Incremental Sparse Bayesian Learning for Parameter Estimation of Superimposed Signals

Dmitriy Shutin, Wei Wang, Thomas Jost

Abstract—This work discuses a novel algorithm for joint sparse estimation of superimposed signals and their parameters. The proposed method is based on two concepts: a variational Bayesian version of the incremental sparse Bayesian learning (SBL)– fast variational SBL – and a variational Bayesian approach for parameter estimation of superimposed signal models. Both schemes estimate the unknown parameters by minimizing the variational lower bound on model evidence; also, these optimizations are performed incrementally with respect to the parameters of a single component. It is demonstrated that these estimations can be naturally unified under the framework of variational Bayesian inference. It allows, on the one hand, for an adaptive dictionary design for FV-SBL schemes, and, on the other hand, for a fast superresolution approach for parameter estimation of superimposed signals. The experimental evidence collected with synthetic data as well as with estimation results for measured multipath channels demonstrate the effectiveness of the proposed algorithm.

I. Introduction

In this paper our goal is to estimate the parameters of the following model

$$\boldsymbol{y} = \sum_{l=1}^{L} \boldsymbol{s}(\boldsymbol{\theta}_l) w_l + \boldsymbol{\xi} = \boldsymbol{S}(\boldsymbol{\Theta})\boldsymbol{w} + \boldsymbol{\xi}, \qquad (1)$$

where $\boldsymbol{y}$ is an N-dimensional signal vector, $\boldsymbol{s}(\boldsymbol{\theta}_l)$, $l = 1, \ldots, L$, is a set $\boldsymbol{S}(\boldsymbol{\Theta}) = [\boldsymbol{s}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{s}(\boldsymbol{\theta}_L)]$ of known basis functions that are nonlinearly parameterized by $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L]$; $\boldsymbol{w} = [w_1, \ldots, w_L]^T$ is a vector of basis weights, and $\boldsymbol{\xi}$ is a random perturbation vector, which is often assumed to follow a circular symmetric normal distribution with zero mean and covariance $\boldsymbol{\Sigma} = \lambda^{-1}\boldsymbol{I}$. Such model is almost ubiquitous in signal processing literature, and appears under different disguises in almost all fields of signal processing, e.g., in array processing, channel estimation, radar, to name just a few.

The estimation of signal parameters $\boldsymbol{\Theta}$ and $\boldsymbol{w}$ has often been solved using Expectation-Maximization (EM) type of algorithms [1]–[3], mainly due to the nonlinearity of (1) with respect to the parameter set $\boldsymbol{\Theta}$. Yet these methods are applicable only when the order L of the model is known and fixed – a requirement that is rarely satisfied in practice. However, introducing sparsity constraints into the parameter estimation step might eliminate this drawback of the EM-based estimation.

Sparse signal processing methods have become a very active area of research in recent years due to their rich theoretical nature and their usefulness in a wide range of applications (see e.g., [4]–[6]). With a few minor variations, the general goal of sparse reconstruction is to optimally estimate the parameters $\boldsymbol{w}$ of the model (1) with fixed design matrix $\boldsymbol{S}(\boldsymbol{\Theta}) \equiv [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L]$. The sparse solution is obtained by imposing specific sparsity constraints on the signal parameter $\boldsymbol{w}$ [4], [6].

Sparse Bayesian learning (SBL) [5], [7], [8] is a family of empirical Bayes techniques that finds a sparse estimate of $\boldsymbol{w}$ by modeling the weights using a hierarchical prior $p(\boldsymbol{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) = \prod_{l=1}^{L} p(w_l|\alpha_l)p(\alpha_l)$, where $p(w_l|\alpha_l)$ is a Gaussian probability density function (pdf) with zero mean and precision parameter α_l, also called the sparsity parameter; larger values of α_l drive the corresponding weight toward zero, thus encouraging a sparse solution. One particular method for SBL recently proposed in the literature is a fast variational SBL (FV-SBL) [8]. The FV-SBL algorithm optimizes the corresponding objective function – the variational lower bound on the model evidence $\log p(\boldsymbol{y})$ – incrementally, i.e., with respect to one basis function at a time. This allows for a very efficient and adaptive implementation of FV-SBL [9] — a feature that is very useful for estimating superimposed signals. Yet due to the nonlinear dependence of (1) on the parameter set $\boldsymbol{\Theta}$, the classical sparse estimation techniques are inapplicable. Obviously, an appropriate sampling or gridding of the parameters $\boldsymbol{\Theta}$ [10], [11] circumvents the nonlinearity problem. This approach, however, does not provide high resolution estimates of the parameters; alternatively, heuristics have to be used to make the gridding adaptive.

Our goal in this paper is to show how SBL technique can be applied to (1) to enable joint sparse signal extraction and superresolution parameter estimation. The proposed technique builds upon two key concepts: variational Bayesian estimation of signal parameters $\boldsymbol{\Theta}$, and an incremental FV-SBL algorithm [8]. Through the use of variational Bayesian techniques both schemes can be jointly realized within the same optimization framework. The first attempts to do so have been proposed in [12], where the authors make a typical assumption on the independence of individual components in (1). Our empirical evidence suggest that this assumption is overly optimistic. The new algorithm is based on the FV-SBL scheme. This allows taking correlations between the linear parameters of the superimposed signals into account. Additionally, the FV-SBL algorithm allows for an adaptive implementation [9], which further accelerates the inference.

Throughout the paper we make use of the following notation. Vectors and matrices are represented as, respectively, boldface lowercase letters, e.g., $\boldsymbol{x}$, and boldface uppercase letters, e.g., $\boldsymbol{X}$. The expression $[\boldsymbol{B}]_{\overline{lk}}$ denotes a matrix ob-

tained by deleting the lth row and kth column from the matrix $\boldsymbol{B}$; similarly, $[\boldsymbol{b}]_{\bar{l}}$ denotes a vector obtained by deleting the lth element from the vector $\boldsymbol{b}$. With a slight abuse of notation we will sometimes refer to a matrix as a set of column vectors; for instance we write $\boldsymbol{a} \in \boldsymbol{X}$ to imply that $\boldsymbol{a}$ is a column in $\boldsymbol{X}$, and $\boldsymbol{X} \setminus \boldsymbol{a}$ to denote a matrix obtained by deleting the column vector $\boldsymbol{a} \in \boldsymbol{X}$. We use $\boldsymbol{e}_l = [0_1, \ldots, 0_{l-1}, 1_l, 0_{l+1}, \ldots, 0_L]^T$ to denote a canonical vector of appropriate dimension. Finally, for a random vector $\boldsymbol{x}$, $\mathrm{CN}(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{B})$ denotes a circular symmetric normal distribution pdf with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{B}$; similarly, for a random variable x, $\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$ denotes a gamma pdf with parameters a and b.

II. SIGNAL MODEL AND ADAPTIVE FAST SPARSE BAYESIAN LEARNING

In Fig. 1 we show the graphical model that captures the dependencies between the parameters of (1). According to the
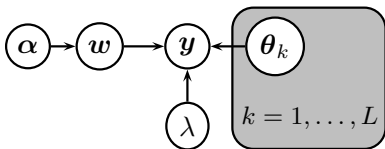

Fig. 1. Graphical model representing (1) with L components.

graph structure, the joint pdf of the graph variables can be factored as

$$p(\boldsymbol{w}, \lambda, \boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{w}, \lambda, \boldsymbol{\theta})p(\boldsymbol{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\lambda)p(\boldsymbol{\Theta}), \quad (2)$$

where $p(\boldsymbol{y}|\boldsymbol{w}, \lambda, \boldsymbol{\theta}) = \mathrm{CN}(\boldsymbol{y}|\boldsymbol{S}(\boldsymbol{\Theta})\boldsymbol{w}, \lambda^{-1}\boldsymbol{I})$, $p(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{l=1}^{L} \mathrm{CN}(w_l|0, \alpha_l^{-1})$, $p(\boldsymbol{\alpha}) \propto \prod_{l=1}^{L} \alpha_l^{-1}$, and $p(\lambda) \propto \lambda^{-1}$, following the standard SBL model assumption [8], [9].[1] The choice of the prior $p(\boldsymbol{\Theta})$ is arbitrary in the context of this work and is generally application specific. The variational inference on this graph aims at estimating a "proxy" pdf $q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta})$ that maximizes the lower bound on the log-evidence $\log p(\boldsymbol{y})$ [13]:

$$\log p(\boldsymbol{y}) \geq \mathop{\mathbb{E}}_{q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta})} \log \frac{p(\boldsymbol{w}, \lambda, \boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{y})}{q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta})} \quad (3)$$

We will assume that $q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta})$ factors as follows

$$q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta}) = q(\boldsymbol{w})q(\lambda)\prod_{l=1}^{L} q(\alpha_l)q(\boldsymbol{\theta}_l), \quad (4)$$

with the variational factors in (4) constrained as: $q(\boldsymbol{w}) = \mathrm{CN}(\boldsymbol{w}|\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{\Phi}})$, $q(\alpha_l) = \mathrm{Ga}(\alpha_l|1, \widehat{\alpha}_l^{-1})$, and $q(\lambda) = \mathrm{Ga}(\lambda|N/2, N\widehat{\lambda}^{-1}/2)$. In case of parameters $\boldsymbol{\Theta}$ we assume $q(\boldsymbol{\theta}_l) = \delta(\boldsymbol{\theta}_l - \widehat{\boldsymbol{\theta}}_l)$.[2] By doing so we restrict ourselves to point

---

[1] In the following we will consider complex measurement data; extensions for real case are trivial. Also, we will use non-informative form of prior $p(\lambda)$ and $p(\alpha_l)$, $\forall l$. This is known as SBL with automatic relevance determination [7].

[2] More complex forms of $q(\boldsymbol{\theta}_l)$ are outside the scope of this paper.

estimates[3] of these parameters. The optimal $q(\boldsymbol{w}, \boldsymbol{\alpha}, \lambda, \boldsymbol{\Theta})$ is then found by maximizing (3) with respect to the parameters $\{\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{\Phi}}, \widehat{\lambda}, \widehat{\alpha}_1, \widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\alpha}_L, \widehat{\boldsymbol{\theta}}_l\}$ by cycling through all factors in a "round-robin" fashion [13].

Should the parameters $\boldsymbol{\Theta}$ be assumed as known and fixed, i.e., $\widehat{\boldsymbol{S}} \equiv \boldsymbol{S}(\widehat{\boldsymbol{\Theta}})$, update expressions for the variational parameters can be easily found [14]:

$$\widehat{\boldsymbol{\Phi}} = \left(\widehat{\lambda}\boldsymbol{S}(\widehat{\boldsymbol{\Theta}})^H \boldsymbol{S}(\widehat{\boldsymbol{\Theta}}) + \mathrm{diag}(\widehat{\boldsymbol{\alpha}})\right)^{-1}, \widehat{\boldsymbol{w}} = \widehat{\lambda}\widehat{\boldsymbol{\Phi}}\boldsymbol{S}(\widehat{\boldsymbol{\Theta}})^H \boldsymbol{y}, \quad (5)$$

$$\widehat{\alpha}_l = \frac{1}{|\widehat{w}_l|^2 + \widehat{\Phi}_{ll}}, \widehat{\lambda} = \frac{N}{\|\boldsymbol{t} - \widehat{\boldsymbol{S}}\widehat{\boldsymbol{w}}\|^2 + \mathrm{Trace}(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{S}}^H \widehat{\boldsymbol{S}})}, \quad (6)$$

where $\widehat{w}_l$ is the lth element of the vector $\widehat{\boldsymbol{w}}$, and $\widehat{\Phi}_{ll}$ is the lth element on the main diagonal of the matrix $\widehat{\boldsymbol{\Phi}}$.

The FV-SBL algorithm is a computationally efficient method to accelerate the convergence of the inference expressions (5) and (6). Essentially, it maximizes the bound (3) incrementally: the variational updates of $q(\alpha_l)$ and $q(\boldsymbol{w})$ for a fixed l are performed successively *ad infinitum* while keeping the other variational factors fixed. The convergence of $q(\alpha_l)$ can then be established analytically, which allows for a significant speed-up [8]; moreover, FV-SBL allows for an adaptive implementation, where basis functions can also be easily added to the model (for more information on the adaptive FV-SBL algorithm the reader is referred to [9]).

One of the key features of variational methods is that the factors in (4) can be updated in any order.[4] This allows incorporating the estimation of $q(\boldsymbol{\Theta})$ in the FV-SBL scheme, as explained in the following.

III. ESTIMATION OF SIGNAL PARAMETERS $\boldsymbol{\Theta}$

Let us begin by considering a variational inference of $q(\boldsymbol{\Theta})$. To this end we define $\boldsymbol{\Theta}_{\bar{l}} = \boldsymbol{\Theta} \setminus \boldsymbol{\theta}_l$. Following the standard variation inference steps (see [13]), it can be shown that the bound on $\log p(\boldsymbol{y})$ with respect to $q(\boldsymbol{\theta}_l)$ can be expressed as $\log p(\boldsymbol{y}) \geq \mathbb{E}_{q(\boldsymbol{\theta}_l)} \log \frac{\tilde{p}(\boldsymbol{\theta}_l)}{q(\boldsymbol{\theta}_l)}$, where $\tilde{p}(\boldsymbol{\theta}_l) \propto \exp\left(\mathbb{E}_{q(\boldsymbol{w}, \lambda, \boldsymbol{\Theta}_{\bar{l}})} \log p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{\Theta}, \lambda)p(\boldsymbol{\Theta})\right)$. This bound is maximized when the Kullback-Leibler divergence between $q(\boldsymbol{\theta}_l)$ and $\tilde{p}(\boldsymbol{\theta}_l)$ is minimal. Since $q(\boldsymbol{\theta}_l)$ is constrained to be a Dirac distribution, the minimum divergence is achieved when $q(\boldsymbol{\theta}_l)$ is aligned with the mode of $\tilde{p}(\boldsymbol{\theta}_l)$. By evaluating $\tilde{p}(\boldsymbol{\theta}_l)$ we find $\widehat{\boldsymbol{\theta}}_l$ as

$$\widehat{\boldsymbol{\theta}}_l = \operatorname*{argmax}_{\boldsymbol{\theta}_l} \Big\{ \log p(\boldsymbol{\theta}_l) - \widehat{\lambda}\|\boldsymbol{r}_l - \widehat{w}_l \boldsymbol{s}(\boldsymbol{\theta}_l)\|^2 - \widehat{\lambda}\sum_{k \neq l} 2\Re\left\{\Phi_{kl}\boldsymbol{s}(\widehat{\boldsymbol{\theta}}_k)^H \boldsymbol{s}(\boldsymbol{\theta}_l)\right\} - \widehat{\lambda}\Phi_{ll}\|\boldsymbol{s}(\boldsymbol{\theta}_l)\|^2 \Big\}, \quad (7)$$

---

[3] As a point estimate we understand maximum likelihood or maximum *a posteriori* estimation; the latter case is automatically obtained when a prior $p(\boldsymbol{\theta}_l) \neq \mathrm{const}$.

[4] Note, however, that the order in which the factors are updated is important since different update orderings might lead to different local optima of the variational lower bound.

where

$$r_l = y - \sum_{k=1, k \neq l}^{L} \widehat{w}_k s(\widehat{\theta}_k), \qquad (8)$$

and $\Re\{\cdot\}$ denotes the real part operator. Finding $\widehat{\theta}_l$ from (7), which requires nonlinear optimization, readily gives the optimal pdf $q(\theta_l)$. Note that the last two terms in (7) account for the correlations between the weights w of the components, effectively penalizing the estimator for θ_l. We are now ready to bring all the pieces of the proposed sparse estimation scheme together.

The proposed algorithm updates the factors in (4) in groups, where an lth group contains factors $\{q(\theta_l), q(\alpha_l), q(w)\}$: starting with $q(\theta_l)$, we then update $q(\alpha_l)$ and $q(w)$ using the FV-SBL scheme. If the estimate of $\widehat{\alpha}_l$ diverges, the corresponding component is removed from the model; otherwise, its parameters are updated, and the next component is considered. The realization of the algorithm includes two steps: the initialization and update which are sequentially carried out and summarized in Algorithms 1 and 2, respectively. Note that

---

Algorithm 1 Initialization

$L \leftarrow 0$, $S(\Theta) \leftarrow []$, $\widehat{\Phi} \leftarrow []$, $\widehat{\alpha} \leftarrow []$, Continue $\leftarrow$ true
while Continue **do**
   Compute r_{L+1} from (8) and $q(\theta_{L+1})$ from (7)
   $\overline{s} \leftarrow s(\widehat{\theta}_{L+1})$
   $\varsigma \leftarrow (\widehat{\lambda}\overline{s}^H \overline{s} - \widehat{\lambda}^2 \overline{s}^H S(\Theta) \widehat{\Phi} S(\Theta)^H \overline{s})^{-1}$
   $\omega^2 \leftarrow \varsigma^2 (\widehat{\lambda}\overline{s}^H y - \widehat{\lambda}^2 \overline{s}^H S(\Theta) \widehat{\Phi}(\Theta)^H y)^2$
   if $\omega^2 > \varsigma$ **then**
     Add a new component $s(\widehat{\theta}_{L+1})$
     Update $q(\alpha_{L+1})$: $\widehat{\alpha}_{L+1} \leftarrow (\omega^2 - \varsigma)^{-1}$
     Update $q(w)$ *using a new basis* $\overline{s}$:

$$X_{L+1} = \widehat{\Phi}^{-1} - \frac{\widehat{\lambda} S(\Theta)^H \overline{s}\,\overline{s}^H S(\Theta)}{\widehat{\alpha}_{L+1} + \overline{s}^H \overline{s}}$$

$$\widehat{\Phi}_{L+1} = \begin{pmatrix} X_{L+1}^{-1} & -\widehat{\lambda}\dfrac{\widehat{\Phi} S(\Theta)^H \overline{s}}{\widehat{\alpha}_{L+1} + \varsigma^{-1}} \\ -\widehat{\lambda}\dfrac{\overline{s}^H S(\Theta)\widehat{\Phi}}{\widehat{\alpha}_{L+1} + \varsigma^{-1}} & (\widehat{\alpha}_{L+1} + \varsigma^{-1})^{-1} \end{pmatrix} \qquad (9)$$

     $S(\Theta) \leftarrow [S(\Theta), s(\widehat{\theta}_{L+1})]$,
     $\widehat{w}_{L+1} \leftarrow \widehat{\lambda}\widehat{\Phi}_{L+1} S(\Theta)^H y$,
     $L \leftarrow L + 1$
   else
     Reject $s(\widehat{\theta}_{L+1})$; Continue $=$ False
   end if
end while

---

the inverse of a Schur complement X_{L+1} in the Algorithm 1 can be computed efficiently using a rank-one update [15]. The variables ω and ς and the test $\omega^2 > \varsigma$ are explained in detail in [8], [9]. Let us point out that the sparsity inducing property of the whole scheme is "encoded" in the test $\omega^2 > \varsigma$ that determines the convergence of $q(\alpha_l)$ update: if the mean of $q(\alpha_l)$ diverges, the component is removed from the model.

Algorithm 2 Update

while Continue **do**
   Compute r_l from (8) and $q(\theta_l)$ from (7)
   $\overline{s} \leftarrow s(\widehat{\theta}_l)$
   $\overline{S}_{\bar{l}} \leftarrow S(\Theta) \setminus \overline{s}_l$, $\widehat{\Phi}_{\bar{l}} = \left[\widehat{\Phi} - \dfrac{\widehat{\Phi} e_l e_l^H \widehat{\Phi}}{e_l^H \widehat{\Phi} e_l}\right]_{\bar{l}\bar{l}}$
   $\varsigma \leftarrow (\widehat{\lambda}\overline{s}_l^H \overline{s}_l - \widehat{\lambda}^2 \overline{s}_l^H \overline{S}_{\bar{l}} \widehat{\Phi}_{\bar{l}} \overline{S}_{\bar{l}}^H \overline{s}_l)^{-1}$
   $\omega^2 \leftarrow (\widehat{\lambda}\varsigma_l \overline{s}_l^H y - \widehat{\lambda}^2 \varsigma_l \overline{s}_l^H \overline{S}_{\bar{l}} \widehat{\Phi}_{\bar{l}} \overline{S}_{\bar{l}}^H y)^2$
   if $\omega^2 > \varsigma$ **then**
     $S(\Theta) \leftarrow [\overline{S}_{\bar{l}}, \overline{s}]$
     Update $q(\alpha_l)$: $\widehat{\alpha}_l \leftarrow (\omega^2 - \varsigma)^{-1}$;
     Update $q(w)$ *using* $\overline{s}$ *and* $\widehat{\alpha}_l$
     Compute $\widehat{\Phi}$ as in (9), $\widehat{w} \leftarrow \widehat{\lambda}\widehat{\Phi} S(\Theta)^H y$
   else
     Remove the component $\overline{s}_l$
     $S(\Theta) \leftarrow \overline{S}_{\bar{l}}$; $L \leftarrow L - 1$
     Update $q(\alpha)$: $\widehat{\alpha}_{\bar{l}} \leftarrow [\widehat{\alpha}]_{\bar{l}}$
     Update $q(w)$:$\widehat{\Phi} \leftarrow \widehat{\Phi}_{\bar{l}}$, $\widehat{w} \leftarrow \widehat{\lambda}\widehat{\Phi} S(\Theta)^H y$
   end if
end while

IV. SIMULATION RESULTS

Here we study the performance of the proposed estimation scheme using synthetic data generated according to model (1) as well measured data.

For simplicity we consider a Single-Input-Single-Output channel with zero Doppler shift; thus, each component is characterized by a delay $\theta_l = \{\tau_l\}$ and a complex gain w_l, i.e., $y = \sum_{l=1}^{L} w_l s(\tau_l) + \xi$. The channel is synthesized in frequency domain with the following parameters: $L = 4$, $N = 1537$, signal bandwidth is $f_B = 120\,\text{MHz}$; the signal was sampled at the Nyquist rate and the carrier frequency is assumed to be $5.2\,\text{GHz}$. The delays of synthetic multipath components are set to $17.5\,\text{ns}$, $40.83\,\text{ns}$, $59.33\,\text{ns}$, and $91.67\,\text{ns}$; corresponding complex amplitudes are selected as $w_l = e^{j\varphi_l}$, $l = 1, \ldots, 4$, where φ_l is a random variable drawn from a uniform distribution. As as replica of the transmitted signal $s(t)$ we use the actual measured calibration data of the Medav RUSK-DLR channel sounder [16]. The calibration data is obtained by directly connecting the transmitter to the receiver and recording the received signal. Its sampled version is then used to construct a vector $s(\cdot)$, whose shifted versions are used in synthesizing the channel, as well as in the estimation step.

In Fig.2 we show the estimated impulse response and transfer function for 15dB SNR. Observe that the estimated responses closely follow the measured data with only four components. Let us stress that depending on the actual noise realization, the algorithm tends to overestimate the number of components. In Fig. 3(a) we plot distributions of estimated sparsity parameters for all detected components collected over 1000 Monte Carlo runs with different noise realizations. Note that in the worst case the algorithm identifies up to 8 components, all of which are very close to the true ones. This
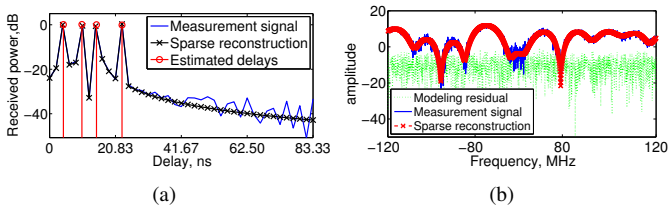
Fig. 2.   Estimated synthetic channel in a) time domain and b) frequency domain for 15dB SNR.
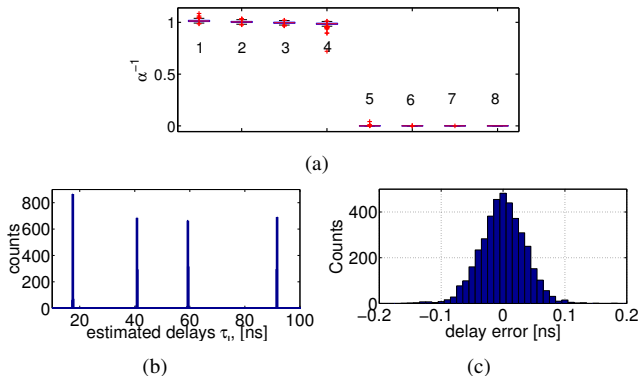

Fig. 4.   Estimated channel response using a) proposed algorithm and b) using SAGE algorithm.

Fig. 3.   a)Estimated sparsity parameters $\widehat{\alpha}$ and b) delays for synthetic scenarios. c) Error distribution for estimated τ_l when 4 components are found.

can also be seen in Fig. 3(b), where we plot the distribution of all estimated delays. The inverse sparsity parameters of these artifact components are also quite small, which means they do not contribute to the model. In the case when the algorithm identifies exactly 4 components we can compute the error between the true and the estimated delay. In Fig 3(c) we plot the histogram of estimated delay errors. Note that the estimation error is smaller than 1% of the used sampling period (≈ 8.3ns).

A. Estimation results for measured multipath channels.

Here we consider the estimation of the actual measured multipath channels using the proposed algorithm. The data was collected during a recent measurement campaign [16] performed at German Aerospace Center in Oberpfaffenhofen, Germany. The measurements parameters coincide with those used in simulations. As the actual channel parameters cannot be known for a measured channel, we qualitatively compare the performance of the proposed scheme to that of the SAGE algorithm [3]. As the latter scheme requires knowing the number of components L, we first estimate it using the proposed method, and then use SAGE with same model order. The estimation results are summarized in Fig. In total $L = 31$ path has been identified. Despite some similarities, the SAGE algorithm tends to miss weak components. Also, it tends to cluster multipaths around areas of high power, which often indicates estimation artifacts [12].

V. Conclusion

In this work an adaptive fast variational Sparse Bayesian Learning (FV-SBL) algorithm has been used for parameter es-
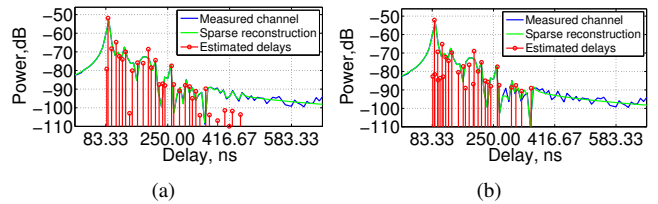
timation of superimposed signals. Using variational framework both superresolution parameter estimation and sparse signal extraction can be done jointly by minimizing the common objective function. Thus, the proposed scheme "frees" the classical EM-based parameter estimation from specifying a model order. Simulation results obtained with synthetic and measured data demonstrate the effectiveness of the proposed estimation scheme. However, more detailed analysis of experimental data is needed.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Mag.*, pp. 67–94, July 1996.

[2] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Trans. on Acoustics, Speech, and Sig. Proc.*, vol. 36, no. 4, pp. 477–489, April 1988.

[3] B. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal on Sel. Areas in Comm.*, vol. 17, no. 3, pp. 434–450, March 1999.

[4] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[5] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, November 2008.

[6] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, april 2006.

[7] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Res.*, vol. 1, pp. 211–244, June 2001.

[8] D. Shutin, T. Buchgraber, S. R. Kulkarni, and H. V. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *IEEE Trans. on Sig. Proc.*, vol. 59, no. 12, pp. 6257 –6261, Dec. 2011.

[9] ——, "Fast adaptive variational sparse Bayesian learning with automatic relevance determination," in *IEEE Int. Conf. on Acoustics Speech and Sig. Proc.*, Prague, Czech Republic, May 2011, pp. 2180–2183.

[10] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[11] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. on Sign. Proc.*, vol. 53, no. 8, pp. 3010–3022, 2005.

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

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[14] C. M. Bishop and M. E. Tipping, "Variational relevance vector machines," in *Proc. 16th Conf. Uncer. in Artif. Intell.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 46–53.

[15] W. W. Hager, "Updating the inverse of a matrix," *SIAM Review*, vol. 31, no. 2, pp. pp. 221–239, 1989.

[16] W. Wang and T. Jost, "A low-cost platform for time-variant wireless channel measurements with application to positioning," *IEEE Trans. on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1597 –1604, june 2012.