Block-sparse beamforming for spatially extended sources in a Bayesian formulation

Xenaki, Angeliki; Fernandez Grande, Efren; Gerstoft, Peter

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/1.4962325

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Block-sparse beamforming for spatially extended sources in a Bayesian formulation

Angeliki Xenaki[a] and Efren Fernandez-Grande
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark (DTU), Kongens Lyngby, 2800 Denmark

Peter Gerstoft
Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA

(Received 14 April 2016; revised 28 July 2016; accepted 11 August 2016; published online 20 September 2016)

Direction-of-arrival (DOA) estimation refers to the localization of sound sources on an angular grid from noisy measurements of the associated wavefield with an array of sensors. For accurate localization, the number of angular look-directions is much larger than the number of sensors, hence, the problem is underdetermined and requires regularization. Traditional methods use an ℓ_2-norm regularizer, which promotes minimum-power (smooth) solutions, while regularizing with ℓ_1-norm promotes sparsity. Sparse signal reconstruction improves the resolution in DOA estimation in the presence of a few point sources, but cannot capture spatially extended sources. The DOA estimation problem is formulated in a Bayesian framework where regularization is imposed through prior information on the source spatial distribution which is then reconstructed as the maximum *a posteriori* estimate. A composite prior is introduced, which simultaneously promotes a piecewise constant profile and sparsity in the solution. Simulations and experimental measurements show that this choice of regularization provides high-resolution DOA estimation in a general framework, i.e., in the presence of spatially extended sources. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4962325]

[ZHM] Pages: 1828–1838

I. INTRODUCTION

The problem of direction-of-arrival (DOA) estimation in acoustic imaging refers to localizing sound sources from noisy measurements of the wavefield with an array of sensors. Since the spatial distribution of sources is unknown, the number of steering directions is typically much greater than the number of sensors on the array. Hence, DOA estimation is an ill-posed (underdetermined) inverse problem and regularization is essential to find a stable solution.

Optimization methods solve the DOA estimation problem as a least-squares parameter estimation problem and use constraints to regularize it. Traditional methods for DOA estimation[1] use an ℓ_2-norm constraint on the model parameters to achieve smooth solutions at the expense of low resolution. Regularization with the ℓ_2-norm is also typical in near-field acoustical holography (NAH) methods for sound field reconstruction.[2,3] Since there are usually only a few sources, regularization with the ℓ_1-norm[4,5] attracts increasing attention[6–16] as it provides high-resolution DOA reconstruction due to its sparsity promoting characteristics. In underwater acoustics, using multiple constraints to account for simultaneous characteristics of the solution as sparsity and smoothness (i.e., ℓ_1-norm and ℓ_2-norm) is shown to improve the resolution of shallow-water source localization with matched-field processing.[17] The inclusion of an ℓ_2-norm regularizer allows spatially extended sources to appear in the DOA map, that otherwise would be ignored by the ℓ_1-norm penalty alone.[18]

In a probabilistic formulation, regularization is imposed in the form of prior information on the model parameters. Bayesian estimation theory provides a systematic way to include prior constraints to the data fitting term. In fact, under the assumption of additive Gaussian noise, the maximum *a posteriori* (MAP) estimate is the optimal solution of a regularized least-squares problem. In statistics, the best known example for sparse signal reconstruction is the least absolute shrinkage and selection operator[19,20] (LASSO), which is the application of an ℓ_1-norm constraint to least-squares. In acoustic source localization, LASSO has superior performance compared to traditional methods, providing high-resolution DOA maps even in the presence of coherent sources or single-snapshot data.[21,22]

The LASSO in its standard form does not promote block sparsity (i.e., grouping of correlated locations), hence, it is not suitable for recovering multipole or spatially extended sources (e.g., edge-noise in aeroacoustic applications[23] or submerged objects in sonar imaging[24]). For such problems, which involve certain structural constraints rather than pure sparsity, a weighed LASSO[25,26] (WL) formulation is introduced. WL applies the ℓ_1-norm constraint on the model parameters weighed by a structured matrix $\mathbf{D}$, which reflects some assumed structure or geometry in the model. WL has a wide range of applications dictated by the choice of $\mathbf{D}$. Particularly, when $\mathbf{D}$ is a discrete differential operator, the corresponding WL problem is equivalent to a total variation (TV)

[a]Electronic mail: anxe@dtu.dk

regularization problem.[27] TV-norm regularization promotes piecewise constant solutions and is extensively used in image processing for denoising since it preserves the sharp edges of large features in the image while suppressing the smaller-scale noisy components.[28] To the best of our knowledge, TV-norm regularization has not been introduced in acoustics even though it is a common technique in ultrasound imaging.[29,30]

We derive the optimization problem for regularized DOA estimation in a statistical framework and describe the reconstructed DOA map as the MAP estimate in Bayesian inference. We give an overview of the qualitative characteristics of the MAP estimates for different regularizing prior distributions. We show that the fused LASSO[31] (FL), which is a special form of WL combining an ℓ_1-norm and a TV-norm constraint, outperforms established methods for DOA estimation in a generalized framework comprising spatially extended sources. This is verified both with simulations and experimental measurements.

In this paper, vectors are represented by bold lowercase letters and matrices by bold uppercase letters. The superscripts "^{T}" and "^{H}" denote the transpose and the Hermitian (i.e., conjugate transpose) operators, respectively, on vectors and matrices. The superscript "$^+$" denotes the generalized inverse operator on a matrix $\mathbf{A} \in \mathbb{C}^{R\times Q}$. The ℓ_p-norm of a vector $\mathbf{x} \in \mathbb{C}^Q$ is defined as $\|\mathbf{x}\|_p = (\sum_{q=1}^{Q} |x_q|^p)^{1/p}$. By extension, the ℓ_0-norm is defined as $\|\mathbf{x}\|_0 = \sum_{q=1}^{Q} 1_{x_q\neq 0}$. The infinity norm is defined as the maximum vector element in absolute value, $\|\mathbf{x}\|_\infty = \max\{|x_1|, \ldots, |x_Q|\}$. We use the CVX toolbox for disciplined convex optimization which uses interior point solvers to obtain the global solution of a well-defined optimization problem and is available in the MATLAB environment.[32]

II. SINGLE SNAPSHOT DOA ESTIMATION

We consider the simple one-dimensional (1D) problem with a uniform linear array (ULA) of sensors and the sources residing in the plane of the array. We assume that the sources are in the far-field of the array, i.e., plane wave propagation, and narrowband processing with a known sound speed. In this case, the location of a source is characterized by the direction of arrival of the associated plane wave, $\theta \in [-90^\circ, 90^\circ]$ with respect to the array axis (broadside angle).

The propagation delay from the ith potential source to each of the M array sensors is described by the forward (or replica) vector $\mathbf{a}(\theta_i) = e^{jk\sin\theta_i \mathbf{r}_s}$, where $k = 2\pi/\lambda$ is the wavenumber of the associated plane wave with wavelength λ and $\mathbf{r}_s = [0, 1, \ldots, M-1]^T d$ is the vector with the sensors' positions in a uniform configuration with intersensor distance d. The time convention $e^{j\omega t}$ is implied and neglected for simplicity. Hence, the forward vector is related to the DOA of the ith source as

$$\mathbf{a}(\theta_i) = \frac{1}{\sqrt{M}}\left[1, e^{j2\pi d/\lambda 1\sin\theta_i}, \ldots, e^{j2\pi d/\lambda (M-1)\sin\theta_i}\right]^T. \quad (1)$$

The normalization $1/\sqrt{M}$, such that $\|\mathbf{a}\|_2 = 1$, is to simplify the analysis.

Discretizing the half-space of interest, $\theta \in [-90^\circ, 90^\circ]$, into N angular directions the DOA estimation problem can be expressed with the linear model,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{y} \in \mathbb{C}^M$ is the complex-valued data vector from the measurements at the M sensors, $\mathbf{x} \in \mathbb{C}^N$ is the unknown vector of the complex plane wave amplitudes at all N directions on the angular grid of interest and $\mathbf{n} \in \mathbb{C}^M$ is the additive noise vector. The sensing matrix,

$$\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_N)], \quad (3)$$

maps the signal $\mathbf{x}$ to the observations $\mathbf{y}$ and has as columns the vectors [Eq. (1)] at all steering directions.

In the following, the complex wave amplitudes in $\mathbf{x}$ are modeled with deterministic amplitude and random phase uniformly distributed in $[0,2\pi)$ (alternatively they can be modeled as random variables following a complex Gaussian distribution $\mathbf{x} \sim \mathcal{CN}(0, \sigma_x^2\mathbf{I})$). The noise is generated as independent and identically distributed (iid) complex Gaussian. The array signal-to-noise ratio (SNR) is used, defined as

$$\mathrm{SNR} = 10\log_{10}\frac{\mathrm{E}\left\{\|\mathbf{A}\mathbf{x}\|_2^2\right\}}{\mathrm{E}\left\{\|\mathbf{n}\|_2^2\right\}}\ (\mathrm{dB}). \quad (4)$$

The problem of DOA estimation is to recover the source vector $\mathbf{x} \in \mathbb{C}^N$, given the sensing matrix $\mathbf{A}_{M\times N}$ and an observation vector $\mathbf{y} \in \mathbb{C}^M$. Usually, we are interested in a fine resolution on the angular grid to achieve precise localization (even if there are only a few sources $K < M$ generating the acoustic field) such that $M \ll N$ and the problem in Eq. (2) is underdetermined, i.e., has infinitely many solutions. A way to regularize this ill-posed inverse problem is constraining the possible solutions with prior information.

III. DOA ESTIMATION VIA BAYESIAN INFERENCE

DOA estimation as a regularized parameter estimation problem[1] can be interpreted in a statistical Bayesian setting by treating both the unknowns $\mathbf{x}$ and the observations $\mathbf{y}$ as stochastic (random) processes, and selecting a prior distribution on the vector $\mathbf{x}$ to impose the desired regularization.

Concisely, Bayes theorem connects the posterior distribution $p(\mathbf{x}|\mathbf{y})$ of the model parameters $\mathbf{x}$ conditioned on the data $\mathbf{y}$ with the data likelihood $p(\mathbf{y}|\mathbf{x})$, the prior distribution of the model parameters $p(\mathbf{x})$, and the marginal distribution of the data $p(\mathbf{y})$,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (5)$$

The MAP estimate, omitting $p(\mathbf{y})$ as it is independent of the model $\mathbf{x}$, is

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{MAP}} &= \underset{\mathbf{x}}{\mathrm{argmax}}\ \ln p(\mathbf{x}|\mathbf{y}) \\ &= \underset{\mathbf{x}}{\mathrm{argmin}}\ [-\ln p(\mathbf{y}|\mathbf{x}) - \ln p(\mathbf{x})]. \end{aligned} \quad (6)$$

The MAP estimate [Eq. (6)] can be used for DOA reconstruction combining information both from the observed data through the likelihood distribution and from prior knowledge on the joint distribution of the model parameters (i.e., the complex wave amplitudes), which is by definition independent from the observations.

The data likelihood is based on the noise model, which is assumed complex Gaussian with iid real and imaginary parts $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ [see Eq. (2)]. Hence, the data likelihood is also complex Gaussian distributed $p(\mathbf{y}|\mathbf{x}) \sim \mathcal{CN}(\mathbf{Ax}, \sigma^2\mathbf{I})$,

$$p(\mathbf{y}|\mathbf{x}) \propto \mathrm{e}^{-\left(\|\mathbf{y}-\mathbf{Ax}\|_2^2\right)/\sigma^2}. \tag{7}$$

The assumption of a Gaussian likelihood [Eq. (7)] results in a least-squares estimation problem,

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}}{\operatorname{argmin}} \ \|\mathbf{y} - \mathbf{Ax}\|_2^2 - \sigma^2 \ln p(\mathbf{x}), \tag{8}$$

where the prior distribution $p(\mathbf{x})$ functions as a regularization term, which conditions the inverse problem [Eq. (8)]. We can introduce a composite prior (comprising Q priors) to improve the estimation accuracy

$$p(\mathbf{x}) \propto \prod_{i=1}^{Q} p_i(\mathbf{x}). \tag{9}$$

For now, we employ a general expression for the priors $p_i(\mathbf{x})$ through the weighed multivariate generalized complex Gaussian distribution,[33]

$$p_i(\mathbf{x}) \propto \mathrm{e}^{-\left(\|\mathbf{D}_i\mathbf{x}\|_{p_i}/\nu_i\right)^{p_i}}, \tag{10}$$

where $\mathbf{D}_i \in \mathbb{R}^{P\times N}$ is a weighting matrix that can be used to enforce structural or geometric constraints on the solution, $\nu_i \in \mathbb{R}_+$ is the scaling parameter and $p_i \in \mathbb{R}_+$ is the shape parameter.

Inserting Eqs. (9) and (10), the regularized least-squares problem (8) is expressed as

$$\hat{\mathbf{x}}_{\mathrm{GG}}(\mu_1, \ldots, \mu_Q) = \underset{x}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \sum_{i=1}^{Q} \mu_i \|D_i x\|_{p_i}^{p_i}, \tag{11}$$

where $\mu_i = \sigma^2/\nu_i^{p_i} \geq 0$ is the regularization parameter of the ith prior, which controls the relative importance between the data fit and the ith regularization term. The characteristics of the MAP estimate depend on the choice of the number of regularization terms Q, the shape parameters p_i, the weighting matrices $\mathbf{D}_i$, and the regularization paremeters μ_i.

A. Maximum likelihood

The MAP estimate in its simplest form corresponds to the maximum likelihood (ML) estimate, i.e., the vector $\mathbf{x}$ that best fits the data,[1]

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{ML}} &= \underset{\mathbf{x}}{\operatorname{argmin}} \ \|\mathbf{y} - \mathbf{Ax}\|_2^2 \\ &= \mathbf{A}^{+}\mathbf{y} + \mathbf{x}_{\mathrm{null}},\end{aligned} \tag{12}$$

where $\mathbf{x}_{\mathrm{null}} \in \mathrm{N}(A)$ is a null solution, i.e., $\mathbf{Ax}_{\mathrm{null}} = 0$.

The ML estimate [Eq. (12)] neglects prior information, i.e., considering Eq. (11) with $\mu_i = 0$ for all $i \in [1,\ldots,Q]$. The resulting optimization problem (12) corresponds to an unregularized least-squares problem, which is convex with analytic solution, hence, simple to solve. However, the corresponding MAP estimate $\hat{\mathbf{x}}_{\mathrm{ML}}$ depends solely on optimizing the data fit, assuming no specific prior knowledge on the model $\mathbf{x}$. Thus, it tends to be unstable,[34] especially for underdetermined problems $M < N$ that $\mathbf{A}$ has a non-trivial null space, i.e., $\mathbf{x}_{\mathrm{null}} \neq 0$.

B. CBF

Traditional methods in DOA estimation assume an iid multivariate complex Gaussian prior distribution $\mathbf{x} \sim \mathcal{CN}(0, \nu^2\mathbf{I})$ for the model parameters, which promotes smooth solutions[35] [i.e., Eq. (11) with $Q = 1$, $\mathbf{D} = \mathbf{I}_N$, and $p = 2$],

$$\begin{aligned}\hat{\mathbf{x}}_{\ell_2}(\mu) &= \underset{\mathbf{x}}{\operatorname{argmin}} \ \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \mu\|\mathbf{x}\|_2^2 \\ &= \mathbf{A}^H(\mathbf{AA}^H + \mu\mathbf{I}_M)^{-1}\mathbf{y}.\end{aligned} \tag{13}$$

Problem (13) is an ℓ_2-norm regularized least-squares problem, which is convex and has an analytic solution, thus, it is simple to solve. The estimate [Eq. (13)] has improved robustness to noise compared to Eq. (12) as the ℓ_2-norm constraint penalizes the energy in the solution (forcing $\mathbf{x}_{\mathrm{null}} = 0$).

Conventional beamforming (CBF)[1] is related to the ℓ_2-norm estimate for large μ,[22]

$$\hat{\mathbf{x}}_{\mathrm{CBF}} = \lim_{\mu\to\infty}(\mu\hat{\mathbf{x}}_{\ell_2}(\mu)) = \mathbf{A}^H\mathbf{y}. \tag{14}$$

In principle, CBF combines the sensor outputs coherently to enhance the signal at a specific look direction from the ubiquitous noise. CBF is robust to noise but suffers from low resolution and the presence of sidelobes.

C. LASSO

Many DOA estimation scenarios involve few spatially localized sources (there are only $K \ll N$ sources). In such cases, the solution vector $\mathbf{x}$ is inherently sparse, i.e., comprises only a few non-zero components, and an appropriate regularizer is the ℓ_0-norm, which counts the number of non-zero entries in the vector, to find a sparse solution. However, the resulting ℓ_0-norm minimization problem is a non-convex combinatorial problem, which becomes computationally intractable even for moderate dimensions. Compressive sensing (CS)[4,5] offers a remedy by proving that for sufficiently sparse signals, $K \ll N$, and sensing matrices $\mathbf{A}$ with sufficiently incoherent columns, the ℓ_0-norm minimization problem is equivalent (at least in the noiseless case) to its convex relaxation, the ℓ_1-norm minimization problem.[36,37]

By replacing the ℓ_0-norm with the convex ℓ_1-norm, the problem can be solved efficiently with convex optimization even for large dimensions.[32,38]

The ℓ_1-norm minimization problem can be formulated in the Bayesian framework[19,20,39] assuming that the coefficients of the solution $\mathbf{x}$ are iid and follow a Laplacian-like distribution (for complex random variables),[40]

$$p(\mathbf{x}) \propto \prod_{i=1}^{N} \mathrm{e}^{-[\sqrt{(\mathrm{Re}\{x_i\})^2+(\mathrm{Im}\{x_i\})^2}]/\nu} = \mathrm{e}^{-\|\mathbf{x}\|_1/\nu}. \tag{15}$$

Assuming a Laplacian prior [Eq. (15)], the MAP estimate is the solution to an ℓ_1-norm regularized least-squares problem [i.e., Eq. (11) with $Q=1$, $\mathbf{D}=\mathbf{I}_N$, and $p=1$), and is known as the least absolute shrinkage and selection operator[19] (LASSO),

$$\hat{\mathbf{x}}_{\mathrm{LASSO}}(\mu) = \underset{\mathbf{x}}{\mathrm{argmin}} \, \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\|\mathbf{x}\|_1, \tag{16}$$

since the ℓ_1 regularizer shrinks the coefficients of $\mathbf{x}$ toward zero as the regularization parameter $\mu=\sigma^2/\nu$ increases; see Fig. 1.

Opposed to a Gaussian prior, which promotes smooth solutions with minimum energy, the Laplacian-like prior distribution encourages sparse solutions since it concentrates more mass near 0 and in the tails. Thus, the parsimonious LASSO estimate improves significantly the resolution in DOA estimation in the presence of only a few sources as shown in Ref. 22.

D. WL

The LASSO optimization problem, Eq. (16), encourages sparse solutions with few non-zero coefficients through the ℓ_1-norm regularization term. By extension, the ℓ_1-norm can be used to enforce more general structural or geometric constraints on the solution by replacing the sparsity constraint $\|\mathbf{x}\|_1$ with $\|\mathbf{D}\mathbf{x}\|_1$ for a structured matrix $\mathbf{D}$. The resulting optimization problem is known as the WL problem[25]

$$\hat{\mathbf{x}}_{\mathrm{WL}}(\mu) = \underset{\mathbf{x}}{\mathrm{argmin}} \, \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\|\mathbf{D}\mathbf{x}\|_1. \tag{17}$$

The generalized formulation performs well in applications where the solution exhibits spatial correlation, e.g., DOA tracking for moving sources by an adaptive update of a diagonal weighting matrix $\mathbf{D}$, which reflects the evolution of the source probability distribution[21] and recovery of continuous sources by promoting block sparsity.[31]

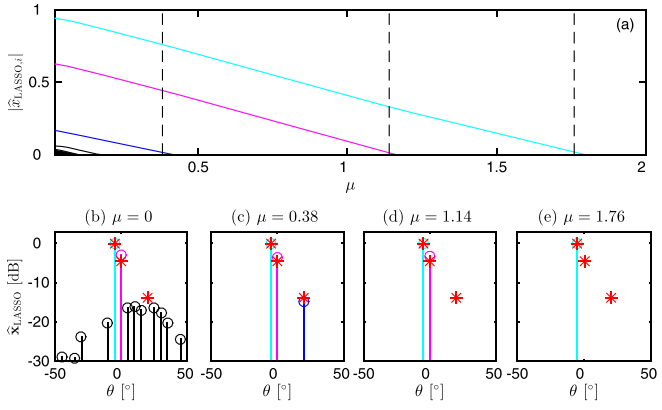

FIG. 1. (Color online) Three sources (*) at DOAs $[-5,0,20]^\circ$ with amplitudes [1,0.6,0.2] (or [0,−4,−14] dB), respectively, reconstructed on an angular grid $[-90{:}5{:}90]^\circ$ with a ULA of $M=20$ sensors separated by $d=\lambda/2$ at SNR = 20 dB (as in Ref. 22). (a) The amplitude of each component of the LASSO estimate (16) versus μ. The ℓ_1-norm regularizer shrinks the coefficients of $\mathbf{x}$ toward zero as the regularization parameter μ increases. The solution vector $\hat{\mathbf{x}}_{\mathrm{LASSO}}$ for (b) $\mu=0$, (c) $\mu=0.38$, (d) $\mu=1.14$, (e) $\mu=1.76$, denoted in (a) with dashed vertical lines.

E. TV-norm regularization

A particular formulation of WL [Eq. (17)], which is of interest in applications concerned with DOA estimation of extended sources,[23,24] uses the band matrix $\mathbf{D}=\mathbf{D}_{\mathrm{TV}} \in \mathbb{R}^{(N-1)\times N}$ defined as

$$\mathbf{D}_{\mathrm{TV}} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{18}$$

to promote block sparsity, i.e., sparsity not directly on the coefficients of the solution but rather on their successive differences (few kinks in the solution vector). From a Bayesian perspective, this formulation [Eq. (11) with $Q=1$, $\mathbf{D}=\mathbf{D}_{\mathrm{TV}}$, and $p=1$] assumes a Laplacian-like prior distribution on the coefficients differences of the model [compare with Eq. (15)],

$$\begin{aligned} p(\mathbf{x}) &\propto \prod_{i=1}^{N-1} \mathrm{e}^{-[\sqrt{(\mathrm{Re}\{x_{i+1}-x_i\})^2+(\mathrm{Im}\{x_{i+1}-x_i\})^2}]/\nu} \\ &= \mathrm{e}^{-\|\mathbf{D}_{\mathrm{TV}}\mathbf{x}\|_1/\nu}. \end{aligned} \tag{19}$$

The resulting MAP estimate is

$$\begin{aligned} \hat{\mathbf{x}}_{\mathrm{TV}}(\mu) &= \underset{\mathbf{x}}{\mathrm{argmin}} \, \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\sum_{i=1}^{N-1}|x_{i+1}-x_i| \\ &= \underset{\mathbf{x}}{\mathrm{argmin}} \, \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\|\mathbf{D}_{\mathrm{TV}}\mathbf{x}\|_1. \end{aligned} \tag{20}$$

Note that the regularization term on the coefficient differences is a 1D discrete TV norm.[27] TV-norms are essentially ℓ_1-norms of derivatives, $\|\nabla\mathbf{x}\|_1$, and are used to promote piecewise constant solutions.[41] We use the subscript "$_{\mathrm{TV}}$" to indicate this correspondence.

We demonstrate the block-sparsity characteristics of Eq. (20) as a function of the regularization parameter μ with a simulation example. We consider a ULA with $M=20$ sensors and spacing $d=\lambda/2$. The source distribution consists of two extended sources covering the DOA intervals ($[-30, -20]^\circ$,$[0, 20]^\circ$) with amplitudes [−6,0] dB re max. Each continuous source is modeled as a series of monopoles with equal complex amplitude (deterministic amplitude and uniformly distributed phase). The noise variance is chosen such

that SNR = 20 dB. To facilitate the demonstration, the sensing matrix **A** is defined on a coarse angular grid [–90:5:90]°, i.e., $N = 37$.

Figure 2 depicts the solution path (20), i.e., the course of each of the 37 source coefficients as a function of μ, along with DOA maps for selected values of μ. Since we are interested in block sparsity, it is natural to inspect the solution path from right to left, i.e., for decreasing values of μ. For large values of μ (e.g., $\mu = 9$), the TV problem [Eq. (20)] is over-regularized and all the coefficients are fused into one group [Fig. 2(e)]. For smaller μ, the grouped coefficients gradually unfuse into more groups [two groups for $\mu = 5$, Fig. 2(d) and three groups for $\mu = 1$, Fig. 2(c)]. Finally, for $\mu = 0$, i.e., the unregularized solution [Eq. (12)], all solution coefficients are ungrouped into their individual amplitudes.

The solution path in Fig. 2 follows the edge-preserving and scale-dependent properties of TV;[28] the edge location of the solution-features is generally preserved for appropriate values of the regularization parameter but the dynamic range is reduced according to the size of the features [see Fig. 2(c)]. Hence, TV regularization promotes a piecewise constant solution such that the solution coefficients are closely related to their neighbors, while single noisy peaks are suppressed. Note that all components of the solution are active along the full path (compare with Fig. 1), i.e., TV-norm regularization does not promote sparsity on individual coefficients of the solution. To achieve pure sparsity apart from structural sparsity an additional ℓ_1-norm constraint is required.

TV-norm regularization promotes piecewise constant solutions by penalizing the first derivative through the band matrix [Eq. (18)]. Even though we limit this study to recovering spatial source distributions characterized by block-sparsity through such piecewise constant fit, the WL framework accommodates to solutions fitting a more general polynomial trend, e.g., smooth spatial source distributions without discontinuities at the boundaries. For example, defining the weighing matrix **D** in Eq. (17) as a discrete nth-order derivative operator[25] promotes solutions with $(n-1)$-degree polynomial trend (e.g., linear,[42] quadratic, cubic).

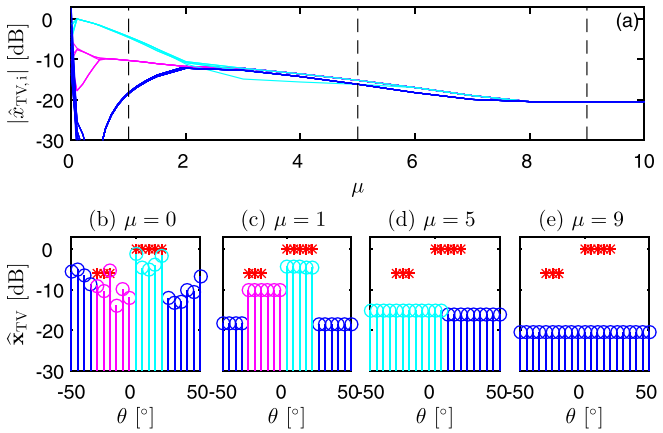

FIG. 2. (Color online) (a) The amplitude of each component of the TV-norm regularized least-squares solution for varying values of the regularization parameter μ. The TV-norm regularizer fuses the coefficients of **x** into groups of equal amplitude as the regularization parameter μ increases. The solution vector $\hat{\mathbf{x}}_{TV}$ for (b) $\mu = 0$, (c) $\mu = 1$, (d) $\mu = 5$, (e) $\mu = 9$, denoted in (a) with dashed vertical lines.

F. FL

The FL[31] combines a LASSO prior [Eq. (15)] with a TV-norm prior [Eq. (19)] to promote simultaneously sparsity in the coefficients (i.e., few active coefficients in the solution) and sparsity in the difference of successive coefficients (i.e., flatness in the coefficient profile)

$$\hat{\mathbf{x}}_{FL}(\mu_{TV}, \mu_1) = \underset{\mathbf{x}}{\operatorname{argmin}} [\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mu_{TV}\|\mathbf{D}_{TV}\mathbf{x}\|_1 + \mu_1\|\mathbf{x}\|_1]. \quad (21)$$

The combination of an ℓ_1-norm and a TV-norm regularizer makes FL suitable for applications that involve block sparsity or mixtures of spikes and flat plateaus, e.g., promoting coefficient sparsity (point sources) along with block sparsity (extended sources).

The FL optimization problem, Eq. (21), can be cast into a WL formulation, Eq. (17), by defining $\mu = \mu_{TV}$ and $\mathbf{D} \in \mathbb{R}^{P \times N}$ as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{TV} \\ (\mu_1/\mu_{TV})\mathbf{I}_N \end{bmatrix}, \quad (22)$$

where $\mathbf{D}_{TV}$ is the band matrix in Eq. (18), hence, $P = (N-1) + N = 2N - 1$.

1. FL solution path

We demonstrate the behavior of the FL estimate [Eq. (21)], as a function of the regularization parameters μ_{TV} and μ_1 when both point and continuous sources are present (Fig. 3). For that, we employ the setup of Sec. III E on a refined 1° angular grid and we replace the continuous source covering the DOA interval $[-30,-20]°$ with two point sources at DOAs $[-30,-20]°$ with amplitudes [−3,−6] dB re max, respectively, and random uniformly distributed phase.

Figure 3 depicts the FL estimate on a grid of values [0,0.1,1] for the regularization parameters $[\mu_{TV}, \mu_1]$. For $\mu_{TV} = \mu_1 = 0$, we obtain the least-squares solution [Eq. (12), Fig. 3(a)], which is characterized by low resolution while it fails to reconstruct both point sources as they are buried under the maximum sidelobe level. The LASSO estimates [Eq. (16), Figs. 3(d) and 3(g)] are sparse, allowing only few peaks in the DOA map. Even though point sources are favored and reconstructed more accurately with the standard LASSO, the continuous source is estimated as a collection of much fewer monopoles than the actual distribution (only 6 point sources reconstructed out of 21 within the range of the continuous source). Contrarily, the TV-norm estimates [Eq. (20), Figs. 3(b) and 3(c)] promote a piecewise constant profile. In this case, the continuous source is reconstructed rather accurately, but the point sources are fused into groups with neighboring DOAs. When $\mu_{TV} > 0$ and $\mu_1 > 0$ simultaneously, the resulting FL solution [Eq. (21), Figs. 3(e), 3(f), 3(h), and 3(i)] combines the characteristics

of the LASSO and TV priors, each to an extent determined by the corresponding regularization parameter. For example, for $\mu_{\mathrm{TV}} > \mu_1$ group-sparsity outbalances coefficient-sparsity in the FL estimate (few groups, no spikes), Fig. 3(f), and vice versa for $\mu_1 > \mu_{\mathrm{TV}}$ (many groups, few spikes), Fig. 3(h). For the considered setup, choosing $\mu_1 = 1$ and $\mu_{TV} = 0.1$ gives the best reconstruction, Fig. 3(h).

The solution path in Fig. 3 shows the qualitative characteristics of the FL solution as a function of μ_{TV} and μ_1. It does not intend to provide an exhaustive search over a wide range of values for the regularization parameters. Depending on the relative importance between block and coefficient sparsity for a specific application, the regularization parameters can be tuned accordingly through the solution path, following a similar procedure to the one described in Fig. 3. After a range of relevant values for the regularization parameters is determined through this procedure, a refined search on the regularization parameters may improve the reconstruction accuracy.

2. Dual path

The dual equivalent of an optimization problem[38] is often solved instead of the primal optimization problem either to facilitate the practical implementation due to reduced computational complexity[15] or to offer a more intuitive insight in the optimization process, e.g., for regularization parameter selection.[22]

In Ref. 22, the LASSO problem [Eq. (16)] is solved in the dual domain, which yields $\|2\mathbf{A}^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})\|_\infty \leq \mu$. This primal-dual correspondence dictates that for every active element $\hat{x}_i \neq 0$ of the primal solution $\hat{\mathbf{x}}$, the corresponding element $2\mathbf{a}_i^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})$ of the dual vector $2\mathbf{A}^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})$ has amplitude equal to μ. Thus, gradually reducing μ from an initial value of $\mu = 2\|\mathbf{A}^H\mathbf{y}\|_\infty$, the number of elements in the dual vector with amplitude equal to μ increases. For a given sparsity level, this procedure finds the locations (indexes) of the active components through the dual path (i.e., the evolution of the coefficients of the dual vector for several values of the regularization parameter μ).

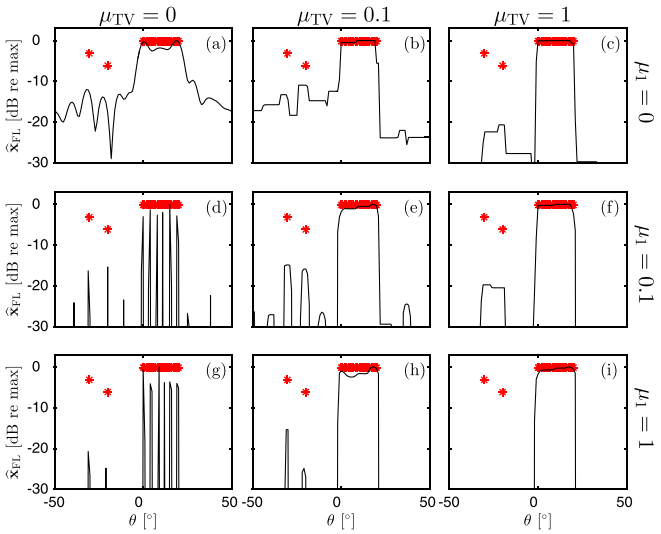

FIG. 3. (Color online) The FL solution vector (21) for a set of values of the regularization parameters μ_1 and μ_{TV}.

When $\mathbf{D}_{P\times N}$ has $\mathrm{rank}(\mathbf{D}) = P$ (hence, $P \leq N$), the WL problem can be reduced to the standard LASSO problem through a transformation of variables and solved through the dual path. However, when $\mathrm{rank}(\mathbf{D}) < P$ such a transformation is not possible.[25] Hence, for the FL formulation the corresponding dual problem[43] (see the Appendix) offers no advantage. The FL dual is an underdetermined problem without analytic solution, which can be solved with convex optimization over $P = 2N - 1$ dual optimization variables, i.e., it has greater computational complexity than the primal (N primal optimization variables).

Due to the large number of dual variables, it is not straightforward to select the appropriate regularization parameter from the FL dual path. Hence, we focus on the properties of the primal path (i.e., the evolution of the coefficients of the primal vector for several values of the regularization parameter μ); see Sec. III F 1.

3. Maximum number of resolvable DOAs

The maximum number of resolvable DOAs through the standard LASSO problem [Eq. (16)], cannot exceed the rank of the sensing matrix, $\mathrm{rank}(\mathbf{A}) = \min\{M, N\} = M$, Ref. 25,

$$K_{\mathrm{LASSO}} = \|\hat{\mathbf{x}}_{\mathrm{LASSO}}\|_0 \leq M, \tag{23}$$

while K_{LASSO} is even lower for exact reconstruction.[15] The limit imposed on the maximum number of resolvable DOAs by the number of sensors on the array makes standard LASSO impractical for applications involving continuous sources due to misleading source representation.

For FL the M limit refers to the number of non-zero coefficient blocks rather than the number of non-zero coefficients,[31]

$$K_{\mathrm{FL}} = \|\mathbf{D}\hat{\mathbf{x}}_{\mathrm{FL}}\|_0 \leq M, \tag{24}$$

increasing the total number of resolvable DOAs. Hence, FL offers the capacity to reconstruct a larger number of DOAs than the standard LASSO problem (as long as at least some of them are grouped in blocks).

4. Spatial coherence of extended sources

The effect of source spatial coherence to the FL reconstruction is depicted in Fig. 4. Figure 4(a) shows the FL estimate for coherent sources based on the simulation setup of Sec. III E on a 1° grid, where each of the two sources have equal complex amplitude along their extent (e.g., vibrating surfaces). Figure 4(b) considers that each monopole on each of the two considered sources is driven by an incoherent mechanism such that is modeled with constant amplitude but random, uniformly distributed phase (e.g., aeroacoustic noise on wind-blades).

The more complicated the phase pattern is in the distributed arrangement of an extended source, the less robust the amplitude estimation, as is the case when the monopoles are modeled with random phase, Fig. 4(b). This is due to the fact that the TV regularization term $\|\mathbf{D}_{\mathrm{TV}}\mathbf{x}\|_1$ applies the

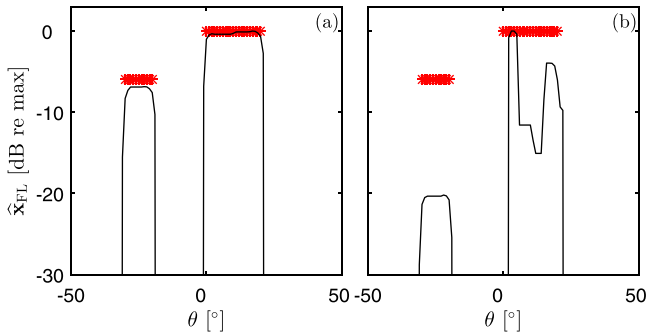

FIG. 4. (Color online) FL reconstruction ($\mu_{\mathrm{TV}} = \mu_1 = 1$) of spatially extended (a) coherent and (b) incoherent sources.

multiplexing matrix $\mathbf{D}_{\mathrm{TV}}$ to the complex vector $\mathbf{x}$ rather than to its amplitude $|\mathbf{x}|$. However, replacing the TV constraint in Eq. (21) with $\| \mathbf{D}_{\mathrm{TV}}|\mathbf{x}| \|_1$ leads to a non-convex problem, and its solution is out of the scope of this paper.[44]

G. Elastic net

Another optimization method proposed to achieve group selection of correlated variables (thus, promote block sparsity) is the elastic net[45] (EN), which combines an ℓ_2-norm and an ℓ_1-norm regularizer to achieve a balance between smoothness and sparsity[17]

$$\hat{\mathbf{x}}_{\mathrm{EN}}(\mu_2, \mu_1) = \underset{\mathbf{x}}{\operatorname{argmin}}[\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mu_2\|\mathbf{x}\|_2^2 + \mu_1\|\mathbf{x}\|_1]. \tag{25}$$

Problem (25) can be reformulated as

$$\hat{\mathbf{x}}_{\mathrm{EN}}(\mu_2, \mu_1) = \underset{\mathbf{x}}{\operatorname{argmin}}\ \|\mathbf{y}_{\mathrm{EN}} - \mathbf{A}_{\mathrm{EN}}\mathbf{x}\|_2^2 + \mu_1\|\mathbf{x}\|_1, \tag{26}$$

where $\mathbf{y}_{\mathrm{EN}} = [\mathbf{y}; 0_{(N-M)\times 1}]$ and $\mathbf{A}_{\mathrm{EN}} = [\mathbf{A}; \mu_2\mathbf{I}_{(N-M)\times N}]$. Note that Eq. (26) is a standard LASSO formulation where the ℓ_2-norm regularization is incorporated into the sensing matrix $\mathbf{A}_{\mathrm{EN}} \in \mathbb{C}^{N\times N}$.

In a Bayesian interpretation, the EN problem combines a Gaussian and a Laplacian-like prior. The relative importance of each prior is controlled by the regularization parameters $\mu_2 = \sigma^2/\nu^2$ and $\mu_1 = \sigma^2/\nu$, which are both dependent on the relative noise level. Regularization with the ℓ_1-norm often improves accuracy in least-squares estimation over ℓ_2-norm regularization, trading-off decreased variance for increased bias.[14] EN optimization was introduced in order to alleviate the bias-variance trade-off through accurate regularization. EN optimization is useful for applications that involve groups of highly correlated variables and improves the resolution compared to ℓ_2-norm regularization as depicted in the solution path of Fig. 5. However, EN regularization does not promote a piecewise constant coefficient profile but rather a smoothed solution; compare the corresponding estimates in Figs. 3 and 5.

H. Regularization parameter selection

The characteristics of the MAP estimate [Eq. (11)] and, by extension, the quality of the source reconstruction depend strongly on the regularization parameters μ_i as indicated in Figs. 1, 2, 3, and 5. The choice of the regularization parameters $\mu_i = \sigma^2/\nu_i$ is dictated by the noise level relative to the source strengths, which are, normally, unknown. This makes the regularization parameter selection non-trivial and partly explains the wide application of the parameter-free CBF method [Eq. (14)].

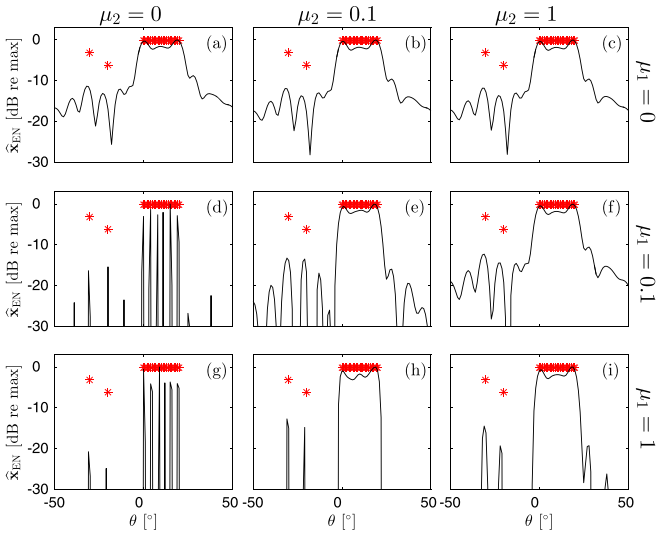

FIG. 5. (Color online) The EN solution vector (25) for a set of values of the regularization parameters μ_1 and μ_2.

There are several methods for regularization parameter selection for the ℓ_2-norm regularized least-squares problem.[2,34] For problems with sparsifying regularization terms, as the LASSO and TV-norm regularization, the regularization parameter is indicated by either the solution path or the dual path kinks.[22,25] However, the accuracy of the existing methods depends on the configuration of the individual problem (i.e., the spatial source distribution). Regularization parameter tuning becomes more involved with multiple regularization terms as the solution path involves a two-dimensional search (see Sec. III F 1) while the dual path is non-informative (see Sec. III F 2).

A study on regularization parameter selection methods would deserve a dedicated analysis. An interesting approach is based on a hierarchical Bayes model, which offers evidence-based regularization parameter selection (i.e., maximizing a type-II likelihood).[46–48] The focus of this paper is to compare qualitatively different regularization methods; thus, for a given SNR we present the results with *ad hoc* regularization parameters.

I. Comparison of regularization methods

Figure 6 compares the MAP estimates in Eqs. (14), (16), (20), (21), and (25) for the simulation setup of Sec. III E, on a refined 1° angular grid. To facilitate the comparison, all regularization parameters are set to 1, $\mu = \mu_1 = \mu_2 = \mu_{\mathrm{TV}} = 1$. The CBF solution [Fig. 6(a)] is smooth, spreading energy along the whole angular spectrum, which is characteristic of a Gaussian prior. Contrary, the LASSO solution [Fig. 6(b)] is sparse, comprising only a few non-zero peaks, as imposed from the Laplacian-like prior. The TV-norm solution

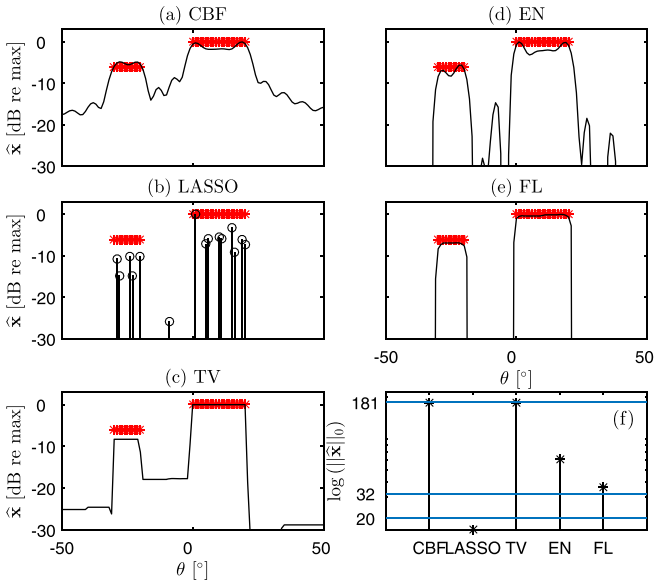

FIG. 6. (Color online) Comparison of the MAP estimate (11) conditioned with different priors setting $\mu_i = 1$ for all $i \in [1,\ldots,Q]$. (a) CBF [Eq. (14)], (b) LASSO [Eq. (16)], (c) TV [Eq. (20)], (d) EN [Eq. (25)], and (e) FL [Eq. (21)]. (f) Sparsity level of each MAP estimate, where the dimensions of the problem $M = 20$ and $N = 181$ as well as the sparsity level of the true solution $\|\mathbf{x}\|_0 = 32$ are also indicated.

[Fig. 6(c)] has the characteristic piecewise constant profile with reduced dynamic range (Laplacian-like prior for the model's coefficients difference). The EN [Fig. 6(d)] and the FL [Fig. 6(e)] solutions are based on composite priors. Hence, the EN solution combines the smoothness and sparsity characteristics of CBF and LASSO, respectively, while the FL solution combines the piecewise constant profile and sparsity characteristics of TV and LASSO accordingly. Even though all estimates indicate concentration of energy along the regions of DOAs covered by the assumed extended sources, the FL solution recovers almost perfectly the exact solution.

The sparsity level of each MAP estimate (number of non-zero coefficients in the solution) is depicted in Fig. 6(f). CBF and TV estimates are not sparse, attributing energy to all N coefficients in the DOA grid, while the LASSO, EN, and FL solutions, which incorporate an ℓ_1-norm regularizer, exhibit a degree of sparsity, hence, improved resolution. The sparsity of the LASSO solution does not exceed M as dictated by Eq. (23), thus, LASSO cannot recover accurately spatially extended sources. The EN and FL solutions allow the degree of sparsity to exceed M.

The comparison indicates that FL, which is designed for problems that exhibit block sparsity,[31] is a suitable optimization problem for DOA estimation and sound source reconstruction involving spatially extended sound sources.

IV. EXPERIMENTAL RESULTS

The suitability of FL for reconstructing spatially extended sources due to the block sparsity characteristics of the FL solution is validated with experimental data in anechoic conditions and compared to other regularization methods as in Sec. III I.

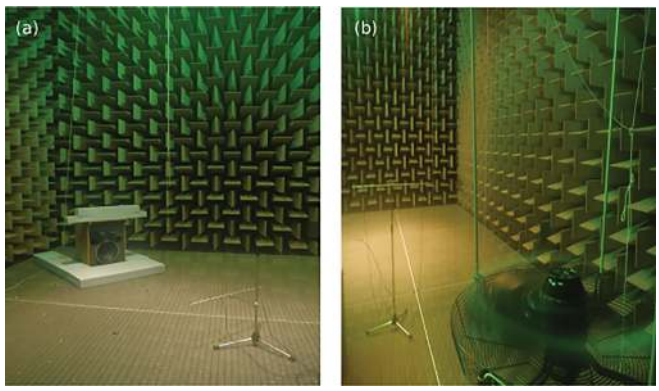

FIG. 7. (Color online) Experimental setup. Measurements with a ULA for DOA estimation of (a) a loudspeaker source and (b) aeroacoustic noise from a fan.

Two spatially extended sources are tested (Fig. 7); a loudspeaker with driver diameter 0.3 m as a coherent piston-like source and a fan with diameter 0.7 m as an incoherent aeroacoustic source. The data are collected from a ULA with $M = 12$ sensors and intersensor spacing $d = 0.075$ m. The sound speed in air is considered as $c = 343$ m/s. Both sources are placed at a range of 2.5 m from the array plane. The loudspeaker is placed parallel to the array plane and centered to the array axis such that it extends over $[-3,3]^\circ$ DOAs, i.e., $\|\mathbf{x}\|_0 = 7$. The fan is placed with its rotational axis parallel to the array plane such that it constitutes a linear source extending over $[-15,0]^\circ$ DOAs, i.e., $\|\mathbf{x}\|_0 = 16$. Placing the sources at larger distances from the array would limit their apparent spatial extent to point sources rather than extended

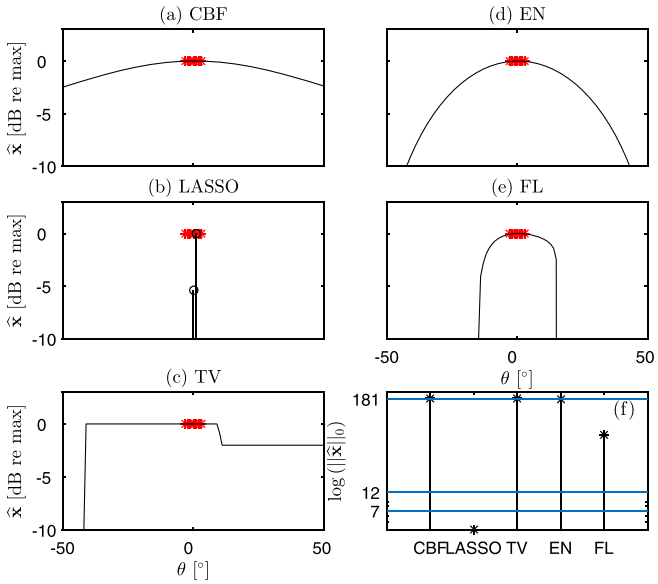

FIG. 8. (Color online) MAP estimates from experimental single-snapshot data for the loudspeaker source at 200 Hz. (a) CBF [Eq. (14)], (b) LASSO [Eq. (16)], (c) TV [Eq. (20)], (d) EN [Eq. (25)], and (e) FL [Eq. (21)]. (f) Sparsity level of each MAP estimate, where the dimensions of the problem $M = 12$ and $N = 181$ as well as the sparsity level of the true solution $\|\mathbf{x}\|_0 = 7$ are also indicated.

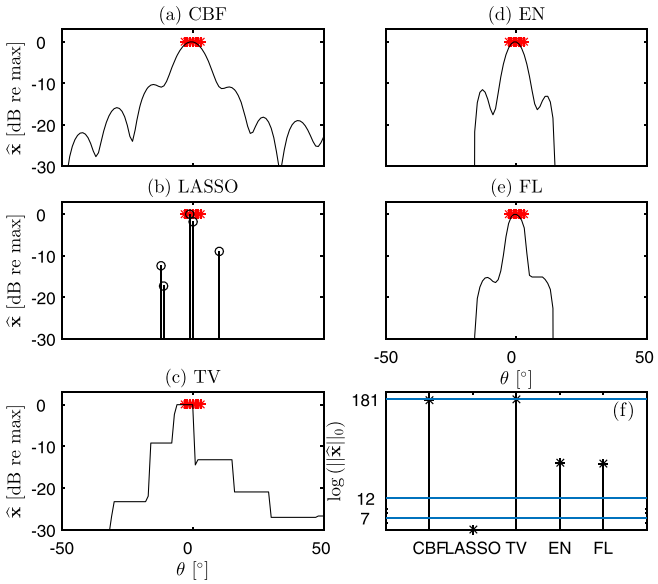

FIG. 9. (Color online) MAP estimates (as in Fig. 8) from experimental single-snapshot data for the loudspeaker source at 2000 Hz.

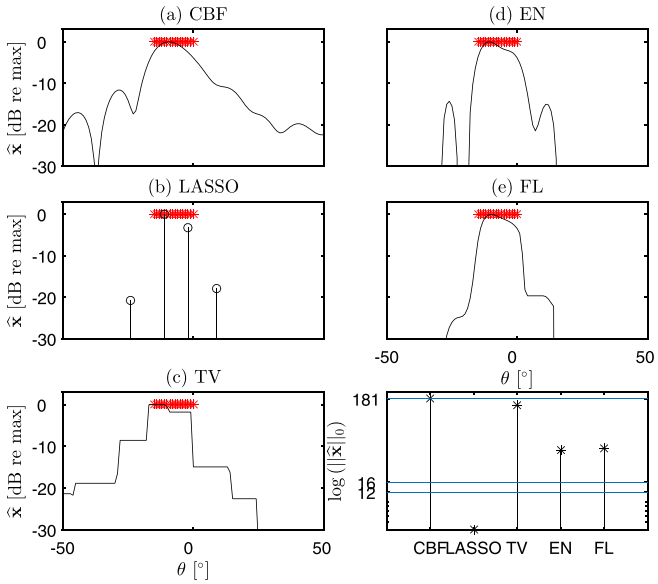

FIG. 11. (Color online) MAP estimates (as in Fig. 8) from experimental single-snapshot data for the fan source at 2000 Hz.

sources. The plane wave model [Eq. (2)] is assumed and possible errors due to wavefront curvature are neglected as near-field processing does not affect significantly the reconstruction (results not shown). The measurements are conducted in an anechoic chamber with free volume of around 1000 m^3, which offers anechoic conditions down to 50 Hz (the large anechoic chamber of the Technical University of Denmark). The data were acquired with a sampling frequency of 8209 Hz and Fourier transformed with 2^{14} samples per recorded second (with zero-padding).

According to Sec. III I, the data from a 10 s recording are post-processed at 200 and 2000 Hz, with CBF [Eq. (14)], LASSO [Eq. (16)], TV-norm regularization [Eq. (20)], FL [Eq. (21)], and EN [Eq. (25)], on a DOA grid $[-90{:}1{:}90]^\circ$ by setting all regularization parameters $\mu = \mu_1 = \mu_2 = \mu_{\mathrm{TV}} = 1$.

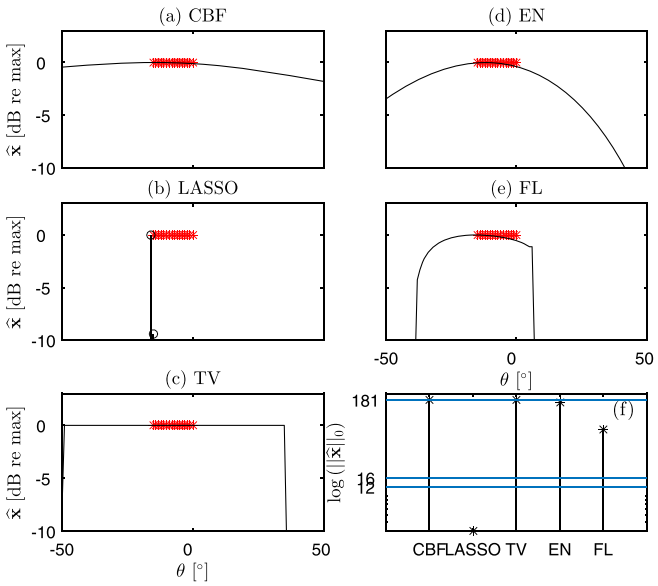

FIG. 10. (Color online) MAP estimates (as in Fig. 8) from experimental single-snapshot data for the fan source at 200 Hz.

The resulting MAP estimates and their corresponding sparsity levels are depicted in Figs. 8 and 9 for the loudspeaker at 200 and 2000 Hz and in Figs. 10 and 11 for the fan at 200 and 2000 Hz. In all cases, the FL estimate captures more accurately the underlying source spatial distribution [see the sparsity profile of the estimates in subplots (f)], even though the regularization parameters are not finely tuned. Fine tuning the regularization parameters μ_1 and μ_{TV} underlines the improved resolution in reconstructing extended sources compared to conventional methods (CBF); see Fig. 12.

V. CONCLUSION

DOA estimation with sensor arrays is essentially an underdetermined problem that requires regularization. In a probabilistic formulation, regularization is imposed in the form of prior information and the DOA map results as the MAP estimate in Bayesian inference. The characteristics of

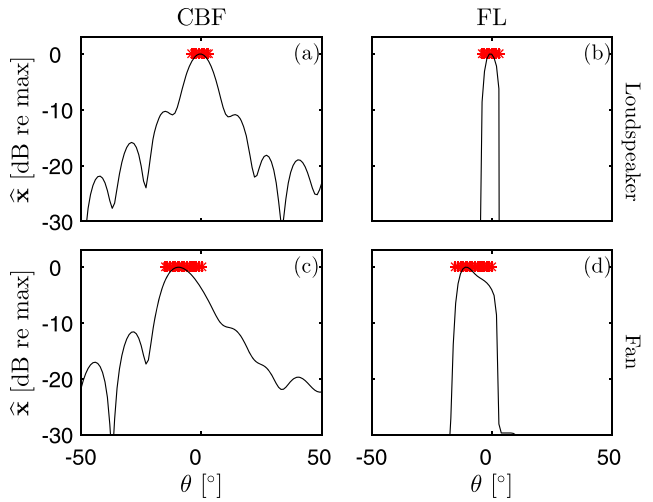

FIG. 12. (Color online) CBF (a),(c) and tuned FL estimates with (b) $\mu_1 = 2.4$, $\mu_{\mathrm{TV}} = 0.8$ and (d) $\mu_1 = 1.4$, $\mu_{\mathrm{TV}} = 0.8$ from experimental single-snapshot data for the loudspeaker and fan source at 2000 Hz.

the DOA estimate are dictated by the choice of the underlying prior distribution on the model parameters: a Gaussian distribution promotes smooth, low-resolution solutions through an ℓ_2-norm regularizer; a Laplacian distribution promotes sparse solutions through an ℓ_1-norm regularizer, improving the resolution significantly in the presence of point sources; a TV-norm regularizer promotes piecewise constant solutions (block sparsity), hence, it is suitable for applications that involve extended sources. The choice of the regularization parameters determines the quality of the reconstruction.

Composite priors are used to combine constraints in the solution controlled by the regularization parameters. For example, EN balances smoothness and sparsity in the estimate by combining an ℓ_2-norm and ℓ_1-norm regularizer, whereas FL combines a TV-norm and an ℓ_1-norm regularizer to impose additional structural constraints on the solution (such as block sparsity) besides pure sparsity. It is shown with simulated and experimental data that FL with proper regularization parameters improves the resolution in DOA estimation and sound source reconstruction in a more general framework (i.e., comprising spatially extended sources) compared to established DOA estimation methods.

APPENDIX: THE WL DUAL PATH

The non-differentiable ℓ_1 constraint in the WL formulation (17) is applied to a linear transformation of $\mathbf{x}$ and is, therefore, difficult to analyze the problem in this form. For that, we introduce the transformation of variables $\mathbf{Dx}=\mathbf{z}$, where $\mathbf{z}\in\mathbb{C}^P$, such that the optimization problem in Eq. (17) is reformulated as[43,49]

$$\min_{\mathbf{x},\mathbf{z}} \left(\|\mathbf{y}-\mathbf{Ax}\|_2^2+\mu\|\mathbf{z}\|_1\right) \text{ subject to } \mathbf{Dx}=\mathbf{z}. \tag{A1}$$

The equality constraints in Eq. (A1) can become implicit in the objective function with the use of Lagrange multipliers $\mathbf{u}\in\mathbb{C}^P$ such that the objective function in Eq. (A1) in an unconstrained form is the Lagrangian,

$$\begin{aligned} L(\mathbf{x},\mathbf{z},\mathbf{u},\mu) &= \|\mathbf{y}-\mathbf{Ax}\|_2^2+\mu\|\mathbf{z}\|_1+\mathrm{Re}[\mathbf{u}^H(\mathbf{Dx}-\mathbf{z})] \\ &= L_1(\mathbf{x},\mathbf{u})+L_2(\mathbf{z},\mathbf{u},\mu), \end{aligned} \tag{A2}$$

where

$$L_1(\mathbf{x},\mathbf{u})=\|\mathbf{y}-\mathbf{Ax}\|_2^2+\mathrm{Re}[\mathbf{u}^H\mathbf{Dx}], \tag{A3}$$

and

$$L_2(\mathbf{z},\mathbf{u},\mu)=\mu\|\mathbf{z}\|_1-\mathrm{Re}[\mathbf{u}^H\mathbf{z}]. \tag{A4}$$

The dual function is obtained by minimizing the Lagrangian [Eq. (A2)] over the primal optimization variables $\mathbf{x}$ and $\mathbf{z}$.

Minimizing $L_1(\mathbf{x},\mathbf{u})$ over $\mathbf{x}$ implies $\partial_{\mathbf{x}}L_1(\mathbf{x},\mathbf{u})=0$ which yields

$$\mathbf{D}^T\mathbf{u}=2\mathbf{A}^H(\mathbf{y}-\mathbf{A}\hat{\mathbf{x}}). \tag{A5}$$

Equation (A5) denotes the relation between the primal and the dual variables, $\mathbf{x}$ and $\mathbf{u}$, respectively. Granted that the vector $\mathbf{D}^T\mathbf{u}$ is orthogonal to the null space of the sensing matrix $\mathbf{A}$, i.e., $\mathbf{D}^T\mathbf{u}\perp\mathrm{N}(\mathbf{A})$ or equivalently $(\mathbf{DW})^H\mathbf{u}=0$, where $\mathbf{W}$ is an orthonormal basis of the null space $N(\mathbf{A})$ obtained from a singular value decomposition, Eq. (A5) yields

$$\mathbf{u}=2(\mathbf{D}^T)^+\mathbf{A}^H(\mathbf{y}-\mathbf{A}\hat{\mathbf{x}}), \tag{A6}$$

and

$$\hat{\mathbf{x}}=\mathbf{A}^+\mathbf{y}+\mathbf{x}_{\text{null}}-\frac{1}{2}(\mathbf{A}^H\mathbf{A})^+\mathbf{D}^T\mathbf{u}, \tag{A7}$$

where $\mathbf{A}^+\mathbf{y}$ represents the minimum norm solution and $\mathbf{x}_{\text{null}}$ represents a null solution such that their sum leads to the unconstrained least squares solution. The infimum of Eq. (A3) is evaluated by inserting successively Eqs. (A5) and (A7),

$$\begin{aligned} L_1(\hat{\mathbf{x}},\mathbf{u}) &= \|\mathbf{y}-\mathbf{A}\hat{\mathbf{x}}\|_2^2+\mathrm{Re}\left[(\mathbf{D}^T\mathbf{u})^H\hat{\mathbf{x}}\right] \\ &= \|\mathbf{y}\|_2^2-\|\mathbf{A}\hat{\mathbf{x}}\|_2^2 \\ &= \|\mathbf{y}\|_2^2-\left\|\mathbf{A}\left(\mathbf{A}^+\mathbf{y}+\mathbf{x}_{\text{null}}-\frac{1}{2}(\mathbf{A}^H\mathbf{A})^+\mathbf{D}^T\mathbf{u}\right)\right\|_2^2 \\ &= \|\mathbf{y}\|_2^2-\|\tilde{\mathbf{y}}-\tilde{\mathbf{D}}^H\mathbf{u}\|_2^2, \end{aligned} \tag{A8}$$

where, to simplify notation, we have introduced $\tilde{\mathbf{y}}=\mathbf{AA}^+\mathbf{y}$ and $\tilde{\mathbf{D}}=1/2\mathbf{DA}^+$. Note that, since $\mathbf{x}_{\text{null}}\in N(\mathbf{A})$, $\mathbf{Ax}_{\text{null}}=0$.

Minimizing $L_2(\mathbf{z},\mathbf{u},\mu)$ over $\mathbf{z}$ implies $\partial_{\mathbf{z}}L_2(\mathbf{z},\mathbf{u},\mu)\ni 0$, where the subdifferential operator $\partial_{\mathbf{z}}$ is a generalization of the partial differential operator for functions that are not differentiable everywhere (Ref. 38, p. 338). The subgradient for the ℓ_1-norm is the set of vectors,

$$\partial_{\mathbf{z}}\|\mathbf{z}\|_1=\{\mathbf{s}:\ \|\mathbf{s}\|_\infty\le 1,\mathbf{s}^H\mathbf{z}=\|\mathbf{z}\|_1\}, \tag{A9}$$

which implies

$$\begin{aligned} s_i &= \frac{z_i}{|z_i|}, \quad z_i\neq 0, \\ |s_i| &\le 1, \quad z_i=0, \end{aligned} \tag{A10}$$

i.e., for every active element $z_i\neq 0$ of the vector $\mathbf{z}\in\mathbb{C}^P$ (or equivalently $x_{i-1}\neq x_i$), the corresponding element of the subgradient is a unit vector in the direction of z_i. For every null element $z_i=0$ the corresponding element of the subgradient has magnitude less than or equal to one. Thus, the magnitude of the subgradient is uniformly bounded by unity, $\|\mathbf{s}\|_\infty\le 1$.

Since Eq. (A2) is convex, the global minimum is attained if $0\in\partial_{\mathbf{z}}L_2(\mathbf{z},\mathbf{u},\mu)$, which leads to the necessary and sufficient condition

$$\mu^{-1}\mathbf{u}\in\partial_{\mathbf{z}}\|\mathbf{z}\|_1. \tag{A11}$$

Then, from Eqs. (A10) and (A11), the coefficients u_i of the dual vector $\mathbf{u}\in\mathbb{C}^P$ have amplitude such that

$$\begin{aligned} |u_i| &= \mu, \quad \hat{z}_i\neq 0, \\ |u_i| &\le \mu, \quad \hat{z}_i=0, \end{aligned} \tag{A12}$$

i.e., whenever a component of $\hat{\mathbf{z}}$ becomes non-zero, the corresponding component of the dual vector hits the boundary identified with the regularization parameter, $\|\mathbf{u}\|_\infty\le\mu$. Thus,

$$L_2(\hat{\mathbf{z}}, \mathbf{u}, \mu) = \begin{cases} 0, & \|\mathbf{u}\|_\infty \le \mu \\ -\infty, & \text{otherwise}, \end{cases} \tag{A13}$$

which corresponds to a box constraint for the dual variable $\mathbf{u}$.

From Eqs. (A8) and (A13), the dual problem is

$$\begin{aligned} &\max_{\mathbf{u}} \ \|\mathbf{y}\|_2^2 - \|\tilde{\mathbf{y}} - \tilde{\mathbf{D}}^H \mathbf{u}\|_2^2 \\ &\quad \text{subject to } \|\mathbf{u}\|_\infty \le \mu, \\ &\quad \text{and } (\mathbf{DW})^H \mathbf{u} = 0. \end{aligned} \tag{A14}$$

[1]H. L. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)* (Wiley-Interscience, New York, 2002), Chaps. 1–10.
[2]Y. Kim and P. Nelson, "Optimal regularisation for acoustic source reconstruction by inverse methods," J. Sound Vib. **275**(3), 463–487 (2004).
[3]S. Paillasseur, J.-H. Thomas, and J.-C. Pascal, "Regularization for improving the deconvolution in real-time near-field acoustic holography," J. Acoust. Soc. Am. **129**(6), 3777–3787 (2011).
[4]M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Springer, New York, 2010), pp. 1–359.
[5]S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Springer, New York, 2013), pp. 1–589.
[6]J.-J. Fuchs, "Linear programming in spectral estimation: Application to array processing," in *IEEE Int. Conf. on Acoust., Speech, and Signal Process* (1996), Vol. 6, pp. 3161–3164.
[7]M. Cetin, D. M. Malioutov, and A. S. Willsky, "A variational technique for source localization based on sparse signal reconstruction perspective," in *IEEE Int. Conf. on Acoust., Speech, and Signal Process* (2002), Vol. 3, pp. 2965–2968.
[8]D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," Signal Process. **86**(3), 624–638 (2006).
[9]P. Simard and J. Antoni, "Acoustic source identification: Experimenting the ℓ_1 minimization approach," Appl. Acoust. **74**(7), 974–986 (2013).
[10]D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005).
[11]G. F. Edelmann and C. F. Gaumond, "Beamforming using compressive sensing," J. Acoust. Soc. Am. **130**(4), 232–237 (2011).
[12]S. Fortunati, R. Grasso, F. Gini, M. S. Greco, and K. LePage, "Single-snapshot DOA estimation by using compressed sensing," EURASIP J. Adv. Signal Process. **2014**, 120 (2014).
[13]C. Yardim, P. Gerstoft, W. S. Hodgkiss, and J. Traer, "Compressive geoacoustic inversion using ambient noise," J. Acoust. Soc. Am. **135**(3), 1245–1255 (2014).
[14]A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," J. Acoust. Soc. Am. **136**(1), 260–271 (2014).
[15]A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," J. Acoust. Soc. Am. **137**(4), 1923–1935 (2015).
[16]G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, and R. Gribonval, "Near-field acoustic holography using sparse regularization and compressive sampling principles," J. Acoust. Soc. Am. **132**(3), 1521–1534 (2012).
[17]P. A. Forero and P. A. Baxley, "Shallow-water sparsity-cognizant source-location mapping," J. Acoust. Soc. Am. **135**(6), 3483–3501 (2014).
[18]W. Mantzel, J. Romberg, and K. Sabra, "Compressive matched-field processing," J. Acoust. Soc. Am. **132**(1), 90–102 (2012).
[19]R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Statist. Soc. B **58**(1), 267–288 (1996).
[20]T. Park and G. Casella, "The Bayesian lasso," J. Amer. Statist. Assoc. **103**(482), 681–686 (2008).
[21]C. F. Mecklenbräuker, P. Gerstoft, A. Panahi, and M. Viberg, "Sequential Bayesian sparse signal reconstruction using array data," IEEE Trans. Signal Process. **61**(24), 6344–6354 (2013).
[22]P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, "Multiple and single snapshot compressive beamforming," J. Acoust. Soc. Am. **138**(4), 2003–2014 (2015).
[23]T. F. Brooks and W. M. Humphreys, "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays," J. Sound Vib. **294**(4), 856–879 (2006).
[24]M. Palmese and A. Trucco, "Acoustic imaging of underwater embedded objects: Signal simulation for three-dimensional sonar instrumentation," IEEE Trans. Instrum. Meas. **55**(4), 1339–1347 (2006).
[25]R. Tibshirani and J. Taylor, "The solution path of the generalized lasso," Ann. Stat. **39**(3), 1335–1371 (2011).
[26]M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," Inv. Prob. **23**(3), 947–968 (2007).
[27]L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Phys. D (Amsterdam, Neth.) **60**(1), 259–268 (1992).
[28]D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," Inv. Prob. **19**(6), 165–187 (2003).
[29]C. Li, N. Duric, P. Littrup, and L. Huang, "*In vivo* breast sound-speed imaging with ultrasound tomography," Ultrasound Med. Biol. **35**(10), 1615–1628 (2009).
[30]E. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**(17), 4777–4807 (2008).
[31]R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," J. R. Statist. Soc. B **67**(1), 91–108 (2005).
[32]M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," version 2.0 beta, available at http://cvxr.com/cvx (Last viewed September 13, 2013).
[33]S. Nadarajah, "A generalized normal distribution," J. Appl. Stat. **32**(7), 685–694 (2005).
[34]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**(4), 1976–1988 (2011).
[35]J. Antoni, "A Bayesian approach to sound source reconstruction: Optimal basis, regularization, and focusing," J. Acoust. Soc. Am. **131**(4), 2873–2890 (2012).
[36]E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," IEEE Signal Proc. Mag. **25**(2), 21–30 (2008).
[37]R. G. Baraniuk, "Compressive sensing," IEEE Signal Proc. Mag. **24**(4), 118–121 (2007).
[38]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004), pp. 1–684.
[39]M. Yuan and Y. Lin, "Efficient empirical Bayes variable selection and estimation in linear models," J. Am. Statist. Assoc. **100**(472), 1215–1225 (2005).
[40]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. Proc. **15**(5), 1551–1563 (2007).
[41]A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," Radon Series Comp. Appl. Math **9**(227), 263–340 (2010).
[42]E. Fernandez-Grande and L. Daudet, "Acoustic imaging based on Laplacian sparsity," in *Proceedings of 22nd Int. Cong. on Acoust. (ICA)* (2016), number ABS-567.
[43]C. F. Mecklenbräuker, P. Gerstoft, and E. Zöchmann, "c–LASSO and its dual for sparse signal estimation from array data," Signal Process. **130**, 204–216 (2016).
[44]M. Cetin, W. C. Karl, and A. S. Willsky, "Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," Opt. Eng. **45**(1), 017003 (2006).
[45]H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Statist. Soc. B **67**(2), 301–320 (2005).
[46]D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," IEEE Trans. Signal Process. **52**(8), 2153–2164 (2004).
[47]A. Pereira, Q. Antoni, and J. Leclere, "Empirical Bayesian regularization of the inverse acoustic problem," Appl. Acoust. **97**, 11–29 (2015).
[48]P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA estimation," IEEE Sig. Proc. Lett. **23**, 1469–1473 (2016).
[49]S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," SIAM Rev. **51**(2), 339–360 (2009).