

Variational Bayesian Learning for Decentralized Blind Deconvolution of Seismic Signals Over Sensor Networks

DMITRIY SHUTIN¹, (Senior Member, IEEE) and BAN-SOK SHIN¹ (MEMBER, IEEE)

¹Institute of Communications and Navigation, German Aerospace Center (e-mail: dmitriy.shutin@dlr.de, ban-sok.shin@dlr.de)

Corresponding author: Dmitry Shutin (e-mail: dmitriy.shutin@dlr.de).

ABSTRACT This work discusses a variational Bayesian learning approach towards decentralized blind deconvolution of seismic signals within a sensor network. Blind seismic deconvolution is cast into a probabilistic framework based on Sparse Bayesian learning developed for blind image deconvolution. The posterior distribution of the signals of interest is then approximated using a variational Bayesian method. Depending on a particular form of selected variational factors, the scheme is shown to generalize the state-of-the-art distributed seismic blind deconvolution algorithm. The algorithm operates by repeatedly alternating between two stages: (i) estimation of seismic source wavelet and (ii) reflectivity estimation. The wavelet is computed in a distributed fashion using Alternating Direction Method of Multipliers. Based on this estimate each sensor then locally obtains a sparse reflectivity estimate. Numerical evaluation with synthetic seismic data shows that the proposed method outperforms existing deconvolution algorithms in the high signal-to-noise ratio (SNR) regime. In the low SNR regime a higher sensitivity of sparse Bayesian learning regarding model mismatches in the estimated source wavelet is observed. Also, processing of seismic data generated with an acoustic wave equation shows that the proposed method is able to recover the original reflectivities more accurately, with more distinct support of the reflectivities, as compared to existing methods despite present model mismatches. The algorithm is also applied to the estimation of real seismic data, and shows improved performance as compared to a state-of-the-art estimation method.

INDEX TERMS Blind deconvolution, sparse Bayesian learning, decentralized estimation, seismic signal processing, seismic networks

I. INTRODUCTION

A. MOTIVATION

Seismic surveys such as reflection or refraction methods use multiple seismic sensors or geophones to record the seismic waves reflected at underground layers. Based on received waves the subsurface of a given area can be reconstructed using signal processing and imaging techniques [1]. However, most of these algorithms assume a centralized operation, when all measurement data are available at one central entity. In this work we instead focus on a *distributed* approach towards subsurface reconstruction realized over a network of “intelligent” geophones that are equipped with computing and communication capabilities. The latter allows for a cooperative processing of measured data. Such an architecture is particularly relevant for autonomous exploration of a planet’s subsurface such as that on Mars or Moon [2], where an autonomous operation is required and a central processing

entity might not be available.

A common survey methodology in seismic exploration is a reflectivity survey. In such a survey a seismic source such as a sledge hammer or an air gun is activated. The generated waves are reflected at underground layers and then recorded at multiple geophones that are placed over an area of interest. This measurement process can be modeled by a single-input-multiple-output system where a source signal is convolved with impulse responses of multiple reflectivity paths. Each such reflectivity path describes a structure profile of a path that a seismic wave travels from the source to the receiver. The goal is to recover the hidden impulse response of the reflectivity profile by deconvolving the source signal and the reflectivity profile from geophone measurements. However, the seismic source (also called wavelet) is usually unknown, which poses challenges for the deconvolution. Hence, blind deconvolution algorithms are needed that recover both the

source signal and the reflectivity profile.

In the past, a significant research work has been dedicated to blind deconvolution of signals for images processing (see e.g., [3]) as well as specifically for seismic signals (see e.g., [4]–[6] and references therein). The methods in general follow an iterative approach, where at first a model of an excitation signal is estimated, followed by the inference of the signal – the seismic reflectivity profile. In particular, sparse multichannel blind deconvolution (SMBD) methods have shown promising performance results for the reflectivity estimation. The recent work of [5] uses an iterative algorithm that alternates between wavelet and reflectivity estimation. The wavelet estimation is conducted in the frequency domain whereas the reflectivity estimation is done in the time domain to exploit sparsity in the reflectivity’s impulse response. The sparsity is achieved through an ℓ_1 penalty, which leads to a LASSO-type optimization [7]. Although optimization of LASSO-type functionals does lead to a sparse estimate, such methods were found to be inferior to sparse Bayesian learning (SBL) [8], which results in a sparser support [9], and provides uncertainty estimates of the parameters.

Bayesian learning has been applied to blind deconvolution problems in the past. However, most works do not consider seismic signals and focus on the problem of image deconvolution as in [10], [11]; the transfer of the results for image processing into the seismic domain is not straightforward and can pose significant challenges, despite formal similarities between the models. In [12], sparse multichannel deconvolution of seismic traces is conducted using block sparse Bayesian learning techniques and the expectation-maximization algorithm. However, this work assumes knowledge of the source wavelet and therefore is not a blind approach. Also typical blind deconvolution methods adopt a centralized operation. This contribution instead, builds upon and extends these ideas for a distributed operation in a network of geophones.

B. MAIN CONTRIBUTIONS

The main contribution of this work consists of deriving a decentralized blind deconvolution algorithm within a variational Bayesian framework and analyzing its performance for the deconvolution of seismic signals. In particular, the estimation of the source signal – the wavelet – is conducted in a distributed fashion such that this algorithm is applicable to sensor networks where no central entity is handling the estimation procedure.

The proposed methodology follows a blind deconvolution approach of [5] and [10] and extends our previous work [13], where the wavelet estimation is implemented in a distributed fashion followed by a local reflectivity estimation. In [13] the wavelet estimation is implemented using the alternating directions method of multipliers (ADMM) algorithm [14]. Based on the estimated wavelet each sensor then solves an ℓ_1 -regularized problem locally to obtain a sparse estimate of the impulse response of its reflectivity. In this work, we cast the proposed blind iterative estimation scheme into a

variational Bayesian framework, which on the one hand, provides a formal basis for the algorithm in [13], while on the other hand, allows for application of empirical Bayesian techniques to fix (or rather learn) the regularization parameters both for the wavelet as well as for the reflectivity estimation. Based on the variational Bayesian formulation, we derive a distributed wavelet estimation that enables each sensor in the network to obtain a wavelet estimate by exchanging data with neighboring sensors. For the reflectivity estimation, we employ SBL locally at each sensor. SBL techniques also facilitate analytical inference of model parameters and account for posterior variance of the reflectivity estimate, which is difficult for the ℓ_1 -type constraints due to non-smoothness of the posterior function. Note that the use of non-variational approaches, such as EM-based estimation of SBL parameters with approximate message passing (AMP), as in e.g., [15], also provide analytical inference expressions for some of the parameters.¹ Yet due to the dimensionality of the electrical impedance tomography problem addressed in [15], not all inference expressions (and in particular the E-step of the EM algorithm) are computationally feasible, which requires the use of numerical approximations. The numerical complexity of the considered problem remains, however, tractable; moreover, the computation of the E-step (or of the corresponding variational approximation factor) can be avoided as in e.g., fast marginal likelihood SBL realisations [17], [18]. These aspects of the algorithm implementation are, however, left outside the scope of the paper.

C. NOTATION

Column vectors are represented as boldface lowercase letters, e.g., \mathbf{x} , and matrices as boldface uppercase letters, e.g., \mathbf{X} . Their transpose is denoted by $(\cdot)^\top$. For some positive-semidefinite matrix \mathbf{A} , $\|\mathbf{x}\|_{\mathbf{A}} \triangleq \sqrt{\mathbf{x}^\top \mathbf{A} \mathbf{x}}$ is the weighted ℓ_2 norm of vector \mathbf{x} ; similarly, $\|\mathbf{x}\| \triangleq \sqrt{\mathbf{x}^\top \mathbf{x}}$. Notation $\|\mathbf{x}\|_1$ is used to denote an ℓ_1 norm of the vector \mathbf{x} . The expectation operator is denoted by $\mathbb{E}\{\cdot\}$, or $\mathbb{E}_q\{\cdot\}$ when the context requires to explicitly state the probability density function (pdf) q over which expectation is taken. We denote the pdf of a Gaussian random vector \mathbf{x} with expectation $\mathbf{a} \triangleq \mathbb{E}\{\mathbf{x}\}$ and covariance matrix $\mathbf{B} \triangleq \mathbb{E}\{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^\top\}$ as $\mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{B})$. Also, we denote the pdf of a Gamma random variable x as $\text{Ga}(x; a, b) = b^a x^{a-1} e^{-bx} / \Gamma(a)$, where a is a shape parameter, b is a rate parameter, and $\Gamma(\cdot)$ is the Gamma function. The notation $\delta_{\mathbf{a}}(\mathbf{x})$ is used to denote a Dirac distribution over domain of \mathbf{x} with a support \mathbf{a} . We will also use notation $\log x \propto^e a$ to imply that $x \propto e^a$.

II. SYSTEM MODEL

Consider a network of J interconnected sensors/geophones each measuring a seismic trace at its corresponding position. We describe the network topology by a graph $\mathcal{G} = \{\mathcal{J}, \mathcal{E}\}$ with a set of nodes $\mathcal{J} = \{1, 2, \dots, J\}$ and a set of edges $\mathcal{E} =$

¹Variational Bayesian methods generalize the EM-algorithm for a specific choice of approximating factors [16]

$\{(j, i) | j, i \in \mathcal{J}, j \neq i\}$. With \mathcal{N}_j we denote a neighborhood of the sensor j – a set of all sensors directly connected to sensor j , including sensor j itself. In addition, we assume that the network graph is undirected and strongly connected [19].

It is common to model the acquisition of seismic traces via a convolution of a seismic source wavelet \mathbf{w} and a set of impulse responses $\mathbf{h}_j, j \in \mathcal{J}$, that model the subsurface reflections between the source and the receiver j [1]. Note that the wavelet \mathbf{w} is a common source signal for all sensors, while responses $\mathbf{h}_j, j \in \mathcal{J}$, differ between sensors. The impulse responses $\mathbf{h}_j, j \in \mathcal{J}$, are typically termed *reflectivities*. They naturally characterize the reflections occurring as a seismic wave passes between the source and the receiver via subsurface layers. It is thus the key to a quantification of the subsurface under investigation.

Unfortunately, due to the non-ideal form of the seismic source, the reflectivity profile of the subsurface is blurred, which makes a reliable subsurface reconstruction difficult. Therefore, the goal of seismic deconvolution is to remove the influence of the seismic source in the seismic traces in order to infer “a clean” reflectivity. However, neither source wavelet \mathbf{w} nor the reflectivities $\mathbf{h}_j, j \in \mathcal{J}$ are typically known. Thus, blind approaches are required. In [5] it was suggested to implement a blind deconvolution approach as a sequence of two consecutive steps: a wavelet estimation and a reflectivity estimation. In [13] we cast this approach in a distributed setting, where the wavelet estimation was solved using ADMM algorithm [14], followed by a “local” estimation of the reflectivities $\mathbf{h}_j, j \in \mathcal{J}$. Both are then iteratively repeated until a certain convergence criterion is met.

In the following we will show that the algorithm in [13] can be seen as a solution to a variational Bayesian optimization of a posterior distribution of wavelet and reflectivity signals. Moreover, the variational setting allows one to extend the method in [13] and estimate the regularization parameters within the empirical Bayes approach. This is detailed in the following.

A. BAYESIAN FORMULATION OF THE SPARSE BLIND DECONVOLUTION

Each sensor j acquires a noisy seismic trace $d_j[l]$, which can be described by a noisy convolution between a length L_W discrete time source wavelet $w[l]$ and a length L_R reflectivity $h_j[l], j \in \mathcal{J}$:

$$d_j[l] = w[l] * h_j[l] + n_j[l] = \sum_{\nu=0}^{L_W} h_j[l - \nu]w[\nu] + n_j[l], \quad (1)$$

where $l = 0, \dots, L_W + L_R - 2$. By defining

$$\begin{aligned} \mathbf{d}_j &\triangleq [d_j[0], \dots, d_j[L_W + L_R - 2]]^\top, \\ \mathbf{w} &\triangleq [w[0], w[1], \dots, w[L_W - 1]]^\top, \text{ and} \\ \mathbf{h}_j &\triangleq [h_j[0], h_j[1], \dots, h_j[L_R - 1]]^\top \end{aligned}$$

model (1) can be reformulated in equivalent matrix-vector forms as

$$\mathbf{d}_j = \mathbf{H}_j \mathbf{w} + \mathbf{n}_j = \mathbf{W} \mathbf{h}_j + \mathbf{n}_j = \mathbf{h}_j * \mathbf{w} + \mathbf{n}_j, \quad j \in \mathcal{J}, \quad (2)$$

where $\mathbf{H}_j \in \mathbb{R}^{(L_W+L_R-1) \times L_W}$ is the convolution matrix of the reflectivity \mathbf{h}_j at sensor j , $\mathbf{W} \in \mathbb{R}^{(L_W+L_R-1) \times L_R}$ is the convolution matrix consisting of wavelet \mathbf{w} , and $\mathbf{n}_j \in \mathbb{R}^{L_W+L_R-1}$ are additive noise samples. Note that the products $\mathbf{W} \mathbf{h}_j$ and $\mathbf{H}_j \mathbf{w}$ are absolutely equivalent. The third form $\mathbf{h}_j * \mathbf{w}$ with a slight abuse of a vector notation also indicates a convolution operation. Depending on the context, these forms will be used interchangeably to simplify notation.

We will assume that noise samples $\mathbf{n}_j, j \in \mathcal{J}$, observed by seismic sensors are statistically independent across sensors. Furthermore, we assume that each $\mathbf{n}_j, j \in \mathcal{J}$, follows a normal distribution with zero mean and scaled covariance matrix $\tau_j^{-1} \mathbf{I}$, where $\tau_j, j \in \mathcal{J}$, is a noise precision parameter.

To simplify further notation, let us now define $\mathcal{D} \triangleq \{\mathbf{d}_1, \dots, \mathbf{d}_J\}$, $\mathcal{T} \triangleq \{\tau_1, \dots, \tau_J\}$, and $\mathcal{H} \triangleq \{\mathbf{h}_1, \dots, \mathbf{h}_J\}$ as a collection of variables associated with J sensors. Now, we can begin with formulating the probabilistic structure of the estimation model. Let us point out the following modeling is similar to that proposed in [10], [11] for blind image deconvolution. In the following it is adapted to blind seismic deconvolution in a distributed setting.

1) Probabilistic formulation of the model

Due to the assumption made for the additive noise terms $\mathbf{n}_j, j \in \mathcal{J}$, we can write the likelihood function of seismic traces \mathcal{H} , wavelet \mathbf{w} , and noise precisions \mathcal{T} as follows

$$p(\mathcal{D} | \mathcal{H}, \mathbf{w}, \mathcal{T}) = \prod_{j \in \mathcal{J}} p(\mathbf{d}_j | \mathbf{w}, \mathbf{h}_j, \tau_j) \quad (3)$$

where $p(\mathbf{d}_j | \mathbf{w}, \mathbf{h}_j, \tau_j) = \mathcal{N}(\mathbf{d}_j | \mathbf{H}_j \mathbf{w}, \tau_j^{-1} \mathbf{I})$.

Let us now assume that the wavelet \mathbf{w} likewise follows a Gaussian distribution, with zero mean and circular covariance matrix controlled by an unknown precision λ_w . Specifically, we assume that

$$p(\mathbf{w} | \lambda_w) = \mathcal{N}(\mathbf{w} | \mathbf{0}, \lambda_w^{-1} \mathbf{I}). \quad (4)$$

In general, we can also treat the precision parameter λ_w as an unknown random variable. As such, we also select a prior distribution $p(\lambda_w) = \text{Ga}(\lambda_w; c, d)$ for it. The choice of a Gamma prior is motivated by the conjugacy [20] of the latter to $p(\mathbf{w} | \lambda_w)$, which will eventually simplify the inference of λ_w , as we will see later.

Now, we define a prior distribution for reflectivities $\mathbf{h}_j, j \in \mathcal{J}$. As seismic reflectivities are known to be sparse [21], [22] – the nonzero entries reflect the propagation velocity jumps in the subsurface structure – the statistical model for $\mathbf{h}_j, j \in \mathcal{J}$, should be selected to reflect this. To this end we model the prior $p(\mathbf{h}_j), j \in \mathcal{J}$, with a so-called scale mixture distribution [9]:

$$p(\mathbf{h}_j) = \int p(\mathbf{h}_j | \alpha_j) p(\alpha_j) d\alpha_j, \quad j \in \mathcal{J}, \quad (5)$$

where $p(\alpha_j)$ is called a mixing density and $p(\mathbf{h}_j|\alpha_j)$ belongs to a power-exponential family of distributions [23]. Note that parameters α_j , $j \in \mathcal{J}$, are also treated as unknown random variables.

Finally, we endow the probabilistic model with a prior distribution $p(\mathcal{T}) = \prod_{j \in \mathcal{J}} p(\tau_j)$ for noise precision parameters \mathcal{T} . Similarly to parameter λ_w , we select $p(\tau_j) = \text{Ga}(\tau_j; s_j, t_j)$, $j \in \mathcal{J}$.

Now, the probabilistic model underlying the estimation problem can be summarized with the following joint pdf

$$p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}) = p(\mathbf{w}|\lambda_w)p(\lambda_w) \times \prod_{j \in \mathcal{J}} p(\mathbf{d}_j|\mathbf{w}, \mathbf{h}_j, \tau_j)p(\mathbf{h}_j|\alpha_j)p(\alpha_j)p(\tau_j). \quad (6)$$

where we defined $\mathcal{A} \triangleq \{\alpha_1, \dots, \alpha_J\}$ to simplify notation. The blind deconvolution under sparsity constraints for reflectivities can then be posed as a maximization of the following posterior:

$$p(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}|\mathcal{D}) = \frac{p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})}{p(\mathcal{D})}. \quad (7)$$

In general the inference of the posterior in (7) cannot be done in a closed form and numerical approaches are needed. Here we make use of variational Bayesian inference to approximate the posterior of interest with a proxy distribution $q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})$ which makes analysis tractable.

2) Variational Bayesian inference

Variational Bayesian inference (VBI) techniques are a class of global approximative inference approaches. They attempt to approximate an intractable posterior $p(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}|\mathcal{D})$ with a simpler pdf $q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})$. The latter is in general a “free” pdf function that is chosen to minimize the Kullback-Leibler divergence [16] $\text{KL}(q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})||p(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}|\mathcal{D}))$. Although the latter cannot be computed since the posterior of interest $p(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}|\mathcal{D})$ is unknown, the divergence can be minimized indirectly, by maximizing the lower bound on the log-evidence [16] given as:

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})} \left\{ \log \frac{p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})}{q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})} \right\}. \quad (8)$$

The approximating density $q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})$ can be further constrained to reduce the complexity. Specifically, we will adopt a mean field approximation [16], which implies that

$$q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}) = q(\mathbf{w})q(\lambda_w) \prod_{j=1}^J q(\mathbf{h}_j)q(\tau_j)q(\alpha_j). \quad (9)$$

In other words, we assume independence of all the factors in the approximating distribution. Furthermore, we will constrain individual factors in (9) to parametric pdfs. This will allow us to cast variational optimization into a parametric one. Specific choices of individual factors will be discussed in the following.

B. DISTRIBUTED SPARSE BLIND DECONVOLUTION AS A VARIATIONAL INFERENCE PROBLEM

The maximization of the lower bound in (8) requires specifying (i) the form of the factors in (9) and (ii) the sparsifying prior $p(\mathbf{h}_j)$ in (5). Depending on a particular choice of the factors in (9) different inference schemes can be implemented. In particular, the distributed sparse blind deconvolution (D-SBD) algorithm proposed in [13] can be shown to be an instance of the proposed variational optimization of (8), as we will describe in the following.

Let us assume that $p(\mathbf{h}_j)$ in (5) is selected such that

$$p(\mathbf{h}_j|\alpha_j) = \prod_{i=1}^{L_R} N(h_{i,j}|0, \alpha_{i,j}), \quad p(\alpha_{i,j}) = \frac{\mu^2}{2} e^{-\frac{\mu^2}{2} \alpha_{i,j}}, \quad (10)$$

for some fixed parameter μ and $i \in \{1, \dots, L_R\}$, $j \in \mathcal{J}$. Then a direct integration of (5) shows that $p(\mathbf{h}_j)$ is given by

$$p(\mathbf{h}_j) = \frac{\mu^{L_R}}{2} e^{-\frac{\mu}{2} \|\mathbf{h}_j\|_1}, \quad j \in \mathcal{J},$$

which is a multivariate Laplace distribution for (independent) entries in \mathbf{h}_j . By marginalizing the posterior (7) over \mathcal{A} using this result, we obtain a variational lower bound in the form

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T})} \left\{ \log \frac{p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T})}{q(\mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T})} \right\}. \quad (11)$$

Note that a “direct” approach to a specification of the prior is referred to as Type-I sparse Bayesian signal reconstruction [9].

Now, an instance of the D-SBD algorithm from [13] is obtained by selecting all variational factors as Dirac measures defined over the corresponding domains.

$$q(\mathbf{h}_j) = \delta_{\hat{\mathbf{h}}_j}(\mathbf{h}_j), \quad q(\tau_j) = \delta_{\hat{\tau}_j}(\tau_j), \quad j \in \mathcal{J}, \quad \text{and} \quad (12a)$$

$$q(\lambda_w) = \delta_{\hat{\lambda}_w}(\lambda_w), \quad q(\mathbf{w}) = \delta_{\hat{\mathbf{w}}}(\mathbf{w}). \quad (12b)$$

In [13], however, the factors $q(\lambda_w)$ and $q(\tau_j)$, $j \in \mathcal{J}$, are assumed to be fixed, with the supports of these factors being free parameters of the D-SBD algorithm. With this choice of variational factors, it then becomes easy to evaluate the bound (11) and maximize it with respect to $q(\mathbf{w})$ and $q(\mathbf{h}_j)$, $j \in \mathcal{J}$.

Specifically, it can be shown that with respect to $q(\mathbf{w})$ the bound (11) can be expressed as

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\mathbf{w})} \left\{ \log \frac{\tilde{p}(\mathbf{w})}{q(\mathbf{w})} \right\} = -\text{KL}(q(\mathbf{w})||\tilde{p}(\mathbf{w})) \quad (13)$$

where

$$\log \tilde{p}(\mathbf{w}) \propto^e \mathbb{E}_{q(\mathcal{H})q(\mathcal{T})} \{ \log p(\mathcal{D}|\mathbf{w}, \mathcal{H}, \mathcal{T}) \} + \mathbb{E}_{q(\lambda_w)} \{ \log p(\mathbf{w}|\lambda_w) \}. \quad (14)$$

As we see, the variational bound is maximized, when the divergence $\text{KL}(q(\mathbf{w})||\tilde{p}(\mathbf{w}))$ is minimized. Since $q(\mathbf{w})$ is constrained to a space of Dirac measures, the minimum of $\text{KL}(q(\mathbf{w})||\tilde{p}(\mathbf{w}))$ is obtained when support $\hat{\mathbf{w}}$ of $q(\mathbf{w})$ coincides with the location of the maximum of $\tilde{p}(\mathbf{w})$, or

equivalently, with the maximum of $\log \tilde{p}(\mathbf{w})$. Evaluating $\log \tilde{p}(\mathbf{w})$, we find $\hat{\mathbf{w}}$ as the solution to the following problem:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{j=1}^J \frac{\hat{\tau}_j}{2} \|\mathbf{d}_j - \widehat{\mathbf{H}}_j \mathbf{w}\|^2 + \frac{\hat{\lambda}_w}{2} \|\mathbf{w}\|_2^2, \quad (15)$$

where $\widehat{\mathbf{H}}_j$ is a reflectivity convolution matrix composed of elements of $\hat{\mathbf{h}}_j$ – support of $q(\mathbf{h}_j)$.

Similarly, the bound in (11) with respect to $q(\mathbf{h}_j)$ can be expressed as

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\mathbf{h}_j)} \left\{ \log \frac{\tilde{p}(\mathbf{h}_j)}{q(\mathbf{h}_j)} \right\} = -\operatorname{KL}(q(\mathbf{h}_j) \parallel \tilde{p}(\mathbf{h}_j)),$$

where

$$\log \tilde{p}(\mathbf{h}_j) \propto^e \mathbb{E}_{q(\mathbf{w})q(\mathcal{T})q(\lambda_w)} \{ \log p(\mathcal{D} \mid \mathbf{w}, \mathcal{H}, \mathcal{T}) p(\mathbf{h}_j) \}.$$

Computing the latter, we find that the support of $q(\mathbf{h}_j)$, $j \in \mathcal{J}$, that maximizes the bound is obtained from

$$\hat{\mathbf{h}}_j = \underset{\mathbf{h}_j}{\operatorname{argmin}} \frac{\hat{\tau}_j}{2} \|\mathbf{d}_j - \widehat{\mathbf{W}} \mathbf{h}_j\|^2 + \frac{\mu}{2} \|\mathbf{h}_j\|_1, \quad j \in \mathcal{J}. \quad (16)$$

The estimation of factors $q(\mathbf{h}_j)$, $j \in \mathcal{J}$ and $q(\mathbf{w})$ are then done iteratively until convergence. This will lead to the maximization of the marginalized lower bound in (11). Expressions (15) and (16) coincide with those obtained in [13]. A distributed solution to optimization (15) can be found using the ADMM algorithm [14], while (16) is a least absolute shrinkage and selection operator (LASSO) problem that can be solved individually by each agent, as was proposed in [13].

Constraining the variational factors in (12) to Dirac distributions does simplify numerical evaluation of the variational bound, yet it unnecessarily simplifies the variational approximation. In the following, we propose a more complete treatment of the inference problem with non-degenerate variational factors. Moreover, we include the estimation of $q(\alpha_j)$, $q(\tau_j)$, $j \in \mathcal{J}$, and $q(\lambda_w)$ into the algorithms.

III. BLIND SEISMIC DECONVOLUTION USING SPARSE BAYESIAN LEARNING

We will begin by specifying the model for the sparsity inducing prior $p(\mathbf{h}_j)$, $j \in \mathcal{J}$. Here we will follow a Sparse Bayesian Learning (SBL) approach, where we select

$$p(\mathbf{h}_j \mid \alpha_j) = \prod_{i=1}^{L_R} \mathcal{N}(h_{i,j} \mid 0, \alpha_{i,j}^{-1}) \quad (17)$$

with $\alpha_j = [\alpha_{1,j}, \dots, \alpha_{L_R,j}]^\top$. Note that in contrast to (10), here α_j are precisions rather than variance parameters; within SBL they are referred to as sparsity parameters. As a mixing density $p(\alpha_j)$, $j \in \mathcal{J}$, SBL selects a Gamma prior [8], [9]:

$$p(\alpha_j) = \prod_{i=1}^{L_R} p(\alpha_{i,j}) = \prod_{i=1}^{L_R} \operatorname{Ga}(\alpha_{i,j}; a_j, b_j), \quad j \in \mathcal{J}. \quad (18)$$

Note that for each agent $j \in \mathcal{J}$ there is a single scale parameter a_j and a rate parameter b_j that define the prior $p(\alpha_j)$. Such formulation leads to $p(\mathbf{h}_j)$, $j \in \mathcal{J}$, being Student's t-distributed [8], [9].

Although in general we can again follow a Type-I approach and marginalize the posterior over \mathcal{A} , as we did before, this will lead to an intractable variational inference. Instead, a Type-II approach, where parameters \mathcal{A} are treated as unknown random variables and estimated alongside the other parameters, is computationally simpler, and thus marginalization is done implicitly. The same approach was also proposed for images in [10].

We will now constrain individual factors in (9) to parametric pdfs to cast variational optimization into a parametric one. To this end we select

$$q(\mathbf{h}_j) = \mathcal{N}(\mathbf{h}_j \mid \hat{\mathbf{h}}_j, \hat{\mathbf{S}}_{\mathbf{h}_j}), \quad j \in \mathcal{J}, \quad (19a)$$

$$q(\tau_j) = \operatorname{Ga}(\tau_j \mid \hat{s}_j, \hat{t}_j), \quad j \in \mathcal{J}, \quad (19b)$$

$$q(\alpha_j) = \prod_{i=1}^{L_R} \operatorname{Ga}(\alpha_{i,j} \mid \hat{a}_{i,j}, \hat{b}_{i,j}), \quad j \in \mathcal{J}, \quad (19c)$$

$$q(\lambda_w) = \operatorname{Ga}(\lambda_w \mid \hat{c}, \hat{d}), \quad (19d)$$

$$q(\mathbf{w}) = \delta_{\hat{\mathbf{w}}}(\mathbf{w}), \quad (19e)$$

where $\delta_{\hat{\mathbf{w}}}(\mathbf{w})$ is a Dirac measure with a support $\hat{\mathbf{w}}$ on the \mathbb{R}^{L_w} -dimensional domain of wavelets. Note that in contrast to all other factors, $q(\mathbf{w})$ is selected as a degenerated density, despite the fact that it is not difficult to show that the posterior for the wavelet is Gaussian and thus an optimal variational factor $q(\mathbf{w})$ should be Gaussian as well. Yet since computation of this factor requires decentralized processing, it would imply communication of both mean and covariance matrix of $q(\mathbf{w})$, which in general grows as $\mathcal{O}(L_w^2)$. By constraining this factor to a set of Dirac measures, we thus ignore the impact of wavelet estimation uncertainty on the inference of other parameters and resort to a *maximum a posteriori* (MAP) estimation of its value. This reduces the communication load on the network as only the support $\hat{\mathbf{w}}$ needs to be communicated. This leads to an $\mathcal{O}(L_w)$ communication load. Additionally, this simplifies the evaluation of the bound in (8).

Now, the inference of factors (9) is done in sequential fashion, one factor at a time, with each factor selected so as to maximize the bound in (8) until convergence.

A. ESTIMATION OF $q(\mathbf{w})$

We begin with estimating the wavelet factor $q(\mathbf{w})$. Evaluating (14) with variational factors (19a), (19b) and (19d) (see also Appendix A-A) we find that the divergence $\operatorname{KL}(q(\mathbf{w}) \parallel \tilde{p}(\mathbf{w}))$ in (13) is minimized when support $\hat{\mathbf{w}}$ of $q(\mathbf{w})$ is found as a solution to the optimization problem

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{j=1}^J \frac{\hat{\tau}_j}{2} \left(\|\mathbf{d}_j - \widehat{\mathbf{H}}_j \mathbf{w}\|^2 + \|\mathbf{w}\|_{\hat{\Omega}_j}^2 \right) + \frac{\hat{\lambda}_w}{2} \|\mathbf{w}\|^2 \quad (20)$$

Observe now that the matrix $\widehat{\Omega}_j$ (see (48) in Appendix A-A) appears due to uncertainty in the reflectivity estimate $q(\mathbf{h}_j)$, specifically, due to its covariance \widehat{S}_{h_j} . This uncertainty has an additional “smoothing” effect on the optimization (20) in terms of a weighted ℓ_2 norm of the wavelet estimate. Due to the quadratic form of (20), its solution is then readily obtained as

$$\widehat{\mathbf{w}} = \left(\sum_{j=1}^J \widehat{\tau}_j \widehat{\mathbf{H}}_j^T \widehat{\mathbf{H}}_j + \sum_{j=1}^J \widehat{\tau}_j \widehat{\Omega}_j + \widehat{\lambda}_w \mathbf{I} \right)^{-1} \sum_{j=1}^J \widehat{\tau}_j \widehat{\mathbf{H}}_j^T \mathbf{d}_j, \quad (21)$$

from which the regularization effect of the reflectivity covariance \widehat{S}_{h_j} becomes apparent.

1) Distributed estimation of $q(\mathbf{w})$

As we see, the solution (21) requires cooperation between J sensors. To find this solution in a distributed setting we propose to use the ADMM algorithm [14], similar to that used in [24].

Let us introduce a latent variable $\mathbf{z}_j \in \mathbb{R}^{L_w}$ at each sensor j and reformulate optimization (20) as

$$\min_{\{\mathbf{w}_j, \mathbf{z}_j\}_{j=1}^J} \frac{1}{2} \sum_{j=1}^J \widehat{\tau}_j \|\mathbf{d}_j - \widehat{\mathbf{H}}_j \mathbf{w}_j\|^2 + \mathbf{z}_j^T (\widehat{\tau}_j \widehat{\Omega}_j + \widehat{\lambda}_w \mathbf{I}) \mathbf{z}_j$$

s.t. $\mathbf{w}_j = \mathbf{z}_i, \quad \forall i \in \mathcal{N}_j.$ (22)

Note that in (22) we added the constraint $\mathbf{w}_j = \mathbf{z}_i, \quad \forall i \in \mathcal{N}_j$ that enforces a consensus solution for $\{\mathbf{w}_j\}_{j \in \mathcal{J}}$ over the whole network. Hence, all estimates $\{\mathbf{w}_j\}_{j \in \mathcal{J}}$ will converge to the same solution, i.e., the solution of (20). Applying the ADMM algorithm to problem (22) we obtain the following iterative update equations:

$$\mathbf{w}_j^{[k+1]} = \underset{\mathbf{w}_j}{\operatorname{argmin}} \left\{ \frac{\widehat{\tau}_j}{2} \left\| \mathbf{d}_j - \widehat{\mathbf{H}}_j \mathbf{w}_j \right\|^2 + \frac{\rho}{2} \sum_{i \in \mathcal{N}_j} \left\| \mathbf{w}_j - \mathbf{z}_i^{[k]} + \mathbf{u}_{ij}^{[k]} \right\|^2 \right\} \quad (23)$$

$$\mathbf{z}_j^{[k+1]} = \underset{\mathbf{z}_j}{\operatorname{argmin}} \left\{ \mathbf{z}_j^T (\widehat{\tau}_j \widehat{\Omega}_j + \widehat{\lambda}_w \mathbf{I}) \mathbf{z}_j + \frac{\rho}{2} \sum_{i \in \mathcal{N}_j} \left\| \mathbf{w}_j^{[k+1]} - \mathbf{z}_i + \mathbf{u}_{ij}^{[k]} \right\|^2 \right\}$$

$$= \underset{\mathbf{z}_j}{\operatorname{argmin}} \left\{ \mathbf{z}_j^T (\widehat{\tau}_j \widehat{\Omega}_j + \widehat{\lambda}_w \mathbf{I}) \mathbf{z}_j + \frac{\rho}{2} \sum_{i \in \mathcal{N}_j} \left\| \mathbf{w}_i^{[k+1]} - \mathbf{z}_j + \mathbf{u}_{ij}^{[k]} \right\|^2 \right\} \quad (24)$$

$$\mathbf{u}_{ij}^{[k+1]} = \mathbf{u}_{ij}^{[k]} + \mathbf{w}_j^{[k+1]} - \mathbf{z}_i^{[k+1]}. \quad (25)$$

In the above the variable $\mathbf{u}_{ij}^{[k]}$ is a Lagrange multiplier originating from the employed consensus constraint and ρ is the

ADMM penalty parameter. Note that for the latent variable $\mathbf{z}_j, j \in \mathcal{J}$, in (24) we swapped the order of indices i and j in the last term. This is valid since in the augmented Lagrangian cost function we have a double summation over all receivers $j \in \mathcal{J}$. Therefore, swapping these indices only changes the order of the summation, not the result. We do this in order to allow for a minimization with respect to \mathbf{z}_j in closed form. Solving (23)-(24) we finally obtain

$$\mathbf{w}_j^{[k+1]} = \left(\widehat{\tau}_j \widehat{\mathbf{H}}_j^T \widehat{\mathbf{H}}_j + \rho |\mathcal{N}_j| \mathbf{I} \right)^{-1} \times \left(\widehat{\tau}_j \widehat{\mathbf{H}}_j^T \mathbf{d}_j + \rho \sum_{i \in \mathcal{N}_j} \mathbf{z}_i^{[k]} - \mathbf{u}_{ij}^{[k]} \right) \quad (26a)$$

$$\mathbf{z}_j^{[k+1]} = \left(\widehat{\tau}_j \widehat{\Omega}_j + (\widehat{\lambda}_w + \rho |\mathcal{N}_j|) \mathbf{I} \right)^{-1} \times \left(\rho \sum_{i \in \mathcal{N}_j} \left(\mathbf{w}_i^{[k+1]} + \mathbf{u}_{ij}^{[k]} \right) \right), \quad (26b)$$

$$\mathbf{u}_{ij}^{[k+1]} = \mathbf{u}_{ij}^{[k]} + \mathbf{w}_j^{[k+1]} - \mathbf{z}_i^{[k+1]}. \quad (26c)$$

Above equations enable a distributed estimation of the wavelet \mathbf{w} at each sensor j in the network. To this end, each sensor needs to exchange variables $\mathbf{w}_j^{[k]}, \mathbf{z}_j^{[k]}$ and $\mathbf{u}_{ij}^{[k]}$ with its neighboring sensors in $\mathcal{N}_j \setminus \{j\}$. All other quantities are available for each sensor j .

B. ESTIMATION OF $q(\lambda_w)$

With the factor $q(\lambda_w)$ we proceed in a similar fashion and reformulate the bound (8) as $\log p(\mathcal{D}) \geq \mathbb{E}_{q(\lambda_w)} \{ \log \tilde{p}(\lambda_w) / q(\lambda_w) \}$ where (see also Appendix A-B)

$$\log \tilde{p}(\lambda_w) \propto^e \mathbb{E}_{q(\mathcal{H})q(\mathbf{w})q(\mathcal{T})q(\mathcal{A})} \{ \log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}) \} \propto^e \mathbb{E}_{q(\mathbf{w})} \{ p(\mathbf{w} | \lambda_w) p(\lambda_w) \}. \quad (27)$$

Computing the expectation in (27) we find that

$$\log \tilde{p}(\lambda_w) \propto^e \left(\frac{L_w}{2} + c - 1 \right) \log(\lambda_w) - \frac{\|\widehat{\mathbf{w}}\|^2 + 2d}{2} \lambda_w \quad (28)$$

Now, $q(\lambda_w)$ is found so as to minimize the Kullback-Leibler divergence $\text{KL}(q(\lambda_w) || \tilde{p}(\lambda_w))$.

Inspecting (28) we notice that its right-hand side coincides with the logarithms of a Gamma pdf. As such, by selecting the parameters \widehat{c} and \widehat{d} of $q(\lambda_w)$ as

$$\widehat{c} = \frac{L_w}{2} + c \quad \text{and} \quad \widehat{d} = \frac{1}{2} \|\widehat{\mathbf{w}}\|^2 + d \quad (29)$$

we ensure $\text{KL}(q(\lambda_w) || \tilde{p}(\lambda_w)) = 0$. From the properties of the Gamma distribution we can also compute the mean $\widehat{\lambda}_w$ of $q(\lambda_w)$ as

$$\widehat{\lambda}_w = \frac{\widehat{c}}{\widehat{d}} = \frac{L_w + 2c}{\|\widehat{\mathbf{w}}\|^2 + 2d}. \quad (30)$$

Note, that once support $\widehat{\mathbf{w}}$ of $q(\mathbf{w})$ becomes known to the sensors, (29) can be computed “locally”, i.e., no cooperation is required for this computation.

C. ESTIMATION OF $q(\tau_j)$

With $q(\tau_j)$ we proceed in a similar fashion and simplify (8) as $\log p(\mathcal{D}) \geq \mathbb{E}_{q(\tau_j)} \{\log \tilde{p}(\tau_j)/q(\tau_j)\}$ where

$$\begin{aligned} \log \tilde{p}(\tau_j) &\propto^e \mathbb{E}_{q(\mathcal{H})q(\mathbf{w})q(\mathcal{T})q(\mathcal{A})} \{\log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})\} \\ &\propto^e \mathbb{E}_{q(\mathbf{w})q(\mathcal{H})q(\mathcal{T} \setminus \{\tau_j\})} \{\log p(\mathcal{D}|\mathbf{w}, \mathcal{H}, \mathcal{T})p(\tau_j)\}. \end{aligned} \quad (31)$$

Using the fact that the likelihood $p(\mathcal{D}|\hat{\mathbf{w}}, \mathcal{H}, \mathcal{T})$ can be factored as $p(\mathcal{D}|\hat{\mathbf{w}}, \mathcal{H}, \mathcal{T}) = \prod_{j=1}^J p(\mathbf{d}_j|\hat{\mathbf{w}}, \hat{\mathbf{h}}_j, \tau_j)$, we can evaluate (31) using the same steps as for the factor $q(\lambda_w)$ (see Appendix A-B). As a result,

$$\begin{aligned} \log \tilde{p}(\tau_j) &\propto^e \left(\frac{L_W + L_R - 1}{2} + s_j - 1 \right) \log \tau_j \quad (32) \\ &\quad - \left(\frac{1}{2} \|\mathbf{d}_j - \hat{\mathbf{H}}_j \hat{\mathbf{w}}\|^2 + \frac{1}{2} \hat{\mathbf{w}}^\top \hat{\mathbf{\Omega}}_j \hat{\mathbf{w}} + t_j \right) \tau_j. \end{aligned}$$

Expression (32) is similar in its structure to $\tilde{p}(\lambda_w)$ in (28): it is the logarithm of a Gamma pdf. Therefore, by selecting parameters \hat{s}_j and \hat{t}_j of $q(\tau_j)$ as

$$\hat{s}_j = \frac{L_W + L_R - 1}{2} + s_j \quad \text{and} \quad (33a)$$

$$\hat{t}_j = \frac{1}{2} \|\mathbf{d}_j - \hat{\mathbf{H}}_j \hat{\mathbf{w}}\|^2 + \frac{1}{2} \text{tr}\{\hat{\mathbf{S}}_{h_j} \hat{\mathbf{W}}^\top \hat{\mathbf{W}}\} + t_j \quad (33b)$$

we will ensure that $\text{KL}(q(\tau_j)||\tilde{p}(\tau_j)) = 0$. The mean $\hat{\tau}_j$ of $q(\tau_j)$ used in, e.g., (21), can then be computed as

$$\hat{\tau}_j = \frac{\hat{s}_j}{\hat{t}_j} = \frac{L_W + L_R - 1 + 2s_j}{\|\mathbf{d}_j - \hat{\mathbf{H}}_j \hat{\mathbf{w}}\|^2 + \text{tr}\{\hat{\mathbf{S}}_{h_j} \hat{\mathbf{W}}^\top \hat{\mathbf{W}}\} + 2t_j}. \quad (34)$$

Again, similarly to the estimation of $q(\lambda_w)$ we see that once support $\hat{\mathbf{w}}$ of $q(\mathbf{w})$ becomes known to each sensor j , computing (33) does not require cooperation.

D. ESTIMATION OF $q(\mathbf{h}_j)$ AND $q(\alpha_j)$

Following similar steps we can now estimate factors $q(\mathbf{h}_j)$ and $q(\alpha_j)$ by optimizing bounds

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\mathbf{h}_j)} \{\log \tilde{p}(\mathbf{h}_j)/q(\mathbf{h}_j)\}$$

and

$$\log p(\mathcal{D}) \geq \mathbb{E}_{q(\alpha_j)} \{\log \tilde{p}(\alpha_j)/q(\alpha_j)\}$$

respectively, where

$$\begin{aligned} \log \tilde{p}(\mathbf{h}_j) &\propto^e \mathbb{E}_{q(\mathbf{w})q(\mathcal{T})q(\mathcal{A})} \{\log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})\} \\ &\propto^e \mathbb{E}_{q(\mathbf{w})q(\tau_j)q(\alpha_j)} \{\log p(\mathbf{d}_j|\mathbf{w}, \mathbf{h}_j, \tau_j)p(\mathbf{h}_j|\alpha_j)\} \\ &\propto^e \log p(\mathbf{d}_j|\hat{\mathbf{w}}, \mathbf{h}_j, \hat{\tau}_j) + \log p(\mathbf{h}_j|\hat{\alpha}_j), \end{aligned} \quad (35)$$

$\hat{\alpha}_j = [\hat{\alpha}_{j,1}, \dots, \hat{\alpha}_{j,L_R}]^\top$ is the mean of $q(\alpha_j)$, and

$$\begin{aligned} \log \tilde{p}(\alpha_j) &\propto^e \mathbb{E}_{q(\mathcal{H})q(\mathbf{w})q(\mathcal{T})} \{\log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})\} \\ &\propto^e \mathbb{E}_{q(\mathbf{h}_j)} \{\log p(\mathbf{h}_j|\alpha_j)p(\alpha_j)\}. \end{aligned} \quad (36)$$

From (35) we can see that since both $p(\mathbf{d}_j|\hat{\mathbf{w}}, \mathbf{h}_j, \hat{\tau}_j)$ and $p(\mathbf{h}_j|\hat{\alpha}_j)$ are Gaussian, $\log \tilde{p}(\mathbf{h}_j)$ must be quadratic in

\mathbf{h}_j . As such, $\tilde{p}(\mathbf{h}_j)$ must be Gaussian. The optimal pdf $q(\mathbf{h}_j)$ that will minimize the Kullback-Leibler divergence $\text{KL}(q(\mathbf{h}_j)||\tilde{p}(\mathbf{h}_j))$ should thus be selected so that its moments coincide with those of $\tilde{p}(\mathbf{h}_j)$. These moments can be found by computing the first and second derivative of $\log \tilde{p}(\mathbf{h}_j)$ in (35) with respect to \mathbf{h}_j , which leads to

$$\hat{\mathbf{S}}_{h_j} = \left(\hat{\tau}_j \hat{\mathbf{W}}^\top \hat{\mathbf{W}} + \hat{\mathbf{A}}_j \right)^{-1}, \quad \hat{\mathbf{h}}_j = \hat{\tau}_j \hat{\mathbf{S}}_{h_j} \hat{\mathbf{W}}^\top \hat{\mathbf{d}}_j, \quad (37)$$

where $\hat{\mathbf{A}}_j = \text{diag}\{\hat{\alpha}_{j,1}, \dots, \hat{\alpha}_{j,L_R}\}$. Expression (37) can be recognized as a regularized least squares estimate of the reflectivity, with α_j playing the role of regularization parameters.

To compute $q(\alpha_j)$ we evaluate the expectation in (36), which results in

$$\begin{aligned} \log \tilde{p}(\alpha_j) &\propto^e \frac{1}{2} \log |\mathbf{A}_j| - \frac{1}{2} \hat{\mathbf{h}}_j^\top \mathbf{A}_j \hat{\mathbf{h}}_j \quad (38) \\ &\quad - \frac{1}{2} \text{trace} \left[\mathbf{A}_j \hat{\mathbf{S}}_{h_j} \right] + \sum_{i=1}^{L_R} (a_j - 1) \log \alpha_{i,j} - b_j \alpha_{i,j} \\ &= \sum_{i=1}^{L_R} (a_j - \frac{1}{2}) \log \alpha_{i,j} \quad (39) \\ &\quad + \sum_{i=1}^{L_R} \frac{1}{2} \left([\hat{\mathbf{h}}_j]_i^2 + [\hat{\mathbf{S}}_{h_j}]_{i,i} + 2b_j \right) \alpha_{i,j} \end{aligned}$$

which can be recognized as a sum of logarithms of Gamma pdfs with a shape parameter $(a_j + \frac{1}{2})$ and L_R rate parameters $\frac{1}{2} \left([\hat{\mathbf{h}}_j]_i^2 + [\hat{\mathbf{S}}_{h_j}]_{i,i} + 2b_j \right)$, $i = 1, \dots, L_R$. Thus, optimal $q(\alpha_j) = \prod_{i=1}^{L_R} q(\alpha_{i,j})$ should be selected as a product of Gamma distributions, with the shape and rate parameters of each factor $q(\alpha_{i,j})$ set to

$$\hat{a}_{i,j} = a_j + \frac{1}{2} \quad \text{and} \quad \hat{b}_{i,j} = \frac{1}{2} [\hat{\mathbf{h}}_j]_i^2 + \frac{1}{2} [\hat{\mathbf{S}}_{h_j}]_{i,i} + b_j, \quad (40)$$

respectively. Note that the mean of $q(\alpha_{i,j})$, $i = 1, \dots, L_R$, used in (37), is then computed as

$$\hat{\alpha}_{i,j} = \frac{\hat{a}_{i,j}}{\hat{b}_{i,j}}. \quad (41)$$

E. INITIALIZATION AND SUMMARY OF THE ALGORITHM

Let us now discuss initialization of the proposed inference scheme and summarize the overall algorithm structure. In the following we will refer to the proposed blind deconvolution algorithm with sparse Bayesian learning as VBI+SBL.

We begin with specifying the prior distributions $p(\lambda_w)$, $p(\alpha_j)$ and $p(\tau_j)$, $j \in \mathcal{J}$. For both $p(\alpha_j)$, $j \in \mathcal{J}$, and $p(\lambda_w)$ we use non-informative priors, i.e., we set $a_j = b_j = 0, \forall j \in \mathcal{J}$, $i = 0, \dots, L_R$, and $c = d = 0$. However, in case of noise prior $p(\hat{\tau}_j)$ we set non-zero values to parameters $s_j, t_j, \forall j \in \mathcal{J}$. This is done to better account for the model mismatch in the estimated wavelet $\hat{\mathbf{w}}$ in addition to additive measurement noise \mathbf{n}_j , $j \in \mathcal{J}$. We observed that this improves empirical performance of the VBI+SBL deconvolution scheme,

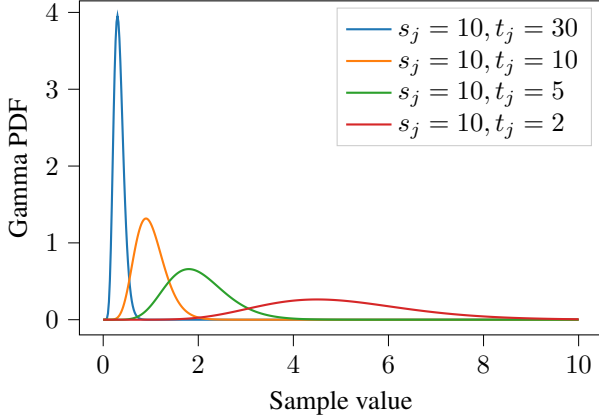


FIGURE 1: Gamma PDF.

as compared to a non-informative choice of $p(\tau_j)$, $j \in \mathcal{J}$. Figure 1 exemplary shows the Gamma pdf $\text{Ga}(\tau_j|s_j, t_j)$ for different prior parameter values. In particular, the ratio s_j/t_j determines the mean of the respective Gamma pdf and this was used as a guiding rule to select appropriate parameter values. Beside prior parameters, we also need to initialize the variational factors in (19). We begin the iterative estimation with a factor $q(\mathbf{w})$, which implies that factors (19a)-(19d), or rather their parameters, have to be specified. For the noise precision and sparsity parameters we set, respectively, $\hat{\tau}_j = 1$ and $\hat{\alpha}_j = \mathbf{1}, \forall j \in \mathcal{J}$, where $\mathbf{1}$ is an all-one vector of an appropriate dimension.

A bit more care is needed for initialization of factors $q(\mathbf{h}_j)$, $j \in \mathcal{J}$. To obtain initial estimates of the reflectivities the authors in [5] proposed to use a peak locator for each seismic trace \mathbf{d}_j . Here we adopt the same approach and set means of $q(\mathbf{h}_j)$, $j \in \mathcal{J}$, to estimated peak locations of corresponding seismic traces. Note, however, that to obtain an appropriate initial estimate of the reflectivities we need to take into account the peak position of the wavelet \mathbf{w} . Since the source wavelet \mathbf{w} is modeled as a causal FIR system, its peak does not lie at the 0-th sample, but is delayed by l_{peak} samples. Therefore, the peaks in the seismic trace \mathbf{d}_j are not aligned to the original peak positions, but rather shifted by l_{peak} samples as well. Hence, if we apply a peak locator on the seismic trace \mathbf{d}_j the original peaks in \mathbf{h}_j will be shifted by l_{peak} in the initial reflectivity estimate $\hat{\mathbf{h}}_j$. This offset will eventually lead to a corrupted wavelet estimate. To enable an appropriate wavelet estimation, after applying the peak locator on the trace \mathbf{d}_j the located peaks are shifted back in time by l_{peak} samples. We agree that in reality this information needs to be appropriately estimated based on preliminary calibration studies. For validation purposes, however, we assume that l_{peak} is known perfectly. The initial estimate of each reflectivity is then modified as follows:

$$\hat{\mathbf{h}}_j \leftarrow [\hat{h}_j(l_{\text{peak}}), \hat{h}_j(1 + l_{\text{peak}}), \dots, \hat{h}_j(L_R + l_{\text{peak}})], \quad (42)$$

where the last l_{peak} entries in $\hat{\mathbf{h}}_j$ are set to 0. Finally, we set the initial covariance matrix of the reflectivities to $\hat{\mathbf{S}}_{h_j} = \mathbf{I}$

and $\hat{\mathbf{\Omega}}_j = \mathbf{I}$.

We now summarize the key steps of the proposed distributed VBI+SBL blind deconvolution scheme in Algorithm 1. As we see, it consists of two estimation stages: (i) a wavelet estimation stage that is conducted in a cooperative fashion by all sensors, and (ii) a reflectivity estimation stage that is done locally at each sensor without data exchange. Furthermore, note that the algorithm has inner and outer iterations, which we index by superscripts $[m], [n]$ and (k) , respectively. The outer iterations correspond to a cycle over variational factors. The inner ones, on the other hand, are iterative updates of some of these factors.

Algorithm 1 Distributed VBI + SBL Blind Deconvolution

```

Set  $\rho, c, d, a_j, b_j, s_j, t_j$ 
Initialize  $\hat{\alpha}_j, \hat{\tau}_j, \hat{\mathbf{S}}_{h_j}, \hat{\mathbf{\Omega}}_j$ 
Initialize reflectivity  $\hat{\mathbf{h}}_j^{(0)}$  via peak locator on  $\mathbf{d}_j$ 
for  $k \leftarrow 1, \dots, K$  do
    Update convolution matrix  $\hat{\mathbf{H}}_j, \forall j \in \mathcal{J}$ , using  $\hat{\mathbf{h}}_j^{(k-1)}$ 
    for  $m \leftarrow 0, M - 1$  do  $\triangleright$  Distributed wavelet stage
         $\mathbf{u}_{ij}^{[m+1]} \leftarrow (26c)$ 
         $\mathbf{z}_j^{[m+1]} \leftarrow (26b)$ 
         $\mathbf{w}_j^{[m+1]} \leftarrow (26a)$ 
    end for
     $\mathbf{w}_j^{(k)} \leftarrow \mathbf{w}_j^{[M]}$ 
    Compute  $\hat{c}, \hat{d} \leftarrow (29)$  and  $\hat{\lambda}_w \leftarrow (30)$ 
    Compute noise precision  $\hat{\tau}_j \leftarrow (34)$ 
    Update convolution matrix  $\hat{\mathbf{W}}_j, \forall j \in \mathcal{J}$ , using  $\mathbf{w}_j^{(k)}$ 
    for  $n \leftarrow 0, N - 1$  do  $\triangleright$  Local reflectivity stage
         $\hat{\alpha}_{i,j} \leftarrow (41)$ 
         $\hat{\mathbf{S}}_{h_j}, \hat{\mathbf{h}}_j^{[n+1]} \leftarrow (37)$ 
    end for
     $\hat{\mathbf{h}}_j^{(k)} \leftarrow \hat{\mathbf{h}}_j^{[N]}$ 
    Compute matrix  $\hat{\mathbf{\Omega}}_j \leftarrow (48)$ 
end for
return Estimated wavelet  $\mathbf{w}_j^{(K)}$  and reflectivity  $\hat{\mathbf{h}}_j^{(K)}$ 

```

IV. NUMERICAL RESULTS

A. SYNTHETIC REFLECTIVITIES

In the following, we evaluate the performance of the proposed algorithm in numerical experiments using synthetic reflectivity data. To this end, we consider a network of $J = 10$ fixed sensors in a line topology where the outer sensors have two neighbours and other sensors have from three to four neighbors depending on the position in the line array. We use synthetically constructed reflectivities to generate seismic traces at the sensors at a sampling frequency of $f_s = 500$ Hz. As a source signal \mathbf{w} we use a Ricker wavelet [25] with dominant frequency $f_R = 40$ Hz, phase shift of 30° and duration $T_W = 0.1$ s. Furthermore, we perturb the generated measurements for each sensor with additive white Gaussian noise with zero mean and variance σ_j^2 , $j \in \mathcal{J}$. To assess the

impact of noise we also define the signal-to-noise ratio (SNR) as $\text{SNR} = 10 \log_{10} \{ \|s_j\|^2 / (L_s \sigma_j^2) \}$, where $s_j \in \mathbb{R}^{L_s}$ is the noise-free seismic trace at sensor j . For each sensor j we assume the same SNR conditions, i.e., $\text{SNR}_j = \text{SNR}$, $\forall j \in \mathcal{J}$. Also, we will average the performance of the algorithms over 100 independent Monte Carlo runs to collect statistics.

The generated true reflectivities are exemplified in Figure 2a together with corresponding seismic traces at SNR = 10 dB in Figure 2b.

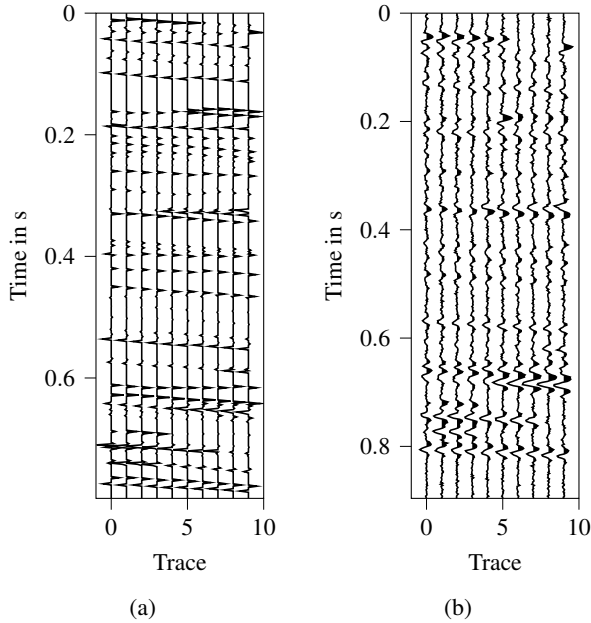


FIGURE 2: (a) Synthetic reflectivities and (b) example of seismic traces at SNR = 10 dB.

As performance metric for the deconvolution algorithm, we will consider two measures: the Pearson correlation coefficient (PCC) and the earth mover’s distance (EMD) metrics

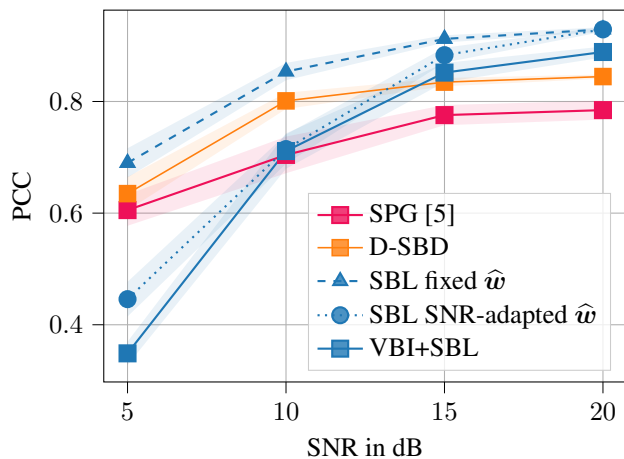


FIGURE 3: Reflectivity PCC performance of all schemes averaged over 100 trials.

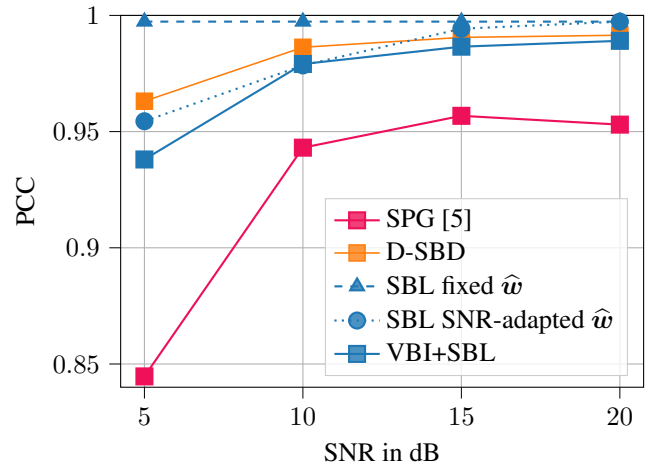


FIGURE 4: Wavelet PCC performance of all schemes averaged over 100 trials.

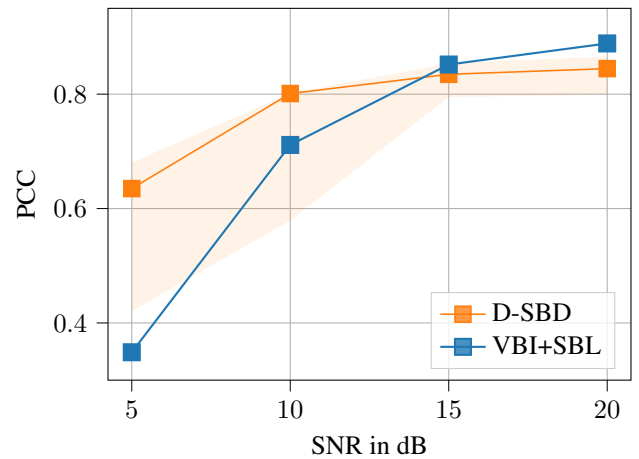


FIGURE 5: Reflectivity PCC performance of VBI+SBL and D-SBD. For D-SBD the minimum and maximum achievable PCC values depending on the ℓ_1 regularization parameter are shown as shadowed region.

[26], [27]. For vectors x and \hat{x} from the same signal space, the PCC is simply defined as $\text{PCC} = |x^\top \hat{x}| / (\|x\| \cdot \|\hat{x}\|)$. A PCC value close to one indicates a high similarity between the two vectors; a value of 0 implies highest dissimilarity (i.e., orthogonality) of vectors. However, since the PCC is not explicitly taking into account the distance between correct and estimated peak location in the signal vectors, we additionally employ the EMD metric, which is a discrete version of the Kantorovich–Rubinstein metric for probability distributions on metric spaces [28]. Since sparse vectors can be regarded as distributions, EMD will thus quantify similarity between sparse vectors more accurately.

In the implementation of the VBI+SBL algorithm we set the ADMM penalty parameter for wavelet estimation to $\rho = 15$ and also set $s_j = 10, t_j = 10, \forall j \in \mathcal{J}$ for the estimate of the noise precision $p(\tau_j)$. We also assume $K = 10$ outer

iterations and $M = N = 10$ inner iterations for the wavelet estimation and reflectivity estimation each.

We will compare the proposed VBI+SBL to (i) a distributed sparse blind deconvolution (D-SBD) algorithm proposed in [13], which uses LASSO for recovery of a sparse signal, and (ii) sparse multichannel blind deconvolution using spectral projected-gradient (SPG) in [5]. Additionally, we include two further cases of the proposed deconvolution using SBL into the analysis. For both cases, we substitute the wavelet estimation with a fixed function. In the first case, we use a wavelet that has been pre-estimated with the D-SBD at the corresponding SNRs, termed *SBL SNR-adapted \hat{w}* in the following comparisons. In the second case, we use a wavelet estimate with a high PCC = 0.99. This approach we term *SBL fixed \hat{w}* . This is done to test how sensitive the VBI+SBL is to mismatches in the wavelet estimation.

Figure 3 shows the PCC performance of reflectivity estimation as a function of SNR and in Figure 4 we show the corresponding PCC of the wavelet estimates. For D-SBD, the ℓ_1 -regularization parameter has been chosen such, that an overall good performance is achieved over the considered SNR range. As we see, VBI+SBL outperforms both D-SBD and SMBD-SPG in the middle to high SNR regime, but degrades as SNR drops. Note that SBL learns optimal regularisation with a factor $q(\alpha_j)$, $j \in \mathcal{J}$. As such, its reflectivity estimate is adapted over SNRs. On the other hand, D-SBD employs a LASSO estimator with a fixed regularization parameter for the ℓ_1 -norm and thus lacks the ability to adapt it to changing SNR values. SMBD-SPG employs the spectral projected gradient that adapts the sparsity threshold based on the noise power. Nevertheless, we observe that it performs worse than the other two methods despite the fact that it conducts the deconvolution based on all the available network data at once in a centralized manner. By fixing the wavelet estimate in VBI+SBL, we see that performance of our proposed method improves: for *SBL SNR-adapted \hat{w}* we observe performance improvement in the middle-to-high SNR regime (see also in Figure 4). For the case of a nearly “perfect” wavelet estimate, VBI+SBL clearly outperforms D-SBD and SMBD-SPG over the whole SNR region. This implies that VBI+SBL is sensitive to a good choice of wavelet model w , yet it does provide a more accurate sparsity estimate as compared to ℓ_1 -based optimization when the wavelet is accurately estimated.

Nonetheless, let us mention that the ℓ_1 -regularization in D-SBD algorithm has been optimized so that an overall good performance is achieved. In Figure 5 we additionally show the impact of this parameter on D-SBD performance. Specifically, we show bounds of achievable PCC performance with respect to variation of the regularization parameter. We can observe that especially in the low SNR regime the PCC performance of D-SBD varies significantly depending on the choice of λ_{ℓ_1} . If λ_{ℓ_1} is selected without care, VBI+SBL performs similar to the D-SBD also in the low SNR regime.

In terms of the EMD performance shown in Figure 6, we can observe a similar behavior of the algorithms as in Figure 3. However, in both cases with a pre-estimated

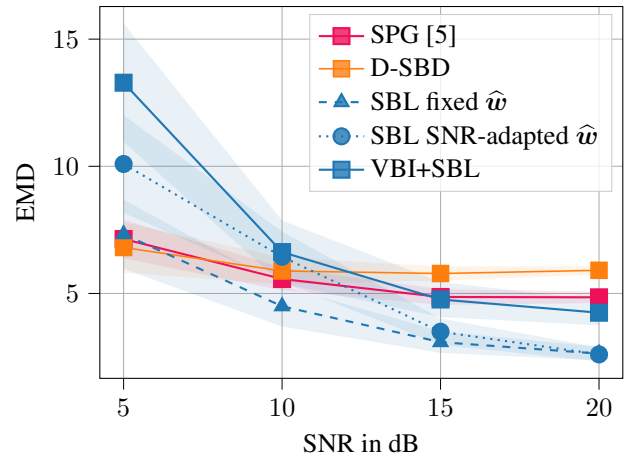


FIGURE 6: EMD performance of reflectivity estimates for all schemes averaged over 100 trials.

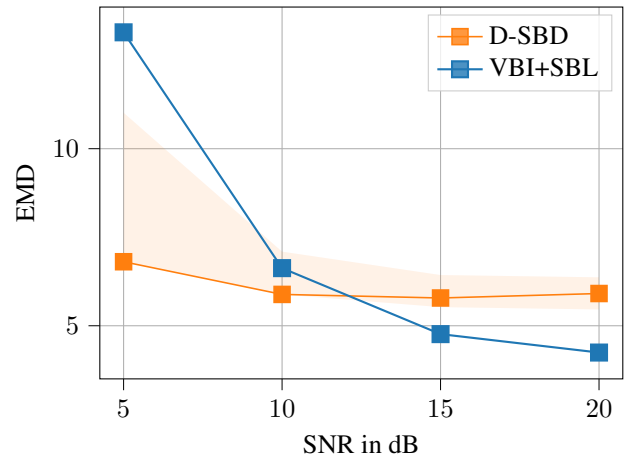


FIGURE 7: EMD performance of VBI+SBL and D-SBD. For D-SBD the minimum and maximum achievable PCC values depending on the ℓ_1 regularization parameter are shown as shadowed region.

wavelet used in VBI+SBL we see a performance increase as compared to D-SBD at the considered SNR range. Furthermore, for higher SNRs all SBL-based schemes outperform D-SBD and SMBD-SPG quite significantly. This again indicates more accurate sparsity estimation results. Also, note that both D-SBD and SMBD-SPG show a very flat EMD performance especially in the high SNR regime where SMBD-SPG outperforms D-SBD. Again here, the learning ability of SBL becomes particularly apparent. If we include maximum and minimum achievable EMD performance of D-SBD depending on the ℓ_1 regularization parameter, we can observe that the EMD varies significantly in the low SNR regime (see Figure 7) and around SNR = 10 dB VBI+SBL falls in the same performance region as D-SBD.

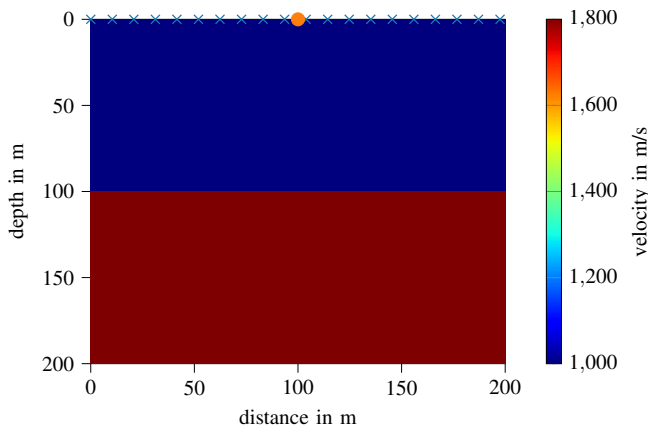


FIGURE 8: Subsurface model with source and receivers positioned on the surface as indicated by crosses.

B. SYNTHETIC DATA GENERATED WITH AN ACOUSTIC WAVE EQUATION

Let us now consider the algorithm performance using data generated based on a physical model of an acoustic wave propagating in a medium. Specifically, we use an acoustic wave equation to generate seismic measurements at the sensors. The acoustic wave equation is given by

$$\frac{\partial^2}{\partial t^2} u(\mathbf{x}, t) = c(\mathbf{x})^2 \Delta u(\mathbf{x}, t) + f(\mathbf{x}, t) \quad (43)$$

where $u(\mathbf{x}, t)$ is the acoustic pressure at position \mathbf{x} and time t , $c(\mathbf{x})$ is the P -wave velocity in the medium and $f(\mathbf{x}, t)$ is the source term. We solve (43) numerically using finite-difference method with absorbing boundary conditions to avoid reflections at the borders of the computational domain [29]. The solution – the computed wave field $u(\mathbf{x}, t)$ – is then sampled at the corresponding sensor positions \mathbf{x}_j , $j \in \mathcal{J}$ to obtain the seismic measurements at the sensors. Note that in this case the model (2) is only approximately valid, in contrast to the previous simulation case; this implies a higher model mismatch.

As subsurface we use a two-layer model in 2D with layer depths $d_1 = 100$ m, $d_2 = 200$ m and wave velocities $v_1 = 1000$ m/s, $v_2 = 1800$ m/s, respectively. We assume $J = 20$ sensors uniformly spaced on a line over a length of 200 m. As source term $f(\mathbf{x}, t)$ we choose a Ricker wavelet with $f_R = 20$ Hz and phase shift of 30° . We place the source on the surface at the center of the sensor line array. The subsurface model together with the source-receiver setup is depicted in Figure 8.

Figure 9a shows the measured seismic traces at SNR = 20 dB as a function of the trace number, where we can clearly see the arrival of the direct wave and the reflections from the interface between the first and second layer. We apply both VBI+SBL and D-SBD algorithms with $K = 10$ outer iterations and $M = N = 10$ inner iterations for wavelet and reflectivity estimation, respectively. Furthermore, for VBI+SBL we set the ADMM penalty parameter for wavelet

estimation to $\rho = 15$ and set the parameters for the estimation of $p(\tau_j)$ to $s_j = 10$, $t_j = 50$ for all sensors and for the estimation of $p(\lambda_w)$ to $c = 1$, $d = 50$. The deconvolution results can be seen in Figure 9b and 9c. One can observe that both SBL and D-SBD successfully recover the reflectivity peaks at the corresponding arrival times. However, VBI+SBL tends to obtain more distinct reflectivity peaks compared to D-SBD. This can be attributed to a better EMD performance of the proposed VBI+SBL algorithm. Especially in the range of 0.3 s to 0.35 s VBI+SBL clearly recovers the significant reflections from the first layer while suppressing unwanted reflections in the measurement data. Figure 10 depicts the same results but with wiggle plots instead of images. Here, it can be more noticeably observed that VBI+SBL recovers a cleaner structure of the seismic signal, as can be seen on the zoomed portion of the reconstructed signal.

C. REAL SEISMIC DATA

For the last evaluation example, we use real seismic data from a reflection survey provided by the National Petroleum Reserve, Alaska (NPR) Legacy Data Archive by USGS (1976), Line ID 31-81 in SEG-Y format available at <http://certmapper.cr.usgs.gov/data/apps/npra/>. The data has been recorded with a sampling frequency of 250 Hz. For our purposes we use measurement data from 100 traces between 1 s and 1.6 s. For the geophone constellation we assume a line array where each geophone has up to six connected neighbors to each of its side. Again, we compare VBI+SBL to D-SBD.

For both VBI+SBL and D-SBD we employ $K = 5$ iterations with $M = N = 10$ iterations for wavelet and reflectivity estimation, respectively. Furthermore, for VBI+SBL we set $\rho = 1$ and $s_j = 100$, $t_j = 0$, $c = 1000$, $d = 0$. These parameter settings were found to give a satisfactory deconvolution result. The original real seismic data and the deconvolved results by D-SBD and VBI+SBL can be seen in Figure 11. For both D-SBD and VBI+SBL one can observe clear deblurring and removal of ringing artifacts of the recorded seismic data. In particular, VBI+SBL obtains a much cleaner image with sharper features compared to D-SBD. This observation coincides with the spikier deconvolution result of VBI+SBL seen in Section IV-B. These results show that VBI+SBL provides satisfactory results also for data from real seismic surveys.

V. CONCLUSION AND FUTURE WORK

This paper presents a distributed probabilistic framework for blind deconvolution of seismic signals under sparsity constraints. The probabilistic formulation of the inference problem allows applying a variational Bayesian inference to efficiently approximate the posterior distribution of the parameters of interest, thus accounting for uncertainty of parameter estimates. Depending on a particular choice of variational factors an instance of the distributed sparse blind deconvolution (D-SBD) algorithm, proposed in [13], can be obtained. Furthermore, the variational formulation also allows for extending the algorithm with an automatic choice

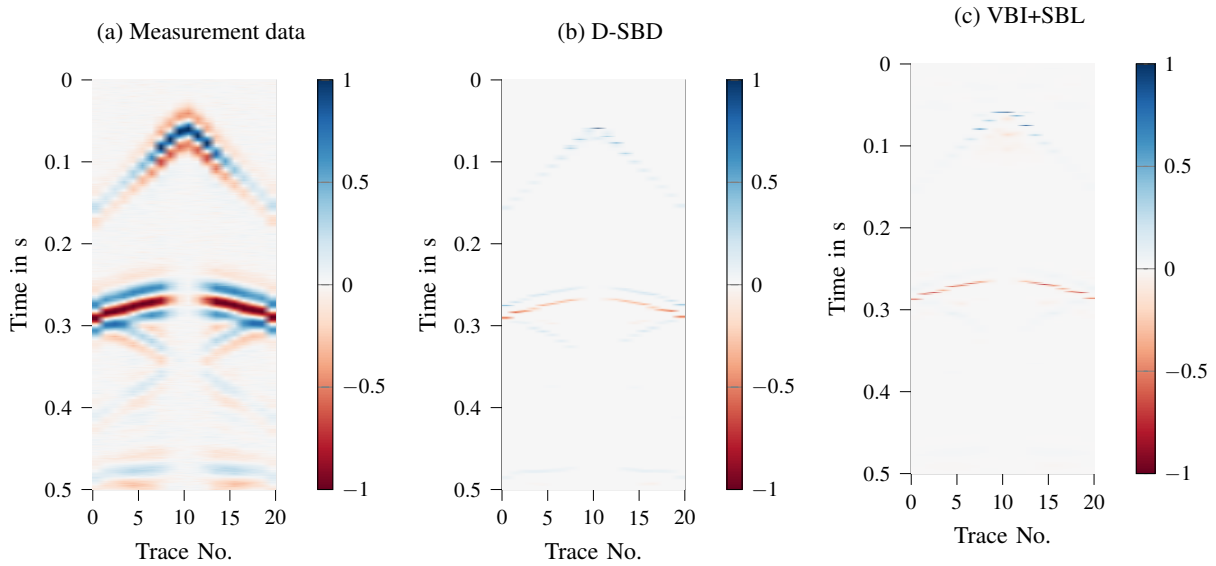


FIGURE 9: Image of (a) measured seismic traces at SNR = 20 dB, (b) estimated reflectivities by D-SBD and (c) estimated reflectivities by VBI+SBL after $K = 10$ iterations.

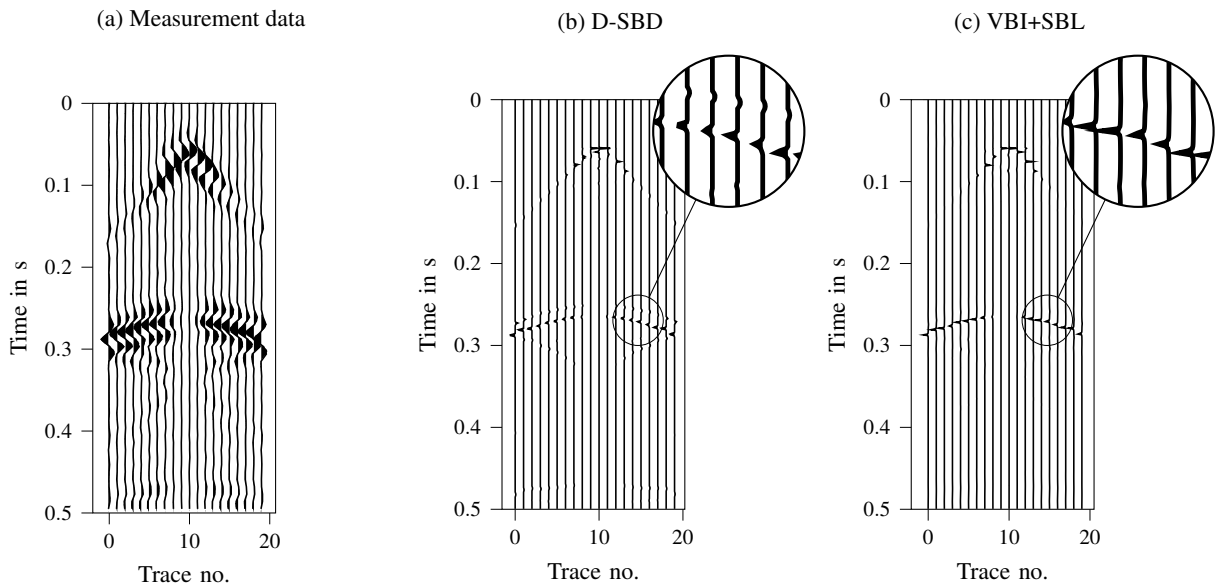


FIGURE 10: Wiggle plots of (a) measured seismic traces at SNR = 20 dB (b) estimated reflectivities by D-SBD and (c) estimated reflectivities by VBI+SBL after $K = 10$ iterations.

of regularization factors that are learned using empirical Bayesian method along with the other parameters within the estimation framework. This represents an extension of the previous method. Specifically, (i) instead of ℓ_1 -penalization used in [13] we apply sparse Bayesian learning for a sparse reconstruction of the reflectivity profile, (ii) we learn unknown additive noise variances at sensors, and (iii) adaptively regularize wavelet estimation using a probabilistic version of ridge regression.

The estimation algorithm includes two key steps repeated iteratively. The first step is a distributed wavelet estima-

tion, which is done cooperatively over the network of sensors/agents using the ADMM algorithm. In the second step, the reflectivity profiles of the subsurface are estimated locally based on the estimate of the wavelet; this computation step does not require any cooperation.

For performance evaluation we compared the proposed VBI+SBL algorithm with D-SBD and SMBD-SPG using Pearson correlation coefficient (PCC) and earth mover's distance (EMD) metrics for estimated wavelet and reflectivity signals. The analysis of the methods in simulated environments have shown that the VBI+SBL learning scheme, which

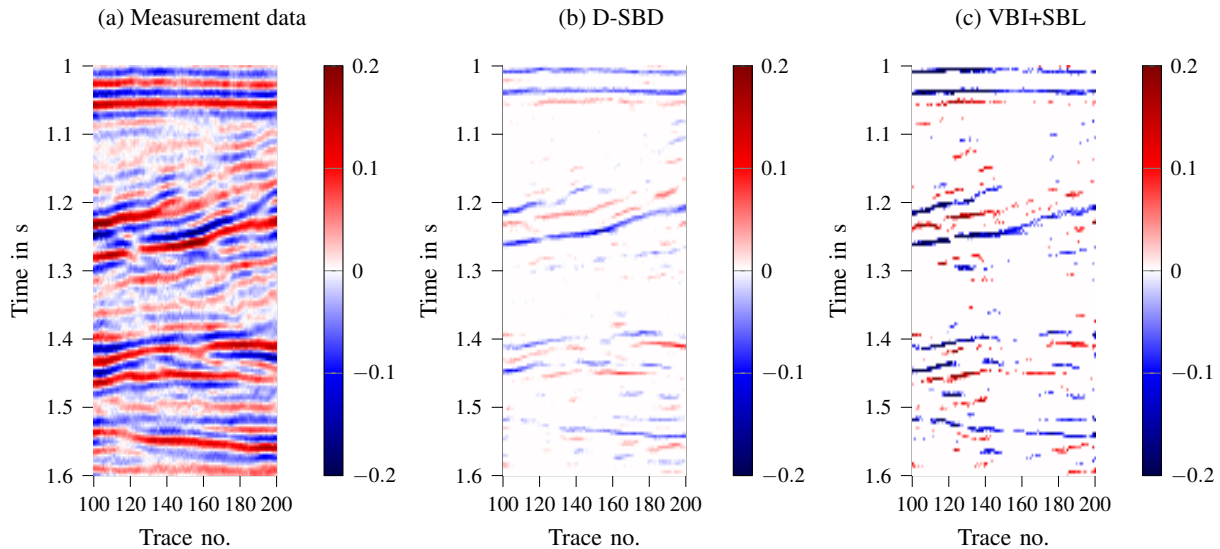


FIGURE 11: Image of (a) measured seismic traces from reflectivity survey, (b) estimated reflectivities by D-SBD and (c) estimated reflectivities by VBI+SBL.

includes estimation of both wavelet and reflectivity signals, as well as regularization parameters performs well in the middle to high SNR regime. In the low SNR regime the coupled estimation of wavelet and reflectivities apparently leads to a propagation of estimation errors between these two stages, which degrades the deconvolution performance. Although, in our evaluations the D-SBD algorithm was found to perform better in the low SNR regime, it requires manual tuning of regularization parameters which is not necessary for VBL+SBL. When not selected properly, D-SBD and VBI+SBL can perform similarly in the low SNR regime. Furthermore, for D-SBD the EMD metric flattens out in the high SNR regime, which indicates that further, more precise tuning of regularization parameters is needed. In case of VBI+SBL this is not observed, since the parameters are automatically tuned.

Furthermore, we also used synthetic seismic data generated by solving the acoustic wave equation. This produces more realistic testing data, yet at the same time the data generated in this way does not exactly follow the signal model used for deconvolution. In general, we see that both methods produce an acceptable performance, with VBI+SBL resulting in more accurate signal sparsity. However, VBI+SBL seems to be more sensitive to the model mismatches. This can be explained by a higher number of available degrees of freedom and higher nonlinearity/non-convexity of the optimization problem. This consecutively leads to a higher sensitivity in the algorithm initialization. Additionally, we tested VBI+SBL with real seismic data obtained in a reflectivity survey. Compared to D-SBD our proposed method achieves a sharper reflectivity image with higher amplitudes showing the effectiveness of SBL for the reflectivity estimation. Furthermore, ringing effects from the

original measurement data are successfully removed. Hence, also for the decentralized blind deconvolution of real seismic data the VBI+SBL provides a satisfactory solution.

Finally, we conclude that in practice it seems reasonable to combine D-SBD and VBI+SBL, where the former can provide an initialization, while our proposed algorithm can be used to fine-tune signal estimates. For future work, we will investigate such combination in more detail.

APPENDIX A COMPUTATION OF THE VARIATION LOWER BOUNDS

To evaluate the variational bound (8) we compute $\log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A})$ and ignore constant terms.

$$\begin{aligned} \log p(\mathcal{D}, \mathbf{w}, \mathcal{H}, \lambda_w, \mathcal{T}, \mathcal{A}) = & \text{const} + & (44) \\ & \frac{L_W}{2} \log \lambda_w - \frac{1}{2} \lambda_w \|\mathbf{w}\|^2 + (c-1) \log(\lambda_w) - d \lambda_w \\ & + \sum_{j \in \mathcal{J}} \left[\frac{L_W + L_R - 1}{2} \log \tau_j - \frac{1}{2} \tau_j \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \right. \\ & \quad \left. - \frac{1}{2} \mathbf{h}_j^T \mathbf{A}_j \mathbf{h}_j + (s_j - 1) \log \tau_j - t_j \tau_j \right. \\ & \quad \left. + \sum_{i=1}^{L_R} \frac{1}{2} \log \alpha_{i,j} + (a_j - 1) \log \alpha_{i,j} - b_j \alpha_{i,j} \right] \end{aligned}$$

To evaluate the variational lower bound, the expectation of (44) is evaluated with respect to the Markov Blanket [16] of the variable being estimated, as will be detailed in the following.

A. VARIATION LOWER BOUND WITH RESPECT TO $q(\mathbf{w})$

For estimation of $q(\mathbf{w})$ we need to evaluate (14). By inserting (6) into (44) and retaining the terms that depend on \mathbf{w} we

obtain

$$\log \tilde{p}(\mathbf{w}) \propto^e \text{const} - \frac{1}{2} \mathbb{E}_{q(\lambda_w)} \{ \lambda_w \} \|\mathbf{w}\|^2 - \frac{1}{2} \sum_{j \in \mathcal{J}} \mathbb{E}_{q(\tau_j)} \{ \tau_j \} \mathbb{E}_{q(\mathbf{h}_j)} \{ \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \} \quad (45)$$

Since $q(\lambda_w)$ and $q(\tau_j)$, $j \in \mathcal{J}$, are selected as Gamma distributions, we can compute

$$\mathbb{E}_{q(\lambda_w)} \{ \lambda_w \} = \frac{\hat{c}}{\hat{d}} = \hat{\lambda}_w, \text{ and } \mathbb{E}_{q(\tau_j)} \{ \tau_j \} = \frac{\hat{s}_j}{\hat{t}_j} = \hat{\tau}_j, j \in \mathcal{J}, \quad (46)$$

which follows from the properties of Gamma pdfs. Now, let us compute $\mathbb{E}_{q(\mathbf{h}_j)} \{ \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \}$. First we note that

$$\mathbb{E}_{q(\mathbf{h}_j)} \{ \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \} = \mathbb{E}_{q(\mathbf{h}_j)} \left\{ \sum_{l=0}^{L_R+L_W-2} \left(d_j[l] - \sum_{\nu=0}^{L_W-1} w[\nu] h_j[l-\nu] \right)^2 \right\} \quad (47)$$

Expanding the square and computing the expectations we find that

$$\mathbb{E}_{q(\mathbf{h}_j)} \{ \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \} = \|\mathbf{d}_j\|^2 - 2\mathbf{d}_j^T \hat{\mathbf{h}}_j * \mathbf{w} + \sum_{l=0}^{L_R+L_W-2} \sum_{\nu=0}^{L_W-1} \sum_{r=0}^{L_W-1} w[\nu] w[r] \mathbb{E}_{q(\mathbf{h}_j)} \{ h_j[l-\nu] h_j[l-r] \}$$

where the expectation in the latter expression can be computed as

$$\mathbb{E}_{q(\mathbf{h}_j)} \{ h_j[l-\nu] h_j[l-r] \} = \hat{h}_j[l-\nu] \hat{h}_j[l-r] + [\mathbf{S}_{h_j}]_{l-\nu, l-r}$$

Finally, we can compute

$$\mathbb{E}_{q(\mathbf{h}_j)} \{ \|\mathbf{d}_j - \mathbf{h}_j * \mathbf{w}\|^2 \} = \|\mathbf{d}_j - \hat{\mathbf{h}}_j * \mathbf{w}\|^2 + \sum_{\nu=0}^{L_W-1} \sum_{r=0}^{L_W-1} w[\nu] w[r] \sum_{l=0}^{L_R+L_W-2} [\mathbf{S}_{h_j}]_{l-\nu, l-r}$$

By introducing a matrix $\hat{\mathbf{\Omega}}_j$ with ν, r -th element defined as

$$[\hat{\mathbf{\Omega}}_j]_{\nu, r} = \sum_{l=0}^{L_W+L_R-2} \begin{cases} [\hat{\mathbf{S}}_{h_j}]_{l-\nu, l-r} & 0 \leq l-\nu, \\ & l-r \leq L_R-1 \\ 0 & \text{otherwise} \end{cases} \quad (48)$$

we finally obtain

$$\log \tilde{p}(\mathbf{w}) \propto^e - \frac{1}{2} \hat{\lambda}_w \|\mathbf{w}\|^2 - \sum_{j \in \mathcal{J}} \frac{\hat{\tau}_j}{2} \|\mathbf{d}_j - \hat{\mathbf{h}}_j * \mathbf{w}\|^2 - \sum_{j \in \mathcal{J}} \frac{\hat{\tau}_j}{2} \mathbf{w}^T \hat{\mathbf{\Omega}}_j \mathbf{w}. \quad (49)$$

B. VARIATION LOWER BOUND WITH RESPECT TO

$q(\lambda_w)$

For estimation of $q(\lambda_w)$ we evaluate (27). After inserting (6) into (44) and retaining the terms depending on λ_w we obtain

$$\log \tilde{p}(\mathbf{w}) \propto^e \frac{L_W}{2} \log \lambda_w - \frac{1}{2} \lambda_w \mathbb{E}_{q(\mathbf{w})} \{ \|\mathbf{w}\|^2 \} + (c-1) \log(\lambda_w) - d \lambda_w \quad (50)$$

By noting that $\mathbb{E}_{q(\mathbf{w})} \{ \|\mathbf{w}\|^2 \} = \|\hat{\mathbf{w}}\|^2$ we obtain

$$\log \tilde{p}(\mathbf{w}) \propto^e \left(\frac{L_W}{2} + c - 1 \right) \log \lambda_w - \left(\frac{1}{2} \|\hat{\mathbf{w}}\|^2 + d \right) \lambda_w$$

REFERENCES

- [1] Hua-Wei Zhou, Practical Seismic Data Analysis, Cambridge University Press, 2014.
- [2] Weijia Sun, Liang Zhao, Yong Wei, and Li-Yun Fu, "Detection of seismic events on Mars: a lunar perspective," Earth and Planetary Physics, vol. 3, no. 4, pp. 1–8, 2019.
- [3] Tom E. Bishop, S. Derin Babacan, Bruno Amizic, Aggelos K. Katsaggelos, Tony Chan, and Rafael Molina, Blind Image Deconvolution: Theory and Applications, chapter Blind image deconvolution: Problem formulation and existing approaches, pp. 1–42, CRC Press, 2007.
- [4] Kenji Nose-Filho, Andre Kazuo Takahata, Renato Lopes, and Joao Marcos Travassos Romano, "Improving sparse multichannel blind deconvolution with correlated seismic data: Foundations and further results," IEEE Signal Process. Mag., vol. 35, no. 2, pp. 41–50, 2018.
- [5] Naveed Iqbal, Entao Liu, James H. McClellan, and Abdullatif A. Al-Shuhail, "Sparse multichannel blind deconvolution of seismic data via spectral projected-gradient," IEEE Access, vol. 7, pp. 23740–23751, 2019.
- [6] Pawan Bharadwaj, Laurent Demanet, and Aime Fournier, "Focused Blind Deconvolution," IEEE Trans. Signal Process., vol. 67, no. 12, pp. 3168–3180, 2019.
- [7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY, 2 edition, Aug. 2009.
- [8] Michael Tipping, "Sparse Bayesian learning and the relevance vector machine," J. Machine Learning Res., vol. 1, pp. 211–244, June 2001.
- [9] R. Giri and B. Rao, "Type I and Type II Bayesian methods for sparse signal recovery using scale mixtures," IEEE Trans. Signal Process., vol. 64, no. 13, pp. 3418–3428, July 2016.
- [10] Bruno Amizic, Rafael Molina, and Aggelos K. Katsaggelos, "Sparse Bayesian blind image deconvolution with parameter estimation," EURASIP Journal on Image and Video Processing, vol. 2012, pp. 1–15, 2012.
- [11] David Wipf and Haichao Zhang, "Revisiting Bayesian blind deconvolution," J. of Machine Learning Res., vol. 15, pp. 3595–3634, 2015.
- [12] Ming Ma, Shangxu Wang, Sanyi Yuan, Jingjing Wang, and Junxiang Wen, "Multichannel spatially correlated reflectivity inversion using block sparse Bayesian learning," Geophysics, vol. 82, no. 4, pp. V191–V199, 2017.
- [13] Ban-Sok Shin and Dmitriy Shutin, "Distributed blind deconvolution of seismic signals under sparsity constraints in sensor networks," in 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), 2020, pp. 1–6.
- [14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends in Machine Learning, vol. 3, no. 1, pp. 1–122, 2010.
- [15] Shengheng Liu, Hancong Wu, Yongming Huang, Yunjie Yang, and Jiabin Jia, "Accelerated structure-aware sparse bayesian learning for three-dimensional electrical impedance tomography," IEEE Trans. Ind. Informat., vol. 15, no. 9, pp. 5033–5041, 2019.
- [16] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
- [17] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in Proc. 9th Int. Workshop Artificial Intelligence and Statistics, Key West, FL, USA, Jan. 2003.
- [18] Dmitriy Shutin, Thomas Buchgraber, Sanjeev R. Kulkarni, and H. Vincent Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," IEEE Trans. Signal Process., vol. 59, no. 12, pp. 6257–6261, Dec. 2011.

- [19] Ali H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.
- [20] Robert Schlaifer Howard Raiffa, *Applied Statistical Decision Theory*, Harvard University Graduate School of Business Administration (Division of Research). Bailey & Swinfen, 1961.
- [21] Ralph A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–35, 1978.
- [22] Ali Gholami and Mauricio D. Sacchi, "A fast and automatic sparse deconvolution in the presence of outliers," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 4105–4116, 2012.
- [23] George E.P. Box and George C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, Ltd, 1992.
- [24] Ban-Sok Shin, Henning Paul, and Armin Dekorsy, "Distributed kernel least squares for nonlinear regression applied to sensor networks," in 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 1588–1592.
- [25] Harold Ryan, "Ricker, ormsby, klander, butterworth – a choice of wavelets," *Hi-Res Geoconsulting*, vol. 19, no. 7, Sept. 1994.
- [26] Ofir Pele and Michael Werman, "Fast and robust earth mover's distances," in 2009 IEEE 12th International Conference on Computer Vision. IEEE, September 2009, pp. 460–467.
- [27] A. Lavrenko, F. Römer, G. Del Galdo, and R. Thoma, "On the earth mover's distance as a performance metric for sparse support recovery," in *IEEE GlobalSIP*, 2016, pp. 1368–1372.
- [28] C. L. Mallows, "A Note on Asymptotic Joint Normality," *The Annals of Mathematical Statistics*, vol. 43, no. 2, pp. 508 – 515, 1972.
- [29] R.W. Clayton and B. Engquist, "Absorbing boundary conditions for acoustic and elastic wave equations," *Bull. Seis. Soc. Amer.*, vol. 67, pp. 1529–1540, 1977.



BAN-SOK SHIN received the Dipl.-Ing. (M.Sc.) and Ph.D. degree (Dr.-Ing.) from the University of Bremen, Bremen, Germany, in 2013 and 2020, respectively. From 2013 till 2019 he was a Research Assistant with the Department of Communications Engineering, University of Bremen, where he was working towards his Ph.D. degree. From 2019 he is with the Institute of Communications and Navigation, German Aerospace Center, where he is currently a member of the Swarm Exploration

Group working on seismic subsurface exploration using a multi-agent network. His research interests include wave-based imaging, distributed inference/estimation, machine learning, and their application to sensor and multi-agent networks.

In 2021, Dr. Shin was awarded with the OHB Award for the best Ph.D. dissertation in the faculty of Electrical Engineering of the University of Bremen.

...



DMITRIY SHUTIN (S'02–M'07–SM'12) received his Master degree in Computer Science in 2000 from Dniepropetrovsk State University, Ukraine, and the Doctoral degree in Electrical Engineering from Graz University of Technology, Graz, Austria, in 2006. During 2001-2006 and 2006-2009, he was a Teaching Assistant and an Assistant Professor, respectively, with the Signal Processing and Speech Communication Laboratory at Graz University of Technology. In 2009 he joined Department of Electrical Engineering at Princeton University, where until 2011 he worked as a Research Associate. In 2011 he joined the Institute of Communications and Navigation, German Aerospace Center, where he is currently a leader of the Swarm Exploration Group. His research interests include machine learning for signal processing, distributed information and signal processing and swarm exploration.

Dr. Shutin was a recipient of the Best Student Paper Award at the 2005 IEEE International Conference on Information, Communications and Signal Processing (ICICSP). In 2009 he was awarded the Erwin Schroedinger Research Fellowship. Since 2012 till 2014 he acted as selected advisor of German air navigation service provider, within the Navigational System Panel of ICAO.