Sequential Bayesian Sparse Signal Reconstruction Using Array Data

Christoph F. Mecklenbräuker, *Senior Member, IEEE*, Peter Gerstoft, *Member, IEEE*,
Ashkan Panahi, *Student Member, IEEE*, and Mats Viberg, *Fellow, IEEE*

Abstract—In this paper, the sequential reconstruction of source waveforms under a sparsity constraint is considered from a Bayesian perspective. Let the wave field, which is observed by a sensor array, be caused by a spatially-sparse set of sources. A spatially weighted Laplace-like prior is assumed for the source field and the corresponding weighted Least Absolute Shrinkage and Selection Operator (LASSO) cost function is derived. After the weighted LASSO solution has been calculated as the maximum a posteriori estimate at time step k, the posterior distribution of the source amplitudes is analytically approximated. The weighting of the Laplace-like prior for time step $k + 1$ is then fitted to the approximated posterior distribution. This results in a sequential update for the LASSO weights. Thus, a sequence of weighted LASSO problems is solved for estimating the temporal evolution of a sparse source field. The method is evaluated numerically using a uniform linear array in simulations and applied to data which were acquired from a towed horizontal array during the long range acoustic communications experiment.

Index Terms—Bayesian estimation, sequential estimation, sparsity, weighted LASSO.

I. INTRODUCTION

COMPRESSIVE sensing and sparse reconstruction is widely used, but their sequential use have received little attention. In a Bayesian framework, the sparseness constraint is obtained using a Laplace-like prior which give rise to the ℓ_1 constraint. In the present contribution the focus is on developing a Bayesian sequential sparse approach by enforcing the prior at each step to be Laplace-like.

Since the late nineties, numerous papers have appeared in the areas *compressive sensing* and *sparse reconstruction* which employ the penalization by a ℓ_1-norm to promote sparsity of the solution, e.g., [1]–[6]. Excellent books on the topic are available [7], [8]. The *least absolute shrinkage and selection operator* (LASSO) is a least squares regression constrained by an ℓ_1-bound of the solution [1]. It is well-known that the LASSO sets many regression coefficients to zero favoring *sparse* solutions.

Research on compressive sensing and sparse reconstruction has focused largely on batch processing. However, many practical settings require the online estimation of sparse signals from noisy data samples that become available sequentially [9]–[11]. Angelosante *et al.* proposed time-recursive algorithms for the estimation of time-varying sparse signals from observations obeying a linear model, and become available sequentially in time [10]. Their time-weighted LASSO is an ℓ_1-norm regularized Recursive Least Squares algorithm. Eiwen *et al.* [11] proposed a sequential compressive channel estimator which is based on modified compressed sensing, which assumes that a part of the signal support is known a priori. The resulting compressive channel tracker uses a modified version of the orthogonal matching pursuit algorithm. An iterative pursuit algorithm is developed in [12] that uses sequential predictions for dynamic compressive sensing. It incorporates prior information using linear MMSE reconstruction.

Here, we extend the Bayesian approach [13]–[15] to *sequential* Maximum A Posteriori (MAP) estimation for sparse signal reconstruction in a problem setting that is similar to [10]. First the LASSO cost function is generalized by weighting the regularization parameters. This allows to update the prior adaptively based on the past history of observations. Based on ideas in [16], we approximate a sequential MAP filter which preserves sparsity in the source field estimates. Two types of approximations are derived in Sections III-D and III-E and their behavior is compared in numerical examples. It is implemented by sequentially minimizing adaptively weighted LASSO cost functions using a single new measurement snapshot in each time step.

The theory is formulated so that it can be applied to sparse source field estimation in higher spatial dimensions. We have earlier demonstrated the sequential approach for a two dimensional array, where an earthquake rupture process evolves sequentially with sparse locations in both time and two dimensional space [17]. High resolution earthquake location has been obtained using compressive sensing [18].

The advantages of the sparse formulation for source localization and direction of arrival (DOA) estimation are:
- Numerical simulations indicate that it is a high-resolution method [3].
- There is no need to estimate the cross-spectral density matrix. Only one snapshot is needed. Eigenvalue based high-

resolution methods require a full or nearly full Cross Spectral Density Matrix (CSDM).

- As new information becomes available, the confidence in the solution is build up and the trajectories are updated.
- The solution is presented in terms of a sparsifying Laplacian distribution which can then be used as a prior in the sequential processing. Relative to more general solutions like particle filters, the solution requires just one parameter per possible source location.

II. ARRAY DATA MODEL AND PROBLEM FORMULATION

Let $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M\}$ be a set of M *hypothetical* source locations on a given fixed grid. Further, let $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{kM})^T$ be the associated complex-valued source vector at time step $k = 1, 2, \ldots$ and frequency ω. We observe time-sampled waveforms on an array of N sensors which are stacked in a vector. We obtain the following linear model which relates the short-time Fourier transformed sensor array output $\boldsymbol{y}_k$ at time step k and at frequency ω to the source vector $\boldsymbol{x}_k$,

$$\boldsymbol{y}_k = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{n}_k. \tag{1}$$

The mth column of the transfer matrix $\boldsymbol{A}$ is the array steering vector $\boldsymbol{a}_m$ for hypothetical source location $\boldsymbol{r}_m$. The transfer matrix $\boldsymbol{A}$ is constructed by sampling all possible source locations, but only very few of these can correspond to real sources. Therefore, the dimension of $\boldsymbol{A}$ is $N \times M$ with $N \ll M$ and $\boldsymbol{x}_k$ is sparse. In our setting, the number of hypothetical source locations M is much larger than the number of sensors N, i.e., $N \ll M$ and the linear model (1) are underdetermined.

We model the nmth element of $\boldsymbol{A}$ by $\mathrm{e}^{-\mathrm{j}\omega\tau_{nm}}$. Here τ_{nm} is the traveltime from hypothetical location $\boldsymbol{r}_m$ to the nth sensor element of the sensor array. The additive noise vector $\boldsymbol{n}_k$ is assumed to be spatially uncorrelated and follows the zero-mean complex normal distribution with diagonal covariance matrix $\sigma^2\boldsymbol{I}$. Further, the noise is assumed to be uncorrelated across time and frequency.

The history of all previous array observations is summarized in $\boldsymbol{Y}_{k-1} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-2}, \boldsymbol{y}_{k-1})$. Given the history $\boldsymbol{Y}_{k-1}$ and the new data $\boldsymbol{y}_k$, we seek the maximum a posteriori (MAP) source estimate $\hat{\boldsymbol{x}}_k$ under a sparsity constraint. To reconstruct a physically meaningful source vector $\mathbf{x}_k$, we enforce its sparsity through the introduction of a prior probability density on $\boldsymbol{x}_k$ which promotes sparsity. A widely used sparseness prior is the Laplace probability density [14], [19] which puts a higher probability mass than the normal distribution both for *large* and *small* absolute deviations.

The prior distribution for each x_{km} is assumed to follow a Laplace distribution with hyperparameter λ_{km}, condensed in vector $\boldsymbol{\lambda}_k = (\lambda_{k1}, \ldots, \lambda_{kM})^T$. A time-recursive expression of the sparse MAP source estimate at step k is obtained using a sequential Bayesian sparse update rule

$$(\hat{\boldsymbol{x}}_k, \boldsymbol{\lambda}_{k+1}) = \boldsymbol{\varphi}(\boldsymbol{y}_k, \boldsymbol{\lambda}_k). \tag{2}$$

Where the function $\boldsymbol{\varphi}$ will be derived in the following. In addition to $\hat{\boldsymbol{x}}_k$, the hyperparameter $\boldsymbol{\lambda}_{k+1}$ for the prior distribution for the next time step $k + 1$ is computed. Due to the sparsity constraint it is necessary to let the hyperparameters carry the

sequential information from each step, and not the extracted parameters as is customary in sequential filtering [20].

Since the conditional distribution for $\boldsymbol{y}_k$ given $\boldsymbol{x}_k$ follows a Gaussian (3) it is not given that the posterior will be Laplace-like. As the posterior for time step k is used as a prior in the next time step $k + 1$ it is required to approximate the posterior with a Laplace-like distribution. The hyperparameter $\boldsymbol{\lambda}_{k+1}$ is determined by matching either the probabilities for large source amplitudes (i.e., the tail of the distribution, see Section III-D) or the expected source amplitudes (Section III-E).

III. BAYESIAN FORMULATION

For the linear model (1), we arrive at the following conditional probability density for the single-frequency observations given the source vector $\boldsymbol{x}_k$,

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{\exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2\right)}{(\pi\sigma^2)^N} \tag{3}$$

where $\|\cdot\|_p$ is the ℓ_p-norm.

We assume a multivariate complex Laplace-like probability density [21] for the source vector $\boldsymbol{x}_k$ at step k, conditioned on the history $\boldsymbol{Y}_{k-1}$,

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \prod_{m=1}^{M}\left(\frac{\lambda_{km}}{\sqrt{2\pi}}\right)^2 \mathrm{e}^{-\lambda_{km}|x_{km}|} \tag{4}$$

$x_{km} = |x_{km}|\mathrm{e}^{\mathrm{j}\phi_{km}}$ is the complex source signal at step k and hypothetical source location $\boldsymbol{r}_m$. Note that (4) defines a joint distribution for $|x_{km}|$ and ϕ_{km} at step k for all $m = 1, \ldots, M$. The associated hyperparameters $\lambda_{km} > 0$ model the prior knowledge on the occurrence of a source at step k, at location $\boldsymbol{r}_m$. Taking the logarithm gives

$$-\ln p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}) = \sum_{m=1}^{M}\lambda_{km}|x_{km}| - 2\sum_{m=1}^{M}\ln\lambda_{km} + M\ln 2\pi. \tag{5}$$

For the posterior probability density function (pdf) $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$, we use Bayes' rule to obtain the weighted LASSO cost function to be minimized, cf. Equation (6) in [16]:

$$
\begin{aligned}
-\ln p(\boldsymbol{x}_k|\boldsymbol{Y}_k) &= -\ln\left(\frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k)}\right) \\
&= \frac{\|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2}{\sigma^2} + \sum_{m=1}^{M}\lambda_{km}|x_{km}| + c_k, \\
&= \frac{\|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2}{\sigma^2} + \|\boldsymbol{\lambda}_k \odot \boldsymbol{x}_k\|_1 + c_k, \quad (6) \\
&= \mathscr{L}_k(\boldsymbol{x}_k) + c_k, \tag{7}
\end{aligned}
$$

in which the notation $\mathscr{L}_k(\boldsymbol{x}_k)$ is introduced and

$$c_k = -2\sum_{m=1}^{M}\ln\lambda_{km} + N\ln\pi\sigma^2 + M\ln 2\pi + \ln p(\boldsymbol{y}_k). \tag{8}$$

Minimizing the weighted LASSO cost function (7) with respect to $\boldsymbol{x}_k$ for given $\boldsymbol{\lambda}_k = (\lambda_{k1}, \ldots, \lambda_{kM})^T$ and σ^2, gives the MAP source estimate $\hat{\boldsymbol{x}}_k$. The $\odot$ in (6) is the element-wise

(Hadamard–) product, making it clear that this is a cost function promoting sparse solutions in which the ℓ_1 constraint is weighted by giving every source amplitude its own hyperparameter λ_{km}. The minimization of (7) constitutes a convex optimization problem.

A. Posterior Probability Density

The purpose of this section is deriving an expression for the posterior pdf $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$. We distinguish the variable $\boldsymbol{x}_k$, a free variable for $p(\boldsymbol{x}_k|\boldsymbol{Y}_k)$, and $\hat{\boldsymbol{x}}_k$, the MAP source estimate which is treated as constant.

The set of active indices $\mathcal{M}_k$ at step k is introduced as the set of m such that $\hat{x}_{km}$ is non zero, i.e.,

$$\mathcal{M}_k = \{m_1, m_2, \ldots, m_R\} := \{m|\hat{x}_{km} \neq 0\}. \tag{9}$$

The true $\boldsymbol{x}_k$ and the estimated $\hat{\boldsymbol{x}}_k$ source amplitude vectors are related by the source vector error $\boldsymbol{e}_k$:

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_k + \boldsymbol{e}_k, \tag{10}$$

using the active indices

$$e_{km} = \begin{cases} x_{km}, & \text{for } m \notin \mathcal{M}_k, \\ x_{km} - \hat{x}_{km}, & \text{for } m \in \mathcal{M}_k. \end{cases} \tag{11}$$

The log-likelihood function (7) is expanded

$$\begin{aligned} \|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2 &= \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2 + \|\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k\|_2^2 \\ &\quad - 2\text{Re}[\boldsymbol{e}_k^H \boldsymbol{A}^H (\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k)] \\ &= \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2 + \|\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k\|_2^2 \\ &\quad - 2\sum_{m\in\mathcal{M}_k} \text{Re}[e_{km}^* \boldsymbol{a}_m^H (\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k)] \\ &\quad - 2\sum_{m\notin\mathcal{M}_k} \text{Re}[x_{km}^* \boldsymbol{a}_m^H (\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k)] \end{aligned} \tag{12}$$

where we separated the sum and used (11). The sum over $m \in \mathcal{M}_k$ is simplified by the following

Theorem 1: For given weighting $\boldsymbol{\lambda}_k$, the MAP source estimate (minimizing solution) $\hat{\boldsymbol{x}}_k$ to (7) is characterized by (cf. [16])

$$\boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k = \lambda_{km} \frac{\hat{x}_{km}}{|\hat{x}_{km}|}, \; for \; m \in \mathcal{M}_k, \tag{13a}$$

$$|\boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k| < \lambda_{km}, \; for \; m \notin \mathcal{M}_k \tag{13b}$$

with the normalized residual vector

$$\hat{\boldsymbol{n}}_k = \frac{2}{\sigma^2}(\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k). \tag{14}$$

The proof is in Appendix A. Continuing from (12), gives

$$\begin{aligned} \|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2 &= \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2 + \|\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k\|_2^2 \\ &\quad - \sigma^2 \sum_{m\in\mathcal{M}_k} \lambda_{km}\text{Re}(\frac{\hat{x}_{km}e_{km}^*}{|\hat{x}_{km}|}) \\ &\quad - \sigma^2 \sum_{m\notin\mathcal{M}_k} \text{Re}(x_{km}^* \boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k). \end{aligned} \tag{15}$$

We expand the LASSO cost function $\mathscr{S}_k(\cdot)$ (7) around the MAP source estimate $\hat{\boldsymbol{x}}_k$ using (15) and then (10):

$$\begin{aligned} \mathscr{S}_k(\boldsymbol{x}_k) &= \frac{\|\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k\|_2^2}{\sigma^2} + \sum_{m=1}^{M} \lambda_{km}|x_{km}| \\ &\quad - \sum_{m\notin\mathcal{M}_k} \text{Re}(x_{km}^* \boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k) + \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2 \\ &\quad + \sum_{m\in\mathcal{M}_k} \lambda_{km}\left[|\hat{x}_{km}| - \text{Re}(\frac{\hat{x}_{km}x_{km}^*}{|\hat{x}_{km}|})\right] \end{aligned} \tag{16}$$

$$\begin{aligned} &= \mathscr{S}_k(\hat{\boldsymbol{x}}_k) + \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2 \\ &\quad + \sum_{m\notin\mathcal{M}_k} |x_{km}|\Big(\lambda_{km} - \zeta_{km}\cos(\phi_{km} - \rho_{km})\Big) \\ &\quad + \sum_{m\in\mathcal{M}_k} |x_{km}|\Big(\lambda_{km} - \lambda_{km}\cos(\phi_{km} - \hat{\phi}_{km})\Big) \end{aligned} \tag{17}$$

where we introduced the polar representations,

$$\boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k = \zeta_{km}e^{j\rho_{km}}, \; \hat{x}_{km} = |\hat{x}_{km}|e^{j\hat{\phi}_{km}} \tag{18}$$

$\zeta_{km} = |\boldsymbol{a}_m^H \hat{\boldsymbol{n}}_k|$ is the magnitude of the projection of the normalized residual vector onto the mth array steering vector. By Theorem 1, $\lambda_{km} > \zeta_{km}$ for $m \notin \mathcal{M}_k$. Equation (17) shows that the posterior pdf depends on *both* the unknown source magnitudes $|x_{km}|$ and phases ϕ_{km}. In the Section III-B, an approximate marginal posterior pdf for the source magnitudes is derived. This allows to approximate the *next* MAP estimate $\hat{\boldsymbol{x}}_{k+1}$ as the minimizer of an updated weighted LASSO cost function.

B. Marginalization With Respect to Source Phases

From (4), it follows that the prior for the source phases is i.i.d. uniformly distributed on $[-\pi, \pi]$. We regard the phases ϕ_{km} as nuisance parameters and marginalize $p(\boldsymbol{x}_k|\boldsymbol{Y}_k) = \exp[-\mathscr{S}_k(\boldsymbol{x}_k) - c_k]$ in (7) inserting (17) and integrating the phases ϕ_{km} out. The source vector errors are assumed small and $\|\boldsymbol{e}_k\|_2^2$ is neglected in (17). Assuming the errors to be small is common in sequential methods. For example the Extended Kalman Filter (EKF) also assumes small errors so that the parameter estimate remains Gaussian. As demonstrated in the sequel, we use this approximation so that the posterior remains Laplacian. This is motivated by the assumption that the posterior pdf is concentrated around its peak point $\hat{\boldsymbol{x}}_k$. Using (17), the approximate marginal posterior pdf depends solely on the source magnitudes

$$\begin{aligned} p(|\boldsymbol{x}_k||\boldsymbol{Y}_k) &= \int \exp[-\mathscr{S}_k(\boldsymbol{x}_k) - c_k]\,\mathrm{d}\phi_{k1}\cdots\mathrm{d}\phi_{kM} \\ &= \int e^{-\mathscr{S}_k(\boldsymbol{x}_k) - c_k + \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2} e^{-\|\boldsymbol{A}\boldsymbol{e}_k\|_2^2}\,\mathrm{d}\phi_{k1}\cdots\mathrm{d}\phi_{kM} \\ &\approx \gamma_k \prod_{m\notin\mathcal{M}_k} e^{-\lambda_{km}|x_{km}|} I_0\left(\zeta_{km}|x_{km}|\right) \\ &\quad \times \prod_{m\in\mathcal{M}_k} e^{-\lambda_{km}|x_{km}|} I_0\left(\lambda_{km}|x_{km}|\right) \end{aligned} \tag{19}$$

where $|\boldsymbol{x}_k|$ is the vector of source magnitudes with elements $|x_{km}|$, γ_k is a normalization constant defined by

$$\gamma_k = \exp(-\mathscr{S}_k(\hat{\boldsymbol{x}}_k) - c_k) = \prod_{m=1}^{M} \gamma_{km}$$

$$\gamma_{km} = \begin{cases} \left[\int_0^\infty e^{-\lambda_{km}x} I_0(\zeta_{km}x)\,dx\right]^{-1} = \sqrt{\lambda_{km}^2 - \zeta_{km}^2} \\ \hspace{4.5cm} \text{for } m \notin \mathcal{M}_k \\ 1 \hspace{4cm} \text{for } m \in \mathcal{M}_k \end{cases}$$

$$(20)$$

and $I_0(z) = \frac{1}{\pi}\int_0^\pi \exp(z\cos\phi)\,d\phi$ is the modified Bessel function of first kind and zero order, cf. [23]:9.6.16. To assess the quality of the approximation in (19), we use the first order Taylor expansion in the vicinity of the MAP estimate $\hat{\boldsymbol{x}}_k$,

$$e^{-\|\boldsymbol{A}\boldsymbol{e}_k\|_2^2} \approx 1 - \|\boldsymbol{A}\boldsymbol{e}_k\|_2^2. \qquad (21)$$

In (19), only the constant term was retained and the quality of the approximation depends on the magnitude of $\|\boldsymbol{A}\boldsymbol{e}_k\|_2^2$. The approximate marginal pdf (19) becomes improper as $\zeta_{km} \to \lambda_{km}$ because it cannot be normalized on the half line $|x_{km}| > 0$ in the limit $\zeta_{km} = \lambda_{km}$. This difficulty occurs when the index m becomes active (see Section III-C). Its form is not Laplace-like (4) and we develop two types of approximations for fitting the posterior distribution at time step k to a Laplace-like prior distribution for time step $k + 1$ in the following Sections III-D and III-E.

C. Active Indices

To estimate the posterior distribution for the active indices we use (19). For any given $\lambda_{km} > 0$, the Laplace-like pdf (4) favors a model with small source amplitude $|x_{km}|$. In the limiting case $\lambda_{km} \to 0$, however, the Laplace-like pdf approaches a uniform pdf for the source amplitude $|x_{km}| \in [0, \infty]$ which cannot be normalized. We express this informally as

$$p(|x_{km}||\boldsymbol{Y}_k) \propto 1 \quad \text{for } \lambda_{km} \to 0. \qquad (22)$$

Thus, when some index m has become active in the current time step k there will be no constraint on the corresponding amplitude and we define the posterior hyperparameter

$$\lambda_{km|k} = 0 \quad \text{for } m \in \mathcal{M}_k, . \qquad (23)$$

Note that this model has no information to see how likely an active index is to become inactive.

D. Asymptotic Fit of Laplace-Like Distribution

Here we are concerned with the λ_{km} for the inactive indices $m \notin \mathcal{M}_k$. The λ_{km} for any $m \notin \mathcal{M}_k$ conveys prior information about how likely the index m is to transition to the set of active indices $\mathcal{M}_{k+1}$ at the next time step $k + 1$. The approximate marginal pdf (19) is close to Laplace-like (4) in terms of asymptotic behavior for large values of x_{km}. The MAP estimate $\hat{\boldsymbol{x}}_k$ is sparse and many $\hat{x}_{km} = 0$, but the variables x_{km} can take any value. In this section, x_{km} is assumed large. The first-order asymptotic expansion is used,

$$I_0(\xi) \approx e^{\xi}/\sqrt{2\pi\xi} \qquad \text{for large } \xi. \qquad (24)$$

For amplitude $|x_{km}|$ we get

$$p(|x_{km}||\boldsymbol{Y}_k) \approx \frac{e^{-(\lambda_{km}-\zeta_{km})|x_{km}|}}{\sqrt{\zeta_{km}|x_{km}|}} \quad \text{for } m \notin \mathcal{M}_k. \qquad (25)$$

The asymptotic expansion (24)–(25) has a small approximation error for $|x_{km}| \gg \zeta_{km}^{-1}$. Taking the logarithm analogously to (5), gives

$$-\ln p(|x_{km}||\boldsymbol{Y}_k) \approx (\lambda_{km} - \zeta_{km})|x_{km}| \\ + \frac{1}{2}\ln\zeta_{km} + \frac{1}{2}\ln|x_{km}|. \qquad (26)$$

Since $\zeta_{km} < \lambda_{km}$ (Theorem 1) and $|x_{km}|$ is large, we neglect $\ln\zeta_{km}$ and $\ln|x_{km}|$ in (26). Finally, all these steps and approximations give the posterior weighted ℓ_1-penalty

$$-\ln p(|\boldsymbol{x}_k||\boldsymbol{Y}_k) \approx \sum_{m \notin \mathcal{M}_k} (\lambda_{km} - \zeta_{km})\,|x_{km}|. \qquad (27)$$

Comparison with (4) and combination with (23) shows that this translates to the posterior weight vector $\boldsymbol{\lambda}_{k|k}$ with elements

$$\lambda_{km|k} = \begin{cases} \lambda_{km}\left(1 - \frac{\zeta_{km}}{\lambda_{km}}\right) & \text{for } m \notin \mathcal{M}_k, \\ 0 & \text{for } m \in \mathcal{M}_k, \end{cases} \qquad (28)$$

The weight $\lambda_{km|k}$ reduces to zero when index m becomes active (see Section III-C). The first line in (28) is always positive as $\zeta_{km} < \lambda_{km}$.

E. Mean Fit of Laplace-Like Distribution

Alternatively to the approach in Section III-D, the expected source amplitudes for the assumed the Laplace-like distribution (4) can be fitted to the expected posterior source amplitudes.

$$\mu_{km|k} = \text{E}\{|x_{km}||\boldsymbol{Y}_k\} = \int_{\mathbb{R}_+^M} a_m p(\boldsymbol{a}|\boldsymbol{Y}_k)\,da_1\cdots da_M. \qquad (29)$$

Proceeding from (19) for $m \notin \mathcal{M}_k$, this gives

$$\mu_{km|k} \approx \gamma_{km}\int_0^\infty a_m e^{-\lambda_{km}a_m} I_0(\zeta_{km}a_m)\,da_m \qquad (30)$$

$$= \frac{\gamma_{km}\lambda_{km}}{(\lambda_{km}^2 - \zeta_{km}^2)^{\frac{3}{2}}} = \frac{\lambda_{km}}{\lambda_{km}^2 - \zeta_{km}^2}. \qquad (31)$$

For a Laplace-like distribution,

$$\mu_{km|k} = \frac{1}{\lambda_{km|k}}. \qquad (32)$$

Combining (31) and (32) gives the elements of the posterior weight vector $\boldsymbol{\lambda}_{k|k}$ for $m \notin \mathcal{M}_k$. The corresponding elements for $m \in \mathcal{M}_k$ are obtained in the limit $\zeta_{km} \to \lambda_{km}$ consistent with (23). The posterior weight vector $\boldsymbol{\lambda}_{k|k}$ becomes

$$\lambda_{km|k} = \begin{cases} \lambda_{km}(1 - \frac{\zeta_{km}^2}{\lambda_{km}^2}) & \text{for } m \notin \mathcal{M}_k, \\ 0 & \text{for } m \in \mathcal{M}_k, \end{cases} \qquad (33)$$

For $m \in \mathcal{M}_k$, the weight $\lambda_{km|k}$ reduces to zero when index m becomes active.

The posterior weight according to (28) is always smaller than posterior weight (33) for $m \notin \mathcal{M}_k$ because $\lambda_{km} > \zeta_{km}$ for $m \notin$
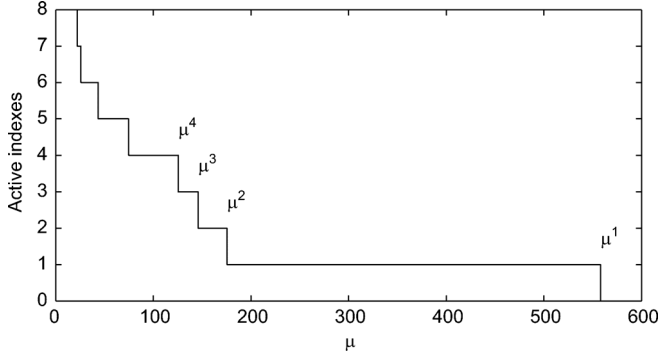
Fig. 1.   Number of active indices versus the regularization parameter μ for the example in Section VI.

$\mathcal{M}_k$ by Theorem 1. We choose (33) rather than (28), because the weighting in the cost function (6) will then play a stronger role.

IV. SIMPLE SOURCE MOTION MODEL

After having estimated the posterior $\boldsymbol{\lambda}_{k|k}$ with elements defined by (33), the next step is to predict $\boldsymbol{\lambda}_{k+1}$ for the Laplace-like prior (4) at time step $k + 1$. Time-varying source signal estimates are obtained by varying the active set (creating new active indices as well as annihilating old ones). Here, a very simple random walk model is assumed: A source occurring at location $\boldsymbol{r}_m$ at time step $k + 1$ is either inherited from the previous time step k with probability $(1 - \alpha)$ or newly generated from an unweighted (uninformed) Laplace-like prior with parameter λ_0' with probability α. Thus, we define the hyperparameter for time step $k + 1$,

$$\boldsymbol{\lambda}_{k+1} := (1 - \alpha)\boldsymbol{\lambda}_{k|k} + \alpha\lambda_0'\boldsymbol{1}, \qquad (34)$$

where $\boldsymbol{1}$ is the all-1 vector. $\lambda_0 = \alpha\lambda_0'$ plays a role which is similar to the *process noise covariance* in the Kalman filter, see Section VI. If λ_0 is large then $\min_m(\lambda_{k+1,m|k})/\max_m(\lambda_{k+1,m|k})$ will be close to one. Then, the history $\boldsymbol{Y}_k$ has only little influence on the future estimate $\hat{\boldsymbol{x}}_{k+1}$. Conversely, for a smaller λ_0, the adaptivity of the weighting through the posterior (28) will be more influential to the future sequential estimates.

V. IMPLEMENTATION OF $\boldsymbol{\varphi}$ USING LASSO PATH

We define the positive μ parameter and the vector of non-negative weighting coefficients $\boldsymbol{w}_k$ via $\boldsymbol{\lambda}_k = \mu\boldsymbol{w}_k$, $\|\boldsymbol{w}_k\|_\infty = 1$, equivalently $w_{km} = \lambda_{km}/\mu$. Thus the weights w_{km} are obtained by the previous data and the source model and normalized so that $\max_m w_{km} = 1$. For the data $\boldsymbol{y}_k$ the source vector $\boldsymbol{x}_k$ is found by minimizing

$$\mathscr{L}_k(\boldsymbol{x}_k) = \frac{\|\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k\|_2^2}{\sigma^2} + \mu \sum_{m=1}^M w_{km}|x_{km}|. \qquad (35)$$

The optimization (35) is convex and is solved assuming a fixed number of sources R. However, μ is unknown and adjusted to reach a desired number of active indices R as described below. The weighted complex-valued LASSO is based on solving LASSO by homotopy [16], [22], as indicated in Fig. 1.

The solution of (35) evolves with μ, due to the uniform continuity of the cost. Furthermore, for sufficiently large μ, $\boldsymbol{x}_k = \boldsymbol{0}$. Thus, a solution with R active indices is found by pursuing the continuum from a large μ. The evolution of μ from large to smaller values is segmented into intervals (μ^{p+1}, μ^p) for $p = 0, 1, \ldots$, in which the set of active indices is constant, for $p = 0$ corresponds to $\mu^0 = \infty$. Varying μ within an interval changes just x_{km} for the active indices $m \in \mathcal{M}_k$. Each interval (μ^{p+1}, μ^p) has an end-point at a candidate point μ^{p+1}, where the active set $\mathcal{M}_k$ changes by *including* or *deleting* an element (potentially multiple elements).

Creation is defined as including an index $l \notin \mathcal{M}_k^p$ into $\mathcal{M}_k^{p+1}$ if $|x_{kl}| > 0$ as μ decreases below μ^{p+1}. Vice versa, *annihilation* is the process of deleting an index $l \in \mathcal{M}_k^p$ as $x_{kl} \to 0$ with decreasing μ. Due to the (rather unlikely) possibility of annihilation, the number of active indices R can be less than p for $\mu \in (\mu^{p+1}, \mu^p)$.

The sequence of μ with source amplitudes $\boldsymbol{x}_k(\mu)$ is found using the optimality conditions in Theorem 1. Starting from μ large, in the first interval (μ^1, ∞), $\boldsymbol{x}_k(\mu) = \boldsymbol{0}$. Then, μ^1 is the smallest value, for which Theorem 1 is satisfied with $\boldsymbol{x}_k(\mu) = \boldsymbol{0}$:

$$\mu^1 = \min(\mu) \quad \text{s.t.} \quad 2|\boldsymbol{a}_m^H \boldsymbol{y}_k| \le \mu w_{km}\sigma^2 \quad \forall m$$
$$= \max_{m=1,\ldots,M} \frac{2|\boldsymbol{a}_m^H \boldsymbol{y}_k|}{w_{km}\sigma^2}. \qquad (36)$$

The LASSO is next solved for $\mu < \mu^1$. Recalling Theorem 1 with a given active set $\mathcal{M}_k$ from the previous candidate solution $\boldsymbol{x}_k(\mu)$, the next candidate point μ^{p+1} is the smallest μ $(< \mu^p)$ with the active indexes unchanged, i.e.,

$$\mu^{p+1} = \min(\mu) \quad \text{s.t.} \quad 2|\boldsymbol{a}_m^H(\boldsymbol{y}_k - \boldsymbol{A}\boldsymbol{x}_k(\mu))|$$
$$\le \mu w_{km}\sigma^2 \ \forall m$$
$$\text{and} \quad x_{km} \ne 0 \quad \forall m \in \mathcal{M}_k^p \qquad (37)$$

where $\boldsymbol{x}_k(\mu)$ is the LASSO solution for given μ and $\mathcal{M}_k^p$ is the set of R active indexes in the interval (μ^{p+1}, μ^p):

$$\mathcal{M}_k^p = \{m_1, m_2, \ldots, m_R\} := \{m | \hat{x}_{km} \ne 0\}. \qquad (38)$$

The contracted dictionary $\boldsymbol{A}_\mathcal{M}$, its Moore-Penrose pseudoinverse $\boldsymbol{A}_\mathcal{M}^+$, contracted source amplitudes $\boldsymbol{x}_\mathcal{M}$, and contracted phase-weighted weights $\boldsymbol{w}_\mathcal{M}$ are defined by

$$\boldsymbol{A}_\mathcal{M} = [\boldsymbol{a}_{m_1} \ldots \boldsymbol{a}_{m_R}] \qquad (39)$$
$$\boldsymbol{A}_\mathcal{M}^+ = (\boldsymbol{A}_\mathcal{M}^H \boldsymbol{A}_\mathcal{M})^{-1} \boldsymbol{A}_\mathcal{M}^H \qquad (40)$$
$$\boldsymbol{x}_\mathcal{M}(\mu) = [x_{km_1}(\mu) \ldots x_{km_R}(\mu)]^T \qquad (41)$$
$$\boldsymbol{w}_\mathcal{M} = \left[w_{km_1} \frac{x_{km_1}}{|x_{km_1}|} \ldots w_{km_R} \frac{x_{km_R}}{|x_{km_R}|} \right]^T. \qquad (42)$$

The projection matrix $\boldsymbol{P}_{\boldsymbol{A}_\mathcal{M}}^\perp$ defines the projection onto the orthogonal complement of the range of $\boldsymbol{A}_\mathcal{M}$,

$$\boldsymbol{P}_{\boldsymbol{A}_\mathcal{M}}^\perp = \boldsymbol{I} - \boldsymbol{Q}_\mathcal{M}\boldsymbol{Q}_\mathcal{M}^H, \quad \text{where } \boldsymbol{Q}_\mathcal{M} = \text{orth}(\boldsymbol{A}_\mathcal{M}) \qquad (43)$$

is a unitary basis of the range of $\boldsymbol{A}_\mathcal{M}$. The following Theorem 2 provides an approximate solution to the minimization problem

	Implementation of $\varphi(\boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ defined via Eq. (2):
	Input: $\boldsymbol{y}_k \in \mathbb{C}^N$, $\boldsymbol{\lambda}_k \in \mathbb{R}_+^M$
	Given constants: $\boldsymbol{A} \in \mathbb{C}^{N \times M}$, $R_{\max} \in \mathbb{N}$, $\epsilon > 0$, $\boldsymbol{\alpha}$, $\lambda_0 > 0$
1:	$\boldsymbol{w}_k := \boldsymbol{\lambda}_k / \|\boldsymbol{\lambda}_k\|_\infty$ (the weighting coefficients are normalized)
2:	$p := 0$
3:	$\boldsymbol{x}_k^0 := \boldsymbol{0}$, $\mathcal{M}_k^0 := \emptyset$, $R := 0$ (there are no active indices)
4:	while $R \leq R_{\max}$ do
5:	Evaluate μ^{p+1} according to $\begin{cases} \text{Eq.(36)} & \text{for } p = 1 \\ \text{Eq.(44)} & \text{for } p > 1 \end{cases}$
6:	$\mu := (1 - \epsilon)\mu^{p+1}$ (we choose μ slightly smaller than μ^{p+1})
7:	$\boldsymbol{x}_k^{p+1} := \arg\min_{\boldsymbol{x}}(-\ln p(\boldsymbol{x}|\boldsymbol{Y}_k))$ using (7) for $\boldsymbol{\lambda}_k = \mu\boldsymbol{w}_k$.
8:	Define active set $\mathcal{M}_k^{p+1}$ according to (9)
9:	$R := \text{card}(\mathcal{M}_k^{p+1})$ (R is the new no. of active indices)
10:	$p := p + 1$
11:	end
12:	$\mathcal{M}_k := \mathcal{M}_k^p$ (this is the active set of the MAP estimate)
13:	$\begin{cases} \hat{\boldsymbol{x}}_k = \boldsymbol{x}_k^p & \text{(choice 1: MAP estimate)} \\ \hat{\boldsymbol{x}}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^+ \boldsymbol{y}_k & \text{(choice 2: bias corrected)} \end{cases}$
14:	Evaluate posterior $\boldsymbol{\lambda}_{k|k}$ using $\begin{cases} \text{choice 3: (28)} \\ \text{choice 4: (33)} \end{cases}$
15:	Evaluate prior $\boldsymbol{\lambda}_{k+1}$ for next time step using (34) with $\boldsymbol{\alpha}$, λ_0.
	Output: $\hat{\boldsymbol{x}}_k \in \mathbb{C}^M$, $\boldsymbol{\lambda}_{k+1} \in \mathbb{R}_+^M$.

(37) for a given set of active indices. The approximation is based on the assumption that the $\boldsymbol{w}_{\mathcal{M}}$ vector (42) is constant inside the interval (μ^{p+1}, μ^p). Equivalently, the phase angles of all elements of $\boldsymbol{x}_{\mathcal{M}}(\mu)$ are assumed to be constant which is a first order approximation in μ.

Theorem 2: The approximate solution to (37) is given by

$$\mu_k^{p+1} := \max\{\mu^+, \ \mu^-\}, \tag{44}$$

where the *creation* μ^+ and *annihilation* μ^- are

$$\mu^+ = \max_{m \leq M} Q\left(\frac{2\boldsymbol{a}_m^H \boldsymbol{P}_{\boldsymbol{A}_{\mathcal{M}}}^\perp \boldsymbol{y}_k}{\sigma^2 w_{km}}, \ \frac{\boldsymbol{a}_m^H \boldsymbol{A}_{\mathcal{M}}^{+H} \boldsymbol{w}_{\mathcal{M}}}{w_{km}}\right) \tag{45}$$

$$\mu^- = \frac{2}{\sigma^2} \max_{l=1,\dots,R} \tilde{x}_{\mathcal{M}l}/\tilde{w}_{\mathcal{M}l}, \tag{46}$$

where $Q(\alpha, \beta)$ is defined in Appendix C, and

$$\tilde{\boldsymbol{x}}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^+ \boldsymbol{y}_k, \tag{47}$$

$$\tilde{\boldsymbol{w}}_{\mathcal{M}} = (\boldsymbol{A}_{\mathcal{M}}^H \boldsymbol{A}_{\mathcal{M}})^{-1} \boldsymbol{w}_{\mathcal{M}} \tag{48}$$

$\tilde{\boldsymbol{x}}_{\mathcal{M}}$ is the bias-corrected estimate.

The proof is in Appendix B. The implementation of the sequential Bayesian sparse signal reconstruction update $\varphi(\boldsymbol{y}_k, \boldsymbol{\lambda}_k)$ in (2) is given concisely in Table I. Line 8 "Define active set" is implemented with a numerical peak finder and subsequent thresholding. Note the bias correction step in line 13 "choice 2" using just the active indexes [30]:

$$\hat{\boldsymbol{x}}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^+ \boldsymbol{y}_k. \tag{49}$$

VI. NUMERICAL EXAMPLE

In this section we investigate the performance of the proposed sequential sparse estimation procedure using computer simulations. For the sequential Bayesian sparse signal reconstruction, we initialize the prior vector $\boldsymbol{\lambda}_1$ with all ones: all hypothetical source locations have the same prior. We then apply (2) sequen-

tially for all time steps $k = 1, 2, 3, \dots$ with the implementation given in Table I with $\epsilon = 0.05$.

We use synthetic data on a uniform linear array with $N = 64$ elements with half-wavelength spacing and using 50 snapshots sequentially ($k = 1, \dots, 50$). The traveltime τ_{nm} for plane waves impinging on the nth sensor is modeled as

$$\omega\tau_{nm} = \pi(n - 1)\cos\theta_m, \tag{50}$$

for a hypothetical source at direction of arrival $\boldsymbol{r}_m = \theta_m = (m - 1)\frac{\pi}{M}$ rad with $M = 180$ and $n = 1, \dots, N$; $m = 1, \dots, M$. The simulated scenario has $R = 9$ far-field sources modeled as emitting plane waves (1). The uncorrelated noise $\boldsymbol{n}$ is $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, i.e., $0\,\text{dB}$ power. Eight sources are stationary at $\theta^T = [45, 60, 76, 99, 107, 120, 134, 162]$ degrees relative to endfire with constant power levels (PL) $[-5, 10, 5, 0, 11, 12, 9, 25]$ dB. The moving source starts at $65°$ and moving $+0.5°$/snapshot and has a power level of $20\,\text{dB}$.

A. Beamforming

In beamforming the Cross Spectral Density Matrix (CSDM) is estimated based on all snapshots $\boldsymbol{y}_k$ for $k = 1, \dots, 50$. The CSDM $\hat{\boldsymbol{R}}$ estimate is singular as it is snapshot deficient, with at least $64 - 50 = 14$ eigenvalues equal zero. For conventional beamforming, the resolution is not sufficient to separate all the sources (Fig. 2). To numerically stabilize the matrix inversion in the adaptive beamforming, diagonal loading is applied at a level of $\epsilon = -10\,\text{dB}$ relative to the sum of the eigenvalues. The stabilized beamformer output B for a given steering vector $\boldsymbol{s}$ is

$$B = \boldsymbol{w}^H \hat{\boldsymbol{R}} \boldsymbol{w} \quad \text{with} \quad \boldsymbol{w} = \frac{[\hat{\boldsymbol{R}} + \epsilon\boldsymbol{I}]^{-1}\boldsymbol{s}}{\boldsymbol{s}^H[\hat{\boldsymbol{R}} + \epsilon\boldsymbol{I}]^{-1}\boldsymbol{s}}. \tag{51}$$

For the stationary case (Fig. 2(a)), the conventional beamformer, $\epsilon = \infty$ in (51), resolves all 8 sources whereas the Minimum Variance Distortionless Response (MVDR) $\epsilon = -10\,\text{dB}$, just resolves 6 sources. For this scenario, we observe that above beamforming finds the source location well for the stationary case (Fig. 2(a)), but the moving source masks the stationary source at $\theta = 76°$ (Fig. 2(b)). The observed bias in the power estimate for the MVDR is due to the snapshot deficient CSDM [24]. More advanced adaptive beamforming might be able to better estimate the source locations in this scenario with higher time resolution, e.g., the multi-rate adaptive beamforming [25], [26]. Multi-rate beamforming uses long time averages to produce the MVDR steer vectors and shorter averages for the data CSDM. It was introduced to track a stable signal in the presence of dynamic interferers [25].

In contrast, the sparse localization works directly with a single snapshot optimization (7) giving higher time-resolution. Successive snapshot localizations then include the information carried forward from previous localizations via the prior and the hyperparameters. This enables finding lower powered sources than if each snapshot was used independently.

B. Sequential LASSO

Going down the LASSO path, Section V, is computationally demanding, especially for the 9 sources involved in the
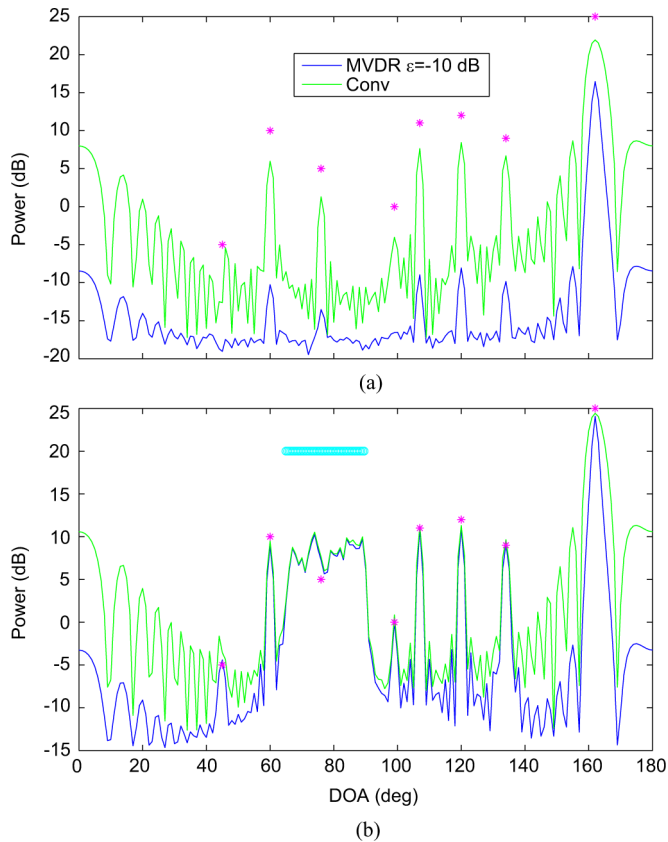
Fig. 2.   MVDR and conventional beamforming for the a) stationary sources (*), and b) stationary (*) and moving (o) sources for the 50 snapshot observation period.
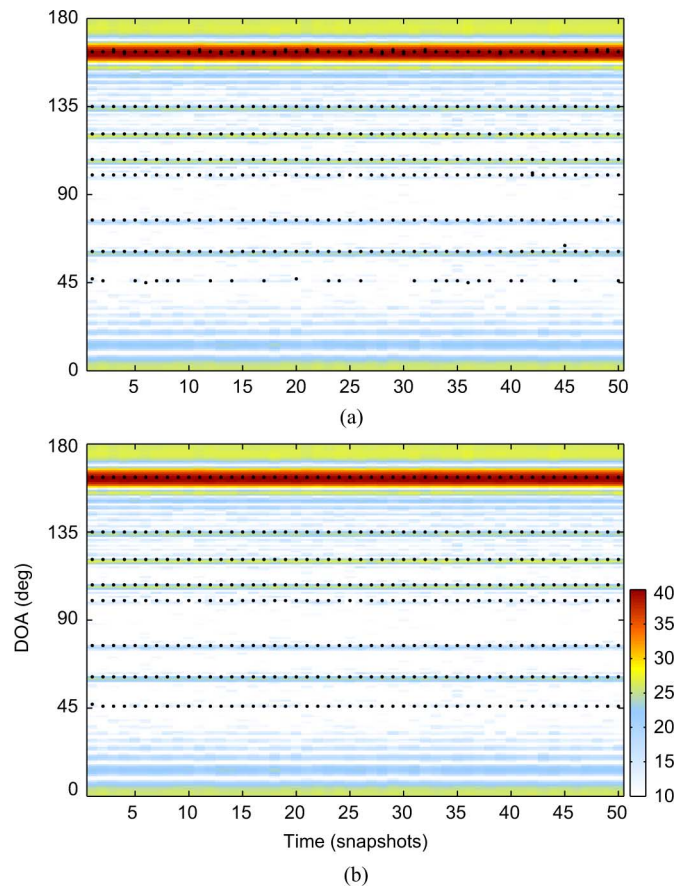

Fig. 3.   Track over 50 time steps for the 8 stationary sources (dots) for a) non-sequential LASSO and b) sequential LASSO with $\lambda_0 = 0.05$. Background (color, dB) shows the conventional beamformer power.
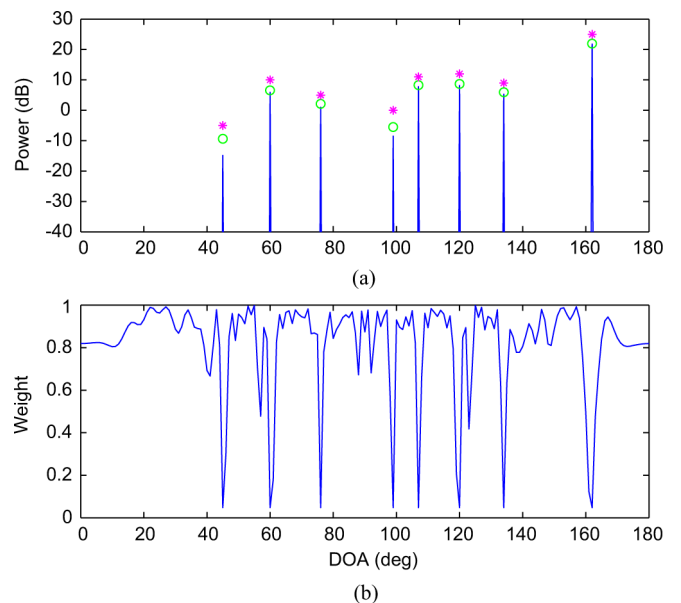

Fig. 4.   a) Output of the sparse optimization (8, solid) with $\lambda_0 = 0.5$. The unbiased (o) and true (*) source power. b) The posterior weights for $t = 10$.

studied scenario. Since the noise level is constant over time, the λ should not change much.

It is not necessary going down the full LASSO path. Therefore, each iteration is started at a μ slightly larger than the expected final μ. Assuming that the active indices do not change, the second term in (35) is proportional to $\mu\lambda_0$. We use $\mu = 400$ for $\lambda_0 = 0.05$ and $\mu = 40$ for $\lambda_0 = 0.5$.

For a given value of μ, the convex optimization (35) is solved using CVX [27]–[29]. We carry out the simulations with the mean fitting approach in Section III-E which corresponds to "choice 4" in line 14 of Table I.

1) Stationary Sources: The posterior weights $\boldsymbol{\lambda}_{k|k}$ defined by (33) are updated after each step and used as prior information to the following snapshot $k + 1$ forming the sequential estimation approach. The non-sequential LASSO results shown in Fig. 3(a) show more variation than those from the proposed sequential Bayesian sparse signal reconstruction shown in Fig. 3(b).

The constraint in (7) that sparsifies the solution causes the power estimates (peak of solid in Fig. 3(a)) to be underestimated especially for the low-power sources (e.g., $\theta = 45°$ and $99°$). Using the unbiased estimate (choice 2 in line 13 of Table I) gives powers within 3 dB of the correct ones.

The posterior weights show how the solution is being guided towards the solution of the active indices in the next step.

2) Moving Sources: For a moving source scenario (Figs. 5 and 6), it will be difficult observing the stationary sources using beamforming. Since the proposed approach works directly on a

snapshot, there is sufficient time resolution to track both moving and stationary sources, cf. Fig. 5(a).

The temporal evolution of the weights is shown in Fig. 6. The weights for the asymptotic fitting technique in Section III-D are
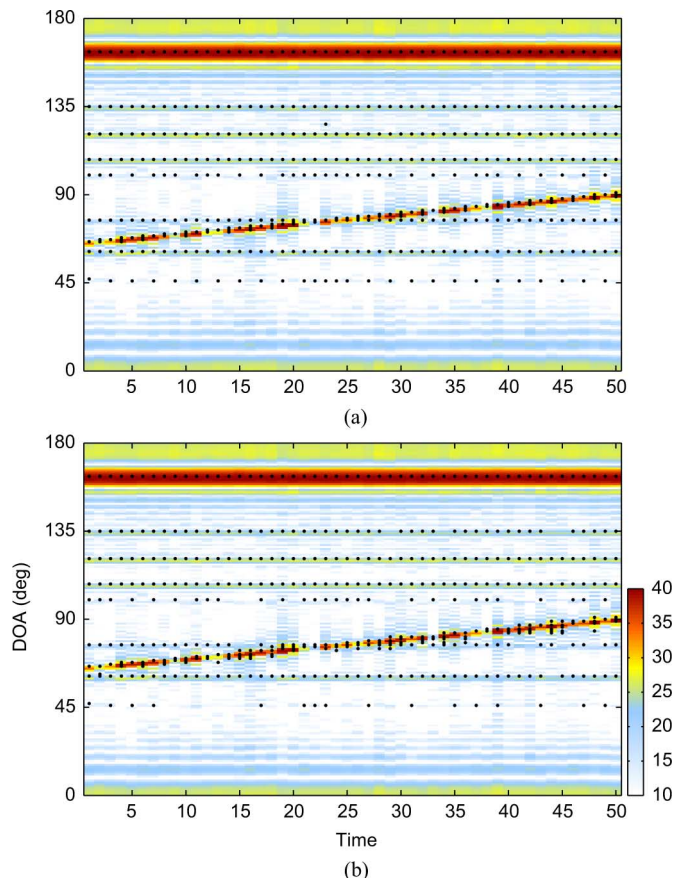
Fig. 5. Track for or the 8 stationary and 1 moving source sources (dots) using sequential LASSO with a) $\lambda_0 = 0.5$ b) $\lambda_0 = 0.05$. Background color shows the conventional beamformer power (dB).
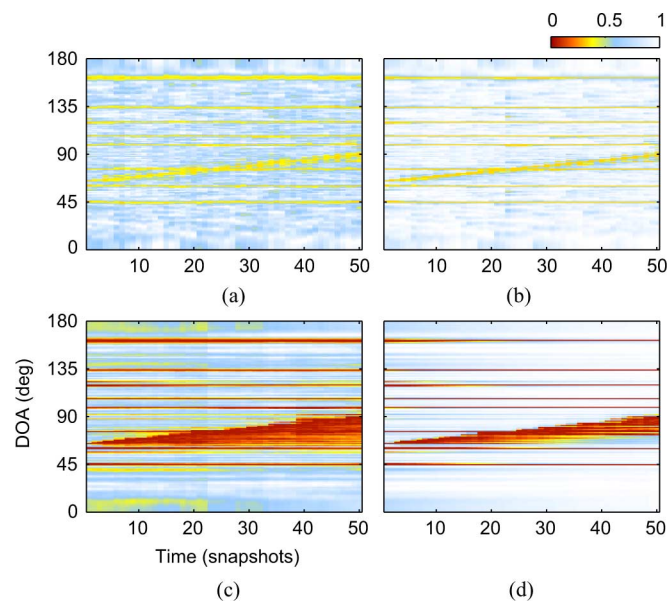

Fig. 6. Evolution of posterior weights $\boldsymbol{\lambda}_{k|k}$ versus k, left: Asymptotic fitting using (25) and right: mean fitting using (31). Large $\lambda_0 = 0.5$ in a) and c). Small $\lambda_0 = 0.05$ in b) and d).

shown in Figs. 6(a) and 6(b). For the asymptotic fitting technique, the difference in the weights between active indices and inactive ones is rather small. On the other hand, the weights for the mean fitting technique in Section III-E are shown in
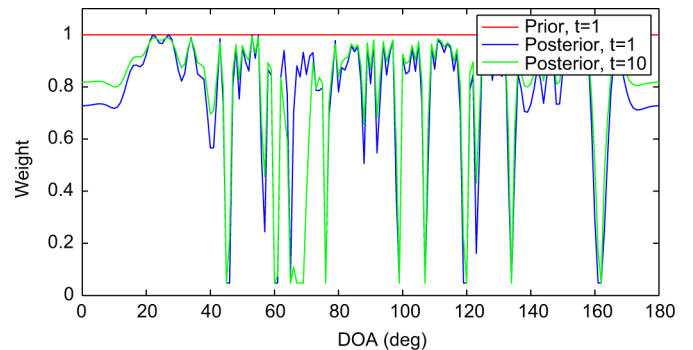

Fig. 7. Evolution of posterior weights $\boldsymbol{\lambda}_{k|k}$ for $\lambda_0 = 0.05$. Prior at $k = 1$ (red), posterior at $k = 1$ (blue), and posterior at $k = 10$ (green).

Figs. 6(c) and 6(d). It is clearly seen that the contrast between the weights for active and inactive indices is much larger.

The value of λ_0 determines how long the weight will remember a track, the non-sequential LASSO results are obtained with $\alpha = 1$ in (34). A large $\lambda_0 = 0.5$ causes the weights to faster forget a previously estimated source position (Figs. 6(a) and 6(c)) and gives more stable track estimates (Fig. 5(a) than for the smaller $\lambda_0 = 0.05$. The weights for the small $\lambda_0 = 0.05$ are shown in Figs. 6(b)–6(d) and the corresponding sequential Bayesian sparse signal reconstruction in Fig. 5(b). A small λ_0 results in a longer memory as seen in Figs. 6(b) and 6(d) which is beneficial for the sparse reconstruction of static source scenarios.

From Fig. 7 it is observed that the posterior weights become close to zero near the active indices. The weights have a broader valley near endfire where the resolution cells are larger. As the moving source moves from $65°$ to $75°$ the weights for the corresponding set of source locations shift.

VII. EXPERIMENT

The data is from a towed horizontal array during the long range acoustic communications (LRAC) experiment [32], [33] from 10:00–10:13 UTC on 16 September 2010 in the NE Pacific in 5-km water depth. Other data periods yield similar results to those shown here. The array was towed at 3.5 knots at a depth of 200 m. The data were sampled at 2000 Hz using a nested array with each configuration having 64 channels [34] with the ultra low frequency (ULF) array hydrophone spacing 3 m ($f_d = 250$ Hz).

Each 4 s snapshot was Fourier transformed with 2^{13} samples without overlapping snapshots. The beamformed time series, Fig. 8(a), is based on single snapshots and performed at one quarter wavelength element spacing, i.e., 125 Hz. The broad arrival at $150–165°$ is from the towship R/V Melville ($180°$ is forward looking). Apparently, the two arrivals at $45°$ and $60°$ come from distant transiting ships, although a log of ships in the area was not kept. Overall, the beam time series shows little change with time.

The sparse sampling was performed assuming $K = 10$ discrete signals and for sequential processing $\lambda_0 = 0.05$.

Fig. 8(b) shows the non-sequential estimates, while the discrete features are well-captured with the processing, there are
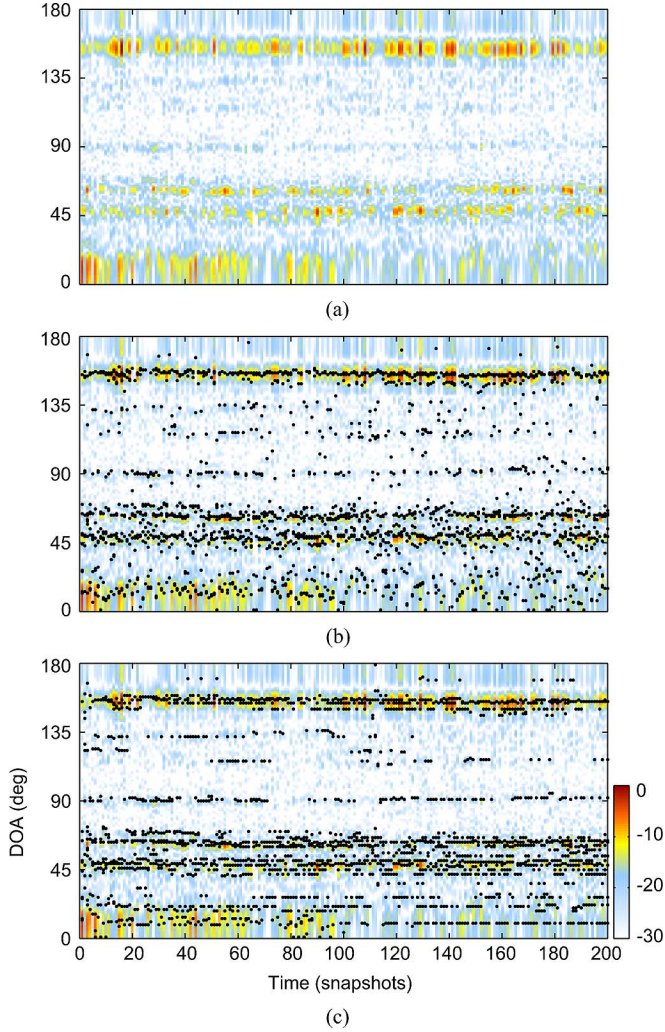
Fig. 8.   Real data example. a) Beamformer output. Tracks for 10 sources (dots) using b) non-sequential LASSO and c) sequential LASSO with $\lambda_0 = 0.05$. Background color shows the conventional beamformer power (dB).

significant fluctuations in the estimates from each snapshot. This is stabilized when using the sequential processing, Fig. 8(c).

VIII. CONCLUSION

A sequential weighted LASSO problem formulation is used to estimate the spatiotemporal evolution of a sparse source field from observed samples of a wave field. The prior is adapted based on the past history of observations. A sequential MAP filter is derived which preserves sparsity in the source field estimates by enforcing the priors at each estimation step to be Laplacian. The sequential information is carried in the hyperparameters.

With a uniform linear array it is demonstrated that the proposed sequential Bayesian sparse signal reconstruction obtains high resolution in both time in space relative to adaptive beamforming. Using towed array data, the estimated tracks become more stable with the sequential estimation.

APPENDIX A

The proof of Theorem 1 is similar to Appendix A in [22], see also [31]. Suppose the active set in (7) is given by $\mathcal{M}_k$ at the

global minimum $\mathscr{L}(\hat{\boldsymbol{x}})$. For an infinitesimal change $\mathrm{d}x_{km}$ at the mth element of $\hat{\boldsymbol{x}}_k$, we have

$$
\mathrm{d}\mathscr{L}_k\big|_{\hat{\boldsymbol{x}}_k} = \begin{cases} \mathrm{Re}\left[\left(-\frac{2}{\sigma^2}\boldsymbol{a}_m^H(\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k) + \lambda_{km}\frac{\hat{x}_{km}}{|\hat{x}_{km}|}\right)\mathrm{d}x_{km}^*\right] \\ \quad \text{for } m \in \mathcal{M}_k \\ \mathrm{Re}\left[-\frac{2}{\sigma^2}\boldsymbol{a}_m^H(\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k)\,\mathrm{d}x_{km}\right] + \lambda_{km}|\mathrm{d}x_{km}| \\ \quad \text{for } m \notin \mathcal{M}_k. \end{cases}
$$
(52)

The variations in $\mathscr{L}$ at the global minimum $\hat{\boldsymbol{x}}_k$ satisfy $\mathrm{d}\,\mathscr{L}_k \geq 0$ for every $m \in \{1, \dots, M\}$ and $\mathrm{d}x_{km} \in \mathbb{C}$. The minimality condition is satisfied for $m \in \mathcal{M}_k$ if and only if

$$
-\frac{2}{\sigma^2}\boldsymbol{a}_m^H(\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k) + \lambda_{km}\frac{\hat{x}_{km}}{|\hat{x}_{km}|} = 0.
$$
(53a)

Alternatively, if $m \notin \mathcal{M}_k$ then

$$
\frac{2}{\sigma^2}|\boldsymbol{a}_m^H(\boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{x}}_k)| < \lambda_{km}.
$$
(53b)

APPENDIX B

The proof of Theorem 2 proceeds similar to Appendix B in [22]. The aim is to find the next singular point μ^{p+1} defined in (37). We rewrite the minimality condition (53a) as

$$
\boldsymbol{A}_{\mathcal{M}}^H(\boldsymbol{y}_k - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{x}_{\mathcal{M}}(\mu)) = \mu\frac{\sigma^2}{2}\boldsymbol{w}_{\mathcal{M}}.
$$
(54)

This is solved by

$$
\boldsymbol{x}_{\mathcal{M}}(\mu) = \boldsymbol{A}_{\mathcal{M}}^+\boldsymbol{y}_k - (\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\mu\frac{\sigma^2}{2}\boldsymbol{w}_{\mathcal{M}}.
$$
(55)

It is assumed that the phase of the solution is approximately constant in one μ-interval [16]. Then also $\boldsymbol{w}_{\mathcal{M}}$ is approximately constant for all $\mu \in (\mu^{p+1},\ \mu^p)$ and $\boldsymbol{x}_{\mathcal{M}}(\mu)$ depends approximately linearly on μ. Then, we have

$$
\boldsymbol{y}_k - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{x}_{\mathcal{M}}(\mu) = \boldsymbol{y}_k - \boldsymbol{A}_{\mathcal{M}}^{+H}\left(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{y}_k - \mu\frac{\sigma^2}{2}\boldsymbol{w}_{\mathcal{M}}\right)
$$
$$
= \boldsymbol{P}_{\boldsymbol{A}_{\mathcal{M}}}^\perp\boldsymbol{y}_k + \mu\frac{\sigma^2}{2}\boldsymbol{A}_{\mathcal{M}}^{+H}\boldsymbol{w}_{\mathcal{M}}
$$
(56)

where $\boldsymbol{P}_{\boldsymbol{A}_{\mathcal{M}}}^\perp$ is the orthogonal projection matrix defined in (43). For the *creation* case, according to (37) and (56), μ^+ is the smallest μ value, for which

$$
2|\boldsymbol{a}_m^H(\boldsymbol{P}_{\boldsymbol{A}_{\mathcal{M}}}^\perp\boldsymbol{y}_k + \mu\frac{\sigma^2}{2}\boldsymbol{A}_{\mathcal{M}}^{+H})| \leq \mu w_{km}\sigma^2
$$
(57)

for every index m. This implies that μ^+ is the largest value in the set of positive solutions of

$$
\mu^+ = \left|\frac{2\boldsymbol{a}_m^H\boldsymbol{P}_{\boldsymbol{A}_{\mathcal{M}}}^\perp\boldsymbol{y}_k}{\sigma^2 w_{km}} + \mu^+\frac{\boldsymbol{a}_m^H\boldsymbol{A}_{\mathcal{M}}^{+H}\boldsymbol{w}_{\mathcal{M}}}{w_{km}}\right|,
$$
(58)

for every m, which is solved by the Q-function in Appendix C. For the *annihilation* case, (37) and (55) is combined. After defining $\tilde{\boldsymbol{x}}_{\mathcal{M}}$ and $\tilde{\boldsymbol{w}}_{\mathcal{M}}$(47)–(48), we express (55) as

$$
\boldsymbol{x}_{\mathcal{M}}(\mu) = \tilde{\boldsymbol{x}}_{\mathcal{M}} - \mu\frac{\sigma^2}{2}\tilde{\boldsymbol{w}}_{\mathcal{M}}.
$$
(59)

Due to (37) annihilation occurs when $x_{\mathcal{M}l}(\mu) = 0$, which gives $\mu = 2\tilde{x}_{\mathcal{M}l}/(\sigma^2 \tilde{w}_{\mathcal{M}l})$ for some $l \in \{1, \ldots, R\}$. Thus, μ^- is the maximum of such annihilation points. This completes the proof.

APPENDIX C

In (45), the function $Q(\cdot, \cdot)$ is introduced. The real-valued function $Q(a, b) = \varrho$ of complex numbers a, b is defined as the positive root of $\varrho = |a + \varrho b|$. The resulting quadratic equation is

$$(|b|^2 - 1)\varrho^2 + 2\operatorname{Re}(a^*b)\varrho + |a|^2 = 0. \tag{60}$$

Its roots are

$$\varrho_\pm = -\frac{\operatorname{Re}(a^*b)}{|b|^2 - 1} \pm \sqrt{\left(\frac{\operatorname{Re}(a^*b)}{|b|^2 - 1}\right)^2 - \frac{|a|^2}{|b|^2 - 1}}. \tag{61}$$

If $|b|^2 > 1$ then $Q(a, b) := \varrho_+$ else $Q(a, b) := \varrho_-$.

ACKNOWLEDGMENT

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[2] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.

[3] D. M. Malioutov, C. Müjdat, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[5] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–120, Jul. 2007.

[7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[8] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[9] D. M. Malioutov, S. Sanghavi, and A. S. Willsky, "Sequential compressed sensing," *IEEE J. Sel. Top. Signal Process.*, vol. 4, pp. 435–444, 2010.

[10] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where rls meets the ℓ_1-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, 2010.

[11] D. Eiwen, G. Tauböck, F. Hlawatsch, and H. G. Feichtinger, "Compressive tracking of doubly selective channels in multicarrier systems based on sequential delay-Doppler sparsity," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2928–2931.

[12] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4967–4972, Sep. 2012.

[13] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[14] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[15] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.

[16] A. Panahi and M. Viberg, "Fast LASSO based DOA tracking," presented at the 4th Int. Workshop Comput. Adv. Multi-Sensor Adap. Process. (IEEE CAMSAP), Puerto Rico, San Juan, Dec. 2011.

[17] P. Gerstoft, C. F. Mecklenbräuker, and H. Yao, "Bayesian sparse sensing of the japanese 2011 earthquake," in *Proc. 46th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 4–7, 2012.

[18] H. Yao, P. Gerstoft, P. M. Shearer, and C. F. Mecklenbräuker, "Compressive sensing of Tohoku-oki M9.0 earthquake: Frequency-dependent rupture modes," in *Geophys. Res. Lett.*, 2011, vol. 38, p. L20310.

[19] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York, NY, USA: Wiley, 1994.

[20] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA, USA: Artech House, 2004.

[21] Z. He, S. Xie, S. Ding, and A. Cichocki, "Convolutive blind source separation in the frequency domain based on sparse representation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1551–1563, Jul. 2007.

[22] A. Panahi and M. Viberg, "Fast candidate points selection in the LASSO path," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 79–82, Feb. 2012.

[23] *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, Eds. New York, NY, USA: Dover, Dec. 1972, 10th printing with corrections.

[24] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, P. Gerstoft, and J. S. Kim, "Null broadening with snapshot-deficient covariance matrices in passive sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 2, pp. 250–261, Apr. 2003.

[25] H. Cox, "Multi-rate adaptive beamforming (MRABF)," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, Boston, MA, USA, Mar. 16–17, 2000, pp. 306–309.

[26] J. Traer and P. Gerstoft, "Coherent averaging of the passive fathometer response using short correlation time," *J. Acoust. Soc. Amer.*, vol. 130, no. 6, pp. 3633–3641, Dec. 2011.

[27] M. Grant and S. Boyd, CVX:Matlab software for disciplined convex programming, ver. 1.21, Apr. 2011 [Online]. Available: http://cvxr.com/cvx

[28] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*. New York, NY, USA: Springer-Verlag Ltd., 2008, pp. 95–110.

[29] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004, ch. 1–7.

[30] M. A. T. Figueiredo, R. D. Nowak, S. J. Wright, and J. Stephen, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[31] M. R. Osborne, B. Presnell, and A. Turlach, "On the LASSO and its dual," *J. Comput. Graph. Statist.*, vol. 9, no. 2, pp. 319–337, Feb. 2000.

[32] H. C. Song, S. Cho, T. Kang, W. S. Hodgkiss, and J. R. Preston, "Long-range acoustic communication in deep water using a towed array," *J. Acoust. Soc. Amer.*, vol. 129, no. 3, pp. EL71–EL75, Mar. 2011.

[33] P. Gerstoft, R. Menon, W. S. Hodgkiss, and C. F. Mecklenbräuker, "Eigenvalues of the sample covariance matrix for a towed array," *J. Acoust. Soc. Amer.*, vol. 132, no. 4, pp. 2388–2396, Oct. 2012.

[34] K. M. Becker and J. R. Preston, "The ONR five octave research array (FORA) at penn state," in *Proc. IEEE OCEANS*, Sep. 22–26, 2003, vol. 5, pp. 2607–2610.

Christoph F. Mecklenbräuker (S'88–M'97–SM'08) received the Dipl.-Ing. degree in electrical engineering from Technische Universität Wien, Austria, in 1992 and the Dr.-Ing. degree from Ruhr-Universität Bochum, Germany, in 1998. His doctoral dissertation received the Gert-Massenberg Prize in 1998. He was with Siemens, Vienna, from 1997 to 2000. From 2000 to 2006, he was senior researcher with the Forschungszentrum Telekommunikation Wien (FTW), Austria. In 2006, he joined the Institute of Telecommunications as full professor with the Technische Universität Wien, Austria. Since 2009 he leads the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility. His research interests include waves, sparsity, vehicular connectivity, ultrawideband radio, and MIMO-techniques. Dr. Mecklenbräuker is a member of the IEEE Signal Processing, Antennas and Propagation, and Vehicular Technology Societies, VDE and EURASIP.

Peter Gerstoft (M'13) received the Ph.D. from the Technical University of Denmark, Lyngby, Denmark, in 1986. From 1987–1992 he was with Ødegaard and Danneskiold-Samsøe, Copenhagen, Denmark. From 1992–1997 he was at Nato Undersea Research Centre, La Spezia, Italy. Since 1997, he has been with the Marine Physical Laboratory, University of California, San Diego. His research interests include modeling and inversion of acoustic, elastic and electromagnetic signals.

Dr. Gerstoft is a Fellow of Acoustical Society of America and an elected member of the International Union of Radio Science, Commission F.

Ashkan Panahi (S'11) received his BSc degree in electrical engineering from Iran University of Science and Technology (IUST) in 2008. He also received his MSc degree in 2010 in the field of communication Eng. from Chalmers University, Sweden, where he is currently pursuing his PhD studies. Being with the signal processing group, signals and systems department, his main research interest includes estimation and inverse problems, sparse techniques and information theory. His current research concerns the application of sparse techniques to the sensor array problems. Ashkan Panahi has been a student member of IEEE since 2011.

Mats Viberg (S'87–M'89–SM'95–F'03) received the PhD degree in Automatic Control from Linköping University, Sweden in 1989. He has held academic positions at Linköping University and visiting Scholarships at Stanford University and Brigham Young University, USA. Since 1993, Dr. Viberg is a professor of Signal Processing at Chalmers University of Technology, Sweden. During 1999–2004 he served as Department Chair. Since May 2011, he holds a position as First Vice President at Chalmers University of Technology.

Dr. Viberg's research interests are in Statistical Signal Processing and its various applications, including Antenna Array Signal Processing, System Identification, Wireless Communications, Radar Systems and Automotive Signal Processing. Dr. Viberg has served in various capacities in the IEEE Signal Processing Society, including chair of the Technical Committee (TC) on Signal Processing Theory and Methods (2001–2003), vice-chair (2009–2010) and chair (2011–2012) of the TC on Sensor Array and Multichannel, Associate Editor of the Transactions on Signal Processing (2004–2005), member of the Awards Board (2005–2007), and member at large of the Board of Governors (2010-). Dr. Viberg has received 2 Paper Awards from the IEEE Signal Processing Society (1993 and 1999 respectively), and the Excellent Research Award from the Swedish Research Council (VR) in 2002. Dr. Viberg is a Fellow of the IEEE since 2003, and his research group received the 2007 EURASIP European Group Technical Achievement Award. In 2008, Dr. Viberg was elected into the Royal Swedish Academy of Sciences (KVA).