# Blind deconvolution of sparse pulse sequences under a minimum distance constraint: a partially collapsed Gibbs sampler method

Georg Kail, Jean-Yves Tourneret, Franz Hlawatsch, Nicolas Dobigeon

# Blind Deconvolution of Sparse Pulse Sequences Under a Minimum Distance Constraint: A Partially Collapsed Gibbs Sampler Method

Georg Kail, Jean-Yves Tourneret, *Senior Member, IEEE*, Franz Hlawatsch, *Fellow, IEEE*, and Nicolas Dobigeon, *Member, IEEE*

***Abstract*—For blind deconvolution of an unknown sparse sequence convolved with an unknown pulse, a powerful Bayesian method employs the Gibbs sampler in combination with a Bernoulli–Gaussian prior modeling sparsity. In this paper, we extend this method by introducing a *minimum distance constraint* for the pulses in the sequence. This is physically relevant in applications including layer detection, medical imaging, seismology, and multipath parameter estimation. We propose a Bayesian method for blind deconvolution that is based on a modified Bernoulli–Gaussian prior including a minimum distance constraint factor. The core of our method is a *partially collapsed Gibbs sampler* (PCGS) that tolerates and even exploits the strong local dependencies introduced by the minimum distance constraint. Simulation results demonstrate significant performance gains compared to a recently proposed PCGS. The main advantages of the minimum distance constraint are a substantial reduction of computational complexity and of the number of spurious components in the deconvolution result.**

***Index Terms*—Bernoulli–Gaussian prior, blind deconvolution, Markov chain Monte Carlo method, partially collapsed Gibbs sampler, sparse deconvolution.**

## I. Introduction

THE problem of blind deconvolution (BD) arises in many applications where some desired signal is to be recovered from a distorted observation, e.g., in digital communications [1]–[5], seismology [6]–[9], biomedical signal processing [10]–[13], and astronomy [14], [15]. The BD problem is ill-posed since different input sequences and impulse responses can provide the same observation. As a consequence, additional assumptions or constraints have to be considered in order to reduce the number of solutions. Examples of constraints that have been considered are monotonicity [16], positivity [17]–[19], and sparsity [20]–[22]. In the Bayesian setting, which will be adopted here, constraints can be modeled through appropriate prior distributions.

In this paper, we propose a Bayesian BD method that is based on a novel *combined sparsity and minimum distance constraint*. We consider a sparse, random sequence $\mathbf{a} = (a_1 \cdots a_K)^{\mathrm{T}} \in \mathbb{C}^K$ convolved with a pulse of unknown shape. For a convenient modeling of the sparsity of $\mathbf{a}$, we use a random binary indicator sequence $\mathbf{b} = (b_1 \cdots b_K)^{\mathrm{T}} \in \{0,1\}^K$, where $b_k = 1$ at each temporal position $k$ where $a_k$ is nonzero and $b_k = 0$ else. The binary indicators thus mark the positions of the weighted replicas of the pulse in the observed sequence. Our goal is to detect the positions of the nonzero $a_k$ (equivalently, to detect the indicators $b_k$) and to estimate their amplitudes (the values of the nonzero $a_k$) as well as the pulse shape.

Most Bayesian BD approaches exploiting sparsity use a Bernoulli–Gaussian prior for the sparse sequence $\mathbf{a}$, i.e., the $b_k$ are independent and Bernoulli distributed and the nonzero $a_k$ are Gaussian distributed [7], [8], [23]–[26]. However, here we will use a modified Bernoulli–Gaussian prior incorporating a hard *minimum distance constraint* that requires the temporal distance between any two nonzero $a_k$ (or two nonzero indicators $b_k = 1$) to be not smaller than some prescribed minimum distance. This is physically relevant in many applications, including layer detection [12], biomedical signal processing [11], [13], seismology, and multipath parameter estimation [27], [28]. We will demonstrate that the minimum distance constraint provides a computationally efficient way to ensure sparsity, avoid overfitting, and improve resolution. To the best of our knowledge, a minimum distance constraint has not been considered previously for BD (prior to our work in [12]).

The proposed BD method is based on a Bayesian strategy using a Markov chain Monte Carlo (MCMC) algorithm, which is a powerful approach for complex problems with a large number of parameters [29], [30]. The Gibbs sampler is a simple and widely used MCMC method with interesting properties for BD [9], [24], [26], [31]; however, it is computationally inefficient when there are strong dependencies among the parameters [23], [32]. Such dependencies are caused by our minimum distance constraint, since a nonzero indicator $b_k$ determines all indicators within a certain neighborhood to be zero. Therefore, we will use a *partially collapsed Gibbs sampler* (PCGS) [33], which is a recently proposed generalization of the Gibbs sampler with significantly faster convergence for strongly

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

J.-Y. Tourneret and N. Dobigeon are with the University of Toulouse, IRIT-INPT/ENSEEIHT, 31071 Toulouse, France (e-mail: Jean-Yves.Tourneret@enseeiht.fr; dobigeon@n7.fr).

Digital Object Identifier 10.1109/TSP.2012.2190066

dependent parameters. The PCGS has been applied to various estimation problems [34] including blind Bernoulli–Gaussian deconvolution [23], [35]. Here, elaborating on our work in [36], we develop a PCGS for BD subject to a minimum distance constraint. Our method tolerates and even exploits the challenging probabilistic structure imposed by the minimum distance constraint.

This paper is organized as follows. In Section II, we describe the signal model, our choice of prior distributions, and the minimum distance constraint. Section III discusses the estimators and detectors underlying the proposed PCGS method. The PCGS is then presented in Section IV. In Section V, the proposed method is extended in order to resolve a scale and shift ambiguity. Finally, simulation results assessing the performance of the proposed method are presented in Section VI.

## II. Signal Model and Parameter Priors

### A. Signal Model

We consider a sparse sequence $a_k$ of length $K$, which is convolved with an unknown pulse $f_k$ and corrupted by additive noise $n_k$, yielding the observed sequence $x_k$. All sequences are complex-valued. The pulse $f_k$ is defined for $k = -Q, \ldots, Q$, where typically $Q \ll K$. The observed sequence can be expressed as

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

where for convenience we have set $f_k \triangleq 0$ for $k \notin \{-Q, \ldots, Q\}$. Like $x_k$, $a_k$ is defined for $k = 1, \ldots, K$; thus, nonzero values of $a_k$ for $k$ near 1 or $K$ correspond to copies of the pulse in $x_k$ that are cut off at the edges. Defining the vectors $\mathbf{x} \triangleq (x_1 \cdots x_K)^{\mathrm{T}}$, $\mathbf{f} \triangleq (f_{-Q} \cdots f_Q)^{\mathrm{T}}$, $\mathbf{a} \triangleq (a_1 \cdots a_K)^{\mathrm{T}}$, and $\mathbf{n} \triangleq (n_1 \cdots n_K)^{\mathrm{T}}$, the signal model (1) can be written as

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

with the $K \times K$ Toeplitz matrix $\mathbf{F} \triangleq \mathrm{toep}(\mathbf{f})$ 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.

Our goal is to estimate the sparse sequence $\mathbf{a}$ along with the pulse shape $\mathbf{f}$ and noise variance $\sigma_n^2$ from the observation $\mathbf{x}$. For this, following [23], [25], [32], [37], it will be convenient to introduce the binary indicator sequence

$$b_k \triangleq \begin{cases} 0, & \text{if } a_k = 0, \\ 1, & \text{if } a_k \neq 0, \end{cases} \quad k = 1, \ldots, K. \tag{3}$$

Let $\mathbf{b} \triangleq (b_1 \cdots b_K)^{\mathrm{T}} \in \{0,1\}^K$ denote the corresponding vector. The number of nonzero $b_k$ (equivalently, nonzero $a_k$) will be denoted as $B$; note that $B = \sum_{k=1}^K b_k = \sum_{k=1}^K b_k^2 = \|\mathbf{b}\|^2$. We can rewrite (2) as

$$\mathbf{x} = \mathbf{F}_b \mathbf{a}_b + \mathbf{n} \tag{4}$$

where $\mathbf{F}_b$ denotes the matrix of size $K \times B$ that is obtained from $\mathbf{F}$ by removing all columns $k$ such that $b_k = 0$ and $\mathbf{a}_b$ denotes the vector of length $B$ that contains the corresponding $a_k$, i.e., all nonzero entries of $\mathbf{a}$.

The indicator sequence $\mathbf{b}$ is subject to a hard *minimum distance constraint* $\mathbf{b} \in \mathcal{C}$, where the "constraint set" $\mathcal{C} \subseteq \{0,1\}^K$ is the set of all $\mathbf{b}$ such that the temporal distance between any two nonzero entries $b_k = 1$ and $b_{k'} = 1$ (equivalently, any $a_k, a_{k'} \neq 0$) satisfies $|k - k'| \geq d_{\min}$, with a given $d_{\min}$. Thereby, the set of possible hypotheses for $\mathbf{b}$ is significantly reduced ($\mathbf{b} \in \mathcal{C}$ rather than $\mathbf{b} \in \{0,1\}^K$). This reduction of the number of hypotheses results in large complexity savings in the deconvolution algorithms to be presented later.

For estimation of the pulse shape vector $\mathbf{f}$, it is convenient to represent $\mathbf{f}$ by a basis expansion

$$\mathbf{f} = \sum_{u=1}^{U} \gamma_u \mathbf{h}_u = \mathbf{H}\boldsymbol{\gamma} \tag{5}$$

where $\boldsymbol{\gamma} \triangleq (\gamma_1 \cdots \gamma_U)^{\mathrm{T}}$ is a random coefficient vector and $\mathbf{H} \triangleq (\mathbf{h}_1 \cdots \mathbf{h}_U)$ is a known $(2Q+1) \times U$ matrix containing the $U$ basis vectors $\mathbf{h}_u$. The basis vectors are fixed; their number $U$ and their shapes (viewed as time-domain signals) express some prior information about the pulse $\mathbf{f}$, e.g., regarding its maximum possible time and frequency supports. Here, we choose the basis functions $\mathbf{h}_u$ as the first $U$ Hermite functions [12], [38], [39], centered at the $(Q+1)$th entry. The first $U$ Hermite functions cover an elliptic region in the time-frequency plane whose area is roughly equal to $U$ [40], [41, p. 26]. Using (5), we can rewrite the signal model in (2) as

$$\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{n} = \mathbf{A}\mathbf{H}\boldsymbol{\gamma} + \mathbf{n} \tag{6}$$

where $\mathbf{A}$ is the $K \times (2Q+1)$ Toeplitz matrix with first column $(a_{Q+1} \ \ a_{Q+2} \cdots a_K \ \ 0 \cdots 0)^{\mathrm{T}}$ and first row $(a_{Q+1} \ \ a_Q \cdots a_1 \ \ 0 \cdots 0)$. Because of (5), estimation of the length-$(2Q+1)$ pulse vector $\mathbf{f}$ reduces to estimation of the length-$U$ coefficient vector $\boldsymbol{\gamma}$. Typically, $U \ll 2Q+1$, so that a parsimonious parametric representation of $\mathbf{f}$ is obtained.

### B. Parameter Priors

Our goal is to estimate the sparse sequence $\mathbf{a}$, along with the unknown pulse coefficients $\boldsymbol{\gamma}$ and noise variance $\sigma_n^2$. The Bayesian methodology adopted in this study requires the specification of prior distributions for all unknown quantities [42, p. 9]. The priors we will use are described in what follows.

*Sparse Sequence:* Rather than specifying the prior probability density function (pdf) $p(\mathbf{a})$ directly, we will specify $p(\mathbf{a}|\mathbf{b})$ and $p(\mathbf{b})$, as previously done in [23], [25], and [32]. To ensure consistency with Section II-A, $p(\mathbf{a}|\mathbf{b})$ must be chosen such that (3) is true and $p(\mathbf{b})$ must be chosen such that $\mathbf{b} \in \mathcal{C}$ is guaranteed. Assuming that different random transitions $b_k \to a_k$ (i.e., for different $k$) are statistically independent, we have

$$p(\mathbf{a}|\mathbf{b}) = \prod_{k=1}^{K} p(a_k|b_k). \tag{7}$$

Furthermore, we assume (note that $b_k = 0$ implies $a_k = 0$)

$$p(a_k|b_k) = \begin{cases} \delta(a_k), & \text{if } b_k = 0 \\ \mathcal{CN}(a_k; 0, \sigma_a^2), & \text{if } b_k = 1 \end{cases} \tag{8}$$

where $\delta(\cdot)$ is the Dirac delta function, $\sigma_a^2$ is a fixed hyperparameter, and $\mathcal{CN}(\cdot;\mu,\sigma^2)$ denotes the circularly symmetric complex Gaussian pdf with mean $\mu$ and variance $\sigma^2$. Consequently, the conditional prior of $\mathbf{a}_b$ is

$$p(\mathbf{a}_b|\mathbf{b}) = \mathcal{CN}\big(\mathbf{a}_b;\mathbf{0},\sigma_a^2\mathbf{I}\big)\,. \tag{9}$$

Here, $\mathcal{CN}(\cdot;\boldsymbol{\mu},\mathbf{C})$ denotes the multivariate circularly symmetric complex Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Note that $\mathbf{a}_b$ depends on $\mathbf{b}$ because its dimension is $B = \|\mathbf{b}\|^2$.

*Indicator Sequence:* For a compact formulation of the prior of $\mathbf{b}$, we will use the indicator function

$$I_{\mathcal{C}}(\mathbf{b}) \triangleq \begin{cases} 0, & \text{if } \mathbf{b} \notin \mathcal{C} \\ 1, & \text{if } \mathbf{b} \in \mathcal{C}. \end{cases}$$

We then define $p(\mathbf{b})$, up to an irrelevant normalization factor, as the product of $I_{\mathcal{C}}(\mathbf{b})$—expressing the minimum distance constraint—and an independent and identically distributed (i.i.d.) Bernoulli pdf $\mathcal{B}(\mathbf{b};\pi_1)$, i.e.,

$$p(\mathbf{b}) \propto \mathcal{B}(\mathbf{b};\pi_1) I_{\mathcal{C}}(\mathbf{b}) = \pi_1^{\|\mathbf{b}\|^2}(1-\pi_1)^{K-\|\mathbf{b}\|^2} I_{\mathcal{C}}(\mathbf{b}). \tag{10}$$

Here, $\propto$ means "proportional to," and the "1-probability" $\pi_1$ is a fixed hyperparameter. Together, $\pi_1$ and $d_{\min}$ determine[1] $\mathrm{E}\{B/K\}$, i.e., the *a priori* mean rate of 1's in $\mathbf{b}$. For $d_{\min} = 1$, the priors (7)–(10) simplify to the classical Bernoulli–Gaussian model. We note that our model can be extended by assigning prior distributions also to $d_{\min}$ and hyperparameters like $\pi_1$ etc. that are assumed fixed here; these hyperparameters can then be estimated along with the model parameters using a hierarchical Bayesian algorithm. Furthermore, the modified Bernoulli prior in (10) can be replaced by a more sophisticated Markovian prior (cf. [43]).

*Pulse Shape:* We choose the prior of $\boldsymbol{\gamma}$ as i.i.d., zero-mean, and circularly symmetric complex Gaussian, i.e.,

$$p(\boldsymbol{\gamma}) = \mathcal{CN}\big(\boldsymbol{\gamma};\mathbf{0},\sigma_\gamma^2\mathbf{I}\big) \tag{11}$$

where the variance $\sigma_\gamma^2$ is a fixed hyperparameter.

*Noise Variance:* The noise $n_k$ is modeled as i.i.d. circularly symmetric complex Gaussian with a constant variance $\sigma_n^2$, which is treated as a random hyperparameter and is estimated jointly with the other unknown parameters. Our stochastic model is thus a hierarchical Bayesian model. For the prior of $\sigma_n^2$, we choose an inverse gamma pdf, i.e.,

$$p\big(\sigma_n^2\big) = \mathcal{IG}\big(\sigma_n^2;\xi,\eta\big) = \frac{\eta^\xi}{\Gamma(\xi)}\frac{1}{(\sigma_n^2)^{\xi+1}}\exp\left(-\frac{\eta}{\sigma_n^2}\right) I_{\mathbb{R}^+}\big(\sigma_n^2\big) \tag{12}$$

where $\Gamma(\xi)$ is the gamma function, $I_{\mathbb{R}^+}(x)$ is the unit step function, and $\xi$ and $\eta$ are fixed hyperparameters. The inverse gamma distribution is convenient (and commonly used in similar contexts) because it is the conjugate prior for the Gaussian likelihood function [42, p. 152]. The same is true for the priors $p(\mathbf{a}|\mathbf{b})$ and $p(\boldsymbol{\gamma})$.

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

### C. Posterior Distribution

The unknown quantities to be estimated are $\mathbf{b}$, $\mathbf{a}$, $\boldsymbol{\gamma}$, and $\sigma_n^2$ or, equivalently, $\mathbf{b}$, $\mathbf{a}_b$, $\boldsymbol{\gamma}$, and $\sigma_n^2$. According to the adopted Bayesian methodology, their estimation is based on the posterior distribution, whose determination involves the likelihood function and the priors [42, p. 9]. The likelihood function of our model is, according to (2), (4), (6), and the i.i.d. Gaussian prior for the $n_k$,

$$\begin{aligned} p\big(\mathbf{x}|\mathbf{a},\boldsymbol{\gamma},\sigma_n^2\big) &= \mathcal{CN}\big(\mathbf{x};\mathbf{F}\mathbf{a},\sigma_n^2\mathbf{I}\big) = \mathcal{CN}\big(\mathbf{x};\mathbf{F}_b\mathbf{a}_b,\sigma_n^2\mathbf{I}\big) \\ &= \mathcal{CN}\big(\mathbf{x};\mathbf{A}\mathbf{H}\boldsymbol{\gamma},\sigma_n^2\mathbf{I}\big)\,. \end{aligned} \tag{13}$$

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

$$\begin{aligned} p\big(\mathbf{b},\mathbf{a},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big) &\propto p\big(\mathbf{x},\mathbf{b},\mathbf{a},\boldsymbol{\gamma},\sigma_n^2\big) \\ &= p\big(\mathbf{x}|\mathbf{a},\boldsymbol{\gamma},\sigma_n^2\big)p(\mathbf{a}|\mathbf{b})p(\mathbf{b})p(\boldsymbol{\gamma})p\big(\sigma_n^2\big) \end{aligned} \tag{14}$$

where the factors in the final expression are given in (7), (8), and (10)–(13).

## III. Monte Carlo Detection-Estimation Method

In this and the next two sections, we will develop a Monte Carlo detection-estimation method for BD. Following [37], our approach is to first *detect* $\mathbf{b} \in \mathcal{C} \subseteq \{0,1\}^K$ (i.e., detect which $b_k$ or, equivalently, $a_k$ are nonzero) and then estimate the corresponding nonzero $a_k$. Without the detection step, sparsity of $\mathbf{a}$ would not be ensured [37]. In addition, we will estimate the unknown pulse coefficients $\boldsymbol{\gamma}$ and noise variance $\sigma_n^2$. We first present the basic detector and estimators and their conceptual relation to optimal detectors and estimators.

### A. Sequence and Component Detectors

*MAP Sequence and Component Detectors:* As a motivation for the block detector to be proposed in Section III-B, we first consider two well-known optimal methods for detecting the indicator sequence $\mathbf{b}$ and their Monte Carlo (sample-based) counterparts. The *MAP sequence detector*

$$\hat{\mathbf{b}}_{\mathrm{MAP}}^{(\mathrm{s})} \triangleq \underset{\mathbf{b}\in\{0,1\}^K}{\arg\max}\, p(\mathbf{b}|\mathbf{x}) = \underset{\mathbf{b}\in\mathcal{C}}{\arg\max}\, p(\mathbf{b}|\mathbf{x}) \tag{15}$$

is optimal in that it minimizes the probability of a sequence error $\mathrm{P}\{\hat{\mathbf{b}} \neq \mathbf{b}\}$ [44, p. 80]. Note that $\hat{\mathbf{b}}_{\mathrm{MAP}}^{(\mathrm{s})} \in \mathcal{C}$. Similarly, the *MAP component detector*

$$\hat{b}_{k,\mathrm{MAP}}^{(\mathrm{c})} \triangleq \underset{b_k\in\{0,1\}}{\arg\max}\, p(b_k|\mathbf{x}), \quad k = 1,\ldots,K \tag{16}$$

minimizes the probability of a component error $\mathrm{P}\{\hat{b}_k \neq b_k\}$; it is also known as "maximum posterior marginal/mode

(MPM) detector" (e.g., [45]). It can be shown that $\hat{\mathbf{b}}^{(\mathrm{c})}_{\mathrm{MAP}} \triangleq (\hat{b}^{(\mathrm{c})}_{1,\mathrm{MAP}} \cdots \hat{b}^{(\mathrm{c})}_{K,\mathrm{MAP}})^{\mathrm{T}} \in \mathcal{C}$.

In principle, both $p(\mathbf{b}|\mathbf{x})$ used in (15) and $p(b_k|\mathbf{x})$ used in (16) can be derived from the joint posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ by marginalization. However, due to the high dimensionality of $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$, we will use a Monte Carlo approach [29, p. 79], [46], i.e., we will generate a sample $\mathcal{S} \triangleq \{(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})\}_{m=1}^{M}$ of realizations $(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})$ from $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ and then perform the detection based on this sample. Note that $\mathcal{S}$ depends on the observation $\mathbf{x}$. The generation of $\mathcal{S}$ will be discussed in Section IV.

*Sample-Based Sequence Detector:* Using the sample $\mathcal{S}$, marginalizations are easily done by ignoring the undesired components of each realization $(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})$. We first consider the marginalization that corresponds to the marginal posterior $p(\mathbf{b}|\mathbf{x})$ underlying the MAP sequence detector $\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathrm{MAP}}$ in (15). Let $q(\mathbf{b})$ denote the relative multiplicity of some $\mathbf{b} \in \{0,1\}^K$ in $\mathcal{S}$, i.e., the number of occurrences of $\mathbf{b}$ in $\mathcal{S}$ normalized by the sample size $|\mathcal{S}| = M$. In particular, $q(\mathbf{b}) = 0$ if $\mathbf{b}$ does not occur in $\mathcal{S}$. Let $\mathcal{B}$ denote the set of the $\mathbf{b} \in \{0,1\}^K$ contained in $\mathcal{S}$ (i.e., for which $q(\mathbf{b}) \neq 0$); note that $\mathcal{B} \subseteq \mathcal{C}$ due to our minimum distance constraint. If the process generating $\mathcal{S}$ does not exclude parts of the support of $p(\mathbf{b}|\mathbf{x})$, i.e., parts of $\mathcal{C}$, then $q(\mathbf{b})$ converges to $p(\mathbf{b}|\mathbf{x})$ as $M$ increases without bound [46, p. 5]. Therefore, the *sample-based sequence detector*

$$\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathcal{S}} \triangleq \underset{\mathbf{b}\in\{0,1\}^K}{\arg\max}\, q(\mathbf{b}) = \underset{\mathbf{b}\in\mathcal{B}}{\arg\max}\, q(\mathbf{b})$$

approximates $\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathrm{MAP}}$ for $M$ sufficiently large. Note that $\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathcal{S}}$ is simply the $\mathbf{b}$ occurring most often in $\mathcal{S}$.

Unfortunately, $M = |\mathcal{S}|$ is usually *not* sufficiently large; it is much smaller than the number $|\mathcal{C}|$ of hypotheses $\mathbf{b} \in \mathcal{C}$ among which $\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathrm{MAP}}$ in (15) selects the best. This means that $q(\mathbf{b})$ corresponds to a very coarse quantization of the probability distribution $p(\mathbf{b}|\mathbf{x})$ of the $|\mathcal{C}|$ admissible hypotheses $\mathbf{b} \in \mathcal{C}$ in steps of $\frac{1}{M}$. This often leads to ties among multiple hypotheses that simultaneously maximize $q(\mathbf{b})$ even though their true posterior probabilities $p(\mathbf{b}|\mathbf{x})$ are not equal (see Section VI-B).

*Sample-Based Component Detector:* The above problem is avoided by the sample-based version of the MAP *component* detector (16), which is given by

$$\hat{b}^{(\mathrm{c})}_{k,\mathcal{S}} \triangleq \underset{b_k\in\{0,1\}}{\arg\max}\, q(b_k), \quad k = 1, \ldots, K.$$

Here, $q(b_k)$ is the relative multiplicity of some $b_k \in \{0,1\}$ in $\mathcal{S}$, i.e., the number of realizations $\mathbf{b}^{(m)}$ in $\mathcal{S}$ that have the given $b_k$ at position $k$, normalized by the sample size $M$. Thus, $\hat{b}^{(\mathrm{c})}_{k,\mathcal{S}}$ is the $b_k \in \{0,1\}$ that the majority of the realizations $\mathbf{b}^{(m)}$ contain at position $k$. If the process generating $\mathcal{S}$ does not exclude parts of the support of $p(b_k|\mathbf{x})$, i.e., of {0,1}, then $q(b_k)$ converges to $p(b_k|\mathbf{x})$ as $M$ increases without bound [46, p. 5]. Therefore, the sample-based component detector $\hat{b}^{(\mathrm{c})}_{k,\mathcal{S}}$ is an approximation of the MAP component detector $\hat{b}^{(\mathrm{c})}_{k,\mathrm{MAP}}$. The fact that $M$ is usually rather small is no problem here since there are only two possible hypotheses for $b_k$. Hence, $q(b_k)$ will be a good approximation of $p(b_k|\mathbf{x})$ and, in turn, $\hat{b}^{(\mathrm{c})}_{k,\mathcal{S}}$ will be a good approximation of $\hat{b}^{(\mathrm{c})}_{k,\mathrm{MAP}}$. Again, $\hat{\mathbf{b}}^{(\mathrm{c})}_{\mathcal{S}} \triangleq (\hat{b}^{(\mathrm{c})}_{1,\mathcal{S}} \cdots \hat{b}^{(\mathrm{c})}_{K,\mathcal{S}})^{\mathrm{T}}$ is in $\mathcal{C}$. This is because each realization $\mathbf{b}^{(m)}$ is in $\mathcal{C}$, and hence, for $k$ and $k'$ such that $|k - k'| < d_{\min}$, none of the realizations $\mathbf{b}^{(m)}$ can contain 1's at both positions $k$ and $k'$. Thus, it is impossible that the majority of the realizations $\mathbf{b}^{(m)}$ contain a 1 at position $k$ and, at the same time, the majority of the realizations $\mathbf{b}^{(m)}$ contain a 1 at position $k'$.

However, the MAP component detector $\hat{b}^{(\mathrm{c})}_{k,\mathrm{MAP}}$ has itself a problem that renders it unsuitable in our context. In regarding only one marginal $p(b_k|\mathbf{x})$ at a time, $\hat{b}^{(\mathrm{c})}_{k,\mathrm{MAP}}$ ignores a significant part of the information contained in the joint posterior $p(\mathbf{b}|\mathbf{x})$, which may yield counterintuitive results. In particular, consider a fixed time interval $\mathcal{K}$ of length $|\mathcal{K}| \leq d_{\min}$, and suppose that for a given observed $\mathbf{x}$, the probability that there is exactly one active indicator $b_k = 1$ in $\mathcal{K}$ is 1. Equivalently, $\sum_{k\in\mathcal{K}} p(b_k = 1|\mathbf{x}) = 1$ since the events $b_k = 1$ are mutually exclusive for $k \in \mathcal{K}$ (there may not be more than one $b_k = 1$ in $\mathcal{K}$ because of the minimum distance constraint). Suppose further that $p(b_k = 1|\mathbf{x})$ is not much larger at any of the possible positions $k \in \mathcal{K}$ than at the respective other positions, so $p(b_k = 1|\mathbf{x}) < 1/2$ for all $k \in \mathcal{K}$. This means that $p(b_k = 0|\mathbf{x}) > p(b_k = 1|\mathbf{x})$ for all $k \in \mathcal{K}$, which implies that $\hat{b}^{(\mathrm{c})}_{k,\mathrm{MAP}} = 0$ for all $k \in \mathcal{K}$, i.e., the MAP component detector does not detect any pulse in $\mathcal{K}$. This is clearly counterintuitive as the probability that there is no pulse in $\mathcal{K}$ is zero.

### *B. The Proposed Block Detector*

In order to mitigate the problems described above, we consider a *block detector* that is a compromise between the sequence detector and the component detector. The sequence $\mathbf{b}$ is split into $J$ nonoverlapping blocks $\boldsymbol{\beta}_j$ of generally different lengths $K_j$, $j = 1, \ldots, J$, i.e., $\mathbf{b} = (\boldsymbol{\beta}_1^{\mathrm{T}} \cdots \boldsymbol{\beta}_J^{\mathrm{T}})^{\mathrm{T}}$. Each block is detected independently of the others. Thus, the *MAP block detector* is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{j,\mathrm{MAP}} &\triangleq \underset{\boldsymbol{\beta}_j\in\{0,1\}^{K_j}}{\arg\max}\; p(\boldsymbol{\beta}_j|\mathbf{x}) \\ &= \underset{\boldsymbol{\beta}_j\in\mathcal{C}_j}{\arg\max}\, p(\boldsymbol{\beta}_j|\mathbf{x}), \quad j = 1, \ldots, J.\end{aligned}$$

Here, $p(\boldsymbol{\beta}_j|\mathbf{x})$ is again a marginal of $p(\mathbf{b}|\mathbf{x})$ and $\mathcal{C}_j$ is the set of all $\boldsymbol{\beta}_j \in \{0,1\}^{K_j}$ conforming to the minimum distance constraint. The MAP block detector minimizes the block error probability $\mathrm{P}\{\hat{\boldsymbol{\beta}}_j \neq \boldsymbol{\beta}_j\}$. The overall detection result is obtained by concatenating all detected blocks $\hat{\boldsymbol{\beta}}_{j,\mathrm{MAP}}$, i.e., $\hat{\mathbf{b}}^{(\mathrm{b})}_{\mathrm{MAP}} \triangleq (\hat{\boldsymbol{\beta}}^{\mathrm{T}}_{1,\mathrm{MAP}} \cdots \hat{\boldsymbol{\beta}}^{\mathrm{T}}_{J,\mathrm{MAP}})^{\mathrm{T}}$. Note that the MAP sequence detector $\hat{\mathbf{b}}^{(\mathrm{s})}_{\mathrm{MAP}}$ and the MAP component detector $\hat{\mathbf{b}}^{(\mathrm{c})}_{\mathrm{MAP}}$ are special cases of the MAP block detector $\hat{\mathbf{b}}^{(\mathrm{b})}_{\mathrm{MAP}}$ corresponding to $J = 1$, $K_j = K$ and $J = K$, $K_j = 1$, respectively.

The sample-based approximation of the MAP block detector $\hat{\boldsymbol{\beta}}_{j,\mathrm{MAP}}$ is given by

$$\hat{\boldsymbol{\beta}}_{j,\mathcal{S}} \triangleq \underset{\boldsymbol{\beta}_j\in\{0,1\}^{K_j}}{\arg\max}\; q(\boldsymbol{\beta}_j) = \underset{\boldsymbol{\beta}_j\in\mathcal{B}_j}{\arg\max}\, q(\boldsymbol{\beta}_j).$$

Here, $q(\boldsymbol{\beta}_j)$ is the relative multiplicity of $\boldsymbol{\beta}_j \in \{0,1\}^{K_j}$ in $\mathcal{S}$, i.e., the number of $\mathbf{b}^{(m)}$ in $\mathcal{S}$ that have the given $\boldsymbol{\beta}_j \in \{0,1\}^{K_j}$ as the $j$th block, normalized by $M$, and $\mathcal{B}_j \subseteq \mathcal{C}_j$ is the set of all $\boldsymbol{\beta}_j \in \{0,1\}^{K_j}$ that are featured by the $\mathbf{b}$ contained in $\mathcal{S}$ (i.e., the $\boldsymbol{\beta}_j$ for which $q(\boldsymbol{\beta}_j) \neq 0$). The overall detection result is again obtained by concatenating all detected blocks, i.e., $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})} \triangleq (\hat{\boldsymbol{\beta}}_{1,\mathcal{S}}^{\mathrm{T}} \cdots \hat{\boldsymbol{\beta}}_{J,\mathcal{S}}^{\mathrm{T}})^{\mathrm{T}}$.

For the definition of the block intervals, we can exploit the typical structure of $q(b_k)$ induced by the sparsity of $\mathbf{b}$. We first calculate $q_k \triangleq q(b_k = 1)$, i.e., the relative multiplicity of $b_k = 1$, at each position $k \in \{1,\ldots,K\}$. We have observed empirically that, for a sparse $\mathbf{b}$, and assuming that the signal model is well matched to the problem, the sequence $q_k$ typically consists of long "zero intervals" and short "nonzero intervals." That is, the intervals of positions $k$ where some realizations contain nonzero indicators are separated by longer intervals where no realization contains a nonzero indicator. We propose to use these zero intervals and nonzero intervals as blocks. Within zero intervals, $q_k \equiv 0$, which means that no $\mathbf{b}^{(m)}$ in $\mathcal{S}$ features a nonzero indicator $b_k = 1$ in the interval considered, and thus the sample-based block detector trivially yields the zero block. Nonzero intervals, on the other hand, have $q_k > 0$ at all positions. They are typically short enough to avoid the problems of the sequence detector. Moreover, they are separated from each other by the zero intervals, which reduces their statistical dependence and thus avoids the problems of the component detector.

The block detector will perform well if the nonzero intervals are short (not much longer than $d_{\min}$) and the zero intervals are long (longer than $d_{\min}$). We note that $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ is not guaranteed to be an element of $\mathcal{S}$, or even of $\mathcal{C}$, because the different blocks are processed independently; only the special cases $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{s})}$ and $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{c})}$ are guaranteed to be in $\mathcal{C}$. However, if $q(b_k)$ has the interval structure described above, $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ is highly likely to satisfy the minimum distance constraint. Also, sparsity of $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ is ensured because all realizations $\mathbf{b}^{(m)}$ satisfy the minimum distance constraint. Conditions for $q(b_k)$ to feature zero and nonzero intervals include a well-matched signal model, time-shift compensation (see Section V), and a close match between $p(\mathbf{b},\mathbf{a},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$ and the process generating $\mathcal{S}$. In the case of MCMC methods—see Section IV—the last condition presupposes convergence of the Markov chain, i.e., a long enough burn-in period. For a given sample $\mathcal{S}$, it is easy to calculate $q(b_k)$ and check if it has the desired interval structure. When it does not, a more sophisticated detector like the sequence detector proposed in [47] may be used. This detector maximizes a specially designed metric that is calculated from all $q_k$ jointly and enforces the minimum distance constraint.

### C. Estimation of Amplitudes, Pulse Coefficients, and Noise Variance

*Amplitudes:* For estimation of the amplitudes $\mathbf{a} \in \mathbb{C}^K$ given the previously detected indicator sequence $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$, we ideally use the minimum mean square error (MMSE) estimator

$$\hat{\mathbf{a}}_{\mathrm{MMSE}} \triangleq \mathrm{E}\big\{\mathbf{a}|\mathbf{x},\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}\big\} = \int \mathbf{a}p\big(\mathbf{a}|\mathbf{x},\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}\big)\mathrm{d}\mathbf{a}.$$

However, calculating $p(\mathbf{a}|\mathbf{x},\mathbf{b}) \propto p(\mathbf{b},\mathbf{a}|\mathbf{x})$ by marginalization of $p(\mathbf{b},\mathbf{a},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$ is not feasible. A solution is again provided by a sample-based estimator. Unfortunately, a sample-based approximation of $p(\mathbf{a}|\mathbf{x},\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})})$ is not available because $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ is not guaranteed to be an element of $\mathcal{S}$. We therefore condition each $a_k \in \mathbb{C}$ only on the respective detected indicator $\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})} \in \{0,1\}$, i.e., we consider the componentwise MMSE estimator

$$\begin{aligned}\hat{a}_{k,\mathrm{MMSE}} &\triangleq \mathrm{E}\big\{a_k|\mathbf{x},b_k=\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}\big\}\\ &= \int a_k p\big(a_k|\mathbf{x},b_k=\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}\big)\mathrm{d}a_k.\end{aligned}$$

Accordingly, we use a sample-based approximation of $p(a_k|\mathbf{x},b_k=\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})})$ rather than of $p(a_k|\mathbf{x},\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})})$. Note that, in contrast to $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$, $\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}$ is guaranteed to occur in $\mathcal{S}$ due to our definition of the block detector in Section III-B. The sample-based version of $\hat{a}_{k,\mathrm{MMSE}}$ is then obtained as follows. For each $k \in \{1,\ldots,K\}$, we partition the set of realization indices $\{1,\ldots,M\}$ into two complementary subsets $\mathcal{M}_k(1)$ and $\mathcal{M}_k(0)$ containing all indices $m$ for which $b_k^{(m)} = 1$ and $b_k^{(m)} = 0$, respectively. Then, $a_k$ is estimated as

$$\hat{a}_{k,\mathcal{S}} \triangleq \frac{1}{\big|\mathcal{M}_k\big(\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}\big)\big|} \sum_{m\in\mathcal{M}_k\left(\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}\right)} a_k^{(m)} \tag{17}$$

where $a_k^{(m)}$ is the $k$th entry of $\mathbf{a}^{(m)}$ and $|\mathcal{M}_k(\cdot)|$ denotes the cardinality of $\mathcal{M}_k(\cdot)$. Note that $|\mathcal{M}_k(\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})})| \neq 0$ because $\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})}$ occurs in $\mathcal{S}$. Furthermore note that $\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})} = 0$ entails $\hat{a}_{k,\mathcal{S}} = 0$, because all realizations $a_k^{(m)}$ with $m \in \mathcal{M}_k(0)$ are zero; thus, (17) has to be calculated only for those $k$ where $\hat{b}_{k,\mathcal{S}}^{(\mathrm{b})} = 1$.

The above sample-based componentwise estimator is computationally efficient and performed well in our simulations. An alternative is provided by the joint conditional MMSE estimator

$$\hat{\mathbf{a}}_{b,\mathrm{MMSE}} \triangleq \mathrm{E}\left\{\mathbf{a}_b|\mathbf{x},\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}_{\mathcal{S}},\sigma_n^2=\widehat{\sigma_{n,\mathcal{S}}^2}\right\}.$$

Using the Gaussianity of $p(\mathbf{a}_b|\mathbf{x},\mathbf{b},\boldsymbol{\gamma},\sigma_n^2)$, one easily obtains

$$\begin{aligned}\hat{\mathbf{a}}_{b,\mathrm{MMSE}} &= \operatorname*{arg\,min}_{\mathbf{a}_b\in\mathbb{C}^B}\left(\frac{\|\mathbf{x}-\hat{\mathbf{F}}_b\mathbf{a}_b\|^2}{\widehat{\sigma_{n,\mathcal{S}}^2}} + \frac{\|\mathbf{a}_b\|^2}{\sigma_a^2}\right)\\ &= \left(\hat{\mathbf{F}}_b^{\mathrm{H}}\hat{\mathbf{F}}_b + \frac{\widehat{\sigma_{n,\mathcal{S}}^2}}{\sigma_a^2}\mathbf{I}\right)^{-1}\hat{\mathbf{F}}_b^{\mathrm{H}}\mathbf{x}\end{aligned} \tag{18}$$

where $\hat{\mathbf{F}}_b$ is the $\mathbf{F}_b$ that corresponds to $\mathbf{b}=\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ and $\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}_{\mathcal{S}}$. Note that $\hat{\mathbf{a}}_{b,\mathrm{MMSE}}$ presupposes prior estimation of $\boldsymbol{\gamma}$ and $\sigma_n^2$, to be discussed below. In our simulations, the performance of $\hat{\mathbf{a}}_{b,\mathrm{MMSE}}$ was consistently—if only slightly—better than that of $\hat{a}_{k,\mathcal{S}}$ in (17). However, the complexity is higher.

*Pulse Coefficients and Noise Variance:* For estimation of the pulse coefficients $\boldsymbol{\gamma}$ and noise variance $\sigma_n^2$, we use the sample-based versions of the respective MMSE estimator, i.e.,

$$\hat{\boldsymbol{\gamma}}_{\mathcal{S}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\gamma}^{(m)}, \qquad \widehat{\sigma_{n,\mathcal{S}}^2} \triangleq \frac{1}{M}\sum_{m=1}^{M}\sigma_n^{2(m)}.$$

## IV. Partially Collapsed Gibbs Sampler

MCMC methods [29], [30] are often used when the analytic expression of a detector or estimator is too complex to be calculated directly. The detector or estimator is approximated by a sample-based scheme (as, e.g., in Section III), where a sample is generated by means of an ergodic Markov chain whose stationary distribution is the target distribution from which the sample realizations are to be drawn. In this section, we first review the Gibbs sampler and the PCGS and discuss their suitability for problems with deterministic constraints such as our minimum distance constraint. Then, we propose a PCGS that exhibits fast convergence in the presence of a minimum distance constraint.

### A. Review of Gibbs Sampler and PCGS

*Gibbs Sampler:* Consider a random vector $\boldsymbol{\vartheta} = (\vartheta_1 \cdots \vartheta_L)^{\mathrm{T}} \in \mathbb{C}^L$, and let $\boldsymbol{\vartheta}_{\sim\ell}$ denote $\boldsymbol{\vartheta}$ without the $\ell$th entry $\vartheta_\ell$. (The generalization to the case where the $\vartheta_\ell$ are themselves vectors is straightforward.) To obtain realizations from the joint distribution $p(\boldsymbol{\vartheta}|\mathbf{x})$—which corresponds to $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ in our problem—the Gibbs sampler iteratively samples each $\vartheta_\ell$ from $p(\vartheta_\ell|\boldsymbol{\vartheta}_{\sim\ell}, \mathbf{x})$ in an arbitrary order. This strategy is known to converge to the target distribution $p(\boldsymbol{\vartheta}|\mathbf{x})$, which is the stationary distribution of the underlying Markov chain [29, p. 378], [46]. After convergence, $L$ such *sampling steps* produce a new realization $\boldsymbol{\vartheta}$ from $p(\boldsymbol{\vartheta}|\mathbf{x})$; this will be referred to as one *iteration* of the Gibbs sampler. Since the initialization may strongly influence the first few realizations, only the realizations after a certain "burn-in period" are used in the sample. The main strengths of the Gibbs sampler are the generality of its formulation and the fact that it circumvents the "curse of dimensionality." However, a known weakness is that statistical dependencies between (some of) the $\vartheta_\ell$ tend to result in slow convergence of the Markov chain to its stationary distribution [23], [32].

*PCGS:* The PCGS is an extension of the Gibbs sampler that allows the following three modifications [33].

- *Marginalization*. Rather than sampling only the entry $\vartheta_\ell$ in step $\ell$, some other entries may be sampled along with $\vartheta_\ell$ instead of being conditioned upon. Let $\mathcal{L}(\ell) \subseteq \{1, \ldots, L\}$, and let the vectors $\boldsymbol{\vartheta}_{\mathcal{L}(\ell)}$ and $\boldsymbol{\vartheta}_{\sim\mathcal{L}(\ell)}$ contain the entries of $\boldsymbol{\vartheta}$ indexed by $\mathcal{L}(\ell)$ and by its complement $\{1, \ldots, L\} \setminus \mathcal{L}(\ell)$, respectively. Then step $\ell$ may sample from $p(\boldsymbol{\vartheta}_{\mathcal{L}(\ell)}|\boldsymbol{\vartheta}_{\sim\mathcal{L}(\ell)}, \mathbf{x})$ instead of $p(\vartheta_\ell|\boldsymbol{\vartheta}_{\sim\ell}, \mathbf{x})$. This can improve the convergence rate significantly, especially when there are strong dependencies between the $\vartheta_\ell$. Note that, in general, some $\mathcal{L}(\ell)$ for different $\ell$ overlap. Within one entire PCGS iteration, some $\vartheta_\ell$ are thus sampled several times.
- *Trimming*. If a $\vartheta_\ell$ is sampled several times in consecutive steps, only the last value is relevant, since the other values are never used. Such unused entries can thus be dropped from the respective sampling distribution. We can formulate this as follows: For any $\ell \in \{1, \ldots, L-1\}$, let the vector $\boldsymbol{\vartheta}_{\mathcal{L}^*(\ell)}$ contain those entries of $\boldsymbol{\vartheta}_{\mathcal{L}(\ell)}$ that are not contained in $\boldsymbol{\vartheta}_{\mathcal{L}(\ell+1)}$, i.e., $\mathcal{L}^*(\ell) = \mathcal{L}(\ell) \setminus \mathcal{L}(\ell+1)$. Then step $\ell$ may sample from $p(\boldsymbol{\vartheta}_{\mathcal{L}^*(\ell)}|\boldsymbol{\vartheta}_{\sim\mathcal{L}(\ell)}, \mathbf{x})$ instead of $p(\boldsymbol{\vartheta}_{\mathcal{L}(\ell)}|\boldsymbol{\vartheta}_{\sim\mathcal{L}(\ell)}, \mathbf{x})$, which may reduce the complexity of the sampling steps. The convergence behavior is not affected. Note that the distributions used for sampling are generally no longer conditional distributions associated with the full joint distribution $p(\boldsymbol{\vartheta}|\mathbf{x})$, but conditional distributions associated with certain marginal distributions of $p(\boldsymbol{\vartheta}|\mathbf{x})$.
- *Permutation*. It is reasonable to choose the (arbitrary) sampling order such that trimming can be performed to a maximum extent. After trimming, permutations are only allowed if they preserve the justification of the trimming already applied.

These modifications do not change the stationary distribution of the Markov chain [33]. The PCGS's flexibility regarding the choice of the sampling distributions makes it applicable to many cases in which the sampling distributions required by the Gibbs sampler cannot be calculated analytically (see [34] and references therein).

*Deterministic Constraints:* Deterministic constraints such as our minimum distance constraint may cause slow convergence of the Gibbs sampler and may even inhibit its convergence altogether. This is because each of the $L$ sampling steps constitutes a jump along one of the axes of the $L$-dimensional hypothesis space. A deterministic constraint may restrict the hypotheses with nonzero probability to disjoint regions such that one region cannot be reached from another by such jumps. Sampling several $\vartheta_\ell$ jointly, as in the PCGS, corresponds to a jump along the linear span of the axes associated with these $\vartheta_\ell$. Thus, there are more configurations of disjoint regions in the hypothesis space between which the sampler can jump. In the Gibbs sampler, the $\vartheta_\ell$ may be grouped into vectors, too, but these vectors must be disjoint. Therefore, the possible directions of the jumps are still orthogonal to each other. The restricted hypothesis space typically demands more freedom for the jumps, which can be provided by the overlapping subvectors $\boldsymbol{\vartheta}_{\mathcal{L}(\ell)}$ that are possible in the PCGS.

### B. The Proposed PCGS

We recall that our goal is to obtain a sample $\mathcal{S} = \{(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})\}_{m=1}^{M}$ from the posterior distribution $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2) p(\mathbf{a}|\mathbf{b}) p(\mathbf{b}) p(\boldsymbol{\gamma}) p(\sigma_n^2)$; from this sample, the unknown parameters $\mathbf{b}$, $\mathbf{a}$, $\boldsymbol{\gamma}$, and $\sigma_n^2$ can be detected or estimated as discussed in Sections III-B and III-C. We now present a PCGS—briefly referred to as "proposed sampler" and abbreviated PS—that exhibits fast convergence in the presence of a minimum distance constraint. For now, we ignore the scale and time shift ambiguity inherent to BD; a modification of the PS that accounts for this ambiguity will be discussed in Section V.

For a given time $k \in \{1, \ldots, K\}$, let $\mathcal{J}(k) \triangleq \{k, k+1, \ldots, k_{\max}(k)\}$ denote a right-hand neighborhood of $k$ whose length is $d(k) = d_{\min}$ except when $k$ is so close to the sequence end point $K$ that less than $d_{\min}$ entries are left, i.e., we set $k_{\max}(k) = k + d(k) - 1 = \min\{k + d_{\min} - 1, K\}$ and $d(k) = \min\{d_{\min}, K - k + 1\} \leq d_{\min}$. Let $\mathbf{b}_{\mathcal{J}(k)} = (b_k \; b_{k+1} \cdots b_{k_{\max}(k)})^{\mathrm{T}}$ denote the corresponding subvector of $\mathbf{b}$, and $\mathbf{b}_{\sim\mathcal{J}(k)}$ the complementary subvector. Analogous definitions apply to $\mathbf{a}$. The neighborhoods $\mathcal{J}(k)$ are special instances

of the parameter index subsets $\mathcal{L}(\ell)$ considered in Section IV-A. One iteration of the PS is stated as follows.

**One PS iteration**

- Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x})$.
- For $k = 1, \ldots, K$,
  — sample $b_k$ from $p(b_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$;
  — sample $a_k$ from $p(a_k|b_k, \mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.
- Sample $\sigma_n^2$ from $p(\sigma_n^2|\mathbf{a}, \boldsymbol{\gamma}, \mathbf{x})$.

Thus, a PS iteration consists of three sampling steps, where the second step is split into $K$ *substeps*. The $k$th substep samples $b_k$ and $a_k$, and is equivalent to jointly sampling $(b_k, a_k)$ from $p(b_k, a_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$. The PS is not a Gibbs sampler because $p(b_k, a_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ is not a conditional distribution associated with $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$. Rather, it is a conditional distribution associated with $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ marginalized with respect to all parameters in $\mathcal{J}(k)$ without $b_k$ and $a_k$. Thus, $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{a}_{\mathcal{J}(k)}$ are not contained in the condition for $(b_k, a_k)$, which is hence sampled regardless of the previous realization of $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{a}_{\mathcal{J}(k)}$. This difference from the Gibbs sampler allows the PS to explore the restricted hypothesis space efficiently.

The PS is a valid PCGS, because it is the trimmed version of a sampler that samples $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$ from $p(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ rather than just $(b_k, a_k)$ from $p(b_k, a_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$. In the untrimmed version of the sampler, all the sampling distributions are conditional distributions associated with the full joint posterior. The trimming is justified because all elements of $\mathcal{J}(k) = \{k, \ldots, k_{\max}(k)\}$ except $k$ itself are also contained in $\mathcal{J}(k+1) = \{k+1, \ldots, k_{\max}(k+1)\}$ (cf. Section IV-A).[2]

The fact that $p(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x})$ and $p(\sigma_n^2|\mathbf{a}, \boldsymbol{\gamma}, \mathbf{x})$ do not contain $\mathbf{b}$ in their conditions is not the result of trimming but reflects the fact that $\boldsymbol{\gamma}$ and $\sigma_n^2$ are conditionally independent of $\mathbf{b}$ when $\mathbf{a}$ is given, i.e., $p(\boldsymbol{\gamma}|\mathbf{b}, \mathbf{a}, \sigma_n^2, \mathbf{x}) = p(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x})$ and $p(\sigma_n^2|\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \mathbf{x}) = p(\sigma_n^2|\mathbf{a}, \boldsymbol{\gamma}, \mathbf{x})$.

For a full validation of the PS as a PCGS, Appendix A derives the PS from a classical Gibbs sampler, using only modifications that are allowed by the PCGS concept.

### C. Sampling Distributions

We will now present closed-form expressions of the sampling distributions involved in the PS. Detailed derivations of these distributions are provided in Appendix B.

*Pulse Coefficients:* The sampling distribution for $\boldsymbol{\gamma}$ is

$$p\left(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x}\right) = \mathcal{CN}(\boldsymbol{\gamma}; \boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \tag{19}$$

with

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \frac{1}{\sigma_n^2}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\mathbf{H}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{x}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \left(\frac{1}{\sigma_n^2}\mathbf{H}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{H} + \frac{1}{\sigma_\gamma^2}\mathbf{I}\right)^{-1}. \tag{20}$$

Thanks to the moderate size of $\boldsymbol{\gamma}$ and its jointly Gaussian posterior, the entries of $\boldsymbol{\gamma}$ can be sampled jointly. (This is the point where we exploit the benefit of the basis expansion, i.e., the lower dimensionality of $\boldsymbol{\gamma}$ compared to $\mathbf{f}$.) Before $\boldsymbol{\gamma}$ is sampled, $\mathbf{A}$ is constructed from the most recent realization of $\mathbf{a}$. Similarly, after $\boldsymbol{\gamma}$ is sampled, $\mathbf{F}$ is updated using the new realization of $\boldsymbol{\gamma}$ before the other parameters are sampled (see below).

*Indicators:* In order to obtain the sampling distribution $p(b_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ analytically, we would have to marginalize the joint posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ with respect to $\mathbf{a}_{\mathcal{J}(k)}$ and $b_{k+1}, \ldots, b_{k_{\max}(k)}$. While the marginalization with respect to $\mathbf{a}_{\mathcal{J}(k)}$ is easily done in closed form, we want to avoid the marginalization with respect to the discrete-valued $b_{k+1}, \ldots, b_{k_{\max}(k)}$. Instead, we sample $\mathbf{b}_{\mathcal{J}(k)}$ from $p(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ and then use the $b_k$ contained in the sampled $\mathbf{b}_{\mathcal{J}(k)}$. The sampling distribution is

$$p\left(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\right) \propto \sigma_{\mathbf{b}}^2 \exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right) p(\mathbf{b}) \tag{21}$$

with

$$\mu_{\mathbf{b}} = \frac{\sigma_{\mathbf{b}}^2 \mathbf{b}_{\mathcal{J}(k)}^{\mathrm{H}} \mathbf{F}_{\mathcal{J}(k)}^{\mathrm{H}} \tilde{\mathbf{x}}^{\mathcal{J}(k)}}{\sigma_n^2}, \qquad \sigma_{\mathbf{b}}^2 = \frac{1}{\frac{\left\|\mathbf{F}_{\mathcal{J}(k)}\mathbf{b}_{\mathcal{J}(k)}\right\|^2}{\sigma_n^2} + \frac{1}{\sigma_a^2}}. \tag{22}$$

Here, $\tilde{\mathbf{x}}^{\mathcal{J}(k)} \triangleq \mathbf{x} - \mathbf{F}_{\sim\mathcal{J}(k)}\mathbf{a}_{\sim\mathcal{J}(k)}$, $\mathbf{F}_{\mathcal{J}(k)}$ consists of the columns of $\mathbf{F}$ indexed by $\mathcal{J}(k)$, and $\mathbf{F}_{\sim\mathcal{J}(k)}$ is $\mathbf{F}$ without these columns. We evaluate (21) for all hypotheses $\mathbf{b}_{\mathcal{J}(k)}$; summing the results yields the normalization constant which makes (21) a valid probability mass function (pmf). In this step, we exploit the minimum distance constraint: since $p(\mathbf{b})$ in (21) contains the factor $I_{\mathcal{C}}(\mathbf{b})$, all hypotheses $\mathbf{b}_{\mathcal{J}(k)}$ that violate the constraint have a probability of zero. Because the length $d(k)$ of $\mathcal{J}(k)$ is at most $d_{\min}$, this applies to all hypotheses $\mathbf{b}_{\mathcal{J}(k)}$ that contain more than one 1. Without the minimum distance constraint, there would be $2^{d(k)}$ hypotheses with potentially nonzero probability. With the constraint, there are only $d(k)+1$, namely one which contains no 1's and $d(k)$ which contain one 1. We only need to evaluate (21) for these $d(k)+1$ hypotheses. This drastic reduction of the number of hypotheses is the key to the high efficiency of the PS.

The $b_k$ are sampled in ascending order $k = 1, \ldots, K$, and the sampling distribution of each $b_k$ is conditioned on the previously sampled $b_1, \ldots, b_{k-1}$ (contained in $\mathbf{b}_{\sim\mathcal{J}(k)}$). Together with the factor $I_{\mathcal{C}}(\mathbf{b})$ in (21), this guarantees that the realization $\mathbf{b}^{(m)}$ obtained after the $K$ substeps is in $\mathcal{C}$. Thus, $\mathbf{b}^{(m)} \in \mathcal{C}$ for all $\mathbf{b}^{(m)}$ in $\mathcal{S}$.

*Amplitudes:* The sampling distribution for $a_k$ is

$$p\left(a_k|b_k, \mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\right) = \begin{cases} \delta(a_k), & \text{if } b_k = 0 \\ \mathcal{CN}\left(a_k; \tilde{\mu}_a, \tilde{\sigma}_a^2\right), & \text{if } b_k = 1 \end{cases} \tag{23}$$

[2] This also explains why we define $\mathcal{J}(k)$ as a one-sided neighborhood: within $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$, $(b_k, a_k)$ is the only entry that cannot be trimmed because it is conditioned upon in the next substep. If we defined $\mathcal{J}(k)$ as a two-sided neighborhood, supported on both sides of $k$, we would not be allowed to trim the entries with indices lower than $k$. This is a direct consequence of the order of sampling steps and substeps we choose for the algorithm, namely with $k$ ascending from 1 to $K$. If we chose a random order of these substeps (which would not violate the PCGS concept), no trimming could be performed at all, regardless of how we define $\mathcal{J}(k)$; i.e., we would have to sample $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$ from $p(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$ in each substep.

with

$$\tilde{\mu}_a = \frac{\tilde{\sigma}_a^2 \mathbf{f}_k^{\mathrm{H}} \tilde{\mathbf{x}}^{\mathcal{J}(k)}}{\sigma_n^2}, \qquad \tilde{\sigma}_a^2 = \frac{1}{\frac{\|\mathbf{f}_k\|^2}{\sigma_n^2} + \frac{1}{\sigma_a^2}}. \tag{24}$$

Here, $\mathbf{f}_k$ is the $k$th column of $\mathbf{F}$. When the sampled $b_k$ equals 1, then $\mathbf{F}_{\mathcal{J}(k)}\mathbf{b}_{\mathcal{J}(k)} = \mathbf{f}_k$ in (22), and thus $\sigma_{\mathbf{b}}^2 = \tilde{\sigma}_a^2$ and $\mu_{\mathbf{b}} = \tilde{\mu}_a$. This means that $\tilde{\mu}_a$ and $\tilde{\sigma}_a^2$ need not be calculated.

*Noise Variance:* The sampling distribution for $\sigma_n^2$ is

$$p\big(\sigma_n^2|\mathbf{a},\boldsymbol{\gamma},\mathbf{x}\big) = \mathcal{IG}\big(\sigma_n^2;\xi+K,\eta+\|\mathbf{x}-\mathbf{F}\mathbf{a}\|^2\big) \tag{25}$$

where, as in (12), $\mathcal{IG}(\cdot;\xi,\eta)$ denotes the inverse gamma pdf with parameters $\xi,\eta$.

### D. An Alternative PCGS

An alternative PCGS, referred to as "alternative sampler" (AS), can be formulated as follows.

**One AS iteration**

- Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{b},\mathbf{a}_b,\sigma_n^2,\mathbf{x})$.
- For $k = 1,\dots,K$, sample $b_k$ from $p(b_k|\mathbf{b}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$.
- Sample $\mathbf{a}_b$ from $p(\mathbf{a}_b|\mathbf{b},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$.
- Sample $\sigma_n^2$ from $p(\sigma_n^2|\mathbf{b},\mathbf{a}_b,\boldsymbol{\gamma},\mathbf{x})$.

This sampler, too, is a valid PCGS. The difference from the PS is that in the sampling substeps for the $b_k$, $\mathbf{a}$ is entirely marginalized out, which is similar to the samplers proposed in [23] and [35]. (Subsequently, $\mathbf{a}_b$ can be sampled jointly because it is of moderate dimension and jointly Gaussian.) At first sight, the AS appears to be a promising alternative to the PS, because it is "more collapsed" than the PS, i.e., its sampling distributions are conditioned on fewer parameters. Therefore, the gain in convergence rate relative to the Gibbs sampler is slightly larger than for the PS. However, the complexity of computing the sampling distributions is much higher. In particular, the simple expression $\sigma_{\mathbf{b}}^2 \exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)$ in (21) is replaced by $|\boldsymbol{\Sigma}_{\mathbf{a}}| \exp(\boldsymbol{\mu}_{\mathbf{a}}^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\boldsymbol{\mu}_{\mathbf{a}})$, with a length-$B$ vector $\boldsymbol{\mu}_{\mathbf{a}}$ and a $B\times B$ matrix $\boldsymbol{\Sigma}_{\mathbf{a}}$ that have to be calculated for each of the $d(k)+1$ hypotheses, each time a $b_k$ is sampled. An efficient update method described in [23] can be used for a recursive calculation of the relevant pmf, thus avoiding the explicit inversion of a $B\times B$ matrix for each hypothesis. However, even in that case, the AS is still significantly more complex than the PS.

### E. Reducing Complexity

The complexity of the PS depends strongly on the number $N$ of hypotheses $\mathbf{b}_{\mathcal{J}(k)}$ (for all $k$) for which the probabilities $p(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$ have to be evaluated in one PS iteration. An approximation of $N$ can be obtained by ignoring the reduced neighborhood lengths near the block boundary $k=K$, i.e., by assuming that all $K$ neighborhoods $\mathcal{J}(k)$, $k=1,\dots,K$ have length $d_{\min}$. Then, each of the $K$ sampling substeps of $b_k$ for $k=1,\dots,K$ requires evaluating $p(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$ in (21) for $d_{\min}+1$ hypotheses, and therefore, $N \approx K(d_{\min}+1)$. We can use two modifications of the PS to reduce $N$ approximately by the factor $d_{\min}+1$, so that $N\approx K$, without changing the sampler results. For simplicity, we assume that $k$ is not near the block boundary $K$, which means that $d(k)=d_{\min}$ and $k_{\max}(k)=k+d_{\min}-1$.

For the first modification, assume that we sample $b_k=1$ at position $k$. This $b_k=1$ will be present in the condition of the sampling distributions of all subsequent indicators $b_{k'}$, $k'=k+1,\dots,K$, and it forces the next $d_{\min}-1$ indicators to be zero. This is because the factor $p(\mathbf{b})$ in (21) assigns zero probability to all hypotheses $\mathbf{b}_{\mathcal{J}(k')}$ in which not all $b_{k+1},\dots,b_{k+d_{\min}-1}$ are zero. Therefore, after sampling $b_k=1$, the subsequent indicators $b_{k+1},\dots,b_{k+d_{\min}-1}$ can be set to zero and the corresponding $d_{\min}-1$ sampling substeps can be skipped. This includes skipping the sampling of $a_{k+1},\dots,a_{k+d_{\min}-1}$ and setting these amplitudes to zero.

The second modification applies to the complementary case, i.e., after sampling $b_k=0$. As a motivation, we note that (21) is proportional to $p(\mathbf{b},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$. Here, $\mathbf{b}$ consists of $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{b}_{\sim\mathcal{J}(k)}$, which are contained in the argument and condition of $p(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$ in (21), respectively. Let $\mathcal{H}_k$ denote the set of the $d_{\min}+1$ hypotheses $\mathbf{b}$ consisting of the respective $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{b}_{\sim\mathcal{J}(k)}$ (the latter is the same for all hypotheses). It can be shown—the case $k'=k+1$ will be elaborated presently—that $p(\mathbf{b},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x})$ is invariant to a shift of the neighborhood $\mathcal{J}(k)$ to $\mathcal{J}(k')$ for a given $k'$ if the nonoverlapping parts of $\mathbf{b}_{\sim\mathcal{J}(k)}$ and $\mathbf{b}_{\sim\mathcal{J}(k')}$ contain only zeros. (In that case, also the nonoverlapping parts of $\mathbf{a}_{\sim\mathcal{J}(k)}$ and $\mathbf{a}_{\sim\mathcal{J}(k')}$ contain only zeros.) This means that, if some $\mathbf{b}$ is contained in both $\mathcal{H}_k$ and $\mathcal{H}_{k'}$ and the nonoverlapping parts of $\mathbf{b}_{\sim\mathcal{J}(k)}$ and $\mathbf{b}_{\sim\mathcal{J}(k')}$ contain only zeros, the probability of this $\mathbf{b}$ is the same in the sampling substeps corresponding to $k$ and $k'$; thus, it has to be calculated only once.

An example for the case $k'=k+1$ (corresponding to hypothesis sets $\mathcal{H}_k$ and $\mathcal{H}_{k+1}$) is shown in Fig. 1. It is assumed that $d_{\min}=3$, so there are $d_{\min}+1=4$ hypotheses in $\mathcal{H}_k$ and also in $\mathcal{H}_{k+1}$. In $\mathcal{H}_k$ [see the left box in Fig. 1(a)], there is one hypothesis $\mathbf{b}$ with all indicators in $\mathcal{J}(k)$—i.e., $b_k,\dots,b_{k+d_{\min}-1}$—equal to zero. The other $d_{\min}$ hypotheses each contain exactly one 1 in $\mathcal{J}(k)$. The indicators at the remaining positions, $b_1,\dots,b_{k-1}$ and $b_{k+d_{\min}},\dots,b_K$, are the same for all $d_{\min}+1$ hypotheses. (Among them, note that $b_{k+d_{\min}}$, in particular, is equal to the respective realization drawn in the previous sampler iteration.)

Similarly, in $\mathcal{H}_{k+1}$ [see the right box in Fig. 1(a)], there is one hypothesis with all indicators in $\mathcal{J}(k+1)$—i.e., $b_{k+1},\dots,b_{k+d_{\min}}$—equal to zero, and each of the other $d_{\min}$ hypotheses contains exactly one 1 in $\mathcal{J}(k+1)$. The remaining indicators $b_1,\dots,b_k$ and $b_{k+d_{\min}+1},\dots,b_K$ are the same for all $d_{\min}+1$ hypotheses. (Among them, in particular, $b_k$ is equal to the respective realization drawn in the previous substep.)

The following can now be verified by inspection [see Fig. 1(b)]: If both $b_{k+d_{\min}}$ drawn in the previous sampler iteration and $b_k$ drawn in substep $k$ [highlighted by boxes in Fig. 1(b)] are zero, then $d_{\min}$ of the hypotheses in $\mathcal{H}_{k+1}$ appeared already in $\mathcal{H}_k$. (In Fig. 1(b), these $d_{\min}=3$ hypotheses are indicated by arrows.) This means that only one

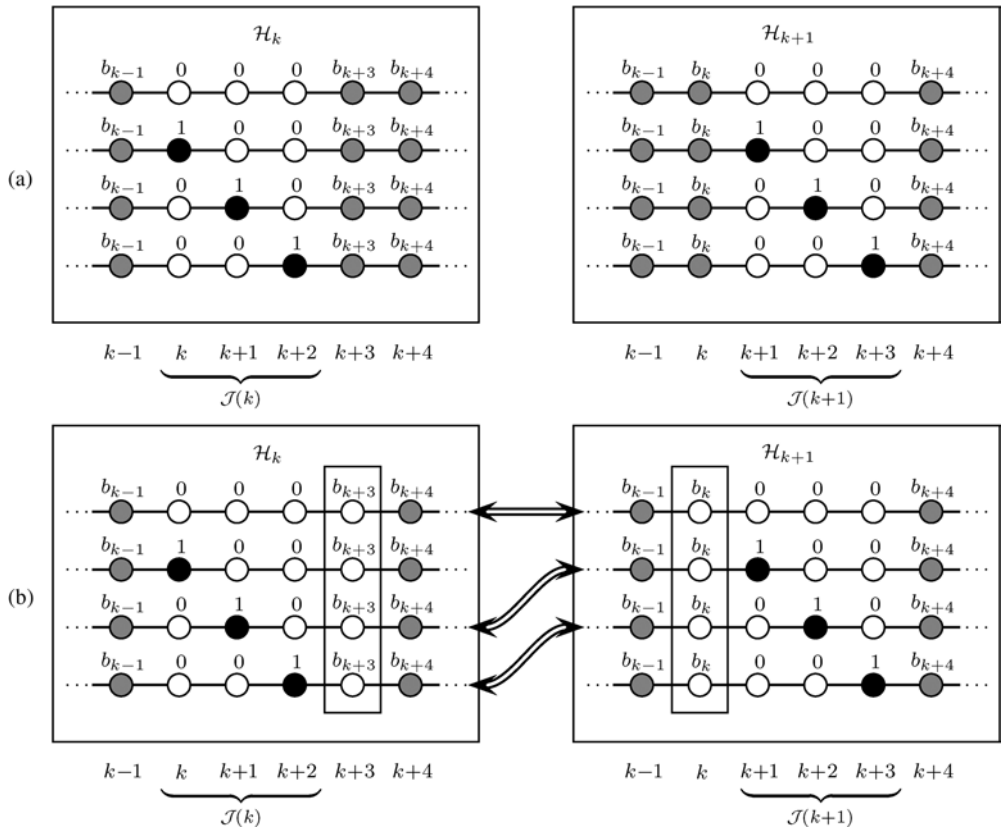


Fig. 1. (a) Hypothesis sets $\mathcal{H}_k$ and $\mathcal{H}_{k+1}$ for $d_{\min} = 3$. White (black) nodes depict zero (nonzero) indicators. Gray nodes may be zero or nonzero, depending on the outcome of the past sampling substeps. (b) The same, under the condition that $b_{k+3}$ in $\mathcal{H}_k$ (left set) and $b_k$ in $\mathcal{H}_{k+1}$ (right set)—both are highlighted by boxes—are zero (white). It can be seen that $d_{\min} = 3$ sequences in $\mathcal{H}_k$ also appear in $\mathcal{H}_{k+1}$, as indicated by the arrows.

hypothesis is new. Therefore, $p(\mathbf{b}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ has to be calculated only for one hypothesis; for all others, the values of $p(\mathbf{b}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ from the previous sampling substep can be reused. (Note, however, that the values to be reused are those of (21), i.e., before normalization.)

Therefore, after sampling $b_k = 0$, (21) is calculated for only one hypothesis in substep $k+1$ (unless $b_{k+d_{\min}} = 1$), which amounts to an average computation of little more than one probability per substep.[3] In the complementary case, i.e., after sampling $b_k = 1$, we apply the first modification and skip $d_{\min} - 1$ substeps. Then, we calculate (21) for $d_{\min}+1$ hypotheses in the following substep $k + d_{\min}$. Again, this amounts to an average computation of approximately one probability per substep. The effects of the two modifications thus complement each other. When both modifications are used, the overall average complexity per PS iteration, $N$, is reduced from about $K(d_{\min}+1)$ to about $K$, the number of substeps.

## V. Resolving the Scale and Shift Ambiguity

### A. Problem Formulation

Assuming for the moment an infinite temporal domain of $x_k$, $a_k$, and $f_k$, we have $f_k * a_k = f'_k * a'_k$ for $a'_k = c a_{k+\kappa}$ and $f'_k = \frac{1}{c} f_{k-\kappa}$, with an arbitrary amplitude scale factor $c \in \mathbb{C} \setminus \{0\}$ and an arbitrary time shift $\kappa \in \mathbb{Z}$. Provided that some $\boldsymbol{\gamma}'$ exists such that $(\mathbf{H}\boldsymbol{\gamma}')_k = \frac{1}{c}(\mathbf{H}\boldsymbol{\gamma})_{k-\kappa}$, this can be expressed as $\mathrm{toep}(\mathbf{H}\boldsymbol{\gamma})\mathbf{a} = \mathrm{toep}(\mathbf{H}\boldsymbol{\gamma}')\mathbf{a}'$. This equality has the following two consequences.

First, the observation $\mathbf{x}$ is invariant to amplitude scalings and time shifts of the *true* $\mathbf{a}$ and $\boldsymbol{\gamma}$. Thus, unless the true parameters are constrained in a way that uniquely defines $c$ and $\kappa$, BD methods are inherently invariant to amplitude scalings and time shifts of the true $\mathbf{a}$ and $\boldsymbol{\gamma}$. This means that the BD result may feature an incorrect amplitude scale and time shift. In practice, this is not necessarily a major problem.

Second, the likelihood function in (13) is invariant to amplitude scalings and time shifts of its arguments:

$$p\big(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2\big) = \mathcal{CN}\big(\mathbf{x}; \mathrm{toep}(\mathbf{H}\boldsymbol{\gamma})\mathbf{a}, \sigma_n^2\mathbf{I}\big) = p\big(\mathbf{x}|\mathbf{a}', \boldsymbol{\gamma}', \sigma_n^2\big)\,.$$

Therefore, the likelihood function is strongly multimodal: instead of a global maximum, there is an equivalence class of combinations $(\mathbf{a}', \boldsymbol{\gamma}', \sigma_n^2)$ with different $c$ and $\kappa$ that maximizes the likelihood. With respect to time shifts, however, the ambiguity is relaxed in our case because our system model does not satisfy the two assumptions made above, namely, infinite temporal support and existence of a $\boldsymbol{\gamma}'$ such that $(\mathbf{H}\boldsymbol{\gamma}')_k = (\mathbf{H}\boldsymbol{\gamma})_{k-\kappa}$. In fact, since $a_k$ and $f_k$ are confined to the time intervals $\{1, \ldots, K\}$ and $\{-Q, \ldots, Q\}$, respectively, some time shifts lead to the loss

[3] The effect of the case $b_{k+d_{\min}} = 1$ on the overall complexity is small, since $\mathbf{b}$ is sparse and thus contains only few nonzero entries. This means that the case $b_{k+d_{\min}} = 1$ occurs very rarely.

of nonzero values $a_k$ or $f_k$ as they are shifted outside their respective interval. In this case, the likelihood changes. Furthermore, a $\boldsymbol{\gamma}'$ satisfying the time shift invariance relation $(\mathbf{H}\boldsymbol{\gamma}')_k = (\mathbf{H}\boldsymbol{\gamma})_{k-\kappa}$ *exactly* typically does not exist; the relation can only be satisfied approximately. This means that the likelihood function is only approximately invariant to time shifts. Therefore, there is generally one specific time shift that fits the data best, although others may be almost as good.

As explained in Section II-C, our BD method is based on the joint posterior distribution $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2)p(\mathbf{a}|\mathbf{b})p(\mathbf{b})p(\boldsymbol{\gamma})p(\sigma_n^2)$, where the likelihood appears as one of the factors. Due to the other factors, the posterior is not invariant to scalings and time shifts. Nevertheless, it usually preserves a multimodal structure similar to that of the likelihood function. This is problematic for both phases of MCMC detection/estimation: the generation of the sample and the sample-based detection/estimation.

### *B. Scale and Shift Ambiguity in the Sampling Phase*

In the sampling phase, the multimodality of the posterior leads to slow convergence of the sampler. In fact, each $\mathbf{a}$ has an "ideal partner" $\mathbf{f}$ with respect to scale and shift, and vice versa. The sampling steps proposed in Section IV-B generate new realizations of $\mathbf{a}$ and $\boldsymbol{\gamma}$ conditioned on the respective other vector, rather than jointly. Therefore, the scale and shift parameters hardly change, i.e., the sampler stays within one mode of the likelihood for many iterations. Following [23], [48], this problem can be avoided by adding two joint sampling steps for $(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$, one for shift compensation and one for scale compensation. Both are inserted after the first PS step, i.e., after sampling $\boldsymbol{\gamma}$ (see Section IV-B), and they use the Metropolis–Hastings (MH) sampling algorithm [29, p. 267]. Based on the current realization $(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$, we first sample a proposal $(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}')$ from some proposal kernel $\pi(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}'|\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$. Then, with some acceptance probability $P$, we replace $(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$ by $(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}')$, i.e., $(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}')$ becomes the current realization of the sampler, whereas with probability $1 - P$, $(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$ remains the current realization. Here, $P$ is determined by $\pi(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}'|\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma})$ and the stationary distribution of the sampler, $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$ (see [29, p. 267] and [48] for details).

The first MH sampling step is for shift compensation and consists of the following substeps.

- Sample $\kappa$ from a uniform distribution on $\{-\kappa_{\max}, \ldots, \kappa_{\max}\}$, denoted $p(\kappa)$, with some fixed $\kappa_{\max} \in \mathbb{N}$.
- Obtain $\mathbf{a}'$ and $\mathbf{b}'$ from $\mathbf{a}$ and $\mathbf{b}$ by means of a circular shift by $\kappa$ steps, i.e., $a'_k = a_{[(k-\kappa-1) \bmod K]+1}$ and $b'_k = b_{[(k-\kappa-1) \bmod K]+1}$.
- Sample $\boldsymbol{\gamma}'$ from $p(\boldsymbol{\gamma}|\mathbf{a}', \sigma_n^2, \mathbf{x})$ [see (19)]. (It can be shown that this corresponds to the proposal kernel $\pi(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}'|\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}) = p(\boldsymbol{\gamma}'|\mathbf{x}, \mathbf{a}', \sigma_n^2)p(\kappa)$.)
- With probability $P_\kappa$, replace $\mathbf{b}$, $\mathbf{a}$, and $\boldsymbol{\gamma}$ by $\mathbf{b}'$, $\mathbf{a}'$, and $\boldsymbol{\gamma}'$, respectively. Here,

$$P_\kappa = \min\left\{1, \frac{p\big(\mathbf{x}|\mathbf{a}', \sigma_n^2\big)\, I_{\mathcal{C}}(\mathbf{b}')}{p(\mathbf{x}|\mathbf{a}, \sigma_n^2)}\right\},$$

where $p\left(\mathbf{x}|\mathbf{a}, \sigma_n^2\right) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}| \exp\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{H}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}}\right)$, with $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ as defined in (20).

This shift compensation method differs from that in [23] and [48] by the support of the uniform distribution of $\kappa$, which is $\{-\kappa_{\max}, \ldots, \kappa_{\max}\}$ instead of $\{-1,0,1\}$. This allows larger jumps, which we observed to be beneficial. Another difference, due to our different priors, is the presence of $I_{\mathcal{C}}(\mathbf{b}')$ in the expression of $P_\kappa$.

The MH sampling step for shift compensation is succeeded by that for scale compensation, which is based on the same concept and consists of the following substeps:

- Generate $c = e^{\alpha + j\phi}$ by sampling $\alpha$ from $\mathcal{N}(\alpha; 0, \sigma_\alpha^2)$ with some fixed $\sigma_\alpha^2$ and $\phi$ from a uniform distribution on $[-\pi, \pi)$. (These pdf's will be denoted by $p(\alpha)$ and $p(\phi)$.)
- The proposal consists of $\mathbf{b}' = \mathbf{b}$, $\mathbf{a}' = c\mathbf{a}$, and $\boldsymbol{\gamma}' = \frac{\boldsymbol{\gamma}}{c}$, where $\mathbf{b}$, $\mathbf{a}$, and $\boldsymbol{\gamma}$ denote the realizations that were accepted in the shift compensation step. (It can be shown that this corresponds to the proposal kernel $\pi(\mathbf{b}', \mathbf{a}', \boldsymbol{\gamma}'|\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}) = p(c) = e^{-\alpha}p(\alpha)p(\phi)$.)
- With probability $P_c$, replace $\mathbf{a}$ by $\mathbf{a}'$ and $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}'$. Here,

$$\begin{aligned} P_c &= \min\left\{1, e^{2\alpha}\frac{p(\mathbf{a}')p(\boldsymbol{\gamma}')}{p(\mathbf{a})p(\boldsymbol{\gamma})}\right\} \\ &= \min\left\{1, \exp\left(2\alpha - \frac{(e^{2\alpha}-1)\|\mathbf{a}\|^2}{\sigma_a^2} \right.\right. \\ &\qquad \left.\left. - \frac{(e^{-2\alpha}-1)\|\boldsymbol{\gamma}\|^2}{\sigma_\gamma^2}\right)\right\}. \end{aligned}$$

After shift/scale compensation, the PS continues with the $b_k$ sampling step, as described in Section IV-B.

### *C. Scale and Shift Ambiguity in the Detection/Estimation Phase*

In the sample-based detection/estimation algorithms of Section III, the scale and shift ambiguity causes severe problems. For example, the estimators are averages over the sample $\mathcal{S} = \{(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})\}_{m=1}^{M}$, which become meaningless if the individual realizations $(\mathbf{b}^{(m)}, \mathbf{a}^{(m)}, \boldsymbol{\gamma}^{(m)}, \sigma_n^{2(m)})$ feature different scales and shifts. A method for achieving identical scales and shifts in all realizations is described in [48].

## VI. Numerical Study

### *A. Simulation Setup*

We will compare the performance of the proposed sampling method (PS) with that of the AS described in Section IV-D. As a performance benchmark, we also consider the following "reference sampler" (abbreviated RS) that does *not* exploit the minimum-distance constraint.

**One RS iteration**

- Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{b}, \mathbf{a}_b, \sigma_n^2, \mathbf{x})$.
- For $k = 1, \ldots, K$, sample $b_k$ from $p(b_k|\mathbf{b}_{\sim k}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.
- Sample $\mathbf{a}_b$ from $p(\mathbf{a}_b|\mathbf{b}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.
- Sample $\sigma_n^2$ from $p(\sigma_n^2|\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \mathbf{x})$.

This sampling algorithm, up to minor modifications, was proposed in [23] for a Bernoulli–Gaussian sequence $a_k$. Note that

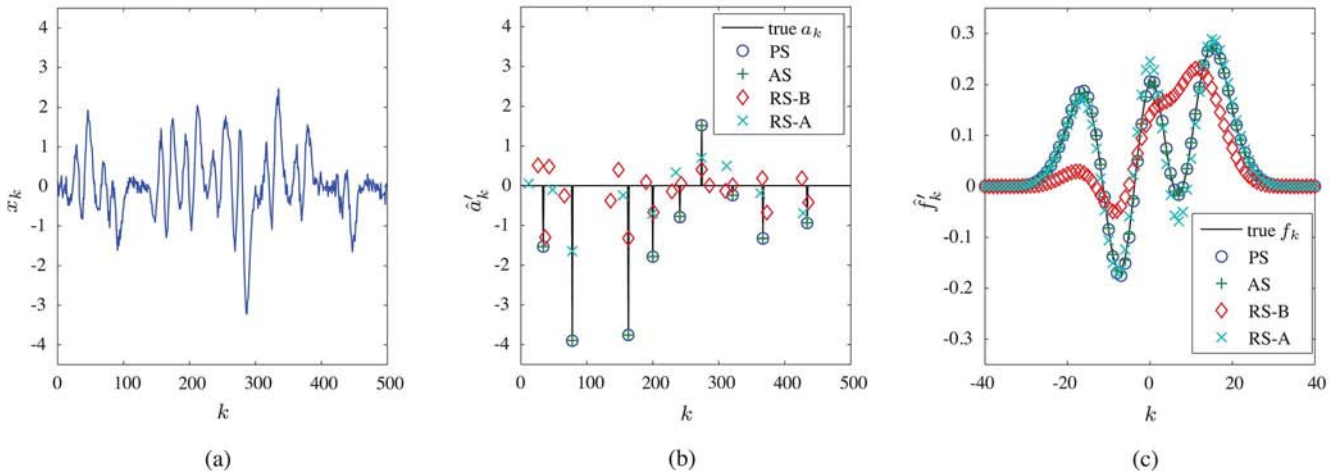


Fig. 2. Results of detection/estimation: (a) Signal $\mathbf{x}$, (b) estimates $\hat{\mathbf{a}}'$, (c) estimates $\hat{\mathbf{f}}'$. The results shown here are obtained after 60 iterations (PS and AS) or 1500 iterations (RS-A and RS-B). The vertical lines in (b) indicate the true $\mathbf{a}$. Real parts are shown.

our signal model is different because the Bernoulli–Gaussian prior of $a_k$ is modified by the minimum-distance constraint. Just as the PS and AS, the RS is a PCGS, not a classical Gibbs sampler, because the sampling distribution for $b_k$ is a conditional distribution associated with $p(\mathbf{b}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$, rather than with the full joint posterior $p(\mathbf{b}, \mathbf{a}_b, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$. The RS differs from the AS in that the AS exploits the minimum distance constraint by sampling $b_k$ from $p(b_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$. Like the AS, the RS requires many matrix inversions. It is much more complex than the PS, even if the recursive inversion method of [23] is used.

We consider two versions of the RS. The first, denoted RS-A, is based on the true signal model (with minimum distance constraint), in which $p(\mathbf{b}) \propto \mathcal{B}(\mathbf{b}; \pi_1) I_{\mathcal{C}}(\mathbf{b})$. Since the RS behaves like a classical Gibbs sampler with respect to dependencies within $\mathbf{b}$, it is not well suited to this prior. Therefore, we consider also a second RS version, denoted RS-B, which (at the cost of a model mismatch) is based on the unconstrained Bernoulli–Gaussian model for which the RS was proposed in [23]. We thus interpret realizations from $p(\mathbf{b})$ as Bernoulli sequences from $\tilde{p}(\mathbf{b}) = \mathcal{B}(\mathbf{b}; \tilde{\pi}_1)$, where $\tilde{\pi}_1$ is the approximation of $\mathrm{E}\{B/K\}$ described in Section II-B. It can be shown that the average distances between 1's in sequences drawn from $p(\mathbf{b})$ and $\tilde{p}(\mathbf{b})$ are approximately identical. Consistently replacing $p(\mathbf{b})$ by $\tilde{p}(\mathbf{b})$ and $\pi_1$ by $\tilde{\pi}_1$ in RS-A, we obtain RS-B.

We generated 100 realizations of $\mathbf{x}$ from parameters randomly drawn according to the priors given in Section II-B, using $K = 1024$, $Q = 64$, $U = 5$, $d_{\min} = 30$, $\pi_1 = 0.15$, $\sigma_a^2 = 10$, $\sigma_\gamma^2 = 2.4$, $\xi = 11$, and $\eta = 0.5$ (the latter two parameters provide a noninformative prior for $\sigma_n^2$). For each realization of $\mathbf{x}$, we generated four Markov chains according to the four sampler methods. Detection and estimation were then performed on each of the four samples as described in Section III. As mentioned in Section V, the time shift and amplitude scale of the estimate $\hat{\mathbf{a}}$ are arbitrary and, indeed, usually irrelevant. Therefore, for performance assessment, we matched the time shift and amplitude scale of each estimate $\hat{\mathbf{a}}$ to the true $\mathbf{a}$, i.e., we calculated the shifted/scaled version $\hat{\mathbf{a}}'$ of $\hat{\mathbf{a}}$ minimizing $\|\hat{\mathbf{a}}' - \mathbf{a}\|^2$, and the corresponding shifted/scaled $\hat{\mathbf{f}}' = \mathbf{H}\hat{\boldsymbol{\gamma}}'$.

### *B. Simulation Results*

As an example, the result of one simulation run—corresponding to one realization of $\mathbf{x}$—is shown in Fig. 2. Here, the detected/estimated sequences $\hat{\mathbf{a}}'$ and pulse shapes $\hat{\mathbf{f}}' = \mathbf{H}\hat{\boldsymbol{\gamma}}'$ of both the PS and AS are seen to coincide with the true $\mathbf{a}$ and $\mathbf{f}$ after only 60 iterations. The results of RS-A and RS-B after 1500 iterations are significantly worse. We note that all methods use the joint conditional estimator $\hat{\mathbf{a}}_{b,\mathrm{MMSE}}$ in (18); however, almost identical results are obtained with the componentwise sample-based estimator $\hat{a}_{k,\mathcal{S}}$ in (17).

To assess the convergence rates of the various samplers, Fig. 3(a) shows the empirical normalized mean-square error (NMSE) of $\hat{\mathbf{a}}'$ versus the number of iterations. The empirical NMSE is defined as the average (over the 100 realizations) of $\|\hat{\mathbf{a}}' - \mathbf{a}\|^2$ normalized by the average of $\|\mathbf{a}\|^2$. The number of iterations indicated on the abscissa equals the total length of the Markov chain. Out of each chain, the last 20% of the iterations were used for detection/estimation. Again, $\hat{\mathbf{a}}_{b,\mathrm{MMSE}}$ was used for amplitude estimation; however, the NMSE obtained with $\hat{a}_{k,\mathcal{S}}$ is effectively equal for AS, RS-A, and RS-B and only about 0.3 dB higher for PS. The values of the NMSE of $\hat{\mathbf{a}}'$ after 60 and 1500 iterations are also given in the first two rows of Table I. It is seen from Fig. 3(a) that RS-A fails to produce satisfactory results, and its error does not decrease with time. This is easily explained by the bad match between model and algorithm: the model features strongly dependent indicators because of the minimum distance constraint, whereas RS-A treats the indicators like a classical Gibbs sampler, which performs poorly in the presence of strong dependencies. This problem is circumvented in RS-B, since the model is adapted to the algorithm by dropping the minimum distance constraint. Indeed, we observe a steady decrease of the error, but at a very low rate: after 1500 iterations, the NMSE has only decreased by less than 2 dB. By contrast, both PS and AS achieve a low error after about 50 iterations. This fast convergence may appear surprising, as the MCMC concept is based on the law of large numbers. It can, however, be explained by the fact that the minimum distance constraint excludes large parts of the parameter space that would otherwise have a nonnegligible posterior

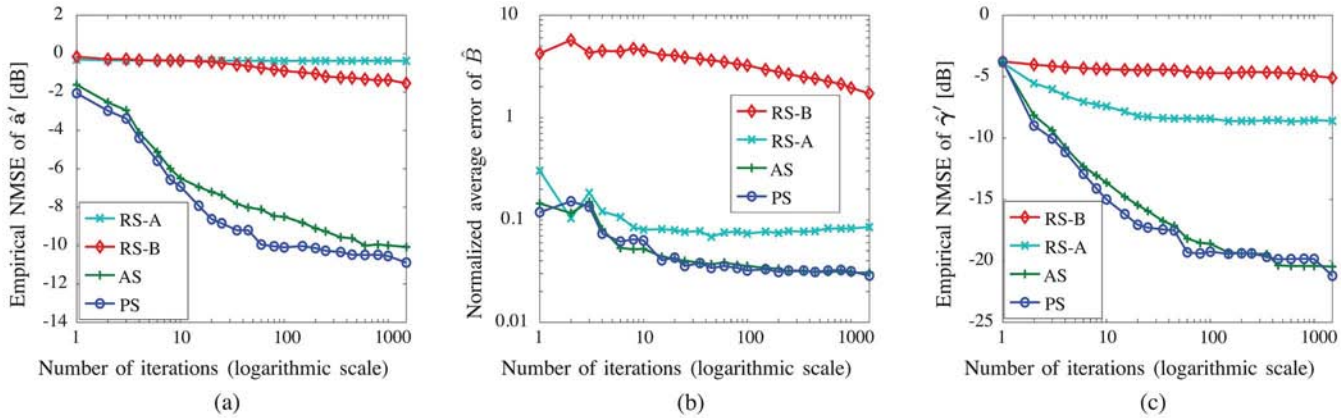


Fig. 3. Detection/estimation performance versus the number of iterations: (a) Empirical NMSE of $\hat{\mathbf{a}}'$, (b) normalized average error of $\hat{B} = \|\hat{\mathbf{b}}\|^2$, (c) empirical NMSE of $\hat{\boldsymbol{\gamma}}'$.

TABLE I
REPRESENTATIVE SELECTION OF SIMULATION RESULTS

| | | PS | AS | RS-A | RS-B |
|---|---|---|---|---|---|
| Empirical NMSE of $\hat{\mathbf{a}}'$ in dB (cf. Fig. 3(a)) | after 60 iterations | $-9.94$ | $-8.12$ | $-0.38$ | $-0.74$ |
| | after 1500 iterations | $-10.89$ | $-10.07$ | $-0.38$ | $-1.55$ |
| Normalized average error of $\hat{B}$ (cf. Fig. 3(b)) | after 60 iterations | $3.6 \cdot 10^{-2}$ | $3.8 \cdot 10^{-2}$ | $7.5 \cdot 10^{-2}$ | 3.5 |
| | after 1500 iterations | $2.9 \cdot 10^{-2}$ | $3.1 \cdot 10^{-2}$ | $8.5 \cdot 10^{-2}$ | 1.7 |
| Computation time | after 20 iterations | 1.63s | 14.15s | 4.48s | 27.26s |

probability. Within the remaining parts of the parameter space, the posterior probability is thus increased, which leads to faster convergence. From the superiority of the results of PS and AS over those of RS-A and RS-B, we can conclude that the use of the minimum distance constraint is highly beneficial.

In Fig. 3(b), we study the sparsity of the estimates $\hat{\mathbf{a}}$ (or, equivalently, $\hat{\mathbf{a}}'$) by assessing the accuracy of $\hat{B} \triangleq \|\hat{\mathbf{b}}\|^2$. More specifically, we show the average (over the 100 realizations) of $|\hat{B} - B|$ normalized by the average of $B$ versus the number of iterations. The values of the normalized average error of $\hat{B}$ after 60 and 1500 iterations are also given in the third and fourth rows of Table I. In RS-B, the absence of a strict sparsity constraint leads to very high values of $\hat{B}$ (up to 5 times the true $B$) in the first iterations. After the first 10 iterations, the error slowly but steadily decreases. In RS-A, a certain level of sparsity is enforced by the minimum distance constraint, leading to a normalized error between 0.07 and 0.09 after about 10 iterations. This error does not decrease with further iterations. The sparsity achieved by RS-A may be meaningless in view of the failed convergence of $\hat{\mathbf{a}}'$. By contrast, both PS and AS reach errors below 0.04 after about 25 iterations.

Fig. 3(c) shows the empirical NMSE of $\hat{\boldsymbol{\gamma}}'$ versus the number of iterations. The results are roughly similar to those in Fig. 3(a). It can be seen that, in general, the error of the estimated pulse shapes is smaller than that of the estimated sparse sequences. Furthermore, RS-A outperforms RS-B because its realizations of $\mathbf{a}$ are more sparse due to the minimum distance constraint. However, this is of little relevance since both RS-A and RS-B fail to estimate $\mathbf{a}$ appropriately.

One reason why RS-B converges so slowly appears to be our choice of a relatively wide class of random pulse shapes in our simulation. To study this issue, we generated 100 realizations of $\mathbf{x}$ from parameters drawn as described above, except that the pulse shape was fixed, namely, the one used in [23]. The estimation methods were not changed, i.e., they still estimated the pulse shape rather than using the true one. Fig. 4 shows the empirical NMSE of $\hat{\mathbf{a}}'$ versus the number of iterations for this case. It is seen that the error of RS-B (and also that of PS and AS) decreases significantly faster than in Fig. 3(a). We can conclude that many of the pulse shapes that we generated randomly are harder to detect and estimate than the pulse shape used in [23].

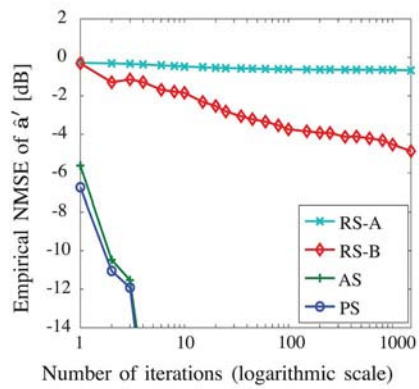


Fig. 4. Empirical NMSE of $\hat{\mathbf{a}}'$ versus the number of iterations, for realizations of $\mathbf{x}$ using the pulse shape described in [23].

Next, we illustrate the rationale of our choice of the block detector $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{b})}$ (see Section III-B). Out of a Markov chain of length 1000 generated by the PS, the last 200 realizations of $\mathbf{b}$ were used as a sample. Within the sample, two realizations appear 22 times and no realization appears more than 22 times. In this case, the sample-based MAP sequence detector $\hat{\mathbf{b}}_{\mathcal{S}}^{(\mathrm{s})}$ in

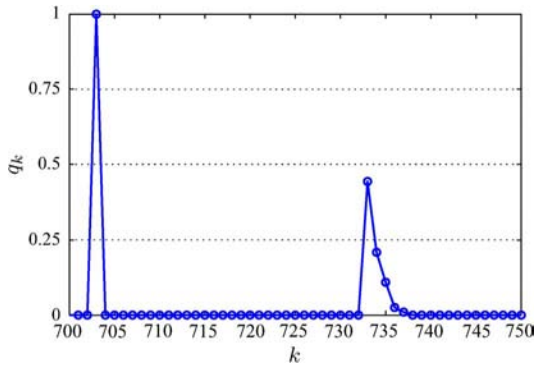


Fig. 5. Relative multiplicity $q_k = q(b_k = 1) \approx p(b_k = 1|\mathbf{x})$ versus $k$.

Section III-A would fail, since the maximum $q(\mathbf{b})(= \frac{22}{200})$ is achieved by two realizations $\mathbf{b}$ simultaneously. As explained in Section III-A, this is due to the coarse quantization of the probability. In another simulation run, again with sample size 200, we obtained a sequence $q_k = q(b_k = 1)$ that is depicted for $k \in \{700, \ldots, 750\}$ in Fig. 5. Within that interval, there are two "zero intervals" (where no realization contains a 1, and thus $q_k = 0$ in the zero intervals); these zero intervals are separated by the "nonzero interval" $\{733, \ldots, 737\}$ (i.e., $q_k > 0$ for all $k \in \{733, \ldots, 737\}$). Within the sample, the following different realizations of $(b_{733}, \ldots, b_{737})$ occurred. Out of the 200 realizations, 89 contain a 1 at position 733, 42 contain a 1 at position 734, 22 contain a 1 at position 735, 4 contain a 1 at position 736, and 2 contain a 1 at position 737. The remaining 41 realizations contain no 1 within $\{733, \ldots, 737\}$. No realization contains more than one 1 within $\{733, \ldots, 737\}$. Thus, at *each* position $k \in \{733, \ldots, 737\}$, less than half of the 200 realizations contain a 1; therefore, the component detector would not detect any 1's. In the block detector, on the other hand, the five positions $k \in \{733, \ldots, 737\}$ are combined into a block $\boldsymbol{\beta} = (b_{733} \cdots b_{737})^{\mathrm{T}}$. According to the multiplicities in the sample listed above, there are six hypotheses $\boldsymbol{\beta}$ with relative multiplicities $q(\boldsymbol{\beta})$ given by 89/200, 42/200, 22/200, 4/200, 2/200, and 41/200. The maximum relative multiplicity $q(\boldsymbol{\beta})$ (89/200) is achieved by the hypothesis $\boldsymbol{\beta}$ with a 1 at position 733. Thus, the MAP block detector would detect a 1 at position 733, which is more intuitive than the result of the component detector.

We finally note that the computational complexities of the PS, AS, RS-A, and RS-B differ significantly. Exemplary computation times for 20 iterations are reported in the last row of Table I for an unoptimized Matlab R2009b 64-bit implementation on a 2.2-GHz Intel Core 2 Duo processor. The processing time of the AS is almost 9 times that of the PS. This high complexity is not justified by a better performance, since (as shown by our simulation results) the performance of the AS is very similar to that of the PS. The processing times of the RS-A and RS-B are about 3 and 17 times that of the PS, respectively.

## VII. Conclusion

We studied Bayesian blind deconvolution of an unknown sparse sequence convolved with an unknown pulse. Our approach extended the conventional Bernoulli–Gaussian prior (modeling sparsity) by a hard minimum distance constraint, which requires any two detected pulse locations to have a certain minimum separation. Such a constraint is physically motivated in many applications and is an effective means of avoiding spurious detected pulses. However, the minimum distance constraint implies strong dependencies, which lead to slow convergence of the classical Gibbs sampler. We demonstrated that this problem can be overcome by a new Monte Carlo blind deconvolution method based on the recently introduced PCGS principle. The proposed method exploits the structure that the constraint imposes on the parameter space to achieve fast convergence and low computational complexity. Our simulation results demonstrated a significant reduction of both the complexity and the number of spurious components compared to a recently proposed PCGS that is not specifically designed for the minimum distance constraint.

The proposed method can be easily generalized to inverse problems with a minimum distance constraint that are not described by a convolution model. This makes it potentially interesting for many signal processing applications including signal segmentation [49], optical coherence tomography [50], electromyography [10], [11], and electrocardiography [13], [51]. Furthermore, the minimum distance constraint itself can be generalized to a wider class of "local deterministic constraints," which can be exploited in an analogous way [36].

## Appendix A
## Validation of the PS

To validate the PS as a PCGS, we will derive it from a classical Gibbs sampler by means of marginalization and trimming (see [33] and Section IV-A). These modifications are allowed because they do not change the stationary distribution of the sampler. One iteration of the Gibbs sampler is given as follows:

- Sample $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x})$.
- For $k = 1, \ldots, K$, sample $(b_k, a_k)$ from $p(b_k, a_k|\mathbf{b}_{\sim k}, \mathbf{a}_{\sim k}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.
- Sample $\sigma_n^2$ from $p(\sigma_n^2|\mathbf{a}, \boldsymbol{\gamma}, \mathbf{x})$.

This is a valid Gibbs sampler, since each parameter is sampled once per iteration and each sampling distribution is conditioned on all other parameters (i.e., all except those being sampled).[4] This concept is not compromised by the fact that some parameters are sampled jointly, namely $b_k$ and $a_k$ as well as all entries of $\boldsymbol{\gamma}$.

The first modification we use is a marginalization, which leads to a sampler in which the second step of the Gibbs sampler is replaced by the following:

- For $k = 1, \ldots, K$, sample $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$ from $p(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.

Here, additional parameters are sampled in each sampling step. The sampler is no longer a Gibbs sampler, since the $b_k$ and $a_k$ (for all $k$) are now sampled several times within each sampler iteration, in varying combinations. The choice of $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$ as a replacement of the tuple $(b_k, a_k)$ sampled by the original Gibbs sampler is key to the fast convergence of the PS. On

[4] As mentioned in Section IV-B, $\mathbf{b}$ does not appear in the conditions of the sampling distributions of $\boldsymbol{\gamma}$ and $\sigma_n^2$ because, for a given $\mathbf{a}$, $\mathbf{b}$ is conditionally independent of $\boldsymbol{\gamma}$ and $\sigma_n^2$.

the one hand, the subvector $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$ is large enough so that the parameters that strongly depend on each other (due to the minimum distance constraint) are sampled jointly. On the other hand, it is small enough to allow efficient sampling as discussed below (22). Note that all sampling distributions are still conditioned on all parameters except those sampled. Thus, they are still conditionals associated with the joint posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$.

Next, we apply trimming to obtain the following modified second step:

- For $k = 1, \ldots, K$, sample $(b_k, a_k)$ from $p(b_k, a_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.

Since some entries of $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$—namely, all except $(b_k, a_k)$—are removed from the argument of the sampling distribution of the respective substeps, these sampling distributions are no longer proportional to the joint posterior $p(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x})$. As previously explained in Section IV-B, the trimming is justified because all elements of $\mathcal{J}(k) = \{k, \ldots, k_{\max}(k)\}$ except $k$ itself are also contained in $\mathcal{J}(k+1) = \{k+1, \ldots, k_{\max}(k+1)\}$. Therefore, out of $(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)})$, the untrimmed sampler uses only $(b_k, a_k)$ in the condition of the next substep, whereas the other entries are ignored. (This is true even for $k = K$, trivially, since $\mathcal{J}(K) = \{K\}$.)

Finally, joint sampling of $(b_k, a_k)$ can be achieved by first sampling $b_k$ and then sampling $a_k$ conditioned on the $b_k$ thus obtained. Thus, the following final version of the second step is obtained:

- For $k = 1, \ldots, K$,
  — sample $b_k$ from $p(b_k|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$;
  — sample $a_k$ from $p(a_k|b_k, \mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$.

Together with the first and third steps of the Gibbs sampler, this is equal to the PS described in Section IV-B.

## Appendix B
## Sampling Distributions

In this appendix, we derive the sampling distributions used by the PS (cf. Section IV-C).

*Pulse Coefficients:* To obtain the sampling distribution for $\boldsymbol{\gamma}$ in (19) and (20), we note that

$$p\big(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x}\big) \propto p\big(\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x}\big) \propto p\big(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2\big)\, p(\boldsymbol{\gamma}).$$

We now insert (13) and (11), and obtain

$$\begin{aligned} p\big(\boldsymbol{\gamma}|\mathbf{a},\sigma_n^2,\mathbf{x}\big) &\propto \exp\!\left(-\frac{\|\mathbf{x}-\mathbf{A}\mathbf{H}\boldsymbol{\gamma}\|^2}{\sigma_n^2}\right)\exp\!\left(-\frac{\|\boldsymbol{\gamma}\|^2}{\sigma_\gamma^2}\right) \\ &= \exp\!\left(-\frac{1}{\sigma_n^2}(\mathbf{x}-\mathbf{A}\mathbf{H}\boldsymbol{\gamma})^{\mathrm{H}}(\mathbf{x}-\mathbf{A}\mathbf{H}\boldsymbol{\gamma})-\frac{1}{\sigma_\gamma^2}\boldsymbol{\gamma}^{\mathrm{H}}\boldsymbol{\gamma}\right). \end{aligned}$$

This can be rewritten in terms of $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ as defined in (20):

$$\begin{aligned} p\left(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x}\right) &\propto \exp\Bigg( -(\boldsymbol{\gamma}-\boldsymbol{\mu}_{\boldsymbol{\gamma}})^{\mathrm{H}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{\mu}_{\boldsymbol{\gamma}}) \\ &\qquad +\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{\mathrm{H}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}} - \frac{\|\mathbf{x}\|^2}{\sigma_n^2}\Bigg) \\ &\propto \exp\left(-(\boldsymbol{\gamma}-\boldsymbol{\mu}_{\boldsymbol{\gamma}})^{\mathrm{H}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{\mu}_{\boldsymbol{\gamma}})\right). \end{aligned}$$

Finally, normalization leads to $p(\boldsymbol{\gamma}|\mathbf{a}, \sigma_n^2, \mathbf{x}) = \mathcal{CN}(\boldsymbol{\gamma}; \boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$, which equals (19).

*Indicators:* In order to derive (21) and (22), we start from $p(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x})$. Noting that $\mathbf{b}$ is composed of $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{b}_{\sim\mathcal{J}(k)}$ and $\mathbf{a}$ is composed of $\mathbf{a}_{\mathcal{J}(k)}$ and $\mathbf{a}_{\sim\mathcal{J}(k)}$, and using (14), we obtain

$$\begin{aligned} &p\big(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)}, \mathbf{a}_{\sim\mathcal{J}(k)}, \boldsymbol{\gamma}, \sigma_n^2, \mathbf{x}\big) \\ &\quad\propto p\big(\mathbf{b}, \mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2|\mathbf{x}\big) \\ &\quad\propto p\big(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2\big)\, p(\mathbf{a}|\mathbf{b})p(\mathbf{b})p(\boldsymbol{\gamma})p\big(\sigma_n^2\big) \\ &\quad\propto p\big(\mathbf{x}|\mathbf{a}, \boldsymbol{\gamma}, \sigma_n^2\big)\, p\big(\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\mathcal{J}(k)}\big)\, p(\mathbf{b}). \end{aligned}$$

In the last step, we used (7) and dropped factors that are constant with respect to $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{a}_{\mathcal{J}(k)}$. Using (13) and again (7) yields

$$\begin{aligned} &p\big(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\ldots\big) \\ &\quad\propto \exp\left(-\frac{\|\mathbf{x}-\mathbf{F}\mathbf{a}\|^2}{\sigma_n^2}\right) p\big(\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\mathcal{J}(k)}\big)\, p(\mathbf{b}) \\ &\quad= \exp\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}-\mathbf{F}_{\mathcal{J}(k)}\mathbf{a}_{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right) \\ &\qquad\times\left[\prod_{k'\in\mathcal{J}(k)} p(a_{k'}|b_{k'})\right] p(\mathbf{b}) \end{aligned} \tag{26}$$

where $\tilde{\mathbf{x}}^{\mathcal{J}(k)} = \mathbf{x} - \mathbf{F}_{\sim\mathcal{J}(k)}\mathbf{a}_{\sim\mathcal{J}(k)}$ and $\mathbf{F}_{\mathcal{J}(k)}$ consists of the columns of $\mathbf{F}$ indexed by $\mathcal{J}(k)$.

We now exploit the minimum distance constraint, i.e., the fact that $\mathbf{a}_{\mathcal{J}(k)}$ can contain at most one nonzero entry. We will consider the cases $\|\mathbf{b}_{\mathcal{J}(k)}\| = 1$ and $\|\mathbf{b}_{\mathcal{J}(k)}\| = 0$ separately. (This means that we may not drop any constant factors until we find a joint expression for both cases.) For the case $\|\mathbf{b}_{\mathcal{J}(k)}\| = 1$, let $\hat{k}$ denote the (unknown) position of the nonzero entry. Noting that $\mathbf{a}_{\mathcal{J}(k)} = a_{\hat{k}}\mathbf{b}_{\mathcal{J}(k)}$ and using (8), we can write (26) as

$$\begin{aligned} &p\big(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\ldots\big) \\ &\quad\propto \exp\!\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}-\mathbf{F}_{\mathcal{J}(k)}\mathbf{b}_{\mathcal{J}(k)}a_{\hat{k}}\right\|^2}{\sigma_n^2}\right)\frac{1}{\pi\sigma_a^2}\exp\!\left(-\frac{|a_{\hat{k}}|^2}{\sigma_a^2}\right) \\ &\qquad\times\left[\prod_{k'\in\mathcal{J}(k)\setminus\{\hat{k}\}}\delta(a_{k'})\right] p(\mathbf{b}) \\ &\quad= \frac{1}{\pi\sigma_a^2}\exp\!\left(-\left[\frac{\left\|\mathbf{F}_{\mathcal{J}(k)}\mathbf{b}_{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}+\frac{1}{\sigma_a^2}\right]|a_{\hat{k}}|^2\right. \\ &\qquad\qquad\left.+2\mathrm{Re}\left\{a_{\hat{k}}^*\frac{\mathbf{b}_{\mathcal{J}(k)}^{\mathrm{H}}\mathbf{F}_{\mathcal{J}(k)}^{\mathrm{H}}\tilde{\mathbf{x}}^{\mathcal{J}(k)}}{\sigma_n^2}\right\}-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right) \\ &\qquad\times\left[\prod_{k'\in\mathcal{J}(k)\setminus\{\hat{k}\}}\delta(a_{k'})\right] p(\mathbf{b}). \end{aligned}$$

Using $\mu_{\mathbf{b}}$ and $\sigma_{\mathbf{b}}^2$ as defined in (22), this can be rewritten as

$$\begin{aligned} &p\big(\mathbf{b}_{\mathcal{J}(k)}, \mathbf{a}_{\mathcal{J}(k)}|\ldots\big) \\ &\quad\propto \frac{1}{\pi\sigma_a^2}\exp\left(-\frac{|a_{\hat{k}}-\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}+\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right) \\ &\qquad\times\left[\prod_{k'\in\mathcal{J}(k)\setminus\{\hat{k}\}}\delta(a_{k'})\right] p(\mathbf{b}). \end{aligned} \tag{27}$$

Integrating out $\mathbf{a}_{\mathcal{J}(k)}$ then yields

$$p\big(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) \propto \frac{1}{\pi\sigma_a^2}\pi\sigma_{\mathbf{b}}^2\exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right)p(\mathbf{b}). \quad (28)$$

For the case $\|\mathbf{b}_{\mathcal{J}(k)}\|=0$, we can rewrite (26) as

$$p\big(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\ldots\big) \propto \exp\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right)\left[\prod_{k'\in\mathcal{J}(k)}\delta(a_{k'})\right]p(\mathbf{b}).$$

Integrating out $\mathbf{a}_{\mathcal{J}(k)}$ gives

$$p\big(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big)\propto\exp\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}\right\|^2}{\sigma_n^2}\right)p(\mathbf{b}). \quad (29)$$

Now according to (22), $\|\mathbf{b}_{\mathcal{J}(k)}\|=0$ implies that $\mu_{\mathbf{b}}=0$ and $\sigma_{\mathbf{b}}^2=\sigma_a^2$. Formally inserting these values into (28) yields (29). Therefore, (28) is valid for both $\|\mathbf{b}_{\mathcal{J}(k)}\|=1$ and $\|\mathbf{b}_{\mathcal{J}(k)}\|=0$. Finally, dropping the constant factors in (28) yields (recall from (22) that $\mu_{\mathbf{b}}$ and $\sigma_{\mathbf{b}}^2$ depend on $\mathbf{b}_{\mathcal{J}(k)}$)

$$p\big(\mathbf{b}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big)\propto\sigma_{\mathbf{b}}^2\exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)p(\mathbf{b})$$

which is (21).

*Amplitudes:* For the derivation of the distribution in (23), we first note that

$$p\big(a_k|b_k,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) \propto p\big(b_k,a_k|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big).$$

The right-hand side can be calculated by marginalizing $p(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})$ with respect to all entries of $\mathbf{b}_{\mathcal{J}(k)}$ and $\mathbf{a}_{\mathcal{J}(k)}$ except $b_k$ and $a_k$: using $\widetilde{\mathcal{J}}(k)$ to denote $\mathcal{J}(k)\setminus\{k\}$, we have

$$p\big(a_k|b_k,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) \propto \sum_{\mathbf{b}_{\widetilde{\mathcal{J}}(k)}}\int p\big(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big)\,\mathrm{d}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}. \quad (30)$$

We will develop this expression separately for the two cases in (23), i.e., for $b_k=1$ and $b_k=0$. For $b_k=1$, the summation collapses:

$$\begin{aligned}p\big(a_k|b_k=1,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) &\propto \sum_{\mathbf{b}_{\widetilde{\mathcal{J}}(k)}}\int p\big(b_k=1,\mathbf{b}_{\widetilde{\mathcal{J}}(k)},\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\\ &\qquad \mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big)\mathrm{d}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}\\ &= \int p\big(b_k=1,\mathbf{b}_{\widetilde{\mathcal{J}}(k)}=\mathbf{0},\mathbf{a}_{\mathcal{J}(k)}|\mathbf{b}_{\sim\mathcal{J}(k)},\\ &\qquad \mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big)\mathrm{d}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}. \quad (31)\end{aligned}$$

This is due to the minimum distance constraint: if $b_k=1$, then the only hypothesis for $\mathbf{b}_{\widetilde{\mathcal{J}}(k)}$ with potentially nonzero probability is $\mathbf{b}_{\widetilde{\mathcal{J}}(k)}=\mathbf{0}$. Our case $b_k=1$, $\mathbf{b}_{\widetilde{\mathcal{J}}(k)}=\mathbf{0}$ is a special case of $\|\mathbf{b}_{\mathcal{J}(k)}\|=1$, namely with $\hat{k}=k$. We now recall that an expression of $p(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\ldots)$ for the case $\|\mathbf{b}_{\mathcal{J}(k)}\|=1$ was given in (27). For $\hat{k}=k$, (27) reads

$$p\big(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\ldots\big) \propto\exp\left(-\frac{|a_k-\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\left[\prod_{k'\in\widetilde{\mathcal{J}}(k)}\delta(a_{k'})\right]p(\mathbf{b})$$

where constant factors have been dropped. Inserting this into (31) yields

$$\begin{aligned}p\big(a_k|b_k=1,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) &\propto \int\exp\left(-\frac{|a_k-\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\\ &\quad\times\left[\prod_{k'\in\widetilde{\mathcal{J}}(k)}\delta(a_{k'})\right]p(\mathbf{b})\mathrm{d}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}\\ &=\exp\left(-\frac{|a_k-\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\exp\left(\frac{|\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)p(\mathbf{b})\\ &\propto\exp\left(-\frac{|a_k-\mu_{\mathbf{b}}|^2}{\sigma_{\mathbf{b}}^2}\right)\end{aligned}$$

or, equivalently, $p(a_k|b_k=1,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})=\mathcal{CN}(a_k;\mu_{\mathbf{b}},\sigma_{\mathbf{b}}^2)$. As explained below (24), due to $\hat{k}=k$, $\mu_{\mathbf{b}}$ and $\sigma_{\mathbf{b}}^2$ equal, respectively, $\tilde{\mu}_a$ and $\tilde{\sigma}_a^2$ as given in (24). Thus, we have verified (23) for $b_k=1$.

In the complementary case $b_k=0$, we can rewrite (26) as

$$\begin{aligned}p\big(\mathbf{b}_{\mathcal{J}(k)},\mathbf{a}_{\mathcal{J}(k)}|\ldots\big) &\propto\exp\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}-\mathbf{F}_{\widetilde{\mathcal{J}}(k)}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}\right\|^2}{\sigma_n^2}\right)\delta(a_k)\\ &\quad\times\left[\prod_{k'\in\widetilde{\mathcal{J}}(k)}p(a_{k'}|b_{k'})\right]p(\mathbf{b}).\end{aligned}$$

Inserting this into (30) yields

$$\begin{aligned}p\big(a_k|b_k=0,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x}\big) &\propto\delta(a_k)\sum_{\mathbf{b}_{\widetilde{\mathcal{J}}(k)}}\int\exp\left(-\frac{\left\|\tilde{\mathbf{x}}^{\mathcal{J}(k)}-\mathbf{F}_{\widetilde{\mathcal{J}}(k)}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}\right\|^2}{\sigma_n^2}\right)\\ &\quad\times\left[\prod_{k'\in\widetilde{\mathcal{J}}(k)}p(a_{k'}|b_{k'})\right]p(\mathbf{b})\mathrm{d}\mathbf{a}_{\widetilde{\mathcal{J}}(k)}\propto\delta(a_k).\end{aligned}$$

After proper normalization, we obtain $p(a_k|b_k=0,\mathbf{b}_{\sim\mathcal{J}(k)},\mathbf{a}_{\sim\mathcal{J}(k)},\boldsymbol{\gamma},\sigma_n^2,\mathbf{x})=\delta(a_k)$, which is (23) for $b_k=0$.

*Noise Variance:* To derive the sampling distribution for $\sigma_n^2$ in (25), we first use

$$p\big(\sigma_n^2|\mathbf{a},\boldsymbol{\gamma},\mathbf{x}\big)\propto p\big(\mathbf{a},\boldsymbol{\gamma},\sigma_n^2|\mathbf{x}\big)\propto p\big(\mathbf{x}|\mathbf{a},\boldsymbol{\gamma},\sigma_n^2\big)\,p\big(\sigma_n^2\big).$$

Inserting (13) and (12), we obtain further

$$\begin{aligned} p\big(\sigma_n^2|\mathbf{a},\boldsymbol{\gamma},\mathbf{x}\big) &\propto \frac{1}{(\sigma_n^2)^K}\exp\left(-\frac{\|\mathbf{x}-\mathbf{F}\mathbf{a}\|^2}{\sigma_n^2}\right)\frac{1}{(\sigma_n^2)^{\xi+1}} \\ &\quad\times\exp\left(-\frac{\eta}{\sigma_n^2}\right) I_{\mathbb{R}^+}\big(\sigma_n^2\big) \\ &= \frac{1}{(\sigma_n^2)^{\xi+K+1}}\exp\left(-\frac{\eta+\|\mathbf{x}-\mathbf{F}\mathbf{a}\|^2}{\sigma_n^2}\right) I_{\mathbb{R}^+}\big(\sigma_n^2\big)\,. \end{aligned}$$

Normalization then leads to $p(\sigma_n^2|\mathbf{a},\boldsymbol{\gamma},\mathbf{x}) = \mathcal{IG}(\sigma_n^2;\xi+K,\eta+\|\mathbf{x}-\mathbf{F}\mathbf{a}\|^2)$, which is (25).

## 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, Jul. 2003.

[10] 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.

[11] D. Ge, E. Le Carpentier, J. Idier, and D. Farina, "Spike sorting by stochastic simulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, pp. 249–259, Jun. 2011.

[12] G. Kail, C. Novak, B. Hofer, and F. Hlawatsch, "A blind Monte Carlo detection-estimation method for optical coherence tomography," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (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 Gauss model with free frequencies," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (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. Topics 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. Biol. Environ. Stat.*, 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, 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, Jun. 2009.

[23] 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.

[24] M. Lavielle, "Bayesian deconvolution of Bernoulli–Gaussian processes," *Signal Process.*, vol. 33, pp. 67–79, Jul. 1993.

[25] 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.

[26] A. Doucet and P. Duvaut, "Bayesian estimation of state-space models applied to deconvolution of Bernoulli–Gaussian processes," *Signal Process.*, vol. 57, pp. 147–161, March 1997.

[27] K. Hausmair, K. Witrisal, P. Meissner, C. Steiner, and G. Kail, "SAGE algorithm for UWB channel parameter estimation," presented at the COST 2100 Management Committee Meeting, Athens, Greece, Feb. 2010.

[28] 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 Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3484–3487.

[29] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer, 2004.

[30] *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London, U.K.: Chapman & Hall, 1996.

[31] S. Yildirim, A. T. Cemgil, and A. B. Ertüzün, "A hybrid method for deconvolution of Bernoulli–Gaussian processes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3417–3420.

[32] 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.

[33] D. A. van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *J. Amer. Statist. Assoc.*, vol. 103, pp. 790–796, Jun. 2008.

[34] T. Park and D. A. van Dyk, "Partially collapsed Gibbs samplers: Illustrations and applications," *J. Comput. Graph. Statist.*, vol. 18, pp. 283–305, Jun. 2009.

[35] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, pp. 2675–2685, May 2010.

[36] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "A partially collapsed Gibbs sampler for parameters with local constraints," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 3886–3889.

[37] 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.

[38] 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.

[39] L. Sörnmo, P. Börjesson, M. Nygards, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. 28, no. 10, pp. 713–717, 1981.

[40] 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.

[41] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces: Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation*. Boston, MA: Kluwer, 1998.

[42] C. P. Robert, *The Bayesian Choice*. New York: Springer, 1996.

[43] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate Poissonian time series. Applications to burst and transient source experiments," in *Proc. EUSIPCO*, Florence, Italy, Sep. 2006.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[45] O. Rabaste and T. Chonavel, "Estimation of multipath channels with long impulse response at low SNR via an MCMC method," *IEEE Trans. Signal Process.*, vol. 55, pp. 1312–1325, Apr. 2007.
[46] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov Chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London, U.K.: Chapman & Hall, 1996, pp. 1–19.
[47] G. Kail, F. Hlawatsch, and C. Novak, "Efficient Bayesian detection of multiple events with a minimum-distance constraint," in *Proc. IEEE SSP*, Cardiff, U.K., Aug.–Sep. 2009, pp. 73–76.
[48] C. Labat and J. Idier, "Sparse blind deconvolution accounting for time-shift ambiguity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 616–619.
[49] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Process.*, vol. 55, pp. 1251–1263, Apr. 2007.
[50] S. C. Sekhar, R. A. Leitgeb, M. L. Villiger, A. H. Bachmann, T. Blu, and M. Unser, "Non-iterative exact signal recovery in frequency domain optical coherence tomography," in *Proc. IEEE ISBI*, Metro Washington, DC, Apr. 2007, pp. 808–811.
[51] C. Lin, G. Kail, J.-Y. Tourneret, C. Mailhes, and F. Hlawatsch, "P and T wave delineation and waveform estimation in ECG signals using a block Gibbs sampler," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 537–540.

**Georg Kail** received the B.Sc. and Diplom-Ingenieur (M.Sc.) degrees in electrical engineering/telecommunications and the Dr. techn. (Ph.D.) degree in signal processing, all from Vienna University of Technology, Vienna, Austria, in 2005, 2008, and 2012, respectively.

In the course of his Master's and Ph.D. studies, he performed research at UTB Zlín, Czech Republic, ETH Zürich, Switzerland, and ENSEEIHT, Toulouse, France, as a short-term visiting researcher. Since 2008, he has been with the Institute of Telecommunications, Vienna University of Technology, where he is currently a Postdoctoral Research and Teaching Assistant. His research interests center around Markov chain Monte Carlo methods and Bayesian detection and estimation under structural constraints.

**Jean-Yves Tourneret** (SM'08) received the Ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT), France, in 1989 and the Ph.D. degree from the National Polytechnic Institute, Toulouse, France, in 1992.

He is currently a Professor at ENSEEIHT and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities center around statistical signal processing, with a particular interest in Bayesian and Markov Chain Monte Carlo methods.

Dr. Tourneret has been involved in the organization of several conferences, including the European Conference on Signal Processing (EUSIPCO) 2002 as the Program Chair, the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2006 (in charge of plenaries), and the Statistical Signal Processing Workshop (SSP) 2012 (for international liaisons). He has been a member of different technical committees, including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society (2001–2007 and 2010–present). He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011.

**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 the 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. 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, a review paper that appeared in the IEEE SIGNAL PROCESSING MAGAZINE, about 190 refereed scientific papers and book chapters, and three patents. He coedited three books. His research interests include signal processing for wireless communications and sensor networks, statistical signal processing, and compressive signal processing.

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 is coauthor of papers that won an IEEE Signal Processing Society Young Author Best Paper Award and a Best Student Paper Award at IEEE ICASSP 2011.

**Nicolas Dobigeon** (S'05–M'08) was born in Angoulême, France, in 1981. He received the Engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, in 2004, and the M.Sc. and Ph.D. degrees in signal processing from the National Polytechnic Institute of Toulouse, in 2004 and 2007, respectively.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been an Assistant Professor with the National Polytechnic Institute of Toulouse (ENSEEIHT, University of Toulouse), within the Signal and Communication Group of the IRIT Laboratory. His research interests are focused on statistical signal and image processing, with particular interest in Bayesian inference and Markov chain Monte Carlo methods.