whenever $\mathbf{b}_{\mathrm{ref}}$ is updated, we also update $q(\mathbf{b}_{\mathrm{ref}})$ and calculate $\mathbf{\Psi}_{b,\mathrm{ref}}$ and $\mathbf{z}$. Finally, to calculate $q(\mathbf{b}_k)$ in each iteration, we calculate τ (in Case 4) or $\mathbf{\Psi}_{b,k}$ (in Cases 1–3) and $\mathbf{g}_k$ (in Cases 2 and 3).

V. BD Based on Modified SMLR

In this section, we present a BD method that uses the MSMLR algorithm proposed in Sections III-B and IV.

A. Review of SMLR-Based BD

We start by briefly reviewing the classical SMLR-based BD method of [33]. The signal model of [33] equals that of [29], i.e., it differs from ours in the aspects listed at the beginning of Section III. We will present the SMLR-based BD method with the simplifications suggested in [38] (i.e., without the ARMA model of the pulse and the estimation of additional parameters) and with slight adaptations to fit our signal model.

The method detects $\mathbf{b}$ and estimates $\mathbf{a}_b$ and $\boldsymbol{\gamma}$, while the noise variance σ_n^2 is assumed fixed at some value $\widehat{\sigma_n^2}$. As shown in Fig. 5, in the ith iteration of the algorithm, the indicator vector $\mathbf{b}^{(i)}$ is generated using the classical SMLR algorithm (reviewed in Section III-A) with initial value $\mathbf{b}^{(i-1)}$ and pulse coefficients $\boldsymbol{\gamma}^{(i-1)}$, i.e.,

$$\mathbf{b}^{(i)} = \hat{\mathbf{b}}_{\mathrm{SMLR}}\left(\mathbf{x}, \mathbf{b}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)}, \widehat{\sigma_n^2}\right). \quad (27)$$

Then, $\mathbf{a}_b^{(i)}$ and $\boldsymbol{\gamma}^{(i)}$ are calculated as the MAP estimators maximizing the joint posterior in (13), i.e.,

$$\mathbf{a}_b^{(i)} = \hat{\mathbf{a}}_{b,\mathrm{MAP}}\Big(\mathbf{x}, \mathbf{b}^{(i)}, \boldsymbol{\gamma}^{(i-1)}, \widehat{\sigma_n^2}\Big) \tag{28}$$

$$\boldsymbol{\gamma}^{(i)} = \hat{\boldsymbol{\gamma}}_{\mathrm{MAP}}\Big(\mathbf{x}, \mathbf{b}^{(i)}, \mathbf{a}_b^{(i)}, \widehat{\sigma_n^2}\Big). \tag{29}$$

Using (13) with (12) and (8) yields

$$\begin{aligned}\hat{\mathbf{a}}_{b,\mathrm{MAP}}\big(\mathbf{x}, \mathbf{b}, \boldsymbol{\gamma}, \sigma_n^2\big) &\triangleq \underset{\mathbf{a}_b \in \mathbb{C}^L}{\arg\max}\, p\big(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2 | \mathbf{x}\big) \\ &= \underset{\mathbf{a}_b \in \mathbb{C}^L}{\arg\min} \left\{ \|\mathbf{x} - \mathbf{F}_b \mathbf{a}_b\|^2 + \frac{\sigma_n^2}{\sigma_a^2} \|\mathbf{a}_b\|^2 \right\} \\ &= \left(\mathbf{F}_b^{\mathrm{H}} \mathbf{F}_b + \frac{\sigma_n^2}{\sigma_a^2} \mathbf{I} \right)^{-1} \mathbf{F}_b^{\mathrm{H}} \mathbf{x}. \end{aligned} \tag{30}$$

Similarly, using (13) with (12) and (10) yields

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_{\mathrm{MAP}}\big(\mathbf{x}, \mathbf{b}, \mathbf{a}_b, \sigma_n^2\big) &\triangleq \underset{\boldsymbol{\gamma} \in \mathbb{C}^M}{\arg\max}\, p\big(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2 | \mathbf{x}\big) \\ &= \underset{\boldsymbol{\gamma} \in \mathbb{C}^M}{\arg\min} \Big\{ \|\mathbf{x} - \mathbf{A}\mathbf{H}\boldsymbol{\gamma}\|^2 + \sigma_n^2 \boldsymbol{\gamma}^{\mathrm{H}} \boldsymbol{\Lambda}^{-1} \boldsymbol{\gamma} \Big\} \\ &= \left(\mathbf{H}^{\mathrm{H}} \mathbf{A}^{\mathrm{H}} \mathbf{A} \mathbf{H} + \sigma_n^2 \boldsymbol{\Lambda}^{-1} \right)^{-1} \mathbf{H}^{\mathrm{H}} \mathbf{A}^{\mathrm{H}} \mathbf{x}. \end{aligned} \tag{31}$$

When evaluating (30), L (the number of 1-entries in $\mathbf{b}$) and $\mathbf{F}_b$ must first be derived from the arguments $\mathbf{b}$ and $\boldsymbol{\gamma}$. Hence, for calculating (28), $L = \|\mathbf{b}^{(i)}\|^2$ and $\mathbf{F}_b$ is formed using $\mathbf{b}^{(i)}$ and $\boldsymbol{\gamma}^{(i-1)}$. Similarly, for calculating (29), $\mathbf{A}$ is formed using $\mathbf{b}^{(i)}$ and $\mathbf{a}_b^{(i)}$.

The steps (27)–(29) are iterated until (27) fails to change $\mathbf{b}^{(i-1)}$, which indicates that a local maximum has been found at $\mathbf{b}^{(i-1)} = \mathbf{b}^{(i)}$. This method can be seen as an approximation of the *iterated conditional mode* (ICM) strategy [48], which cyclically updates all parameters using MAP estimators. In particular, the detection step yielding $\mathbf{b}^{(i)}$ in (27) approximates the MAP detector (in which $\mathbf{a}_b$ has been marginalized out).

The above formulation of the algorithm can also be used for the signal model of [33], with minor modifications because [33] does not use a basis expansion for the pulse shape $\mathbf{f}$, i.e., $\mathbf{f}$ is directly estimated rather than a coefficient vector $\boldsymbol{\gamma}$ (this can be easily accounted for by using the basis matrix $\mathbf{H} \triangleq \mathbf{I}$, so that $\boldsymbol{\gamma} = \mathbf{H}\boldsymbol{\gamma} = \mathbf{f}$), and $\mathbf{a}_b$ and $\boldsymbol{\gamma}$ are assumed real.

B. Proposed MSMLR-Based BD Method

The proposed BD method is stated in Fig. 6 and described in the following. It differs from the classical method reviewed in Section V-A in three respects: (i) it is based on the MSMLR algorithm presented in Sections III-B and IV, (ii) it determines an optimal scale and shift of $\mathbf{a}$ and $\mathbf{f}$, and (iii) it includes estimation of the noise variance σ_n^2 using the prior in (11).

1) Amplitude Scale and Time Shift: Temporarily assuming an infinite k-domain of the sequences x_k, a_k, and f_k, we have

$$f_k * a_k = f'_k * a'_k, \quad \text{with } a'_k = c\, a_{k+\kappa},\ f'_k = \frac{1}{c} f_{k-\kappa}, \tag{32}$$

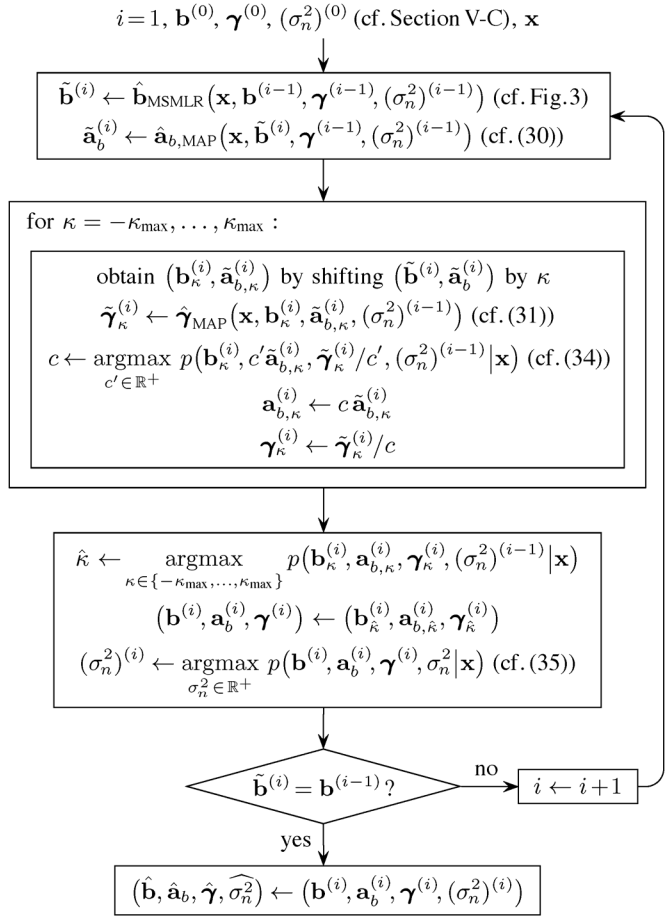

Fig. 6. Flow chart of the proposed MSMLR-based BD method.

where $*$ denotes convolution, $c \in \mathbb{C} \setminus \{0\}$ is an arbitrary amplitude scale factor, and $\kappa \in \mathbb{Z}$ is an arbitrary time shift. We can rewrite (32) as (cf. (5)) $\mathbf{A}\mathbf{f} = \mathbf{A}'\mathbf{f}'$, where (cf. (4)) $\mathbf{f} = \mathbf{H}\boldsymbol{\gamma}$. Assuming that some $\boldsymbol{\gamma}'$ exists such that

$$(\mathbf{H}\boldsymbol{\gamma}')_k = \frac{1}{c} (\mathbf{H}\boldsymbol{\gamma})_{k-\kappa} \tag{33}$$

(equivalently, $\mathbf{H}\boldsymbol{\gamma}' = \mathbf{f}'$), we can finally express relation (32) as $\mathbf{A}\mathbf{H}\boldsymbol{\gamma} = \mathbf{A}'\mathbf{H}\boldsymbol{\gamma}'$. In view of (5), we conclude that different parameter combinations $(\mathbf{a}, \boldsymbol{\gamma})$ and $(\mathbf{a}', \boldsymbol{\gamma}')$ lead to the same observation $\mathbf{x}$. This ambiguity inherent to BD is usually not a problem, since the true amplitude scale and time shift are often irrelevant.

However, while we cannot find the true amplitude scale and time shift, we propose to determine the amplitude scale and time shift that are optimal in terms of maximizing the posterior probability $p(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2 | \mathbf{x})$. Indeed, despite the ambiguity described above, $p(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2 | \mathbf{x})$ is not invariant to amplitude scalings and time shifts, for the following three reasons:

1. Amplitude scalings affect the priors of both $\mathbf{a}$ and $\mathbf{f}$, and time shifts affect the prior of $\mathbf{f}$ (because f_k has a different prior than f_{k+1}, as illustrated in Fig. 1(a)). Because of (13), the priors directly influence $p(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2 | \mathbf{x})$.
2. The assumption of an infinite temporal domain of x_k, a_k, and f_k is not satisfied. Therefore, some time shifts cause certain values a_k or f_k to be shifted outside their respective

temporal domain, which leads to a change of $p(\mathbf{x}|\mathbf{a},\boldsymbol{\gamma},\sigma_n^2)$ (and also of the priors) and, thus, of $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$.

3. A $\boldsymbol{\gamma}'$ satisfying (33) *exactly* typically does not exist, as the shifted vector $\mathbf{f}'$ generally does not lie in the subspace of $\mathbb{C}^{2Q+1}$ spanned by the columns of $\mathbf{H}$ and thus cannot be expressed as $\mathbf{f}'=\mathbf{H}\boldsymbol{\gamma}'$. Therefore, time shifts generally require an approximation of $\mathbf{f}'$, thus changing $p(\mathbf{x}|\mathbf{a},\boldsymbol{\gamma},\sigma_n^2)$ (and also the priors) and, consequently, $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$.

In previous BD methods such as [33], [38], such dependencies of the posterior probability on the scale and shift of $\mathbf{a}$ or $\boldsymbol{\gamma}$ can usually be ignored. Indeed, due to their simpler prior of $\mathbf{f}$, the first reason for the dependency applies only to amplitude scalings but not to time shifts, and the third reason does not apply at all. In our signal model, on the other hand, due to the dependency of $p(f_k)$ on k via the basis representation of $\mathbf{f}$, time shifts have a significant influence on $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$. Our simulations have shown that the proposed basis representation of $\mathbf{f}$, combined with an optimization of the amplitude scale and time shift, yields substantial performance gains relative to [38].

2) Determination of the Optimal Scale and Shift: The algorithmic approach of BD methods such as [33], [38] is not suited to finding the optimal amplitude scale and time shift. Namely, if we fix $\mathbf{a}$ or $\boldsymbol{\gamma}$ each time we maximize $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$ with respect to the other parameters (cf. Fig. 5), then the amplitude scale and time shift are always largely determined by the respective fixed parameter. More specifically, the scale changes only slightly in each maximization and thus approaches its optimal value very slowly, and the time shift, due to its discrete nature, typically does not improve at all.

To resolve this problem, within each iteration of our algorithm, we fix the sequence $\mathbf{a}$ up to a time shift κ and a scale factor c and maximize $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$ with respect to κ, c, and $\boldsymbol{\gamma}$. Thus, the ith iteration now consists of the following steps (cf. Fig. 6): First, $\tilde{\mathbf{b}}^{(i)}$ is generated by means of MSMLR, i.e.,

$$\tilde{\mathbf{b}}^{(i)}=\hat{\mathbf{b}}_{\mathrm{MSMLR}}\Big(\mathbf{x},\mathbf{b}^{(i-1)},\boldsymbol{\gamma}^{(i-1)},\left(\sigma_n^2\right)^{(i-1)}\Big).$$

Then, the $\mathbf{a}_b$ maximizing the joint posterior is calculated (see (28) and (30)):

$$\tilde{\mathbf{a}}_b^{(i)}=\hat{\mathbf{a}}_{b,\mathrm{MAP}}\Big(\mathbf{x},\tilde{\mathbf{b}}^{(i)},\boldsymbol{\gamma}^{(i-1)},\left(\sigma_n^2\right)^{(i-1)}\Big).$$

Next, using $\tilde{\mathbf{b}}^{(i)}$ and $\tilde{\mathbf{a}}_b^{(i)}$, vectors $\mathbf{b}_\kappa^{(i)}$, $\mathbf{a}_{b,\kappa}^{(i)}$, and $\boldsymbol{\gamma}_\kappa^{(i)}$ are calculated for each $\kappa\in\{-\kappa_{\max},\ldots,\kappa_{\max}\}$ as follows.

1) The entries of $\tilde{\mathbf{b}}^{(i)}$ and $\tilde{\mathbf{a}}_b^{(i)}$ are shifted by κ:

$$\left(\mathbf{b}_\kappa^{(i)}\right)_k\triangleq\begin{cases}\left(\tilde{\mathbf{b}}^{(i)}\right)_{k-\kappa}, & 1+\kappa\le k\le K+\kappa\\ 0, & \text{else}\end{cases}$$

$$\left(\tilde{\mathbf{a}}_{b,\kappa}^{(i)}\right)_k\triangleq\begin{cases}\left(\tilde{\mathbf{a}}_b^{(i)}\right)_{k-\kappa}, & 1+\kappa\le k\le K+\kappa\\ 0, & \text{else.}\end{cases}$$

2) The $\boldsymbol{\gamma}$ maximizing the joint posterior is calculated (see (29) and (31)):

$$\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}\triangleq\hat{\boldsymbol{\gamma}}_{\mathrm{MAP}}\Big(\mathbf{x},\mathbf{b}_\kappa^{(i)},\tilde{\mathbf{a}}_{b,\kappa}^{(i)},\left(\sigma_n^2\right)^{(i-1)}\Big).$$

3) The vectors $\tilde{\mathbf{a}}_{b,\kappa}^{(i)}$ and $\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}$ are scaled, i.e.,

$$\mathbf{a}_{b,\kappa}^{(i)}\triangleq c\tilde{\mathbf{a}}_{b,\kappa}^{(i)},\quad \boldsymbol{\gamma}_\kappa^{(i)}\triangleq\frac{\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}}{c},$$

where the scale factor $c\in\mathbb{R}^+$ maximizes the joint posterior in (13). (We use $c\in\mathbb{R}^+$ because (13) is invariant to the phase of c.) Thus,

$$\begin{aligned}c&=\operatorname*{arg\,max}_{c'\in\mathbb{R}^+}\ p\left(\mathbf{b}_\kappa^{(i)},c'\tilde{\mathbf{a}}_{b,\kappa}^{(i)},\frac{\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}}{c'},\left(\sigma_n^2\right)^{(i-1)}\,\middle|\,\mathbf{x}\right)\\&=\operatorname*{arg\,max}_{c'\in\mathbb{R}^+}\left\{p\Big(c'\tilde{\mathbf{a}}_{b,\kappa}^{(i)}\,\Big|\,\mathbf{b}_\kappa^{(i)}\Big)\,p\left(\frac{\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}}{c'}\right)\right\},\end{aligned}$$

where $p(c'\tilde{\mathbf{a}}_{b,\kappa}^{(i)}|\mathbf{b}_\kappa^{(i)})$ is $p(\mathbf{a}_b|\mathbf{b})$ evaluated at $\mathbf{a}_b=c'\tilde{\mathbf{a}}_{b,\kappa}^{(i)}$ and $\mathbf{b}=\mathbf{b}_\kappa^{(i)}$, and $p(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}/c')$ is $p(\boldsymbol{\gamma})$ evaluated at $\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}/c'$. Inserting (8) and (10) yields

$$\begin{aligned}c&=\operatorname*{arg\,min}_{c'\in\mathbb{R}^+}\left\{\frac{(c')^2\,\left\|\tilde{\mathbf{a}}_{b,\kappa}^{(i)}\right\|^2}{\sigma_a^2}+\frac{1}{(c')^2}\left(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}\right)^{\mathrm{H}}\boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}\right\}\\&=\left(\frac{\sigma_a^2}{\left\|\tilde{\mathbf{a}}_{b,\kappa}^{(i)}\right\|^2}\left(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}\right)^{\mathrm{H}}\boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}\right)^{1/4},\end{aligned}\tag{34}$$

with $(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)})^{\mathrm{H}}\boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}=\sum_{m=1}^{M}\left|(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)})_m\right|^2/\lambda_m$.

Finally, the shift $\kappa\in\{-\kappa_{\max},\ldots,\kappa_{\max}\}$ for which the vectors $\mathbf{b}_\kappa^{(i)}$, $\mathbf{a}_{b,\kappa}^{(i)}$, and $\boldsymbol{\gamma}_\kappa^{(i)}$ maximize the joint posterior is determined:

$$\hat{\kappa}\triangleq\operatorname*{arg\,max}_{\kappa\in\{-\kappa_{\max},\ldots,\kappa_{\max}\}}\ p\Big(\mathbf{b}_\kappa^{(i)},\mathbf{a}_{b,\kappa}^{(i)},\boldsymbol{\gamma}_\kappa^{(i)},\left(\sigma_n^2\right)^{(i-1)}\,\Big|\,\mathbf{x}\Big).$$

The corresponding vectors $\mathbf{b}_{\hat{\kappa}}^{(i)}$, $\mathbf{a}_{b,\hat{\kappa}}^{(i)}$, and $\boldsymbol{\gamma}_{\hat{\kappa}}^{(i)}$ are the result of the proposed method for scale and shift determination; they are hereafter denoted simply by $\mathbf{b}^{(i)}$, $\mathbf{a}_b^{(i)}$, and $\boldsymbol{\gamma}^{(i)}$.

3) Estimation of the Noise Variance: The noise variance σ_n^2 is estimated at the end of each iteration. The estimate $\left(\sigma_n^2\right)^{(i)}$ is obtained by maximizing the joint posterior in (13), i.e.,

$$\left(\sigma_n^2\right)^{(i)}=\operatorname*{arg\,max}_{\sigma_n^2\in\mathbb{R}^+}\ p\Big(\mathbf{b}^{(i)},\mathbf{a}_b^{(i)},\boldsymbol{\gamma}^{(i)},\sigma_n^2\,\Big|\,\mathbf{x}\Big).$$

Using (11) and (12) in (13), we obtain

$$\begin{aligned}\left(\sigma_n^2\right)^{(i)}&=\operatorname*{arg\,max}_{\sigma_n^2\in\mathbb{R}^+}\left\{\frac{1}{\sigma_n^{2(\xi+1+K)}}\exp\left(-\frac{\eta+\left\|\mathbf{x}-\mathbf{F}_b^{(i)}\mathbf{a}_b^{(i)}\right\|^2}{\sigma_n^2}\right)\right\}\\&=\frac{\eta+\left\|\mathbf{x}-\mathbf{F}_b^{(i)}\mathbf{a}_b^{(i)}\right\|^2}{\xi+K+1},\end{aligned}\tag{35}$$

where $\mathbf{F}_b^{(i)}$ is formed using $\mathbf{b}^{(i)}$ and $\boldsymbol{\gamma}^{(i)}$.

C. Initialization

We propose the following procedure for obtaining initial values for $\mathbf{b}^{(i)}$, $\boldsymbol{\gamma}^{(i)}$, and $\left(\sigma_n^2\right)^{(i)}$ (i.e., for $i=0$). First, the noise variance is initialized with its prior mean:

$$\left(\sigma_n^2\right)^{(0)}=\mathrm{E}\left\{\sigma_n^2\right\}=\frac{\eta}{\xi-1}.$$

Next, we initialize the pulse shape by fitting it around $k_{\max}\triangleq\operatorname{arg\,max}_{k\in\{1,\ldots,K\}}|x_k|$. More specifically, we set $\mathbf{b}^{(0)}=\boldsymbol{\delta}_{k_{\max}}$ and find $\boldsymbol{\gamma}^{(0)}$ by maximizing $p(\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$

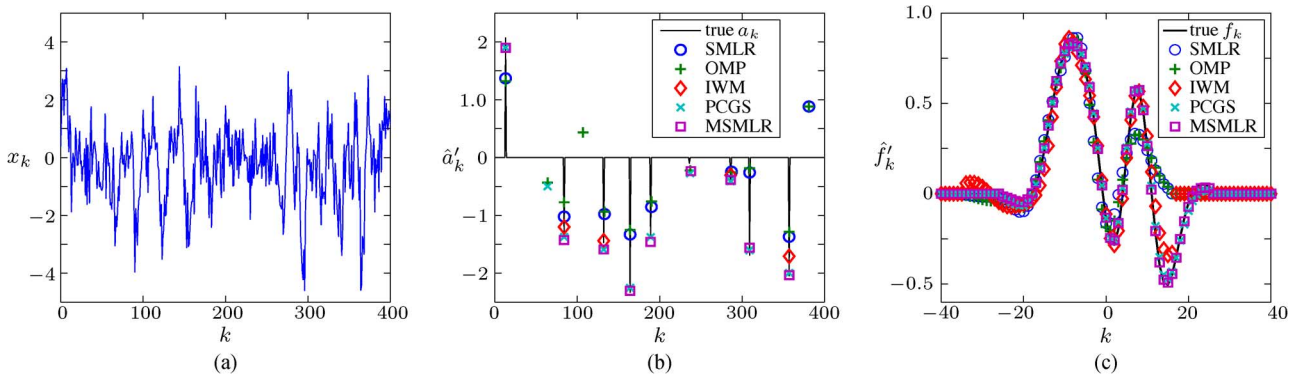

Fig. 7. Results of detection/estimation: (a) Signal $\mathbf{x}$, (b) detected/estimated amplitude sequence $\hat{\mathbf{a}}'$, (c) estimated pulse shape $\hat{\mathbf{f}}'$. The vertical lines in (b) indicate the true $\mathbf{a}$. Real parts are shown. We note that the mismatch of the scales visible in part (b) is a consequence of the minimization of $\|\hat{\mathbf{a}}' - \mathbf{a}\|^2$ (mentioned at the end of Section VI-A) and not a shortcoming of the considered BD methods. In fact, some of the detected pulse locations that visually appear correct in (b) are slightly incorrect. For example, SMLR and OMP detect a large peak at $k = 133$ instead of the true $k = 132$ (detected correctly by the other methods). Through our minimization of $\|\hat{\mathbf{a}}' - \mathbf{a}\|^2$, such location errors and other false detections lead to a smaller scale of $\hat{\mathbf{a}}'$.

with $\boldsymbol{\delta}_{k_{\max}}$, $x_{k_{\max}}$, and $(\sigma_n^2)^{(0)}$ inserted for $\mathbf{b}$, $\mathbf{a}_b$, and σ_n^2, respectively. (Note that because $\mathbf{b} = \boldsymbol{\delta}_{k_{\max}}$, we have $L = 1$ and thus $\mathbf{a}_b$ reduces to a scalar.) Using (31), we obtain

$$\boldsymbol{\gamma}^{(0)} = \hat{\boldsymbol{\gamma}}_{\mathrm{MAP}}\Big(\mathbf{x}, \boldsymbol{\delta}_{k_{\max}}, x_{k_{\max}}, \big(\sigma_n^2\big)^{(0)}\Big)$$
$$= x^*_{k_{\max}} \left(|x_{k_{\max}}|^2 \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}} + \big(\sigma_n^2\big)^{(0)} \boldsymbol{\Lambda}^{-1}\right)^{-1} \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{x}},$$

where $\tilde{\mathbf{x}}$ denotes $\mathbf{x}$ truncated to the entries x_k for $k \in \{\max\{1, k_{\max} - Q\}, \ldots, \min\{K, k_{\max} + Q\}\}$ and $\tilde{\mathbf{H}}$ denotes $\mathbf{H}$ truncated to the rows corresponding to f_k for $k \in \{\max\{-Q, 1 - k_{\max}\}, \ldots, \min\{Q, K - k_{\max}\}\}$ (cf. (4)). The simple initialization $\mathbf{b}^{(0)} = \boldsymbol{\delta}_{k_{\max}}$ is possible because, as shown by simulations, the initialization of $\mathbf{b}^{(i)}$ is less critical than that of $\boldsymbol{\gamma}^{(i)}$, as long as $\mathbf{b}^{(0)} \in \mathcal{C}$.

VI. Numerical Study

We compare the performance of the proposed MSMLR-based BD method with that of the classical SMLR-based BD method reviewed in Sections III-A and V-A. These methods will be briefly referred to as MSMLR and SMLR, respectively. As further performance benchmarks, we also consider orthogonal matching pursuit (OMP) [49], the IWM algorithm proposed in [38], and the partially collapsed Gibbs sampler method proposed in [24] (abbreviated PCGS). We note that IWM according to [38] can be implemented in a wide variety of ways. We used the configurations and parameter values suggested in [38], but did not perform any training or tuning of parameters. Because all these methods use the joint posterior to some extent, their results are influenced by the minimum distance constraint. However, only MSMLR and PCGS exploit the constraint at the algorithmic level.

A. Simulation Setup

We generated several hundred realizations of $\mathbf{x}$ using different values of $d_{\min}$ and of the signal-to-noise ratio (SNR) $\|\mathbf{Fa}\|^2/\|\mathbf{n}\|^2$. The random parameters were drawn from the priors specified in Section II-B, with $K = 1024$, $Q = 25$, $M = 12$, $\pi_1 = 0.15$, $\sigma_a^2 = 10$, and $\eta = 0.5$. For each realization of $\boldsymbol{\gamma}$ and $\mathbf{a}$, the noise $\mathbf{n}$ was scaled such that a given SNR was achieved. The hyperparameter ξ (used by the MSMLR and PCGS) was chosen such that the prior mean of σ_n^2 equals the approximate mean noise power corresponding to the respective SNR, i.e., $\mathrm{E}\{\sigma_n^2\} = \eta/(\xi - 1)$ equals $\mathrm{E}\{\|\mathbf{Fa}\|^2\}/(K \cdot \mathrm{SNR}) = \bar{\pi}_1\sigma_a^2(\sum_{m=1}^M \lambda_m)/\mathrm{SNR}$ (note that $\bar{\pi}_1$ was defined in Footnote 2 in Section II-B). In the OMP, IWM, and SMLR methods, which do not include estimation of σ_n^2, the approximate mean noise power was used for $\widehat{\sigma_n^2}$. The orthonormal Hermite basis functions $\mathbf{h}_m$ used for representing the pulse shape $\mathbf{f}$ according to (4) were time-scaled such that the maximum entry of $\mathbf{h}_1$ was 0.31. For the prior variances of the pulse coefficients γ_m, we used $\lambda_m = 0.48[\cos(\pi(m-1)/M)]^4$. While this particular choice is largely arbitrary, the key characteristic is that the λ_m decrease with increasing m, as explained in Section II-B. The resulting expected squared magnitude of the pulse shape, $\mathrm{E}\{|f_k|^2\} = \sum_{m=1}^M \lambda_m |h_{m,k}|^2$, was shown in Fig. 1(a) (solid line), and some realizations of f_k used in the simulations were shown in Fig. 1(b). Note that the pulses f_k are well concentrated around the time and frequency origins.

For each realization of $\mathbf{x}$, we performed BD using the MSMLR, SMLR, OMP, IWM, and PCGS methods. As mentioned in Section V-B, the time shift and amplitude scale of the estimate $\hat{\mathbf{a}}$ are arbitrary and typically irrelevant. Therefore, for a performance assessment, we matched the time shift and amplitude scale of each estimate $\hat{\mathbf{a}}$ and $\hat{\mathbf{f}}$ to the true $\mathbf{a}$ and $\mathbf{f}$, i.e., we calculated the shifted and scaled version $\hat{\mathbf{a}}'$ of $\hat{\mathbf{a}}$ minimizing $\|\hat{\mathbf{a}}' - \mathbf{a}\|^2$, and the analogously shifted/scaled $\hat{\mathbf{f}}'$.

B. Performance Evaluation

As an example, Fig. 7 shows the results of one simulation run corresponding to a single realization of $\mathbf{x}$ with $d_{\min} = 20$ and $\mathrm{SNR} = 6$ dB. A segment of $\mathbf{x}$ and $\mathbf{a}$ of length 400 is displayed. Since $d_{\min} < 2Q + 1$, pulses may overlap. It can be concluded from Fig. 7(b) that the detected/estimated sequences $\hat{\mathbf{a}}'$ are reasonably accurate for all methods. However, OMP, SMLR, and PCGS produce some false detections, and IWM misses some nonzero positions in a_k (within the displayed segment). Therefore, the result of the proposed MSMLR is clearly the best. Fig. 7(c) shows that all methods estimate the main peak of the pulse shape $\mathbf{f}$ quite accurately, whereas some of the smaller peaks are estimated accurately only by MSMLR, PCGS, and IWM.

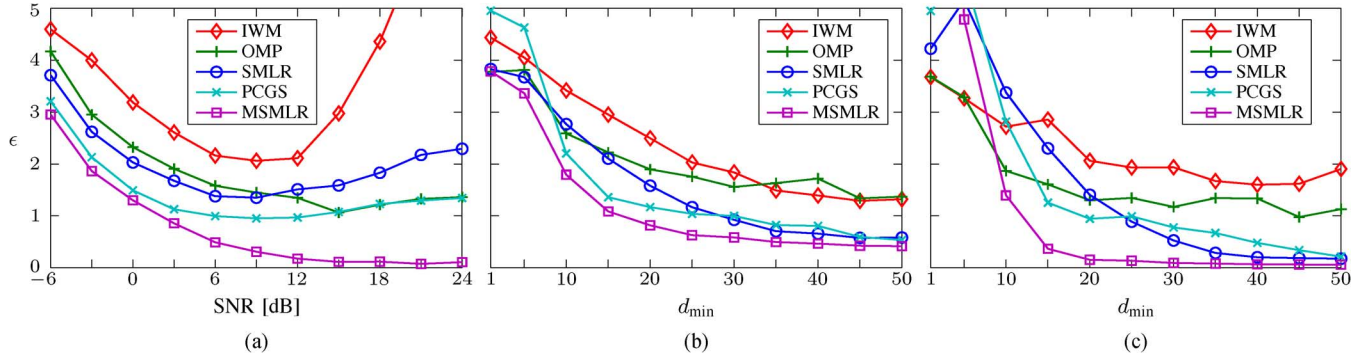

Fig. 8. Average error ϵ: (a) versus SNR, for $d_{\min} = 20$; (b) versus $d_{\min}$, for SNR $= 3$ dB; (c) versus $d_{\min}$, for SNR $= 12$ dB.

Our simulations have shown that the most critical task in the considered BD methods is the detection of $\mathbf{b}$, i.e., of the positions of the nonzero entries in the sparse sequence. If these positions are detected correctly, the other parameters are usually estimated with high accuracy. If the detected $\hat{\mathbf{b}}$ is grossly wrong, the estimate of $\boldsymbol{\gamma}$ may still be good, but the estimate of $\mathbf{a}$ becomes meaningless. Therefore, our performance analysis will focus on the accuracy of $\hat{\mathbf{b}}$. To quantify the difference between $\hat{\mathbf{b}}$ and the true sequence $\mathbf{b}$, we compare the set $\{\hat{k}_l\}_{l=1,\dots,\hat{L}}$ of the positions of the detected nonzero entries contained in $\hat{\mathbf{b}}$ with the set $\{k_l\}_{l=1,\dots,L}$ of the positions of the true nonzero entries contained in $\mathbf{b}$. Let ϖ denote a permutation of the larger of the two sets (or either one of them if $L = \hat{L}$). Furthermore, we define

$$d_l^{(\varpi)} \triangleq \begin{cases} |k_l - \hat{k}_{\varpi(l)}| & \text{if } L \le \hat{L} \\ |k_{\varpi(l)} - \hat{k}_l| & \text{if } L \ge \hat{L}, \end{cases}$$

and we consider the set $\mathcal{L}^{(\varpi)}$ of all l such that $d_l^{(\varpi)} \le \Delta$, with some fixed Δ (in our simulations, we chose $\Delta = 5$). Then, for a given permutation ϖ, the cardinality $L_{\text{correct}} \triangleq |\mathcal{L}^{(\varpi)}|$ is the number of true positions k_l that are detected "almost correctly" (i.e., with position error at most Δ), whereas $L_{\text{missed}} \triangleq L - L_{\text{correct}}$ can be considered as the number of true positions that are missed by the detector and $L_{\text{false}} \triangleq \hat{L} - L_{\text{correct}}$ can be considered as the number of false detections. An error measure that takes into account the deviations $d_l^{(\varpi)}$ for $l \in \mathcal{L}^{(\varpi)}$ and adds penalty terms for missed positions and false detections using the "best" permutation ϖ can then be defined as

$$\epsilon \triangleq \frac{1}{L} \min_{\varpi} \left\{ \sum_{l \in \mathcal{L}^{(\varpi)}} d_l^{(\varpi)} + (L_{\text{missed}} + L_{\text{false}})\Delta \right\}.$$

If ϵ is close to or above Δ, the detection of $\mathbf{b}$ can be considered as failed. For example, a degenerate detector that produces only zeros achieves $\epsilon = \Delta$. However, if L_{false} is large (i.e., $\hat{\mathbf{b}}$ is not sparse), ϵ can be much larger than Δ. We note that ϵ is similar to a simple special case of the optimal subpattern assignment (OSPA) metric [50].

For $d_{\min} = 20$ and different SNRs, Fig. 8(a) shows the error ϵ averaged over 500 realizations. It can be seen that MSMLR performs consistently better than all the reference methods. It is followed by PCGS, with an average ϵ that is larger than that of MSMLR by a factor of 3.4 at SNR $= 9$ dB. The average ϵ of SMLR is larger than that of MSMLR by a factor of 4.6 at SNR $= 9$ dB. The performance gain of MSMLR grows with the SNR. The fact that the errors of SMLR and IWM are larger for high SNR can be explained by increased overfitting: at lower noise levels, the algorithms may more often try to compensate slightly misplaced 1's (or other estimation inaccuracies) from previous iterations by adding more 1's, and the resulting decrease in sparsity leads to a higher ϵ. Compared to SMLR, MSMLR avoids this problem since its modified hypotheses often allow it to correct misplaced 1's more effectively. Figs. 8(b) and (c) show the average ϵ at an SNR of 3 dB and 12 dB, respectively, for different values of $d_{\min}$. In view of the fixed pulse length $2Q + 1 = 51$, the range of $d_{\min}$ considered in Figs. 8(b) and (c) corresponds to different maximum overlap of the pulses in the observed signal, from negligible overlap at $d_{\min} = 50$ to massive overlap at $d_{\min} = 1$. It can be observed in Figs. 8(b) and (c) that MSMLR performs consistently better than all the reference methods except for the case of small $d_{\min}$ and high SNR. We can conclude that overfitting also occurs in MSMLR when $d_{\min}$ is very small. The performance gain of MSMLR over SMLR is largest when $d_{\min}$ is larger than 10 but still small enough to allow for strong overlap of the pulses. The superior performance of MSMLR compared to the other SMLR-type methods (SMLR and IWM) demonstrates that MSMLR is able to cope with and even exploit the minimum distance constraint.

Normalized root mean square (RMS) values of the estimation errors $\|\hat{\mathbf{a}}' - \mathbf{a}\|$ and $\|\hat{\mathbf{f}}' - \mathbf{f}\|$ are shown for $d_{\min} = 20$ and different SNRs in Fig. 9(a) and (b), respectively. As mentioned before, the estimation of $\mathbf{a}$ and $\mathbf{f}$ is not as critical for the considered BD methods as the detection of $\mathbf{b}$. The results are qualitatively similar to those of Fig. 8(a). MSMLR again performs consistently better than all the reference methods.

Fig. 10 studies a particularly difficult scenario where the true pulse has only one peak and is much broader than would be typical for our pulse model (which is assumed by all methods). The results resemble those of Fig. 8(c), but the $d_{\min}$-axis roughly scales with the pulse width, and the estimation errors ϵ are generally increased. MSMLR outperforms the other methods for $d_{\min}$ larger than 16; for smaller $d_{\min}$, all methods produce very large errors (often even larger than $\Delta = 5$). Fig. 10 also shows that two pulses separated by a time distance of 20 (or less) can no longer be resolved visually.

For a rough assessment and comparison of the computational complexities of the five methods, we report the computation

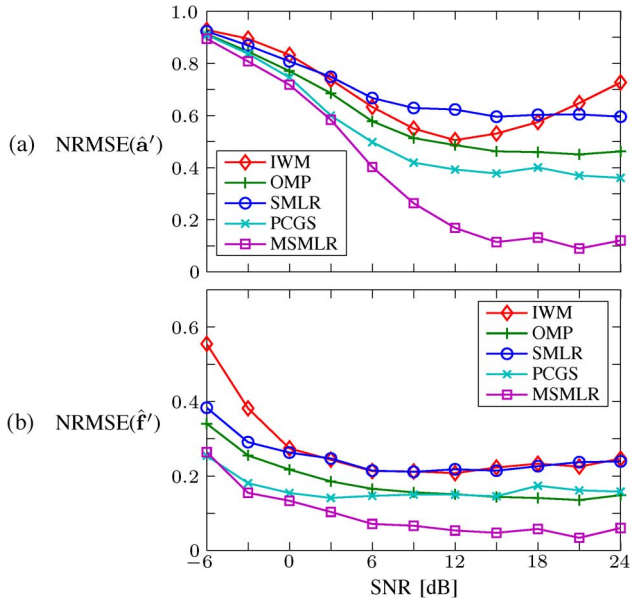

Fig. 9. Normalized RMS value of (a) the amplitude estimation error $\|\hat{\mathbf{a}}' - \mathbf{a}\|$ and (b) the pulse shape estimation error $\|\hat{\mathbf{f}}' - \mathbf{f}\|$ versus SNR, for $d_{\min} = 20$.

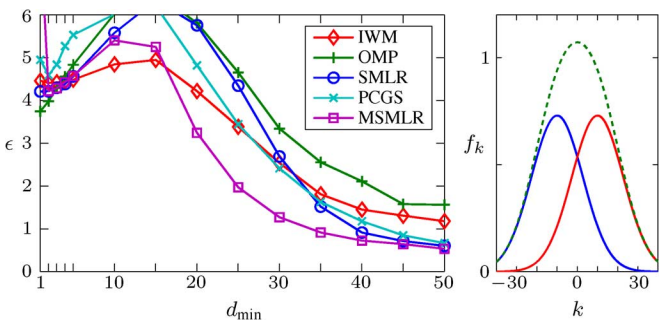

Fig. 10. Left plot: Average error ϵ versus $d_{\min}$ at SNR $= 12$ dB, for the case where the true pulse is Gaussian-shaped with a half-height width of 30 whereas all methods assume the original pulse model. Right plot: Two such pulses shifted relatively to each other by 20 (solid lines) and their superposition (dashed line).

times for one BD task required by nonoptimized MATLAB R2011b 64-bit implementations on a 2.8 GHz Intel Core i7 processor. For $d_{\min} = 20$ and SNR $= 9$ dB, we obtained 1.98 s for MSMLR, 5.00 s for SMLR, 11.68 s for IWM, 0.54 s for OMP, and 4.48 s for PCGS. Hence, within the implementations used, MSMLR is less complex than all the reference methods except OMP.

C. Numerical Validation of M2

The algorithmic modification M2 proposed in Section III-B2 was motivated by a reduction of complexity: on average, M2 reduced the computation time of MSMLR by a factor of about 3.5 in our simulations. To examine how M2 affects the estimation performance, we compare MSMLR (including M2) and SMLR with the two alternative methods given by *SMLR with M2* and *MSMLR without M2*, using the same data that were used for Fig. 8. Fig. 11 shows the error ϵ averaged over 500 realizations at SNR $= 3$ dB and at SNR $= 12$ dB, for different values of $d_{\min}$. For MSMLR, we can see that the results with

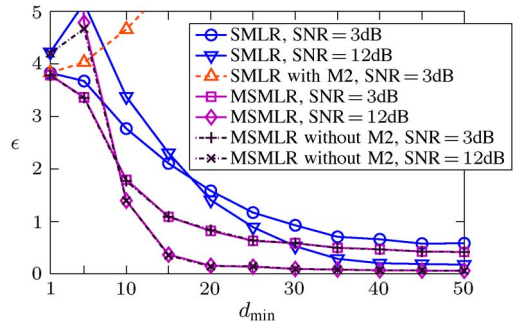

Fig. 11. Average error ϵ versus $d_{\min}$, for MSMLR/SMLR with and without M2 and at two different SNRs (3 dB and 12 dB). SMLR with M2 at SNR $= 12$ dB is not shown because its error is consistently larger than 5.

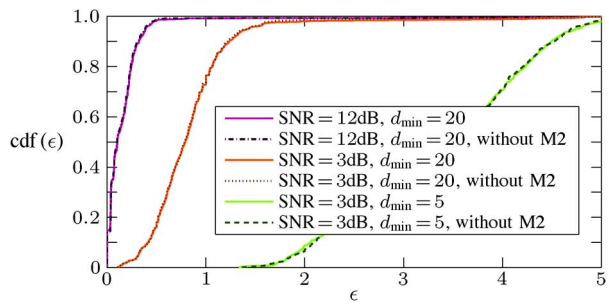

Fig. 12. Empirical cdf of ϵ for MSMLR with and without M2.

and without M2 are nearly indistinguishable for all combinations of SNR and $d_{\min}$, except for $d_{\min} = 1$ at high SNR. We can conclude that the average performance of MSMLR is not affected by M2 at all, as long as there is a minimum distance constraint. On the other hand, SMLR with M2 fails completely. These results corroborate the considerations in Section III-B2: M2 requires that the hypothesis set contains hypotheses with shifted 1's, which is the case for MSMLR with $d_{\min} \geq 2$ but not for SMLR or for MSMLR with $d_{\min} = 1$.

For a more detailed analysis beyond average performance, we also consider the empirical cumulative distribution function (cdf) of ϵ calculated from 500 realizations, for given SNR and $d_{\min}$. Fig. 12 shows three examples of these cdfs for MSMLR with and without M2. We can see that the cdfs with and without M2 are nearly identical. Furthermore, besides the error ϵ of the estimate $\hat{\mathbf{b}}$, we can also use the posterior probability of $\hat{\mathbf{b}}$ as a quality measure. More specifically, we evaluated $\beta \triangleq p(\hat{\mathbf{b}}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})/p(\mathbf{b}, \hat{\boldsymbol{\gamma}}, \widehat{\sigma_n^2}|\mathbf{x})$, i.e., the posterior probability of $\hat{\mathbf{b}}$ normalized by that of the true $\mathbf{b}$. Again, our simulations (not shown) demonstrated that the distributions with and without M2 are nearly identical.

Finally, one may ask how often the estimates $\hat{\mathbf{b}}$ obtained from MSMLR with and without M2 are not only statistically equivalent with respect to performance but in fact identical. Fig. 13 shows the empirical rate ϱ of identical estimates, i.e., the number of realizations that yield identical estimates normalized by the total number of realizations, calculated from 500 realizations, for different SNRs and minimum distances. We can see that the estimates are generally identical when $d_{\min}$ is large, i.e., when pulses are not allowed to overlap significantly. For

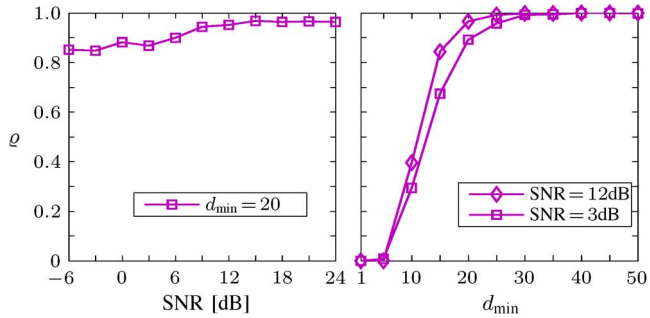

Fig. 13. Empirical rate of identical estimates $\hat{\mathbf{b}}$ for MSMLR with and without M2.

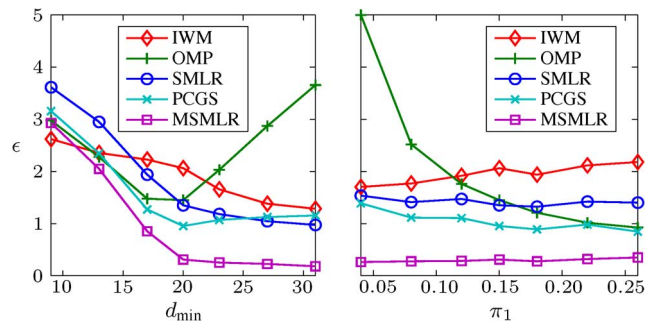

Fig. 14. Average error ϵ for different mismatches of $d_{\min}$ and π_1, at SNR = 9 dB. Left plot: $\pi_1 = 0.15$; all methods erroneously assume that $d_{\min} = 20$. Right plot: $d_{\min} = 20$; all methods erroneously assume that $\pi_1 = 0.15$.

small $d_{\min}$, the estimates are typically not identical although their errors have the same distribution.

D. Robustness Evaluation

Fig. 14 analyzes the robustness of the BD methods to deviations of the parameters $d_{\min}$ and π_1, which determine the sparsity of $\mathbf{a}$, from nominal values $d_{\min} = 20$ and $\pi_1 = 0.15$ assumed by the BD methods. Not surprisingly, we can see that the performance of all BD methods quickly deteriorates when the true minimum distance $d_{\min}$ is smaller than the nominal value $d_{\min} = 20$, i.e., when the BD methods assume a too restrictive minimum distance constraint. On the other hand, it may seem counterintuitive that in some methods—including MSMLR—the error further decreases when the true $d_{\min}$ increases beyond the nominal value. Here, the positive effect of higher sparsity (cf. Figs. 8(b) and (c)) outweighs the negative effect of the mismatch. Within a reasonable range around the nominal value, MSMLR outperforms the reference methods. Regarding the parameter π_1, we see that all methods are fairly robust to deviations from the nominal value.

To further investigate the robustness of the BD methods, we generated 500 realizations using a pulse originally used in [38]. This pulse does not fit our pulse model parameters of Section VI-A because the effective bandwidth of our basis vectors $\mathbf{h}_m$ is too small. The pulse and its projection onto the subspace spanned by the $\mathbf{h}_m$ are shown in Fig. 15(a). All BD methods still assumed the parameter priors and values of Section VI-A. Fig. 15(b) shows (upon comparison with Fig. 8(a)) that this mismatch significantly degrades the performance of all methods. However, MSMLR still tends to

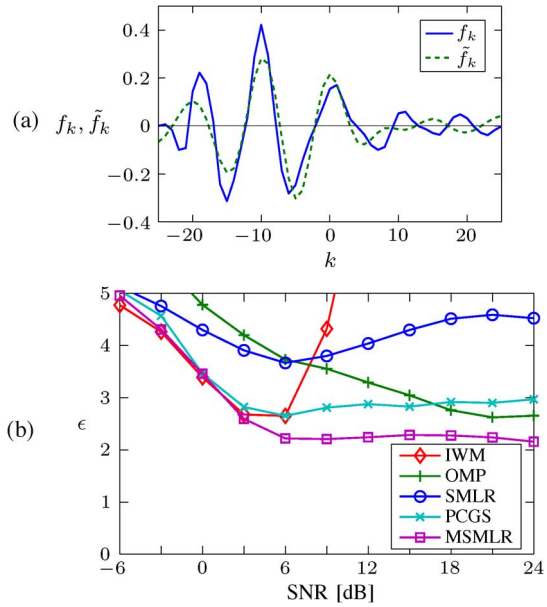

Fig. 15. Pulse model mismatch: (a) Pulse f_k from [38] and its projection $\tilde{f}_k$ onto the subspace spanned by the basis vectors $\mathbf{h}_m$. (b) Average error ϵ versus SNR for $d_{\min} = 20$, for the case where the true pulse is f_k whereas the estimate of f_k obtained by all methods may at best be the projection $\tilde{f}_k$.

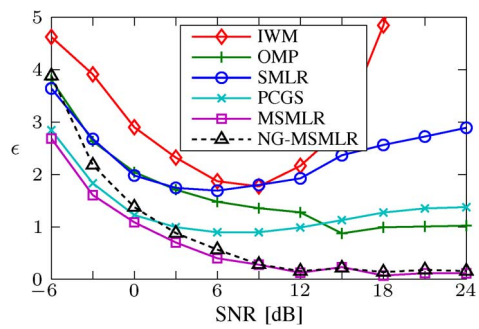

Fig. 16. Average error ϵ versus SNR for $d_{\min} = 20$, for the case where the true prior of $\mathbf{a}_b$ is uniform whereas all methods assume a Gaussian prior. The NG-MSMLR method is discussed in Section VI-E1.

outperform the reference methods, especially for SNR larger than about 3 dB.

Finally, we consider a mismatch of the prior distribution of $\mathbf{a}_b$. Fig. 16 shows the results of simulations where both the real part and the imaginary part of $\mathbf{a}_b$ are uniformly distributed on $[-\sqrt{3/2}\sigma_a, \sqrt{3/2}\sigma_a]^L$, whereas the BD methods still assume the Gaussian prior (8). A comparison with Fig. 8(a) shows that all methods are fairly robust to this mismatch.

E. Further Aspects

1) Non-Gaussian BD: One advantage of the Gaussian prior for $\mathbf{a}_b$ is that it helps avoid numerical problems caused by a badly conditioned matrix $\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b$. However, a minimum distance constraint tends to improve the conditioning of $\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b$, so that the assumption of a Gaussian prior for $\mathbf{a}_b$ may no longer be necessary. In fact, one can use a maximum-likelihood (ML)-like approach in which the factor $p(\mathbf{a}_b|\mathbf{b})$ is removed from the joint posteriors (13) and (15). The resulting MSMLR objective function $q(\mathbf{b})$ is given by (25) with σ_a^2 replaced by 1 and $\mathbf{\Phi}_b$

replaced by chol $\left\{\frac{1}{\sigma_n^2}\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b\right\}$. The $\mathbf{a}_b$ maximizing the modified joint posterior is then no longer $\hat{\mathbf{a}}_{b,\mathrm{MAP}}(\mathbf{x},\mathbf{b},\boldsymbol{\gamma},\sigma_n^2)$ in (30) but $\hat{\mathbf{a}}_{b,\mathrm{ML}}(\mathbf{x},\mathbf{b},\boldsymbol{\gamma},\sigma_n^2) = (\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b)^{-1}\mathbf{F}_b^{\mathrm{H}}\mathbf{x}$. The expression of c in (34) does not have a natural counterpart in this case; we replace it (somewhat arbitrarily) by $c \triangleq \sqrt{(\tilde{\boldsymbol{\gamma}}_\kappa^{(i)})^{\mathrm{H}}\boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{\gamma}}_\kappa^{(i)}/N}$. Apart from these modifications, the algorithm stays the same. The minimum distance constraint (with a sufficiently large $d_{\min}$) then avoids a potential bad conditioning of $\mathbf{F}_b^{\mathrm{H}}\mathbf{F}_b$. This ML-like algorithm will be referred to as non-Gaussian MSMLR or briefly NG-MSMLR. To assess its performance, Fig. 16 shows its average error in simulations where the prior of $\mathbf{a}_b$ is in fact not Gaussian—more specifically, it is uniform as described at the end of Section VI-D—but has the same variance σ_a^2 as that assumed by the BD methods (except by NG-MSMLR, which does not require knowledge of σ_a^2). We can see that the performance of NG-MSMLR is similar to that of MSMLR. At lower SNR, MSMLR obtains better results due to its knowledge of σ_a^2.

2) Nonblind BD: Further simulations showed that MSMLR outperforms the reference methods even in the nonblind case, since the minimum distance constraint still helps avoid overfitting. However, in this simpler estimation problem, the performance advantage of MSMLR is smaller: for example, at SNR $=$ 3 dB and $d_{\min} = 20$, the average error ϵ is 0.68 for MSMLR, 0.86 for PCGS, and 1.00 for SMLR.

VII. Conclusion

We presented an SMLR-type algorithm for Bayesian blind deconvolution of an unknown sparse sequence convolved with an unknown pulse. The algorithm is based on an extended Bernoulli-Gaussian prior that incorporates a hard minimum distance constraint. This prior simultaneously induces sparsity and enforces a prescribed minimum distance between any two detected pulse locations, which is an effective means of avoiding spurious detected pulses.

Compared to the classical SMLR approach to blind deconvolution, the proposed algorithm introduces several modifications that overcome certain known weaknesses of SMLR and result in improved performance and efficiency. Based on the minimum distance constraint, the hypotheses assessed in each iteration are chosen such that the risk of converging to a local optimum is greatly reduced; this is achieved without increasing the number of hypotheses. The minimum distance constraint is also exploited for an efficient calculation of the probabilities associated with the hypotheses. The computational efficiency is further increased through a novel scheduling of different steps of the algorithm. Finally, the explicit determination of an optimal amplitude scale and time shift makes it possible to take into account prior knowledge about the time-frequency concentration of the unknown pulse shape via a basis expansion, which results in improved estimation performance. Our simulation results demonstrated the advantages of the proposed method over several state-of-the-art methods for Bayesian blind deconvolution, regarding both detection/estimation performance and computational complexity.

Possible directions for future research include extensions of the proposed method in which additional hyperparameters (besides σ_n^2) are estimated from the observed sequence $\mathbf{x}$. Estimating π_1, σ_a^2, ξ, η, and/or $d_{\min}$ rather than using fixed values can be expected to result in increased robustness and an easier deployment of the method.

Acknowledgment

The authors would like to thank the anonymous reviewers for numerous helpful comments, which have led to improvements of this paper.

References

[1] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 331–344, Nov. 1980.
[2] J. R. Treichler and B. G. Agee, "A new approach to the multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 1867–1875, Feb. 1983.
[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "Least squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, pp. 2982–2993, Dec. 1995.
[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, pp. 516–525, Feb. 1995.
[5] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky, "Relative optimization for blind deconvolution," *IEEE Trans. Signal Process.*, vol. 53, pp. 2018–2026, Jun. 2005.
[6] J. M. Mendel and J. Goutsias, "One-dimensional normal-incidence inversion: A solution procedure for band-limited and noisy data," *Proc. IEEE*, vol. 74, pp. 401–414, Mar. 1986.
[7] J. Idier and Y. Goussard, "Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, pp. 975–978, Sep. 1990.
[8] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, pp. 377–384, Mar. 1996.
[9] O. Rosec, J.-M. Boucher, B. Nsiri, and T. Chonavel, "Blind marine seismic deconvolution using statistical MCMC methods," *IEEE J. Ocean. Eng.*, vol. 8, pp. 502–414, July 2003.
[10] D. Ge, E. Le Carpentier, J. Idier, and D. Farina, "Spike sorting by stochastic simulation," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 19, pp. 249–259, June 2011.
[11] D. Ge, E. Le Carpentier, and D. Farina, "Unsupervised Bayesian decomposition of multi-unit EMG recordings using Tabu search," *IEEE Trans. Biomed. Eng.*, vol. 56, pp. 1–9, Dec. 2009.
[12] G. Kail, C. Novak, B. Hofer, and F. Hlawatsch, "A blind Monte Carlo detection-estimation method for optical coherence tomography," in *Proc. IEEE ICASSP*, Taipei, Taiwan, Apr. 2009, pp. 493–496.
[13] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 2840–2849, Dec. 2010.
[14] S. Bourguignon and H. Carfantan, "Spectral analysis of irregularly sampled data using a Bernoulli-Gaussian model with free frequencies," in *Proc. IEEE ICASSP*, Toulouse, France, May 2006, pp. 516–519.
[15] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Top. Signal Process.*, vol. 1, pp. 575–585, Dec. 2007.
[16] M.-H. Chen and J. J. Deely, "Bayesian analysis for a constrained linear multiple regression problem for predicting the new crop of apples," *J. Agricult. Biolog. Environ. Statist.*, vol. 1, pp. 467–489, Dec. 1996.
[17] G. A. Rodriguez-Yam, R. A. Davis, and L. L. Scharf, "A Bayesian model and Gibbs sampler for hyperspectral imaging," in *Proc. IEEE SAM*, Washington, DC, USA, Aug. 2002, pp. 105–109.
[18] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. Signal Process.*, vol. 54, pp. 4133–4145, Nov. 2006.
[19] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, pp. 2657–2669, Dec. 2009.
[20] C. Févotte, B. Torrésani, L. Daudet, and S. J. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, pp. 174–185, Jan. 2008.

[21] T. Blumensath and M. E. Davies, "Monte-Carlo methods for adaptive sparse approximations of time-series," *IEEE Trans. Signal Process.*, vol. 55, pp. 4474–4486, Sep. 2007.
[22] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Process.*, vol. 18, pp. 1215–1227, June 2009.
[23] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, pp. 2675–2685, May 2010.
[24] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: A partially collapsed Gibbs sampler method," *IEEE Trans. Signal Process.*, vol. 60, pp. 2727–2743, Jun. 2012.
[25] K. Hausmair, K. Witrisal, P. Meissner, C. Steiner, and G. Kail, "SAGE algorithm for UWB channel parameter estimation," in *Proc. COST 2100 Manag. Committee Meeting*, Athens, Greece, Feb. 2010.
[26] G. Kail, K. Witrisal, and F. Hlawatsch, "Direction-resolved estimation of multipath parameters for UWB channels: A partially collapsed Gibbs sampler method," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 3484–3487.
[27] C. Dossal and S. Mallat, "Sparse spike deconvolution with minimum scale," in *Proc. Signal Process. With Adapt. Sparse Struct. Represent. Workshop*, Rennes, France, Nov. 2005.
[28] C. P. Robert, *The Bayesian Choice*. New York, NY, USA: Springer, 1996.
[29] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 482–488, May 1982.
[30] D. Ge, J. Idier, and E. Le Carpentier, "Enhanced sampling schemes for MCMC based blind Bernoulli-Gaussian deconvolution," *Signal Process.*, vol. 91, pp. 759–772, Apr. 2011.
[31] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *IEEE Trans. Signal Process.*, vol. 44, pp. 2988–2998, Dec. 1996.
[32] D. A. van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *J. Amer. Statist. Assoc.*, vol. 103, pp. 790–796, June 2008.
[33] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation Approach*. New York, NY, USA: Academic, 1983.
[34] C. Andrieu, A. Doucet, and S. J. Godsill, "Bayesian blind marginal separation of convolutively mixed discrete sources," in *Proc. IEEE NNSP*, Cambridge, U.K., Aug.–Sep. 1998, pp. 43–52.
[35] Y. Goussard, M. Demoment, and J. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," in *Proc. IEEE ICASSP*, Albuquerque, NM, USA, Apr. 1990, pp. 1547–1550.
[36] K. F. Kaaresen, "Deconvolution of sparse spike trains by iterated window maximization," *IEEE Trans. Signal Process.*, vol. 45, pp. 1173–1183, May 1997.
[37] C. Y. Chi and J. M. Mendel, "Improved maximum-likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inf. Theory*, vol. 30, pp. 429–435, 1984.
[38] K. F. Kaaresen and T. Taxt, "Multichannel blind deconvolution of seismic signals," *Geophys.*, vol. 63, pp. 2093–2107, Nov. 1998.
[39] M. Allain and J. Idier, "Efficient binary reconstruction for non-destructive evaluation using gammagraphy," *Inverse Problems*, vol. 23, pp. 1371–1393, Aug. 2007.
[40] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 59, pp. 4572–4584, Oct. 2011.
[41] S. Bourguignon and H. Carfantan, "Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data," in *Proc. IEEE SSP*, Bordeaux, France, Jul. 2005, pp. 811–816.
[42] R. Haas and J.-C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Personal Commun.*, vol. 5, pp. 1–18, 1997.
[43] L. Sörnmo, P. Börjesson, M. Nygärds, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. 28, pp. 713–717, 1981.
[44] A. J. E. M. Janssen, "Positivity and spread of bilinear time-frequency distributions," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds. Amsterdam, The Netherlands: Elsevier, 1997, pp. 1–58.
[45] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces: Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation*. Boston, MA, USA: Kluwer, 1998.
[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[47] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[48] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc., Ser. B*, vol. 48, pp. 259–302, 1986.
[49] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.
[50] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, pp. 3447–3457, Aug. 2008.

Georg Kail (M'14) received the B.Sc. and Diplom-Ingenieur (M.Sc.) degrees in electrical engineering/telecommunications and the Dr. techn. (Ph.D.) degree in signal processing from Vienna University of Technology, Vienna, Austria, in 2005, 2008, and 2012, respectively. During his master and doctoral studies, he visited UTB Zlín, Czech Republic, ETH Zürich, Switzerland, and ENSEEIHT, Toulouse, France as a short-term guest researcher. From 2008 to 2013, and again since 2015, he has been with the Institute of Telecommunications, Vienna University of Technology, where he is currently a postdoctoral research assistant. During 2013–2014, he spent eight months as a postdoctoral researcher with the Telecommunications Circuits Laboratory, EPFL, Lausanne, Switzerland. Subsequently, as a recipient of an Erwin Schrödinger Fellowship, he was with the Circuits and Systems Group, Delft University of Technology, Delft, The Netherlands, for one year. His research interests include statistical signal processing with a focus on Bayesian methods and their application to localization tasks.

Franz Hlawatsch (S'85–M'88–SM'00–F'12) received the Diplom-Ingenieur, Dr. techn., and Univ.-Dozent (habilitation) degrees in electrical engineering/signal processing from Vienna University of Technology, Vienna, Austria in 1983, 1988, and 1996, respectively. Since 1983, he has been with the Institute of Telecommunications, Vienna University of Technology, where he is currently an Associate Professor. During 1991–1992, as a recipient of an Erwin Schrödinger Fellowship, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI, USA. In 1999, 2000, and 2001, he held one-month Visiting Professor positions with INP/ENSEEIHT, Toulouse, France, and IRCCyN, Nantes, France. He (co)authored a book, three review papers that appeared in the IEEE SIGNAL PROCESSING MAGAZINE, about 200 refereed scientific papers and book chapters, and three patents. He coedited three books. His research interests include statistical and compressive signal processing methods and their application to sensor networks and wireless communications.

Prof. Hlawatsch was Technical Program Co-Chair of EUSIPCO 2004 and served on the technical committees of numerous IEEE conferences. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2007 and for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2008 to 2011. From 2004 to 2009, he was a member of the IEEE SPCOM Technical Committee. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. He coauthored papers that won an IEEE Signal Processing Society Young Author Best Paper Award and a Best Student Paper Award at IEEE ICASSP 2011.

Clemens Novak received the Dipl.-Ing. degree in electrical engineering and the Dr. techn. degree in communication engineering from Vienna University of Technology, Austria, in 2006 and 2010, respectively. From 2006 to 2010, he was a Research and Teaching Assistant with the Institute of Communications and Radio Frequency Engineering, Vienna University of Technology. From 2010 to 2014, he was with Kapsch TrafficCom, Vienna, Austria. Since 2014, he has been with Frequentis AG, Vienna, Austria.