# Improving the efficiency of deconvolution algorithms for sound source localization

Oliver Lylloff, Efrén Fernández-Grande, Finn Agerkvist, Jørgen Hald, Elisabet Tiana Roig, and Martin S. Andersen

Citation: The Journal of the Acoustical Society of America **138**, 172 (2015); doi: 10.1121/1.4922516
View online: https://doi.org/10.1121/1.4922516
View Table of Contents: https://asa.scitation.org/toc/jas/138/1
Published by the Acoustical Society of America



# Improving the efficiency of deconvolution algorithms for sound source localization

Oliver Lylloff,[a] Efrén Fernández-Grande, and Finn Agerkvist
*Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads 352, DK-2800 Kongens Lyngby, Denmark*

Jørgen Hald and Elisabet Tiana Roig
*Brüel and Kjær Sound and Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark*

Martin S. Andersen
*Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*

(Received 30 September 2014; revised 7 May 2015; accepted 3 June 2015; published online 9 July 2015)

The localization of sound sources with delay-and-sum (DAS) beamforming is limited by a poor spatial resolution—particularly at low frequencies. Various methods based on deconvolution are examined to improve the resolution of the beamforming map, which can be modeled by a convolution of the unknown acoustic source distribution and the beamformer's response to a point source, i.e., point-spread function. A significant limitation of deconvolution is, however, an additional computational effort compared to beamforming. In this paper, computationally efficient deconvolution algorithms are examined with computer simulations and experimental data. Specifically, the deconvolution problem is solved with a fast gradient projection method called Fast Iterative Shrikage-Thresholding Algorithm (FISTA), and compared with a Fourier-based non-negative least squares algorithm. The results indicate that FISTA tends to provide an improved spatial resolution and is up to 30% faster and more robust to noise. In the spirit of reproducible research, the source code is available online. 
[http://dx.doi.org/10.1121/1.4922516]

[ZHM] Pages: 172–180

## I. INTRODUCTION

Beamforming is a well-established signal processing technique that is utilized for locating sound sources using a microphone array. In the near field of a source, sound field reconstruction techniques can provide a much better spatial resolution compared to localization methods, e.g., beamforming. However, at measurement distances further away, these techniques become impractical, and beamforming provides an attractive alternative.[1]

The delay-and-sum (DAS) beamforming algorithm computes a map of the relative amplitude and position of acoustic sources in a given measurement area. However, due to the finite size of the array and finite number of microphones, the beamforming map suffers from a spatial resolution that is proportional to the wavelength (i.e., poor resolution at low frequencies), and the appearance of ghost sources, i.e., non-existent sources.[2] These limitations degrade the overall accuracy of the beamformer and can make the localization of sound sources uncertain.

During the last decade, several techniques have been developed to overcome the spatial resolution limit of beamforming.[3–7] One popular approach is using a procedure known as deconvolution. This method tries to recover the true source distribution from the beamforming map, which can be approximated by a convolution of the acoustic source distribution and the beamformer's response to a point source, i.e., the point-spread function (PSF). Some well-known deconvolution algorithms include CLEAN,[8] initially introduced in astrophysics and later applied for sound source localization by Dougherty and Stoker,[4] DAMAS by Brooks and Humphreys[5] and its extensions,[9] the active set method dubbed NNLS[10] modified for sound source localization by Ehrenfried and Koop,[6] and Richardson-Lucy[11,12] another algorithm introduced in astrophysics that has been applied in the acoustic literature.[6,13]

The general assumption in the above-mentioned references is an uncorrelated noise-source distribution. Examples of methods dealing with correlated noise source distributions can also be found in the literature.[7,14–17] In the present work, uncorrelated noise source distributions are considered.

Deconvolution algorithms require additional computational effort compared to conventional DAS beamforming. Spectral (Fourier-based) procedures are known to reduce the computational effort;[6] however, spectral methods can only be employed if the beamformer's PSF is shift-invariant, i.e., the PSF is the same regardless of the position of the point source in the computational grid.[18] This assumption is only a good approximation if the source region is small compared to the distance between the array and the source. This can be achieved by reducing the coverage angle of the array.

[a]Electronic mail: oliverlylloff@gmail.com

However, some degree of shift-variance will inevitably exist and introduce a mismatch between the convolution model and measured data.

The purpose of this work is to propose and validate a new and efficient deconvolution algorithm for sound source localization based on an accelerated gradient projection algorithm called Fast Iterative Shrikage-Thresholding Algorithm (FISTA).[19] As a reference, the Fourier-based non-negative least squares (FFT-NNLS) algorithm, a sub-gradient projection algorithm known to produce computationally efficient solutions to the sound source localization problem,[6] is used to provide a benchmark of efficiency for the comparison with the proposed deconvolution algorithm. FISTA takes its name from a specific application known as compressed sensing[20] in which a so-called shrinkage operator is utilized. With minor modifications, FISTA can be applied in this work, where the shrinkage operator will not be used, although the acronym is retained throughout.

FISTA has a worst-case convergence rate that is greater than that of ordinary gradient methods, such as FFT-NNLS.[21] Consequently, there is reason to expect that the efficiency of FISTA is greater than FFT-NNLS. With a more efficient deconvolution algorithm, it is possible to increase the dimensions of the problem or number of iterations to gain spatial resolution, calculate more frequency bands to obtain a better spectral accuracy, or simply reduce the computational run time of some demanding applications, such as three-dimensional acoustic imaging.[22]

FISTA belongs to the class of accelerated proximal methods, used for solving constrained optimization problems. The theory behind proximal methods is vast and the authors refer to the monograph by Parikh and Boyd[23] for an extensive review.

This paper is organized as follows: the general beamforming framework and problem formulation are reviewed in Sec. II. The present work is concerned with incoherent sources. However, the theory outlined in Sec. II is valid also for the case of partially coherent sources. The deconvolution algorithms FFT-NNLS and FISTA are stated in Sec. III and examined with computer simulations and experimental measurement data in Secs. IV and V. A discussion and concluding remarks are given in Secs. VI and VII. The source code is available online.[24]

In the following, vectors are denoted by boldface lowercase letters and matrices by boldface uppercase letters.

## II. BEAMFORMING AND CROSS-SPECTRAL FORMULATION

Consider an unknown number of sources, producing stationary noise, located in the $x-y$ plane at a distance of $z_0$ m from a microphone array with $M$ microphones (Fig. 1). The source plane is divided into $S = N \times N$ equidistant grid points $(x, y)$ with grid spacing $d_0/(N-1)$, where $d_0 = z_0 \tan(\theta/2)$ is the width of the source plane at a distance $z_0$ with an array coverage angle $\theta$, referred to as opening angle.

For stationary noise, the cross-spectral matrix is calculated by dividing the captured time data of each microphone into $L$ frames. Each frame is then converted into $N_{\mathrm{FFT}}$ narrow frequency bins by the FFT and averaging is performed over the $L$ frames. For a given frequency bin $\omega$, the cross-spectral matrix is averaged as

$$\mathbf{C}(\omega) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{p}_l(\omega)\mathbf{p}_l(\omega)^H, \tag{1}$$

or in short: $\mathbf{C} = \overline{\mathbf{p}\mathbf{p}^H}$, where $\mathbf{p}(\omega) = [p_1(\omega), p_2(\omega), \ldots, p_M(\omega)]^T$, $(\cdot)^H$ denotes the conjugate transpose, and $\overline{(\cdot)}$ the average over frames. For the sake of brevity, $\omega$ is omitted in the following.

The single-frame DAS output being $\mathbf{v}(\mathbf{r})^H\mathbf{p}/M$ leads to the following expression for the mean-square DAS beamformer output:

$$b(\mathbf{r}) = \frac{1}{M^2}\mathbf{v}(\mathbf{r})^H\overline{\mathbf{p}\mathbf{p}^H}\mathbf{v}(\mathbf{r}) = \frac{1}{M^2}\mathbf{v}(\mathbf{r})^H\mathbf{C}\mathbf{v}(\mathbf{r}), \tag{2}$$

where $\mathbf{v}(\mathbf{r}) = [v_1(\mathbf{r}), v_2(\mathbf{r}), \ldots, v_M(\mathbf{r})]^T$, whose elements are given by

$$v_m(\mathbf{r}) = |\mathbf{r} - \mathbf{r}_m|\frac{e^{-\mathrm{j}k|\mathbf{r}-\mathbf{r}_m|}}{|\mathbf{r}|}. \tag{3}$$

In this formulation, $|\mathbf{r}|$ is the distance from the center of the array to the beamformer focus position $\mathbf{r}$, $|\mathbf{r} - \mathbf{r}_m|$ is the distance from the focus point to the $m$th microphone, and $k$ is the wavenumber.[2,6]

Considering Eqs. (1) and (3) and the expression for a monopole in free space,[25] it follows that the response of the beamformer to a single point source at the focus point $\mathbf{r}$, Eq. (2), equals the mean square pressure at the center of the array from that source.

In the following, $b(\mathbf{r})$ is computed for all grid points $(x, y)$ in the source plane with contributions arranged in a vector $\mathbf{b}$ and referred to as the beamformer map. Furthermore, diagonal removal is applied to $\mathbf{C}$, i.e., diagonal elements $C_{mm}$ are set equal to 0 and the term $1/M^2$ is replaced by $1/M(M-1)$ before computing $b(\mathbf{r})$ from Eq. (2). The main reason for using diagonal removal is suppression of

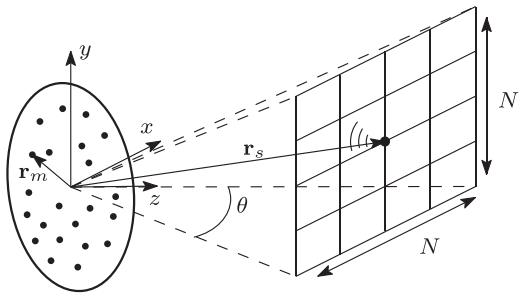



FIG. 1. A microphone array with diameter $D$ is placed in the center of the coordinate system. The $M$ microphone positions are given by $\mathbf{r}_m = (x_m, y_m, z_m)$, beamformer focus positions by $\mathbf{r} = (x, y, z_0)$, and point source positions by $\mathbf{r}_s = (x_s, y_s, z_0)$. The computational grid consists of $N \times N$ equidistant points.

noise in the individual measurement channels, typically flow noise in microphones in a wind tunnel or outdoor.[5]

The procedure described in the preceding text provides an easy way to locate sound sources. However, as mentioned previously, the beamformer map suffers from a poor spatial resolution. To improve on that, deconvolution algorithms can be applied, which requires slight modifications to the former analysis.

Consider the total sound pressure contribution at all $M$ microphones rewritten in vector notation $\mathbf{p} = \mathbf{G}\mathbf{q}$, where

$$\mathbf{G} = \begin{bmatrix} g_1(\mathbf{r}_1) & g_1(\mathbf{r}_2) & \cdots & g_1(\mathbf{r}_S) \\ g_2(\mathbf{r}_1) & g_2(\mathbf{r}_2) & & \\ \vdots & & \ddots & \\ g_M(\mathbf{r}_1) & & & g_M(\mathbf{r}_S) \end{bmatrix} \tag{4}$$

is a normalized propagation matrix with elements $g_m(\mathbf{r}_s) = |\mathbf{r}_s|(e^{-\mathrm{j}k|\mathbf{r}_s-\mathbf{r}_m|}/|\mathbf{r}_s-\mathbf{r}_m|)$ and $\mathbf{q} = [q_1, q_2, \ldots, q_S]^T$ is a vector of source amplitudes in terms of the pressure produced at the array centre, $q_s = (j\omega\rho Q_s/4\pi|\mathbf{r}_s|)$. The elements of $\mathbf{G}$ compensate for the distance scaling of the source terms $q_s$.[2,6] Notice that $\mathbf{G} \in \mathbb{C}^{M\times S}$, $\mathbf{q} \in \mathbb{C}^S$ and $\mathbf{p} \in \mathbb{C}^M$.

Using the vector-notation introduced in Eq. (4), the cross-spectral matrix, stated in Eq. (1), can be modeled as

$$\mathbf{C} = \mathbf{G}\overline{\mathbf{q}\mathbf{q}^H}\mathbf{G}^H, \tag{5}$$

where

$$\overline{\mathbf{q}\mathbf{q}^H} = \begin{bmatrix} \overline{q_1q_1^*} & \overline{q_1q_2^*} & \cdots & \overline{q_1q_S^*} \\ \overline{q_2q_1^*} & \overline{q_2q_2^*} & & \\ \vdots & & \ddots & \\ \overline{q_Sq_1^*} & & & \overline{q_Sq_S*} \end{bmatrix}. \tag{6}$$

There are $S(S+1)/2$ unknown and independent correlation terms. The aim is to estimate these. However, if the acoustic sources are incoherent, the cross terms in $\overline{\mathbf{q}\mathbf{q}^H}$ can be assumed to be negligible compared to the diagonal, which leads to a simplified cross-spectral matrix,

$$\mathbf{C} = \sum_{s=1}^{S} \overline{|q_s|^2} \cdot \mathbf{g}_{\mathbf{r}_s}\mathbf{g}_{\mathbf{r}_s}^H, \tag{7}$$

where $\mathbf{g}_{\mathbf{r}_s}$ is a column vector of $\mathbf{G}$, and the squared term $\overline{|q_s|^2} = \overline{|q_sq_s^*|}$ denotes a power descriptor of the source at the $s$th grid point. That is, sources contribute additively to power descriptors. The assumption of incoherence reduces the problem dimensions from $S(S+1)/2$ to $S$ unknowns. The aim of the present work is to estimate the non-negative power terms $\overline{|q_s|^2}$ by means of deconvolution.

## III. DECONVOLUTION ALGORITHMS

Deconvolution methods are widely used in many field of imaging to improve spatial resolution.[18,26,27] In acoustics, these methods make use of the fact that the beamformer's output can be approximated by a convolution of the acoustic source distribution and the PSF under the assumption that the source distribution can be described as a linear superposition of incoherent point sources. The sound pressure is captured by a microphone array, and the cross-spectral matrix, $\mathbf{C}$, is calculated from Eq. (1). The beamformer map can then be obtained from the real part of Eq. (2). The objective of any deconvolution algorithm is to retrace the underlying source distribution that has been blurred by the beamforming procedure.

The PSF is defined as the beamformer's response to a single unit-power point source. Inserting Eq. (7) into Eq. (2) and assuming a point source at $\mathbf{r}_s$, the PSF is given by

$$\begin{aligned} \mathrm{PSF}(\mathbf{r}|\mathbf{r}_s) &= \frac{1}{M^2}\mathbf{v}(\mathbf{r})^H\left[\mathbf{g}_{\mathbf{r}_s}\mathbf{g}_{\mathbf{r}_s}^H\right]\mathbf{v}(\mathbf{r}) \\ &= \frac{1}{M^2}|\mathbf{v}(\mathbf{r})\mathbf{g}_{\mathbf{r}_s}|^2. \end{aligned} \tag{8}$$

With this expression, the beamformer's output at a single focus point, can be written as

$$b(\mathbf{r}) = \sum_{s=1}^{S} \overline{|q_s|^2} \cdot \mathrm{PSF}(\mathbf{r}|\mathbf{r}_s), \tag{9}$$

or in the case of a shift-invariant PSF,

$$b(\mathbf{r}) = \sum_{s=1}^{S} \overline{|q_s|^2} \cdot \mathrm{PSF}(\mathbf{r}-\mathbf{r}_s), \tag{10}$$

which corresponds to a linear convolution of the source power descriptors, $\overline{|q_s|^2}$, and the shift-invariant PSF.

By computing $\mathrm{PSF}(\mathbf{r}|\mathbf{r}_s)$ for all combinations of $(\mathbf{r}, \mathbf{r}_s)$ in the grid $(x, y)$ and arranging each resulting PSF map column-wise in a matrix $\mathbf{A}$ (with dimensions $N^2 \times N^2$), a linear system of equations arises

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{11}$$

where $\mathbf{b}$ contains the beamformer map and $\mathbf{x} = [\overline{|q_1|^2}, \overline{|q_2|^2}, \ldots, \overline{|q_S|^2}]^T$ is the source distribution of power descriptors. The aim is to recover $\mathbf{x}$ given $\mathbf{A}$ and $\mathbf{b}$. If the PSF is assumed to be shift invariant, the matrix corresponds to a discretized two-dimensional (2D) blur operator, and hence Eq. (11) is an ill-posed discrete inverse problem.[28] Assuming that the vector $\mathbf{b}$ contains Gaussian noise and using the prior information that $\mathbf{x}$ is nonnegative, the following optimization problem arises

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2, \\ &\text{subject to} && x_i \geq 0, \end{aligned} \tag{12}$$

where the non-negativity constraint on $\mathbf{x}$ is due to the fact that the source power descriptors can only attain non-negative values. The optimization problem in Eq. (12) is also known as a non-negative least-squares (NNLS) problem (not to be confused with the NNLS algorithm proposed by Lawson and Hanson[10]).

The convolution formulation in Eq. (10) makes it possible to utilize the convolution theorem to perform efficient calculations in the wavenumber domain via the discrete spatial Fourier transform,

$$\mathbf{B} = \mathbf{X} * \mathbf{PSF} = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{X}) \cdot \mathcal{F}(\mathbf{PSF})], \tag{13}$$

where $\mathbf{B}$, $\mathbf{x} = \mathrm{vec}(\mathbf{X})$, and $\mathbf{PSF}$ are matrices with the beamformer map, the source distribution of power descriptors, and the point-spread function respectively. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the 2D Fourier transform and its inverse and $(*)$ denotes a convolution product. The efficiency advantage of the convolution formulation is due to the fact that the Fourier transform can be performed via a circular convolution with the FFT in $\mathcal{O}(2N^2 \log N)$ time for a $N \times N$ computational grid, compared to a $\mathcal{O}(N^4)$ complexity of a matrix-vector multiplication, $\mathbf{Ax}$. However, replacing a linear convolution with a circular convolution will result in a wrap-around (replicated aperture) error if the input is not periodic. Applying sufficient zero padding, extending the $N \times N$ grid to at least $(2N-1) \times (2N-1)$ by padding with zeros, is necessary in prevent the wrap around effect.[29]

The NNLS problem in Eq. (12) can be rewritten to take advantage of the convolution formulation,

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\mathcal{F}^{-1}[\mathcal{F}(\mathbf{X}) \cdot \mathcal{F}(\mathbf{PSF})] - \mathbf{B}\|_{\mathrm{Fro}}^2, \\ &\text{subject to} \quad X_{i,j} \geq 0, \end{aligned} \tag{14}$$

where $\|\cdot\|_{\mathrm{Fro}}$ is the Frobenius norm. This approximation reduces the computational run time significantly because it removes the need for an explicit formulation of $\mathbf{A}$. The often large dimensions of the problem makes first-order algorithms, such as the steepest descent algorithm, the only practical option for computationally efficient solutions.[30]

To introduce the following deconvolution algorithms, an equivalent formulation of the NNLS problem, Eq. (12), is given by

$$\begin{aligned} \text{minimize} \quad F(\mathbf{x}) &= f(\mathbf{x}) + g(\mathbf{x}) \\ &= \frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2 + I_+(\mathbf{x}), \end{aligned} \tag{15}$$

where $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2$ is a smooth and unconstrained quadratic function and $g(\mathbf{x}) = I_+(\mathbf{x})$, a non-smooth indicator function given by the optimization constraints

$$I_+(\mathbf{x}) = \begin{cases} 0 & x_i \in \mathbb{R}_+ \\ \infty & \text{otherwise}, \end{cases} \tag{16}$$

where $\mathbb{R}_+^n = \{\mathbf{x} \in \mathbb{R}^n | x_i \geq 0, i = 1, \ldots, n\}$ is the non-negative orthant.

The projected gradient descent algorithm for the solution of Eq. (12) is given by

$$\mathbf{x}^{k+1} = \mathcal{P}_+(\mathbf{x}^k - t_k \nabla f(\mathbf{x}^k)), \tag{17}$$

where $\mathcal{P}_+$ is the Euclidean projection of $\mathbf{x}$ onto the non-negative orthant $\mathbb{R}_+^n$.[23] Given a start guess on the solution $\mathbf{x}^0$ (a common choice is $\mathbf{x}^0 = \mathbf{0}$ but other are possible, as will be discussed later), the projected gradient descent algorithm generates a sequence of solutions by taking steps $t_k$ in the negative gradient direction, i.e., the steepest descent direction, followed by a projection that sets all variables $x_i < 0$ to zero. The algorithm is terminated by a stopping criterion defined by the user. A common and computationally inexpensive choice is to stop after a fixed number of iterations.

Based on the projected gradient descent algorithm, various methods have been proposed by the aeroacoustic community for the solution of the NNLS problem, Eq. (12). The FFT-NNLS algorithm, in particular, has been shown to provide computationally efficient solutions comparable to DAMAS2 and Richardson-Lucy.[6,13] The FFT-NNLS algorithm for the solution of Eq. (12) is stated in the following text.

### A. FFT-NNLS

Given a start vector $\mathbf{x}^0$ and a convex function $f$, repeat for $k \geq 1$ until the stopping criterion is satisfied

(1) *Search direction*: Compute subgradient $\mathbf{d}^k$ for all $x_i$

$$d_i^k = \begin{cases} -\dfrac{\partial}{\partial x_i} f\left(x_i^k\right), & x_i > 0 \\ \mathbf{max}\left\{0, -\dfrac{\partial}{\partial x_i} f\left(x_i^k\right)\right\}, & x_i = 0. \end{cases}$$

(2) *Step size*: $t_k = [(\mathbf{Ad}^k)^T(\mathbf{Ax}^k - \mathbf{b})]/[(\mathbf{Ad}^k)^T(\mathbf{Ad}^k)]$.
(3) *Update*: $\mathbf{x}^{k+1} = \mathcal{P}_+(\mathbf{x}^k + t_k \mathbf{d}^k)$.

FFT-NNLS is a projected subgradient method with a line search. Compared to Eq. (17), the computation of the subgradient $\mathbf{d}^k$ ensures that only the subgradients $d_i^k$ within the feasible set is updated; this results in a more efficient algorithm. Furthermore the convolution formulation can be used to compute the matrix-vector products $\mathbf{Ad}^k$ and $\mathbf{Ax}^k$ efficiently; this makes it suitable for the solution of the alternative problem formulation in Eq. (14).

A more recently proposed algorithm for solving inverse problems arising in signal/image processing is known as FISTA.[19] When applied to the problem in Eq. (12), FISTA takes the following form.

### B. FISTA

Given a start vector $\mathbf{x}^0$ and a convex function $f$ with a Lipschitz continuous gradient $\nabla f$ and a Lipschitz constant $L > 0$ defined by $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$, set $\mathbf{y}^1 = \mathbf{x}^0$, $t_1 = 1$, and repeat for $k \geq 1$ until the stopping criterion is satisfied,

(1) *Update* $\mathbf{x}^k = \mathcal{P}_+(\mathbf{y}^k - \nabla f(\mathbf{y}^k)/L)$.
(2) *Intermediate step* $t_{k+1} = \frac{1}{2}(1 + \sqrt{1 + 4t_k^2})$.
(3) *Update* $\mathbf{y}^{k+1} = \mathbf{x}^k + [(t_k - 1)/t_{k+1}](\mathbf{x}^k - \mathbf{x}^{k-1})$.

The two main differences between FFT-NNLS and FISTA are the computation of step sizes and the introduction of an auxiliary vector $\mathbf{y}$ in FISTA.

In FISTA, the step size $1/L$ is fixed and calculated in advance of the main loop from the Lipschitz constant of $\nabla f$.

This can be interpreted as minimizing a local quadratic upper approximation of $f$, around $x$, where the exact local minimizer of the quadratic approximation of $f$ is achieved with a step size of $1/L$. For the problem in Eq. (12), the Lipschitz constant is equal to the largest eigenvalue of the Hessian $\nabla^2 f = \mathbf{A}^T\mathbf{A}$, and this can be estimated using, e.g., the power method.[31] If the estimated $L$ is too conservative, the step size will be small, and the method will be less efficient.

The primary computational cost of the algorithms is due to the matrix product with $\mathbf{A}$ in the evaluation of the objective function $f(\mathbf{x}^k) = \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|_2^2$ and its gradient $\nabla f(\mathbf{x}^k) = \mathbf{A}^T\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|_2$. The main efficiency advantage of FISTA comes from the fact that only two matrix products are required, while FFT-NNLS requires one more for the step size calculation.

Additionally, the auxiliary vector $\mathbf{y}$ acts as a momentum term, where the previous iterate is taken into account to improve the efficiency.

To the authors' knowledge, no convergence proof exists for FFT-NNLS. The theoretical convergence rate of the ordinary projected gradient algorithm is $\mathcal{O}(1/k)$. On the contrary, the convergence rate of FISTA is $\mathcal{O}(1/k^2)$.[19]

Preliminary tests have shown that it is possible to reduce the number of iterations by utilizing a warm start strategy using the solution in one frequency band as the initialization of the next. However, this is not addressed in the present paper.

## IV. SIMULATIONS

In this section, computer simulations of two point sources are considered. The performance of FFT-NNLS and FISTA is evaluated and compared.

Consider two unit-power point sources placed in the $x-y$ plane at a distance of $z_0 = 5$ meters from a planar 60-channel microphone array with irregular microphone positioning and a diameter of 1 m. Gaussian white noise is added with a signal-to-noise ratio of 30 dB at the microphone array.

The beamformer resolution limit due to Rayleigh's criterion is given by

$$R = \frac{az_0\lambda}{\cos^3(\phi)D}, \tag{18}$$

where $a \approx 1.22$ and $\lambda$, $\phi$, and $D$ are the wavelength, the array off-axis angle, and the array aperture diameter, respectively.[2]

At a frequency of 1500 Hz and with a microphone array opening angle of $\theta = 2\phi = 30°$, the resolution limit is about $R \approx 1.54$ m. Placing the point sources $\Delta x = 1.33$ m apart, they are barely separated in the beamformer map (Fig. 2) with $N = 100$.

Applying the deconvolution algorithms, FFT-NNLS and FISTA, with a starting guess $\mathbf{x}^0 = \mathbf{0}$, 5000 iterations, and sufficient zero-padding to the beamformer map, results in two deconvolved maps shown in Fig. 3. The point source positions are well confined and distinguishable, and FISTA provides a slightly better spatial resolution than FFT-NNLS.

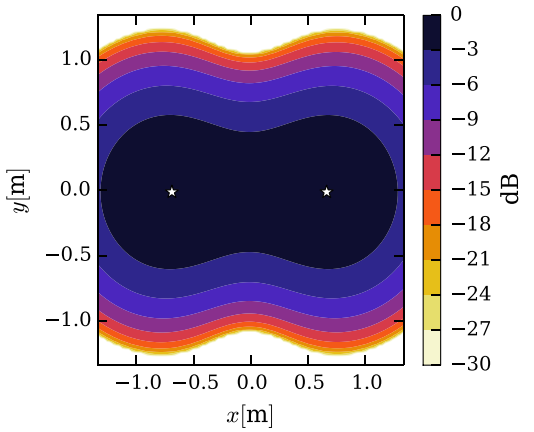
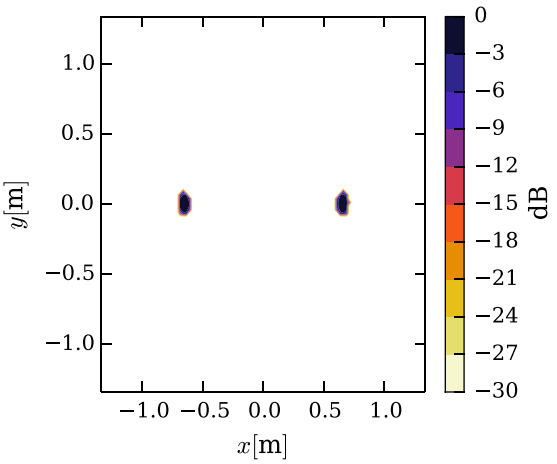
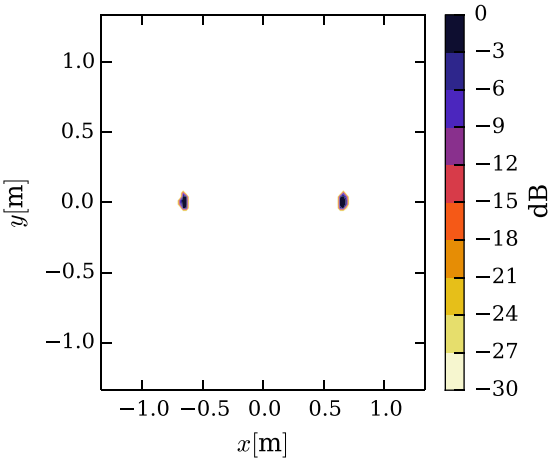



FIG. 2. (Color online) Beamformer map at $f = 1500$ Hz. Star symbols indicate position of the simulated point sources.

The convergence behavior of the deconvolution algorithms is assessed by the ratio




(a) FFT-NNLS.




(b) FISTA.

FIG. 3. (Color online) Deconvolved maps. The scale is normalized to the maximum of each independent map.

$$\frac{f(\mathbf{x}^k) - f(\mathbf{x}^*)}{f(\mathbf{x}^*)}, \tag{19}$$

where $\mathbf{x}^*$ is the optimal solution and $f(\mathbf{x}^k) = \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|_2^2$ is the objective function evaluated at iteration $k$. The optimal solution is two delta functions each of unit power at the positions indicated by the star symbols in Fig. 2. However, to compare the convergence behavior with the experimental study in Sec. V, where the optimal solution is not known, $\mathbf{x}^*$ is approximated by the solution obtained with FISTA after $3K$ iterations, where $K = 5000$ is the maximum number of iterations.

The convergence ratio $[f(\mathbf{x}^k) - f(\mathbf{x}^*)]/f(\mathbf{x}^*)$ is shown in Fig. 4. Initially, FFT-NNLS converges faster to $\mathbf{x}^*$ than FISTA. Between 100 and 300 iterations, FISTA and FFT-NNLS have similar convergence rates and beyond 300 iterations FFT-NNLS starts to produce sudden peaks in the convergence plot. Eventually, after 1000 iterations, FFT-NNLS stagnates completely and FISTA continues to decrease at a fairly constant rate. The random peaks generated by FFT-NNLS are believed to be due to numerical rounding errors in the computation of the step sizes. The stagnation in convergence is due to step sizes that are smaller than the smallest double-precision float point value, i.e., $2^{-52}$.

The time spent by FFT-NNLS and FISTA on an Intel Core 2 Duo, 3.0 GHz processor is at the order of 20 ms per iteration; however, FISTA is about 30% faster than FFT-NNLS. In particular, FFT-NNLS takes 24 ms per iteration and FISTA 17 ms per iteration, which is a total of 120 and 85 s after 5000 iterations, respectively. This means that FISTA can perform more iterations than FFT-NNLS in the same amount of time, e.g., a run time of 20 s, with the described hardware, will result in about 1175 iterations with FISTA but only about 833 iterations with FFT-NNLS. The additional time used by FFT-NNLS is due to the computation of step sizes in each iteration. The fixed step size in FISTA requires no additional time per iteration. The efficiency gain obtained with FISTA is problem independent, which means that FISTA will be 30% faster than FFT-NNLS irrespective of the size or conditioning of the problem.

Ideally the deconvolved maps should converge to two delta functions with unit power; however, due to the approximation of a shift-invariant PSF, some degree of mismatch will exist and affect the convergence of both algorithms.

The power of the sources after applying the deconvolution algorithms is assessed by a cross-sectional plot of the deconvolved maps. Figure 5 shows that FISTA provides a better approximation to a unit-power delta function than FFT-NNLS after 5000 iterations.

The total power in the deconvolved maps (Fig. 3) equals 1.89, slightly lower than the expected power of 2 of the two sources. The deviation is due to zero-padding of the beamformer map (Fig. 2). At this particular frequency, the main lobes of the beamformer map is cut away at the edges and power is removed, thus deconvolution cannot recover the initial source power.

Unlike FFT-NNLS, FISTA continues to converge beyond 1000 iterations and obtains a lower convergence ratio after about 2000 iterations. It is therefore expected that the solution provided by FISTA will be more alike two delta functions.

Although a low value in the convergence ratio is not necessarily a guarantee of high spatial resolution, it is visually clear that FISTA provides a better approximation to the original source distribution.

## V. EXPERIMENTAL RESULTS

Measurements have been conducted in an anechoic room with a volume of about 1000 $m^3$ to assess the performance of the deconvolution algorithms FFT-NNLS and FISTA in an experimental setting (Fig. 6). The microphone array consisted of 60 1/4 in. microphones, Brüel and Kjær type 4935, arranged in a pseudo-random pattern with a diameter of about 1 m. Two loudspeakers (Brüel and Kjær Omnisource type 4295) were placed 2.7 m from the center of the microphone array, driven independently with random noise generated by two noise generators, to ensure mutually incoherent sources. The loudspeakers are manufactured to have radiation characteristics of a monopole with a

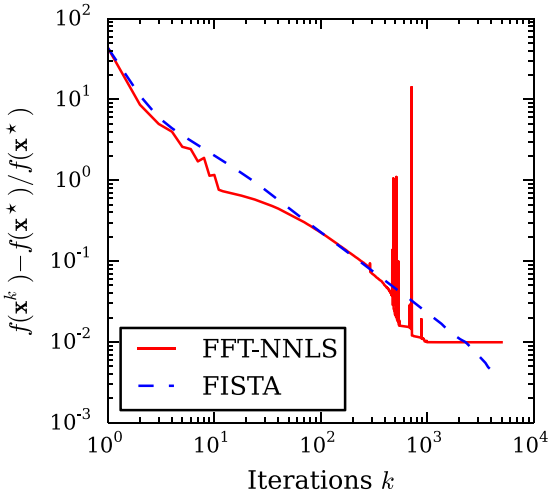



FIG. 4. (Color online) Normalized objective function values as a function of iterations.

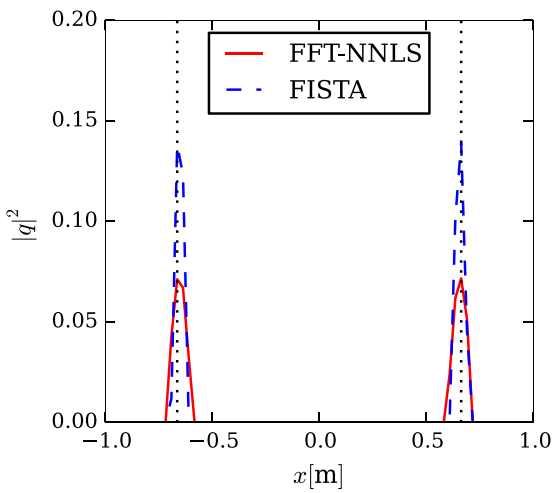



FIG. 5. (Color online) Cross section of deconvolved maps at $y = 0$ after 5000 iterations. Dotted lines represent the positions of the point sources.

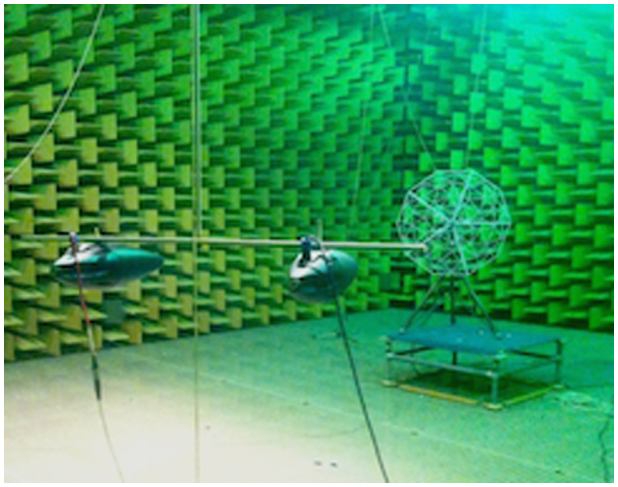


FIG. 6. (Color online) Measurement setup in an anechoic room. In the foreground, two loudspeakers are seen and in the background a microphone array is seen.

near-omnidirectional response and thus are useful for the comparison with the simulation study. The microphone array and loudspeakers were controlled by a Brüel and Kjær PULSE analyzer and a PC outside the anechoic room.

The microphone signals were recorded in 20 s segments, and the cross-spectral matrices were calculated from frames of 1 s Hanning windows with 50% overlap and averaged as described in Eq. (1).

Cross-spectral matrices of dimensions $60 \times 60$ were stored for each frequency bin up to the Nyquist frequency $f = 8192$ Hz with a spectral resolution of 1 Hz.

In the following, two identical measurement setups are considered with and without an external noise source.

### A. Without external noise source

The beamformer map at $f = 1200$ Hz is shown in Fig. 7 with an array opening angle of $\theta = 40°$. The level difference of the two sources is approximately 10 dB at this particular frequency, thus only one of the sources is clearly visible in the beamformer map. The signal-to-noise ratio at the microphone array is about 30 dB.

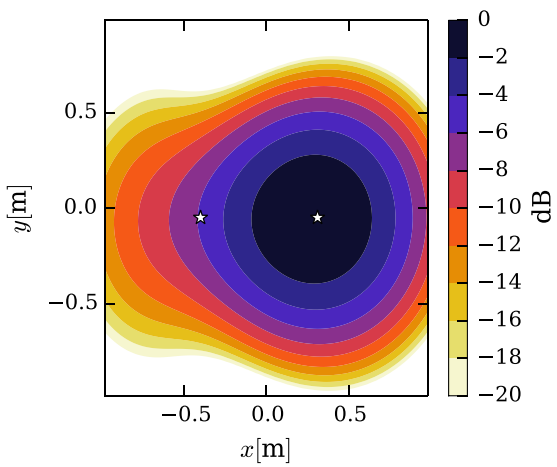
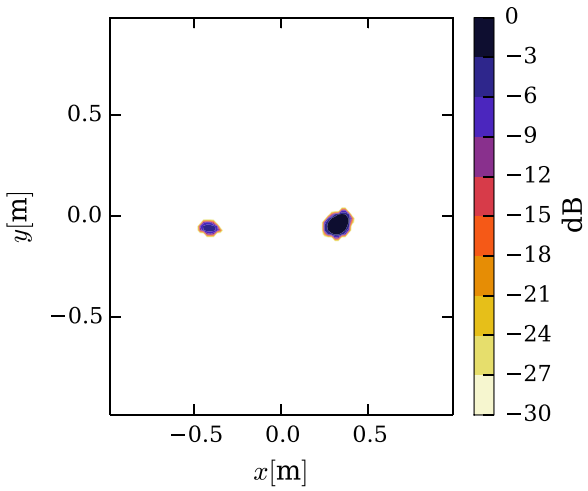



FIG. 7. (Color online) Beamformer map at $f = 1200$ Hz. Star symbols indicate approximate position of the loudspeakers.

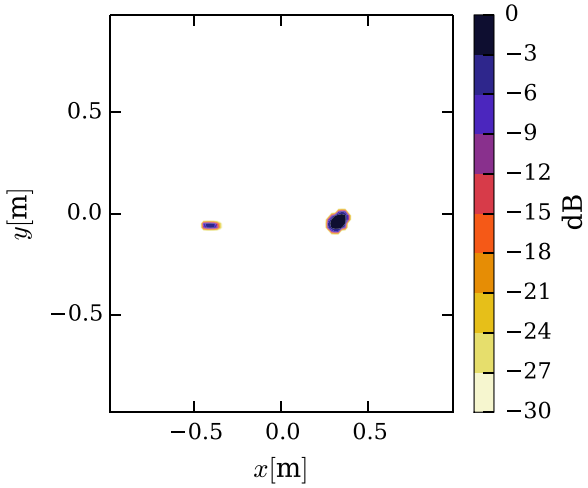
Applying the deconvolution algorithms, FFT-NNLS and FISTA, to the beamformer map with 5000 iterations results in the deconvolved maps shown in Fig. 8.

Both sources are well-resolved by FFT-NNLS and FISTA. The spatial resolution is increased significantly by both algorithms. FISTA provides a slightly better resolution with more confined point sources than FFT-NNLS.

As in the simulation study, the convergence ratio $[f(\mathbf{x}^k) - f(\mathbf{x}^*)]/f(\mathbf{x}^*)$, shown in Fig. 9, serves as a measure of convergence rate of the two algorithms. Initially, the convergence rate of FFT-NNLS is better than FISTA. After about 100 iterations FFT-NNLS slows down and exhibits random peaks, which was also seen in the simulation study, and FISTA obtains a better convergence rate. The convergence of FFT-NNLS stagnates just after 4000 iterations while FISTA continues to decrease.




(a) FFT-NNLS.




(b) FISTA.

FIG. 8. (Color online) Deconvolved maps. The scale is normalized to the maximum of each independent map.

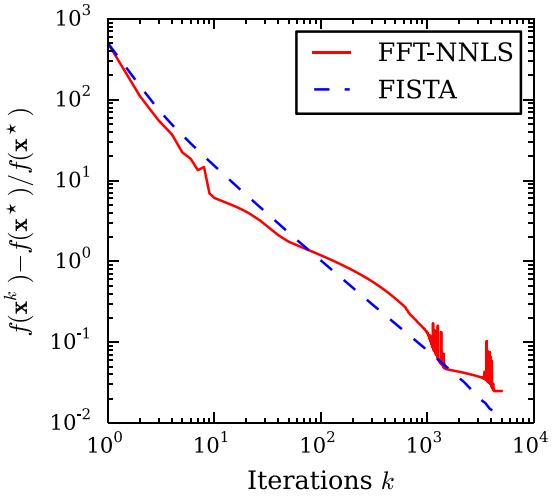



FIG. 9. (Color online) Normalized objective function values as a function of iterations.

### B. With external noise source

To test the performance of the deconvolution algorithms in high background noise conditions, an external noise source (a fan-type reference sound source) is added to the measurement setup located on the side between the array and loudspeakers (outside the field of view in Fig. 6). The beamformer map at $f = 1200$ Hz is shown in Fig. 10, where the array signal-to-noise ratios are $-5$ and 5 dB for the left and right source, respectively.

Applying FFT-NNLS and FISTA to the beamformer map with 5000 iterations results in the deconvolved maps shown in Fig. 11. Similar to the case without an external noise source, the spatial resolution is significantly better than the beamformer map.

The convergence rates of FFT-NNLS and FISTA, shown in Fig. 12, are very alike. As in the previous investigations FFT-NNLS is faster initially. After about 100 iterations and up until 1000 iterations, the two algorithms have almost identical convergence rates. Random peaks are once again seen in the convergence of FFT-NNLS and a total stagnation is observed after 1000 iterations. The greater

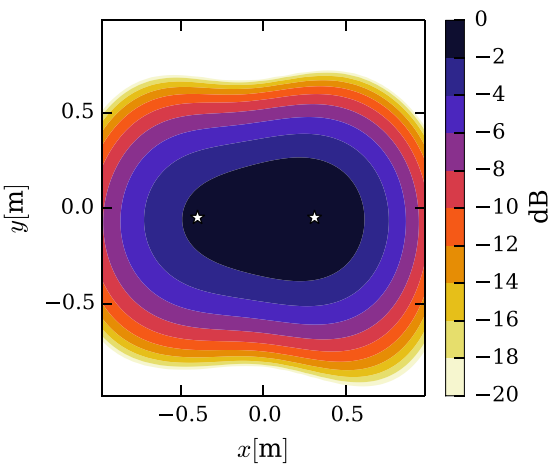
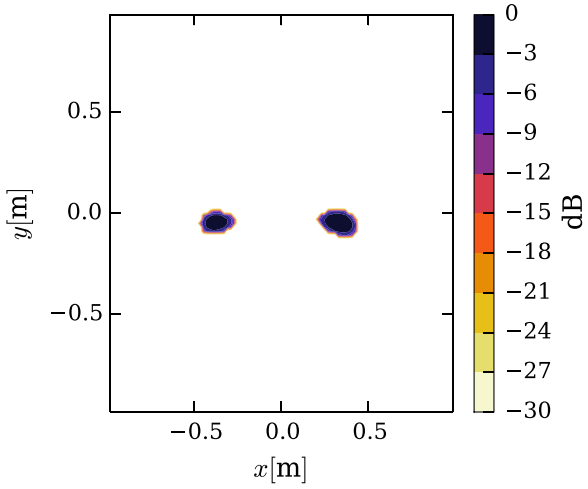
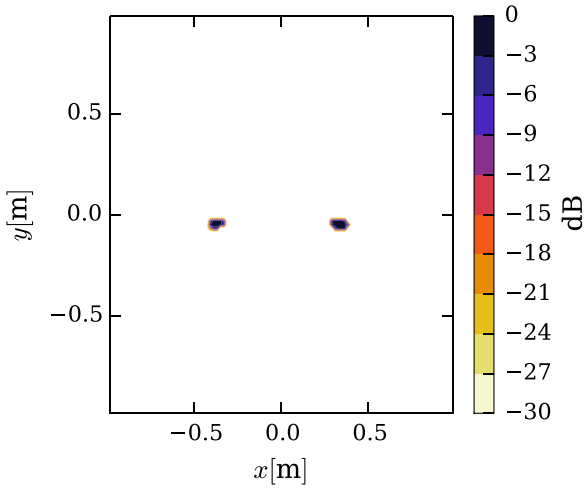



FIG. 10. (Color online) Beamformer map at $f = 1200$ Hz with external noise source. Star symbols indicate approximate position of the loudspeakers.




(a) FFT-NNLS.




(b) FISTA.

FIG. 11. (Color online) Deconvolved maps. The scale is normalized to the maximum of each independent map.

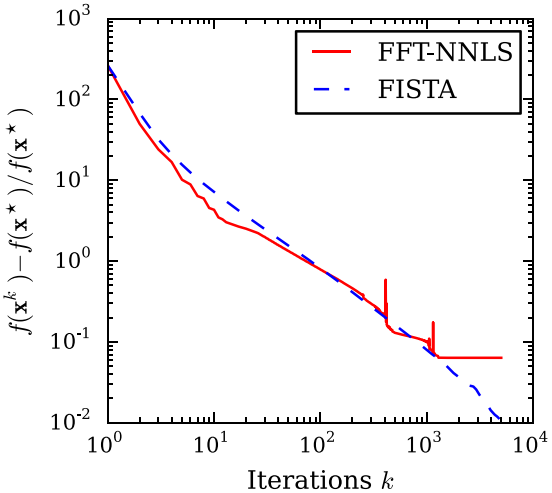



FIG. 12. (Color online) Normalized objective function values as a function of iterations.

convergence rate of FISTA above 1000 iterations is assumed to be the reason for the improved spatial resolution seen in Fig. 11(b).

## VI. DISCUSSION

It has been assumed that the PSF is shift-invariant. This implies an inherent limitation of convergence and, consequently, spatial resolution of the deconvolution algorithms. One approach to overcome this limitation is to apply a coordinate transformation to the equidistant grid, such that the PSF is forced to be shift-invariant.[22] However, the transformed grid does not have a uniform grid density, which results in a decreasing spatial resolution from the center of the grid toward the edges.

The mathematical framework behind FISTA provides an accessible option for adding further constraints. One popular constraint is found in compressed sensing that uses a 1-norm regularization constraint.[20] For future investigations, it will be of interest to extend the capabilities of the present formulation of FISTA to include a sparsity constraint on the solution to see whether the efficiency could be further improved. Additionally, the problem of localizing coherent point sources efficiently is also of interest and the framework around FISTA can possibly provide the necessary tools.

## VII. CONCLUSION

An examination of two deconvolution algorithms, FFT-NNLS and FISTA, has been carried out. It has been seen that they provide an improved spatial resolution over beamforming for sound source localization.

A comparison of the efficiency of the algorithms has shown that FISTA can reach the same solutions as FFT-NNLS in less time; approximately 30% faster. Furthermore, due to the increased and stable convergence rate beyond approximately 100 iterations, FISTA tends to provide solutions with a better spatial resolution than FFT-NNLS. Both algorithms have proved to be robust to high background noise conditions although FISTA presented a better convergence rate than FFT-NNLS after about 1000 iterations.


[1]J. Hald, "Beamforming and wavenumber processing," in *Handbook of Signal Processing in Acoustics*, edited by D. Havelock, S. Kuwano, and M. Vorländer (Springer, New York, 2008), Chap. 9, pp. 131–144.

[2]J. Hald and J. J. Christensen, *Technical Review: Beamforming* (Brüel and Kjær, Nærum, 2005), pp. 1–48.

[3]M. R. Bai, J. Ih, and J. Benesty, *Acoustic Array Systems: Theory, Implementation, and Application* (Wiley and Sons, New York, 2013), Chap. 4.

[4]R. P. Dougherty and R. W. Stoker, "Sidelobe suppression for phased array aeroacoustic measurements," in *Proceedings of the 4th AIAA/CEAS Aeroacoustics Conference*, Toulouse, France (June 1998), AIAA paper 1998-2242.

[5]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**, 856–879 (2006).

[6]K. Ehrenfried and L. Koop, "Comparison of iterative deconvolution algorithms for the mapping of acoustic sources," AIAA J. **45**, 1584–1595 (2007).

[7]A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," J. Acoust. Soc. Am. **136**, 260–271 (2014).

[8]J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," Astron. Astrophys. Supp. Ser. **15**, 417–426 (1974).

[9]R. Dougherty, "Extensions of DAMAS and benefits and limitations of deconvolution in beamforming," in *Proceedings of the 11th AIAA/CEAS Aeroacoustics Conference*, Monterey, CA (May 23–25, 2005), AIAA Paper 2005-2961.

[10]C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice Hall, Englewood Cliffs, NJ, 1974), Chap. 23.3.

[11]W. H. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. **62**, 55–59 (1972).

[12]L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79**, 745–754 (1974).

[13]E. Tiana-Roig and F. Jacobsen, "Deconvolution for the localization of sound sources using a circular microphone array," J. Acoust. Soc. Am. **134**, 2078–2089 (2013).

[14]T. F. Brooks and W. M. Humphreys, "Extension of DAMAS Phased Array Processing for Spatial Coherence Determination (DAMAS-C)," in *Proceedings of the 12th AIAA/CEAS Aeroacoustics Conference*, Cambridge, MA (May 8–10, 2006), AIAA Paper 2006–2654.

[15]C. Bahr and L. Cattafesta, "Wavespace-based coherent deconvolution," in *Proceedings of the 18th AIAA/CEAS Aeroacoustics Conference* (33rd AIAA Aeroacoustics Conference), Colorado Springs, CO (June 4–6, 2012), AIAA Paper 2012-2227.

[16]T. Yardibi, J. Li, P. Stoica, and L. N. Cattafesta, "Sparsity constrained deconvolution approaches for acoustic source mapping," J. Acoust. Soc. Am. **123**, 2631–2642 (2008).

[17]T. Yardibi, J. Li, P. Stoica, N. S. Zawodny, and L. N. Cattafesta, "A covariance fitting approach for correlated acoustic source mapping," J. Acoust. Soc. Am. **127**, 2920–2931 (2010).

[18]M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Institute of Physics, Bristol, UK, 1998), Chaps. 2 and 3.

[19]A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**, 183–202 (2009).

[20]A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, edited by D. P. Palomar and Y. C. Eldar (Cambridge University Press, New York, 2010), pp. 42–88.

[21]Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course* (Kluwer Academic, New York, 2004), Chap. 2.2.

[22]A. Xenaki, F. Jacobsen, and E. Fernandez-Grande, "Improving the resolution of three-dimensional acoustic imaging with planar phased arrays," J. Sound Vib. **331**, 1939–1950 (2012).

[23]N. Parikh and S. Boyd, "Proximal algorithms," Found. Trends Optimizat. **1**, 127–239 (2014).

[24]The source code is available at http://www.staff.dtu.dk/efgr (Last viewed 6/11/2015).

[25]F. Jacobsen and P. M. Juhl, *Fundamentals of General Linear Acoustics* (Wiley, London, 2013), Chap. 9.

[26]W. Wallace, L. H. Schaefer, and J. R. Swedlow, "A working person's guide to deconvolution in light microscopy," Biotechniques **31**, 1076–1097 (2001).

[27]J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," Publ. Astron. Soc. Pac. **114**, 1051–1069 (2002).

[28]P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms* (SIAM, Philadelphia, 2010), Chap. 7.

[29]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography. I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).

[30]J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 1998), Chap. 3.

[31]G. H. Golub and C. F. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, MA, 2013), Chap. 8.2.