Multiple snapshot grid free compressive beamforming

Yongsung Park
Department of Naval Architecture and Ocean Engineering, Seoul National University, Seoul, 151-744, South Korea

Youngmin Choo
Department of Defense System Engineering, Sejong University, Seoul, 143-747, South Korea

Woojae Seong[a)]
Department of Naval Architecture and Ocean Engineering and Research Institute of Marine System Engineering, Seoul National University, Seoul, 151-744, South Korea

(Received 29 December 2017; revised 6 April 2018; accepted 12 April 2018; published online 29 June 2018)

Compressive sensing (CS) based estimation technique utilizes a sparsity promoting constraint and solves the direction-of-arrival (DOA) estimation problem efficiently with high resolution. In this paper a grid free CS based DOA estimation technique is proposed, which uses sequential multiple snapshot data. Conventional CS technique suffers from a basis mismatch issue, while grid free CS technique is relieved of basis mismatch problem. Moreover, when the DOAs are stationary, multiple snapshot processing provides stable estimates over fluctuating single snapshot processing results. For multiple snapshot processing, the generalized version of total variation norm (group total variation norm) is implemented to impose a common sparsity pattern of multiple snapshot solution vectors in a continuous angular domain. Furthermore, an extended version is proposed using the singular value decomposition technique to mitigate computational complexity resulting from a large number of multiple snapshots. Data from SWellEx-96 are used to examine the proposed method. From the experimental data, it was observed that the present method not only offers high resolution even when the sources are coherent, but also the basis mismatch in the conventional CS method can be avoided. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5042242

[JFL] Pages: 3849–3859

I. INTRODUCTION

One main purpose of array signal processing is spatio-temporal filtering to focus on the identification, localization, or estimation of the characteristics of the target signal[1–3] of which the direction-of-arrival (DOA) estimation is a fundamental problem of the array signal processing.[1] In ocean acoustics, acoustic signals emitted from various sound sources arrive at sensor array, which consists of a limited number of sensors, with a few different DOAs. By using the relation between the received signals at the sensors and the DOAs of the sources, the DOA estimation problem can be represented in the form of an underdetermined linear system.

Since the DOA estimation problem is usually referred to as estimating a few (K sparse) DOAs in the limited numbers of received sensor data, this DOA estimation problem can be considered as a sparse signal reconstruction problem.[4] The received signals at M sensors of the array $\mathbf{y} \in \mathbb{R}^M$ can be represented as a sum of the plane waves having different DOAs. The DOA estimation problem is to find a sparse vector $\mathbf{x} \in \mathbb{R}^N$, given $\mathbf{y}$, which is related by the underdetermined linear system $\mathbf{y} = \mathbf{A}_{M\times N}\mathbf{x}$ with $M < N$. The sensing matrix $\mathbf{A}$ is made up of N columns (*a priori* bases), which are the plane waves having different DOAs in a finite discrete angular domain. The DOA estimation problem, accordingly, is to find a linear combination of K sparse bases in *a priori* N bases of the sensing matrix $\mathbf{A}$.

Compressive sensing (CS)[4,5] is a signal processing technique which solves the sparse signal reconstruction problems by using the concept of sparsity, where the sparsity represents the number of non-zero elements of $\mathbf{x}$. Various works employed the CS technique, using an l_1-norm minimization for regularization, in the DOA estimation problems.[6–14] CS based DOA estimation techniques, compared to conventional DOA estimation techniques, showed high resolution and robustness to coherent arrivals, which is fatal for classical super-resolution algorithms, e.g., the minimum variance distortion-less response (MVDR)[15] and the multiple signal classification (MUSIC).[16]

The conventional CS based estimation techniques employ finite discrete grids to estimate the parameters of interest, which actually exist in a continuous parameter space[10–13] and one of the major drawbacks of the conventional CS based estimation techniques is basis mismatch, which occurs when the exact K support (the position of non-zero components of solution vector) does not locate on the discretized grid due to inadequate N *a priori* bases of the sensing matrix. The basis mismatch is a very sensitive issue for the conventional CS DOA estimation techniques, which results in incorrect determination of the number and DOA of the sources.[17,18] Finer angular grids can mitigate the basis mismatch, but it can trigger numerical instability issues.

[a)]Electronic mail: wseong@snu.ac.kr

To overcome the basis mismatch, recent works utilize a sparsity promoting measure, the total-variation (TV) norm (or atomic norm), which is a continuous analog of l_1-norm,[19] and solves the estimation problem directly in the continuous parameter space.[19–25] We refer to this technique as grid free CS technique. In underwater acoustic signal processing, the grid free CS technique has shown its superiority in its high resolution and robustness to the basis mismatch compared to the conventional CS technique.[26,27]

The goal of DOA estimation can be to estimate the DOAs which are stationary across sequential multiple snapshot data. For multiple snapshot processing, the conventional CS DOA estimation technique has provided high resolution, outperforming the widely used classical super-resolution techniques such as MVDR and MUSIC.[11,13] In the case of stationary DOAs across multiple snapshots, it is reasonable that multiple snapshot processing should achieve more stable estimates than single snapshot processing under noisy conditions. In the spectral parameter estimation problem, the grid free CS processing with multiple measurement vectors has shown better performance in accuracy and stability than the grid free CS processing with single measurement vector.[28–30]

The present multiple snapshot grid free CS DOA estimation technique is an extended version of the scheme of Xenaki and Gerstoft[26] on grid free CS technique used for single snapshot DOA estimation, except that we improve the technique to cover the multiple snapshot DOA estimation problem. In doing so, we introduce a sparsity promoting measure, the group total-variation (gTV) norm,[28] and the multiple snapshot DOA estimation problem can be expressed in the form of gTV norm minimization problem. The gTV norm is a generalized version of the TV norm, which imposes a common sparsity pattern to the multiple solution vectors in a continuous space. On the other hand, the gTV norm minimization problem is defined in infinite dimensional space, given the multiple solution vectors in a continuous space. It is difficult to solve the primal problem by itself; however, by introducing a Lagrange multiplier, the primal problem can be solved in its dual domain with finite decision.[28]

Finally, the present multiple snapshot grid free CS DOA estimation technique is applied to real ocean data from the shallow water evaluation cell experiment 1996 (SWellEx-96).[31,32]

II. SINGLE SNAPSHOT GRID FREE CS DOA ESTIMATION

A. System framework for single snapshot DOA estimation

The DOA estimation problem estimates the DOAs of a few (K sparse) sources from measurements of a limited number of sensors M in an array. For simplicity, the sources are located in the far-field from a uniform linear array where the sources and the array are defined in the same plane, then plane waves impinge the array with DOAs $\theta \in [-90°, 90°]$ with respect to the array axis. Narrowband processing with a known sound speed is considered for the signal processing. Since a signal model $\mathbf{y} \in \mathbb{C}^M$ can be considered as a weighted superposition of plane waves, this model is represented as a sum of steering vectors corresponding to different DOAs:

$$y_m = \sum_{k=1}^{K} x_k a_m(\theta_k) + n_m = \sum_{k=1}^{K} x_k e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot \sin\theta_k} + n_m, \quad m = 1, 2, \ldots, M, \tag{1}$$

where y_m is the sound pressure received at the mth sensor, $\mathbf{a}(\theta) \in \mathbb{C}^M$ is the steering vector, which is characterized by the phase factor of the associated plane wave for the DOA θ, i.e., $\mathbf{a}(\theta) = e^{j\cdot(2\pi/\lambda)\cdot d\cdot[0,\ldots,M-1]^T\cdot \sin\theta}$, θ_k is the DOA of the kth source, $\mathbf{n} \in \mathbb{C}^M$ is the vector of additive Gaussian white noise of the mth sensor, $x_k \in \mathbb{C}$ is the complex amplitude of the kth source, λ is the wavelength, d is the inter-sensor spacing, and j is the imaginary unit ($j^2 = -1$).

B. Single snapshot grid free compressive beamforming

In the grid free CS, the K-sparse signal x has its support T (the position of non-zero components), which is restricted to the continuous interval and can be expressed as

$$x(t) = \sum_{k=1}^{K} x_k \delta(t - t_k), \tag{2}$$

where $t_k = \sin\theta_k$ is the support of the signal x, i.e., the sine value of the corresponding θ_k, on the continuous interval $\mathbb{T} = [-1, 1]$ (with $T \subset \mathbb{T}$ the set of the DOAs of all K sources) and $\delta(t - t_k)$ is a Dirac measure at t_k.

Expressed differently, the formulation (1) can be in the form of x on the continuous interval $\mathbb{T}$,

$$y_m = \int_{\mathbb{T}} x(t) e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} dt + n_m = \sum_{k=1}^{K} x_k e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t_k} + n_m, \quad m = 1, 2, \ldots, M. \tag{3}$$

Note that the measurement vector of the sound pressure received at the M sensors has the form of $\mathbf{y} = \mathcal{F}_M x + \mathbf{n}$, where $\mathcal{F}_M$ is the linear mapping matrix for inverse Fourier transform which maps the continuous variable x to the measurement vector $\mathbf{y}$.

CS based estimation techniques estimate the signal by minimizing sparsity-enforcing norm. In this problem, the support T of the signal x is arbitrarily located within a continuous interval, so grid free measure is required. Unlike the conventional CS method, which uses l_1-norm to impose sparsity, a continuous counterpart of l_1-norm known as the TV norm is introduced,[28]

$$\|x\|_{\mathrm{TV}} = \sup_{\|f\|_\infty \le 1, f \in C(\mathbb{T})} \mathrm{Re}\left[\int_{\mathbb{T}} \overline{f(t)} x(t) dt\right]. \tag{4}$$

Note that TV norm can be interpreted as being the continuous analog to the l_1-norm of the amplitudes, i.e., $\|x\|_{\mathrm{TV}} = \sum_{k=1}^{K} |x_k|$.

The single snapshot grid free CS DOA estimation problem with additive Gaussian white noise can be solved by the following convex program,[26]

$$\underset{\tilde{x}}{\text{minimize}} \;\; \|\tilde{x}\|_{\mathrm{TV}} \;\; \text{subject to} \begin{cases} \mathbf{y} = \mathcal{F}_M \tilde{x} + \mathbf{n} \\ \|\mathbf{n}\|_2 \le \varepsilon_s, \end{cases} \tag{5}$$

where $\tilde{x}$ is an estimate of the true solution, x.

The TV norm minimization [Eq. (5)] locates the position of non-zero components of $\tilde{x}$ within a continuous interval with infinite precision, i.e., $\tilde{x}$ is an infinite dimensional variable. The conventional CS method may overcome this problem by discretizing the continuous interval into a finer grid and solving an l_1-norm minimization problem. However, this method invokes high computational time and unstable estimation. Different approach to solve the problem exactly without the discretization has been proposed.[20,26] The method is to recast the Lagrange dual problem of the primal problem [Eq. (5)] as a semi-definite programming (SDP) and then obtain the support of the primal solution from the dual solution. The dual function $g(\mathbf{c},\xi)$ is the infimum of the Lagrangian for Eq. (5) and has the form[33]

$$\begin{aligned} g(\mathbf{c},\xi) &= \inf_{\tilde{x}} \|\tilde{x}\|_{\mathrm{TV}} + \mathrm{Re}\left[\mathbf{c}^H(\mathbf{y} - \mathcal{F}_M\tilde{x} - \mathbf{n})\right] \\ &\quad + \xi(\mathbf{n}^H\mathbf{n} - \varepsilon_s^2), \end{aligned} \tag{6}$$

where $\mathbf{c} \in \mathbb{C}^M$ is the vector of the Lagrange dual variables related to the equality constraints $\mathbf{y} = \mathcal{F}_M\tilde{x} + \mathbf{n}$ and $\xi \in \mathbb{R}^+$ is a Lagrange multiplier related to the inequality constraint $\|\mathbf{n}\|_2 \le \varepsilon_s$. The Lagrange dual problem[33] of Eq. (5) is formulated by maximizing the dual function $g(\mathbf{c},\xi)$,

$$\underset{\mathbf{c}}{\text{maximize}} \;\; \mathrm{Re}\left[\mathbf{c}^H\mathbf{y}\right] - \varepsilon_s\|\mathbf{c}\|_2 \;\; \text{subject to} \;\; \|\mathcal{F}_M^H\mathbf{c}\|_\infty \le 1. \tag{7}$$

The dual problem can be recast as SDP,

$$\begin{aligned} &\underset{\mathbf{c}}{\text{maximize}} \;\; \mathrm{Re}\left[\mathbf{c}^H\mathbf{y}\right] - \varepsilon_s\|\mathbf{c}\|_2 \\ &\text{subject to} \begin{bmatrix} \mathbf{Q} & \mathbf{c} \\ \mathbf{c}^H & 1 \end{bmatrix} \succcurlyeq 0, \\ &\sum_{i=1}^{M-j} Q_{i,i+j} = \begin{cases} 1, & j = 0 \\ 0, & j = 1, 2, \ldots, M\text{–}1, \end{cases} \end{aligned} \tag{8}$$

where ε_s is a noise parameter to be determined separately. The notation $\succcurlyeq$ denotes positive semi-definite, and hereafter, we will use the same notation. The choice of the noise parameter ε_s serves to adjust the relative impact between sparsity of the solution and data fitting with respect to the l_2-norm, i.e., $\|\mathbf{y} - \mathcal{F}_M\tilde{x}\|_2$. Large ε_s results in a very sparse solution, where data fit is poor, by permitting a large amount of noise. On the other hand, when ε_s becomes small, the sparsity of the solution is relieved and the solution becomes better fit to the data, but has the potential of following the noise too closely. In general, $\mathbf{n}$ is unknown so selecting a proper value for ε_s is critical depending on the purpose.[27]

Herein, the CVX program[34] (available in MATLAB) is an efficient tool to solve SDP [Eq. (8)]. The vector of the Lagrange dual variables $\mathbf{c}$ can be obtained from the solution to Eq. (8). Lemma 3.1 of Ref. 20 has shown that the support T of the primal solution $\tilde{x}$ can be obtained from the dual solution $\mathbf{c}$. The support T of the primal solution for Eq. (5) is related to the maximum modulus of trigonometric polynomial $(\mathcal{F}_M^H\mathbf{c})(t) = \sum_{m=0}^{M-1} c_m \cdot e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t}$, and can be determined by finding points where $(\mathcal{F}_M^H\mathbf{c})(t)$ becomes one, i.e., $\tilde{T} = \{t_j || (\mathcal{F}_M^H\mathbf{c})(t_j)| = 1\}$. With this estimated support $\tilde{T}$, we can obtain the amplitudes of $\tilde{x}$ by solving the system of equations $\sum_{t\in\tilde{T}} \tilde{x}(t) e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} = y_m$ for all $m = 1, \ldots, M$, using the method of least squares. The mathematical theorems and proofs underlying the single snapshot grid free CS DOA estimation technique are guaranteed in the prior study of Xenaki and Gerstoft.[26]

Figure 1 shows simulation results of the single snapshot grid free CS DOA estimation technique in a noisy environment. There are two equal strength sources at $[-5.5, 5.5]°$ along $L = 200$ snapshots. The additive noise is Gaussian white noise and the SNR is 0 dB for each snapshot. A uniform linear array consists of $M = 10$ sensors with inter-sensor spacing $d/\lambda = 1/2$. Single snapshot processing with grid free CS is compared to the conventional beamforming (CBF) result as shown in Fig. 1(a). The CBF map suffers from low resolution and merges two peaks for some snapshots. The single snapshot grid free CS technique results in enhanced resolution with two distinctive DOAs. However, due to the presence of noise, the estimated DOAs are

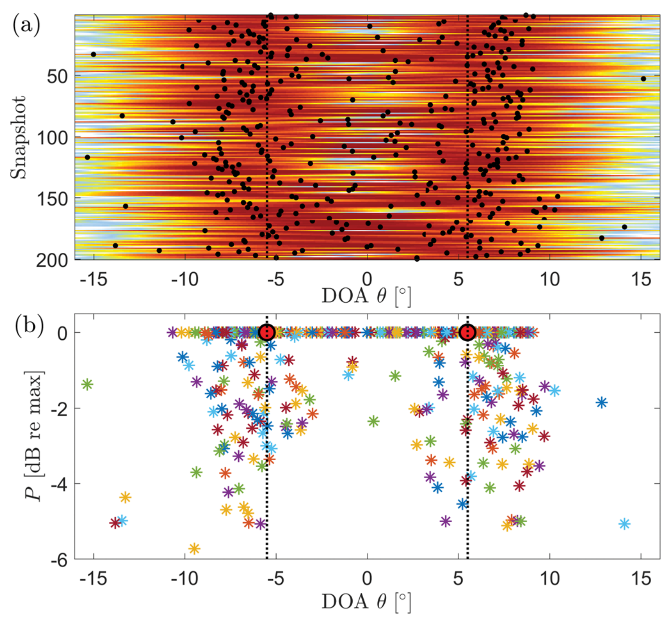

FIG. 1. (Color online) DOA estimation simulation under noisy condition. Two equal strength sources are at $[-5.5, 5.5]°$ along $L = 200$ snapshots. (a) Single snapshot grid free CS DOA estimation (dots) and CBF (background color) and (b) magnitude distribution of the results of single snapshot grid free CS DOA estimation technique (*) and the result of multiple snapshot grid free CS DOA estimation technique (○). The dotted vertical lines indicate the DOAs at $[-5.5, 5.5]°$.

scattered around the true DOAs along the snapshots, as shown in Fig. 1(b). While the true DOAs are stationary throughout the multiple snapshots, the estimated DOAs from single snapshot processing with grid free CS are not consistent and do not correspond to the true DOAs. Note that the solutions of the single snapshot processing should share a common sparsity profile, which is determined by the true stationary DOAs. In that case, multiple snapshot processing is an intuitive way to improve the estimation ability given a solution with a common sparsity profile. The multiple snapshot technique with grid free CS aims to overcome the problem of basis mismatch and estimate the DOAs which are stationary across the multiple snapshots by processing all the multiple snapshots jointly.

III. MULTIPLE SNAPSHOT GRID FREE CS DOA ESTIMATION

A. System framework for multiple snapshot DOA estimation

Multiple snapshot DOA estimation problem can be considered as an extension of single snapshot problem, where DOAs of sources are the same throughout sequential multiple snapshots. The measured signal is modeled as a matrix $\mathbf{Y} \in \mathbb{C}^{M \times L}$,

$$Y_{m,l} = \sum_{k=1}^{K} x_{k,l} a_m(\theta_k) = \sum_{k=1}^{K} x_{k,l} e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot \sin\theta_k}, \quad m = 1,2,\ldots,M, \quad l = 1,2,\ldots,L, \tag{9}$$

where $Y_{m,l}$ is the sound pressure received at the mth sensor for the lth snapshot, $x_{k,l} \in \mathbb{C}$ is the complex amplitude of the kth source for the lth snapshot, and L is the number of snapshots. To be clear, the lth column of $\mathbf{Y}$ contains a single snapshot measurement for the lth snapshot, and we use $\mathbf{y}_l$ to denote the lth column of $\mathbf{Y}$, i.e., $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L]$.

In the grid free CS manner, the sparse signals for the multiple snapshots are modeled by group of K-sparse signals, which are restricted to the continuous interval and share a common support T, and can be expressed as

$$X_l(t) = \sum_{k=1}^{K} x_{k,l}\delta(t - t_k), \quad l = 1,2,\ldots,L, \tag{10}$$

where $t_k = \sin\theta_k$ is the common support of $X_l(t)$ for all $l = 1,2,\ldots,L$, i.e., the sine value of the corresponding θ_k, on the continuous interval $\mathbb{T} = [-1,1]$ (with $T \subset \mathbb{T}$ the set of the DOAs of all K sources) and $\delta(t - t_k)$ is a Dirac measure at t_k.

Formulation (9) is expressed in terms of $X_l(t)$ on the continuous interval $\mathbb{T}$,

$$Y_{m,l} = \int_{\mathbb{T}} X_l(t) e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} dt = \sum_{k=1}^{K} x_{k,l} e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t_k}, \quad m = 1,2,\ldots,M, \; l = 1,2,\ldots,L. \tag{11}$$

Note that the measurement matrix $\mathbf{Y}$ of the sound pressure, received at the M sensors with L number of snapshots, has the form of $\mathbf{Y} = [\mathcal{F}_M X_1, \ldots, \mathcal{F}_M X_L]$.

B. Group total-variation norm

For the single snapshot processing, TV norm is used to impose a sparsity on DOAs for each snapshot and the single snapshot grid free CS is applied to the snapshots separately. However, for the multiple snapshot processing, sparsity has to be enforced over all the snapshots simultaneously. Each $X_l(t)$ for all $l = 1,2,\ldots,L$ is K-sparse and shares a common support T, so that our goal is to promote a common sparsity pattern over $X_l(t)$ for all $l = 1,2,\ldots,L$. This can be done by introducing a new grid free measure, which is called gTV norm,[28]

$$\|X\|_{\text{gTV}} = \sup_{F:\mathbb{T}\to\mathbb{C}^L,\, \|F(t)\|_2 \le 1,\, t\in\mathbb{T}} \sum_{l=1}^{L} \text{Re}\left[\int_{\mathbb{T}} \overline{F_l(t)} X_l(t) dt\right]. \tag{12}$$

Note that when a matrix $\mathbf{X} \in \mathbb{C}^{K \times L}$ is composed of $x_{k,l}$ for all $k = 1,2,\ldots,K$ and $l = 1,2,\ldots,L$, gTV norm can be interpreted as being the continuous analog to the $l_{2,1}$-norm of $\mathbf{X}$, i.e., $\|X\|_{\text{gTV}} = \sum_{k=1}^{K} \|\mathbf{x}_{k:}\|_2$, where $\mathbf{x}_{k:}$ denotes the kth row of $\mathbf{X}$.

C. Primal problem

Our multiple snapshot grid free CS DOA estimation technique utilizes gTV norm minimization approach, as a computational method, to achieve sparsity-enforcement in a continuous domain and to process all the snapshots jointly. The DOA estimation problem can be solved by the following convex program,

$$\underset{\tilde{X}}{\text{minimize}} \;\; \|\tilde{X}\|_{\text{gTV}} \text{ subject to } \mathbf{Y} = \left[\mathcal{F}_M \tilde{X}_1, \ldots, \mathcal{F}_M \tilde{X}_L\right], \tag{13}$$

where $\tilde{X}$ is an estimate of the true solution X.

Since the optimization variable X is infinite dimensional (which comes from the continuous domain), solving the primal problem [Eq. (13)] may seem challenging. The conventional CS method can overcome this problem by discretizing the continuous interval into a finer grid and solving an $l_{2,1}$-norm minimization problem.[11,13] However, this method also invokes high computational time and unstable estimation, as in the case of the single snapshot processing. Similar to the single snapshot grid free CS DOA estimation technique, the multiple snapshot problem can be solved directly without discretization. We recast the Lagrange dual problem of the primal problem [Eq. (13)] as a SDP and then obtain the support of the primal solution from the dual solution.

D. Dual problem

To formulate the dual problem for the primal problem [Eq. (13)], the dual function $g(\mathbf{C})$ is the infimum of the Lagrangian[33] for Eq. (13), and has the form[28]

$$g(\mathbf{C}) = \inf_{\tilde{X}} \|\tilde{X}\|_{\mathrm{gTV}} + \left\langle \mathbf{C}, \mathbf{Y} - \left[\mathcal{F}_M \tilde{X}_1, \ldots, \mathcal{F}_M \tilde{X}_L\right]\right\rangle$$
$$= \langle \mathbf{Y}, \mathbf{C}\rangle + \inf_{\tilde{X}}\left(\sup_{\|F(t)\|_2 \le 1} \sum_{l=1}^{L} \mathrm{Re}\left[\int_{\mathbb{T}} \overline{F_l(t)}\tilde{X}_l(t)dt\right] - \sum_{l=1}^{L}\mathrm{Re}\left[\int_{\mathbb{T}} \overline{\mathcal{F}_M^H \mathbf{c}_l(t)}\tilde{X}_l(t)dt\right]\right), \quad (14)$$

where $\mathbf{C} \in \mathbb{C}^{M\times L}$ is the matrix of the Lagrange dual variables related to the equality constraints $\mathbf{Y} = [\mathcal{F}_M\tilde{X}_1, \ldots, \mathcal{F}_M\tilde{X}_L]$ and the operator $\langle \mathbf{Y}, \mathbf{C}\rangle$ is the sum of the column-wise inner product of $\mathbf{Y}$ and $\mathbf{C}$, i.e., $\langle \mathbf{Y}, \mathbf{C}\rangle = \sum_{l=1}^{L}\langle \mathbf{y}_l, \mathbf{c}_l\rangle$. Note that the second term in Eq. (14) is trivial $(-\infty)$, unless $\sum_{l=1}^{L} |\mathcal{F}_M^H \mathbf{c}_l(t)|^2 \le 1$ for all t. Therefore, the dual function is

$$g(\mathbf{C}) = \begin{cases} \langle \mathbf{Y}, \mathbf{C}\rangle, & \sup_t \sum_{l=1}^{L} |\mathcal{F}_M^H \mathbf{c}_l(t)|^2 \le 1 \\ -\infty, & \text{otherwise.}\end{cases} \quad (15)$$

The Lagrange dual problem of Eq. (13) is formulated by maximizing the dual function $g(\mathbf{C})$,[33]

$$\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y}, \mathbf{C}\rangle \text{ subject to } \sup_t \sum_{l=1}^{L} |\mathcal{F}_M^H \mathbf{c}_l(t)|^2 \le 1. \quad (16)$$

By Proposition 2.4 of Ref. 28, the dual problem can be recast as SDP,

$$\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y}, \mathbf{C}\rangle \text{ subject to } \begin{bmatrix} \mathbf{Q} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{I}\end{bmatrix} \succcurlyeq 0,$$
$$\sum_{i=1}^{M-j} Q_{i,i+j} = \begin{cases} 1, & j = 0 \\ 0, & j = 1, 2, \ldots, M\text{–}1,\end{cases} \quad (17)$$

where $\mathbf{I}$ is the identity matrix of dimensions $L \times L$. Herein, we use the CVX program[34] to solve SDP [Eq. (17)]. The matrix of the Lagrange dual variables $\mathbf{C}$ can be obtained from the solution to Eq. (17). Lemma 3.5 of Ref. 28 has shown that the support $\tilde{T}$ of the primal solution $\tilde{X}$ can be obtained from the dual solution $\mathbf{C}$. The support $\tilde{T}$ of the primal solution for Eq. (13) is related to the maximum modulus of trigonometric polynomial $\sum_{l=1}^{L} |(\mathcal{F}_M^H \mathbf{c}_l)(t)|^2 = \sum_{l=1}^{L} |\sum_{m=0}^{M-1} c_{m,l} \cdot e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t}|^2$, and can be determined by finding points where $\sum_{l=1}^{L} |(\mathcal{F}_M^H \mathbf{c}_l)(t)|^2$ becomes one, i.e., $\tilde{T} = \{t_j | \sum_{l=1}^{L} |(\mathcal{F}_M^H \mathbf{c}_l)(t_j)|^2 = 1\}$. Note that this polynomial is a generalized version of the support-locating polynomial of the single snapshot case. With this estimated support $\tilde{T}$, we can obtain the amplitudes of $\tilde{X}$ by solving the system of equations $\sum_{t\in\tilde{T}} \tilde{X}_l(t) e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} = Y_{m,l}$ for all $m = 1, \ldots, M$, and $l = 1, \ldots, L$, using the method of least squares.

E. Grid free CS DOA estimation with noise

The multiple snapshot grid free CS DOA estimation problem with additive Gaussian white noise $\mathbf{N} \in \mathbb{C}^{M\times L}$ is equivalent to the following convex program,

$$\underset{\tilde{X}}{\text{minimize}}\ \|\tilde{X}\|_{\mathrm{gTV}}$$
$$\text{subject to } \begin{cases} \mathbf{Y} = \left[\mathcal{F}_M\tilde{X}_1, \ldots, \mathcal{F}_M\tilde{X}_L\right] + \mathbf{N} \\ \|\mathbf{N}\|_f \le \varepsilon_m,\end{cases} \quad (18)$$

where the operator $\|\mathbf{N}\|_f$ is the Frobenius norm. To solve the infinite dimensional primal problem [Eq. (18)], we convert this problem into the Lagrange dual problem of Eq. (18),

$$\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y}, \mathbf{C}\rangle - \varepsilon_m \|\mathbf{C}\|_f$$
$$\text{subject to } \sup_t \sum_{l=1}^{L} |\mathcal{F}_M^H \mathbf{c}_l(t)|^2 \le 1, \quad (19)$$

and recast the Lagrange dual problem as SDP,

$$\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y}, \mathbf{C}\rangle - \varepsilon_m \|\mathbf{C}\|_f$$
$$\text{subject to } \begin{bmatrix} \mathbf{Q} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{I}\end{bmatrix} \succcurlyeq 0,$$
$$\sum_{i=1}^{M-j} Q_{i,i+j} = \begin{cases} 1, & j = 0 \\ 0, & j = 1, 2, \ldots, M\text{–}1.\end{cases} \quad (20)$$

Details of the derivation of the dual problem [Eq. (19)] and the SDP [Eq. (20)] are summarized in the Appendix. By solving the finite dimensional SDP [Eq. (20)], we can obtain the matrix of the Lagrange dual variables $\mathbf{C}$. The support $\tilde{T}$ of the primal solution $\tilde{X}$ can be obtained from the dual solution $\mathbf{C}$ from the relation such that $\tilde{T} = \{t_j | \sum_{l=1}^{L} |(\mathcal{F}_M^H \mathbf{c}_l)(t_j)|^2 = 1\}$. With this estimated support $\tilde{T}$, we can obtain the amplitudes of $\tilde{X}$ by solving the system of equations $\sum_{t\in\tilde{T}} \tilde{X}_l e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} = Y_{m,l}$ for all $m = 1, \ldots, M$, and $l = 1, \ldots, L$, using the method of least squares.

Similar to the case of the choice of the noise parameter ε_s for the single snapshot case, the multiple snapshot grid free CS DOA estimation technique requires proper choice of the noise parameter ε_m. The role of the noise parameter ε_m is similar to that of ε_s except that it is defined for the noise matrix $\mathbf{N}$. The choice of the noise parameter ε_m controls the relative impact between sparsity of the solution and data fitting with respect to the Frobenius norm, i.e., $\|\mathbf{Y} - [\mathcal{F}_M\tilde{X}_1, \ldots, \mathcal{F}_M\tilde{X}_L]\|_f$. For the simulation data, we know the magnitude of $\mathbf{N}$, and so we can compute the noise parameter ε_m. However, for the experimental data, this will not be possible so we have to select proper value of ε_m, depending on the purpose.

IV. SUPPORT DETECTION WITH SINGULAR VALUE DECOMPOSITION (SVD)

The proposed grid free CS DOA estimation technique processes sequential multiple snapshot data jointly, so that its computational cost increases with number of snapshots. To reduce computational cost, SVD technique can be employed in the same manner as in Ref. 11. We apply SVD of multiple snapshot data $\mathbf{Y} \in \mathbb{C}^{M\times L}$ to bring it into the singular vector space and then truncate the data matrix leaving S largest singular values. The SVD of the data matrix reduces the original $M \times L$ dimensional problem to $M \times S$ dimensional problem ($L > S$). The SVD of $\mathbf{Y}$ has the following representation: $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ is an $M \times M$ unitary matrix, $\mathbf{\Sigma}$ is a diagonal $M \times L$ matrix with non-negative real numbers (singular values) on the diagonal, and $\mathbf{V}$ is an $L \times L$ unitary matrix. To take S largest singular vectors, a reduced $M \times S$ dimensional matrix $\mathbf{Y}^{\text{SVD}}$ has the form

$$\mathbf{Y}^{\text{SVD}} = \mathbf{U}\mathbf{\Sigma}\mathbf{D}_S = \mathbf{Y}\mathbf{V}\mathbf{D}_S, \tag{21}$$

where $\mathbf{D}_S = [\mathbf{I}_S \; \mathbf{0}]^T$. Here, $\mathbf{I}_S$ is a $S \times S$ identity matrix, and $\mathbf{0}$ is a $S \times (L - S)$ matrix of zeros. With a reduced $M \times S$ dimensional matrix $\mathbf{Y}^{\text{SVD}}$, we propose the SVD version of Eq. (18),

$$\begin{aligned} &\underset{\tilde{X}^{\text{SVD}}}{\text{minimize}} \;\; \|\tilde{X}^{\text{SVD}}\|_{\text{gTV}} \\ &\text{subject to} \begin{cases} \mathbf{Y}^{\text{SVD}} = \left[\mathcal{F}_M \tilde{X}_1^{\text{SVD}}, \ldots, \mathcal{F}_M \tilde{X}_L^{\text{SVD}}\right] + \mathbf{N}^{\text{SVD}} \\ \|\mathbf{N}^{\text{SVD}}\|_f \le \varepsilon^{\text{SVD}}, \end{cases} \end{aligned} \tag{22}$$

where $[\mathcal{F}_M \tilde{X}_1^{\text{SVD}}, \ldots, \mathcal{F}_M \tilde{X}_L^{\text{SVD}}] = [\mathcal{F}_M \tilde{X}_1, \ldots, \mathcal{F}_M \tilde{X}_L]\mathbf{V}\mathbf{D}_S$ and $\mathbf{N}^{\text{SVD}} = \mathbf{N}\mathbf{V}\mathbf{D}_S$. The proposed scheme also requires careful choice of the noise parameter ε^{SVD}, and the role of the noise parameter ε^{SVD} is similar to that of ε_s [Eq. (5)] or ε_m [Eq. (18)].

Procedure to estimate the DOAs of the sources is as follows:

(1) Solve the finite dimensional SDP, which is equivalent to the Lagrange dual problem of the primal problem [Eq. (22)],

$$\begin{aligned} &\underset{\mathbf{C}_{\text{SVD}}}{\text{maximize}} \;\; \langle \mathbf{Y}_{\text{SVD}}, \mathbf{C}_{\text{SVD}} \rangle - \varepsilon^{\text{SVD}} \|\mathbf{C}_{\text{SVD}}\|_f \\ &\text{subject to} \begin{bmatrix} \mathbf{Q} & \mathbf{C}_{\text{SVD}} \\ \mathbf{C}_{\text{SVD}}^H & \mathbf{I} \end{bmatrix} \succcurlyeq 0, \\ &\sum_{i=1}^{M-j} Q_{i,i+j} = \begin{cases} 1, & j = 0 \\ 0, & j = 1, 2, \ldots, M\text{–}1. \end{cases} \end{aligned} \tag{23}$$

(2) Obtain the support $\tilde{T}^{\text{SVD}}$ of the primal solution $\tilde{X}^{\text{SVD}}$ from the dual solution of the SDP [Eq. (23)]. The location of the support can be attained from the support-locating polynomial where its magnitude becomes one, i.e., $\tilde{T}^{\text{SVD}} = \{t_j | \sum_{l=1}^{L} |(\mathcal{F}_M^H \mathbf{c}_l^{\text{SVD}})(t_j)|^2 = 1\}$.

(3) Estimate the amplitudes of the corresponding sources by solving the system of equations $\sum_{t \in \tilde{T}^{\text{SVD}}} \tilde{X}_l(t) e^{j\cdot(2\pi/\lambda)\cdot d\cdot(m-1)\cdot t} = Y_{m,l}$ for all $m = 1, \ldots, M$ and $l = 1, \ldots, L$, using the method of least squares.

Note that in the final procedure, $\tilde{X}_l(t)$ and $\mathbf{Y}$ are not $\tilde{X}_l^{\text{SVD}}(t)$ and $\mathbf{Y}^{\text{SVD}}$, respectively. The proposed SVD version scheme is intended to reduce computational complexity in detecting common support of the true solution X.

The SVD scheme requires the choice of number of large singular values (S). Unlike MUSIC which can be unstable depending on S, the present suggested method shows a robust estimate of the sparse solution regardless of S. MUSIC is sensitive to S as it requires a matrix inversion. However, CS based algorithm does not require the matrix inversion. The number of large singular values S is related to the magnitude of the noise, contained in the measurement matrix $\mathbf{Y}$, and the support of the solution is preserved regardless of S. The CS based algorithm seeks the sparsity of the solution in the singular vector space to find the support of the solution, so that CS based algorithm can overcome the sensitivity to the choice of the number of large singular values.[11]

V. SIMULATION RESULTS

In this section, we present several simulation results of the multiple snapshot grid free CS DOA estimation technique. In general, performance in a noiseless environment of multiple snapshot grid free CS technique is more or less similar to single snapshot grid free CS performance.

Performance in a noisy environment is examined via simulation environment treated in Fig. 1. The result of single snapshot grid free CS technique does not locate the DOAs, which were set as constant, due to existing noise. However, the multiple snapshot grid free CS technique locates the DOAs which are constant throughout the multiple snapshots by processing all snapshots jointly, and the estimated DOAs are in agreement with true DOAs, as shown in Fig. 1(b).

For the purpose of quantifying performance enhancement of multiple snapshot scheme, ensemble root mean

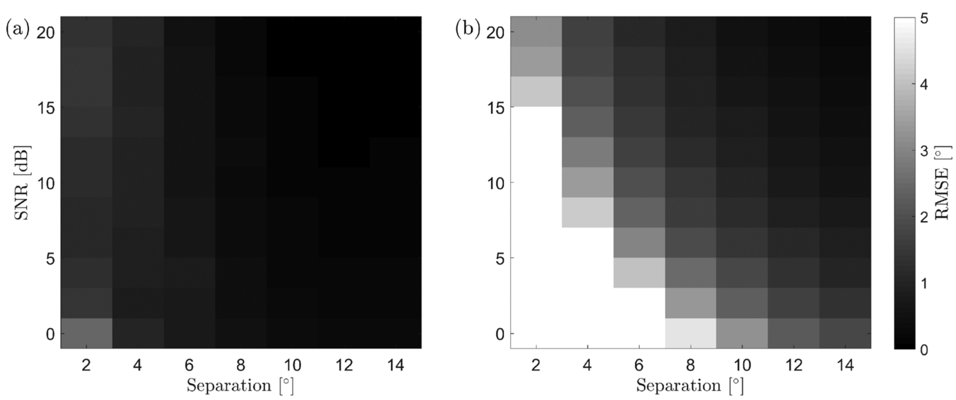

FIG. 2. RMSE [°] comparison between (a) multiple snapshot ($L = 50$) grid free CS DOA estimation and (b) single snapshot grid free CS DOA estimation. Each RMSE is averaged over 100 trials for each separation between two DOAs of sources and SNR condition.

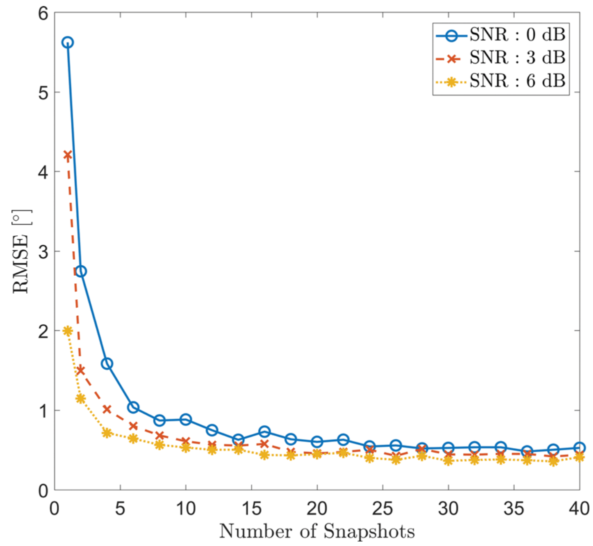

FIG. 3. (Color online) RMSE versus number of snapshots (L) for multiple snapshot grid free CS DOA estimation technique with three different SNR values of [0, 3, 6] dB. Each RMSE is averaged over 100 trials for each condition.

squared error (RMSE) for multiple snapshot grid free CS technique is compared to that of single snapshot grid free CS technique. Once one method estimates the DOAs, the estimated DOAs ($\tilde{\theta}_k$) are matched with the true DOAs (θ_k^{true}) to minimize the RMSE, and the RMSE is computed,

$$\text{RMSE} = \sqrt{\text{E}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\tilde{\theta}_k - \theta_k^{\text{true}}\right)^2\right]}. \tag{24}$$

Figure 2 compares RMSE of multiple snapshot grid free CS and those of single snapshot grid free CS. Darker colors indicate that a method is more likely to have the RMSE near zero, the true DOAs can be estimated. A uniform linear array with $M = 10$ sensors with inter-sensor spacing $d/\lambda = 1/2$. There are two same strength sources with different separation and SNR, and $L = 50$ snapshots. We let one source have DOA at $\theta = 0°$ and the other move from 2° to 14°. We repeated 100 trials for each condition and averaged over these RMSEs. Note that the multiple snapshot grid free CS technique deals with 50 snapshots jointly for each trial, so 100 RMSEs are generated. On the other hand, the single snapshot grid free CS technique generates 5000 (50 snapshots $\times$ 100 trials) RMSEs. The RMSE results of our multiple snapshot scheme [Fig. 2(a)] show better performance than single snapshot scheme [Fig. 2(b)] under various separation and SNR conditions. With our multiple snapshot grid free CS scheme, high resolution capability and accurate estimation under noisy scenarios are achieved.

We investigate the RMSE according to changes of snapshot numbers. As the number of snapshots increases, estimation performance is enhanced and RMSE for multiple

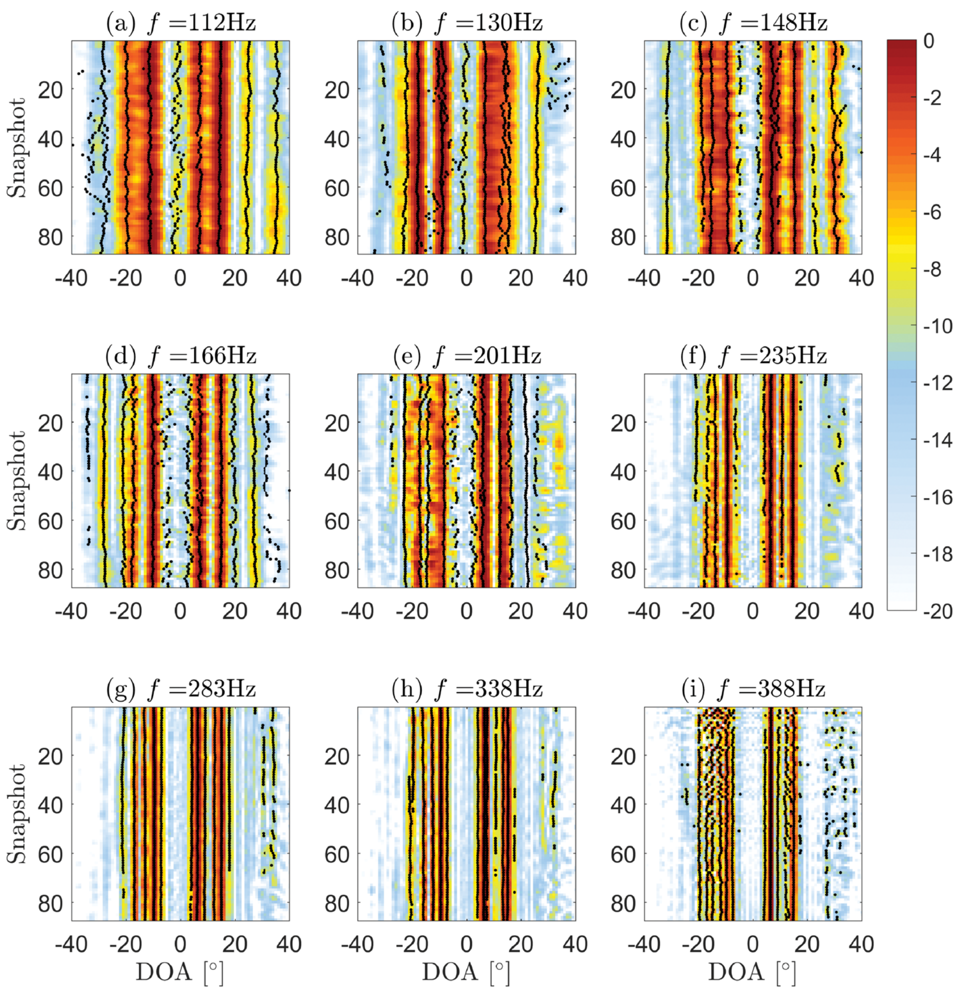

FIG. 4. (Color online) Single snapshot DOA estimation with experimental data for nine frequencies [112, 130, 148, 166, 201, 235, 283, 338, 388] Hz with single snapshot grid free CS DOA estimation (dots) and CBF (background color).

snapshot grid free CS technique decreases. This is confirmed by the numerical simulation results shown in Fig. 3. The array configuration is the same as in Fig. 2. There are two same strength sources with a separation angle of 8° and we consider three cases with different SNR values of 0, 3, and 6 dB. We repeated 100 trials for each condition and averaged over these RMSEs. The plot implies that using a larger number of snapshots allows our multiple snapshot grid free CS technique to estimate DOAs with a smaller bias and robust to noise.

The maximum resolvable DOAs with our multiple snapshot grid free CS technique is determined by the number of sensors, and up to $M-1$ sources can be estimated. This comes from the support locating process which deals with M-dimensional variable $\mathbf{c}_l$ in Eq. (17).

VI. EXPERIMENTAL DATA RESULTS

The proposed grid free CS DOA estimation method for multiple snapshot data is applied to experimental data. The data are from the SWellEx-96 event S5 (Refs. 31 and 32) collected from a vertical uniform linear array from 23:15 to 00:30 on 10–11 May 1996 west of Point Loma, CA and are the same as in Ref. 13. The array has $M = 64$ sensors with inter-sensor spacing $d = 1.875$ m and was arranged from a depth of 94.125 to 212.25 m at a water depth of 216.5 m. Duration data of 1.5 min were collected at sampling frequency of 1500 Hz and were divided into 87 snapshots of 2.7 s duration (2^{12} samples) with 63% overlap. The target source signal transmitted a set of nine frequencies [112, 130, 148, 166, 201, 235, 283, 338, 388] Hz.

Figure 4 shows the results of a single snapshot processing using CBF and single snapshot grid free CS technique. The single snapshot grid free CS DOA estimation technique improves resolution and eliminates ambiguity due to sidelobes and noise. Note that several significant DOAs are stationary along several single snapshots.

In the multiple snapshot approach, suggested multiple snapshot grid free CS DOA estimation scheme is compared to CBF, MVDR, and conventional CS DOA estimation scheme, as shown in Fig. 5. The present grid free CS based method provides high resolution DOA estimation with noticeable peaks compared to CBF and MVDR. The estimated DOAs from multiple snapshot grid free CS scheme are consistent with those from single snapshot grid free CS scheme. The data set is from a shallow water experiment and contains multiple coherent sources due to the multipath propagation. Due to coherent sources, MVDR shows poor performance even with sufficient number of sensors ($M = 64$) and snapshots ($L = 87$). On the other hand, our grid free CS based method works well for coherent sources.

We additionally compare grid free CS DOA estimation scheme to conventional CS DOA estimation scheme[13] for both single snapshot processing with the same data of Fig. 4 and multiple snapshot processing with the same data of Fig. 5. Conventional CS based DOA estimation scheme has discrete angular grids and the grids are defined on a discrete angular spectrum $[-90° : 1° : 90°]$. Comparison figures for single snapshot case are omitted, because for both single and multiple snapshot cases, there is no significant difference between the results, as shown in Fig. 5. This is because the number of sensors in the array is large enough to resolve the DOAs. Large number of sensors in the array decreases coherence between the steering vectors which are allocated to discrete angular grid in the conventional CS method and decreased coherence reduces bias in the estimates of DOAs.

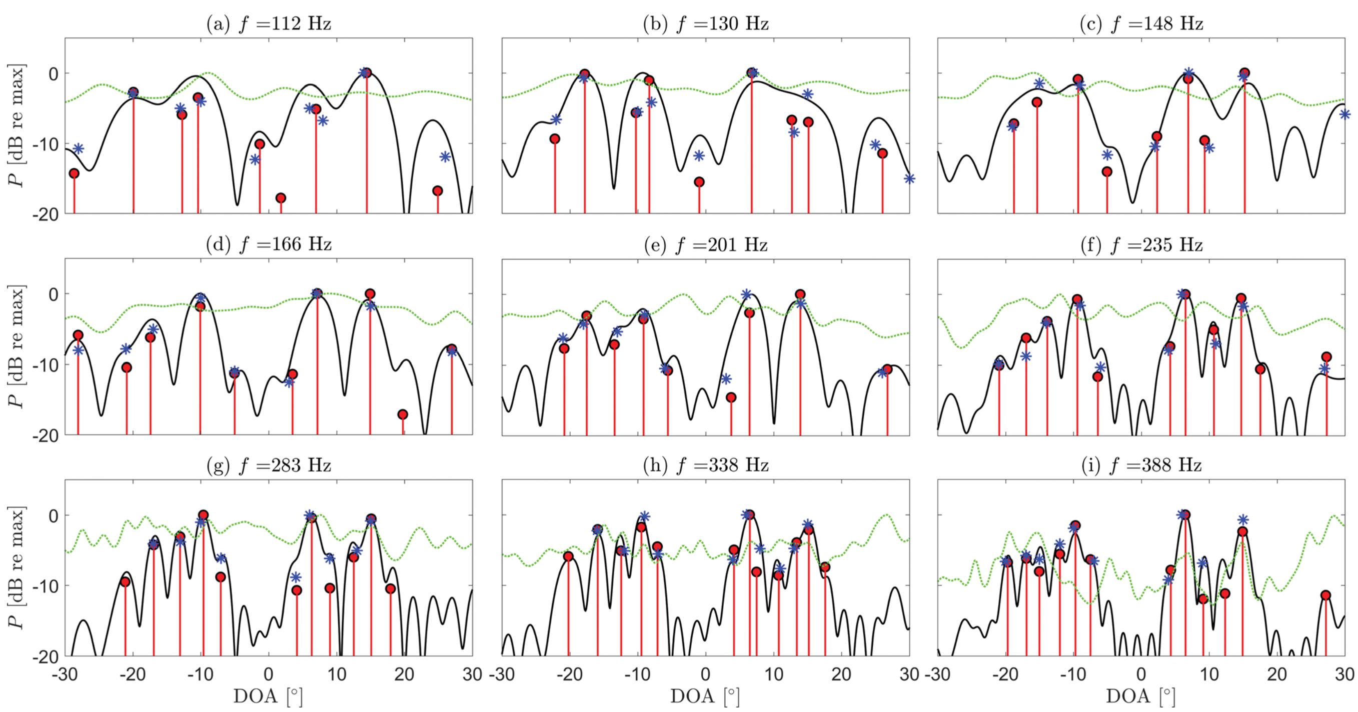

FIG. 5. (Color online) Multiple snapshot DOA estimation for nine frequencies with multiple snapshot grid free CS DOA estimation (○), multiple snapshot conventional CS DOA estimation (∗), CBF (solid line), and MVDR (dotted line).

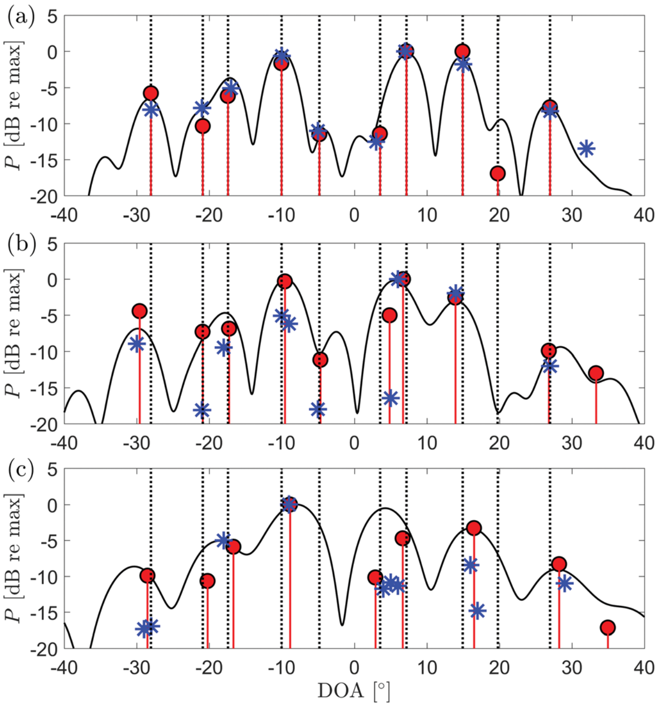

FIG. 6. (Color online) Multiple snapshot DOA estimation with different numbers of sensors for $f = 166$ Hz case in Fig. 4. Multiple snapshot grid free CS DOA estimation (○), multiple snapshot conventional CS DOA estimation (*), and CBF (solid line), for three different numbers of sensors: (a) $M = 64$, (b) $M = 48$, and (c) $M = 32$. The dotted vertical lines indicate estimated DOAs of multiple snapshot grid free CS for $M = 64$.

On the other hand, when the number of sensors is small, the coherence is higher leading to an increase in the bias.

Figure 6 shows the results of multiple snapshot processing using CBF, conventional CS, and grid free CS DOA estimation methods for three different sensor numbers ($M = 64$, 48, 32) for $f = 166$ Hz. We used the same data in Fig. 5(d) and the first M upper sensors. Assuming that 64 sensor result using multiple snapshot grid free CS DOA estimation method is correct, we investigate how the 10 DOAs are estimated for different numbers of sensors. As the number of sensors becomes smaller, performance degrades. Note that CBF merges some of the closely located DOAs, and the conventional CS DOA estimation method with a finite discrete grid suffers from the basis mismatch due to decrease in the number of sensors accompanied by an increase of coherence. By contrast, the present grid free CS DOA estimation method produces more stable estimates over all three cases. Except for two weak strength DOAs ($-4.83°$ and $19.77°$), remaining DOAs are estimated within small errors.

VII. CONCLUSION

We suggested a grid free CS algorithm based DOA estimation technique which can jointly process sequential multiple snapshot data. The group total variation norm minimization was implemented to impose sparsity on the multiple snapshot solution in a continuous angular domain. The sensitive problem of the conventional CS technique, basis mismatch, was overcome by using the continuous model, and multiple snapshot processing provided a more stable estimation than the single snapshot processing. The experimental results from the SWellEx-96 demonstrated advantages of the present model, including (1) capability to avoid the basis mismatch, (2) more stable DOA estimation than the grid free CS single snapshot model and the conventional CS multiple snapshot model, (3) robustness against coherent sources, (4) as well as high resolution. Some interesting areas of research for future works include finding the general theoretical bound of the present model to determine the resolvable minimum separation condition.

ACKNOWLEDGMENTS

This work was supported by the Agency for Defense Development in Korea under Contract No. UD160015DD, the Ministry of Trade, Industry & Energy (MOTIE) (project code: 10045337), and the Civil Military Cooperation Center under the project "Development of Towed Interferometric Synthetic Aperture Sonar," Grant No. 15-CM-SS-01.

APPENDIX: DUAL PROBLEM OF THE GRID FREE CS DOA ESTIMATION WITH NOISE

The multiple snapshot grid free CS DOA estimation problem with additive Gaussian white noise $\mathbf{N} \in \mathbb{C}^{M\times L}$ is equivalent to the following convex problem,

$$\begin{aligned}&\underset{\tilde{X}}{\text{minimize}} \ \ \|\tilde{X}\|_{\text{gTV}}\\ &\text{subject to} \begin{cases}\mathbf{Y} = \left[\mathcal{F}_M\tilde{X}_1,\ \ldots,\ \mathcal{F}_M\tilde{X}_L\right] + \mathbf{N}\\ \|\mathbf{N}\|_f \le \varepsilon_m,\end{cases}\end{aligned} \tag{A1}$$

where the operator $\|\mathbf{N}\|_f$ is the Frobenius norm. To formulate the dual problem of the primal problem [Eq. (A1)], the dual function $g(\mathbf{C}, \xi)$ is the infimum of the Lagrangian[33] for Eq. (A1) and has the form

$$\begin{aligned}g(\mathbf{C},\xi) &= \inf_{\tilde{X}} \|\tilde{X}\|_{\text{gTV}} + \left\langle \mathbf{C}, \mathbf{Y} - \left[\mathcal{F}_M\tilde{X}_1,\ldots,\mathcal{F}_M\tilde{X}_L\right] - \mathbf{N}\right\rangle\\ &\quad + \xi(tr(\mathbf{N}^H\mathbf{N}) - \varepsilon_m^2)\\ &= \langle \mathbf{Y},\mathbf{C}\rangle - \langle \mathbf{N},\mathbf{C}\rangle + \xi(tr(\mathbf{N}^H\mathbf{N}) - \varepsilon_m^2)\\ &\quad + \inf_{\tilde{X}}\left(\|\tilde{X}\|_{\text{gTV}} - \sum_{l=1}^{L}\langle \mathbf{c}_l, \mathcal{F}_M\tilde{X}_l\rangle\right),\end{aligned} \tag{A2}$$

where $\mathbf{C} \in \mathbb{C}^{M\times L}$ is the matrix of the Lagrange dual variables related to the equality constraints $\mathbf{Y} = [\mathcal{F}_M\tilde{X}_1, \ldots, \mathcal{F}_M\tilde{X}_L] + \mathbf{N}$, $\xi \in \mathbb{R}^+$ is a Lagrange multiplier related to the inequality constraint $\|\mathbf{N}\|_f \le \varepsilon_m$, the operator $\langle \mathbf{Y}, \mathbf{C}\rangle$ is the sum of the column-wise inner product of $\mathbf{Y}$ and $\mathbf{C}$, i.e., $\langle \mathbf{Y}, \mathbf{C}\rangle = \sum_{l=1}^{L}\langle \mathbf{y}_l, \mathbf{c}_l\rangle$, and the symbol tr denotes a trace.

Minimizing $g(\mathbf{C}, \xi)$ over the unknown noise $\mathbf{N} \in \mathbb{C}^{M\times L}$,

$$\frac{\partial g(\mathbf{C},\xi)}{\partial \mathbf{N}} = -\mathbf{C} + 2\xi\mathbf{N} = 0, \tag{A3}$$

yields the optimal noise matrix, $\mathbf{N}_{\text{opt}} = \mathbf{C}/(2\xi)$. The dual function for the optimal noise matrix $\mathbf{N}_{\text{opt}}$ becomes

$$g(\mathbf{C},\xi)|_{\mathbf{N}_{\text{opt}}} = \langle \mathbf{Y},\mathbf{C}\rangle - \frac{\langle \mathbf{C},\mathbf{C}\rangle}{2\xi} + \xi\left(\frac{\langle \mathbf{C},\mathbf{C}\rangle}{4\xi^2} - \varepsilon_m^2\right) + \inf_{\tilde{X}}\left(\|\tilde{X}\|_{\text{gTV}} - \sum_{l=1}^{L}\langle \mathbf{c}_l, \mathcal{F}_M \tilde{X}_l\rangle\right). \quad \text{(A4)}$$

Maximizing $g(\mathbf{C},\xi)$ over the Lagrange multiplier related to the inequality constraint ξ,

$$\frac{\partial g(\mathbf{C},\xi)|_{\mathbf{N}_{\text{opt}}}}{\partial \mathbf{N}} = \frac{\langle \mathbf{C},\mathbf{C}\rangle}{4\xi^2} - \varepsilon_m^2 = 0, \quad \text{(A5)}$$

yields the optimal value, $\xi_{\text{opt}} = \|\mathbf{C}\|_f/(2\varepsilon_m)$.

Finally, the dual function for the optimal values $\mathbf{N}_{\text{opt}}$ and ξ_{opt} becomes

$$\begin{aligned} g(\mathbf{C})|_{\mathbf{N}_{\text{opt}},\xi_{\text{opt}}} &= \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f + \inf_{\tilde{X}}\left(\|\tilde{X}\|_{\text{gTV}} - \sum_{l=1}^{L}\langle \mathbf{c}_l, \mathcal{F}_M\tilde{X}_l\rangle\right) \\ &= \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f + \inf_{\tilde{X}}\left(\|\tilde{X}\|_{\text{gTV}} - \sum_{l=1}^{L}\langle \mathcal{F}_M^H\mathbf{c}_l, \tilde{X}_l\rangle\right) \\ &= \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f + \inf_{\tilde{X}}\left(\sup_{\|F(t)\|_2\le 1}\sum_{l=1}^{L}\text{Re}\left[\int_{\mathbb{T}}\overline{F_l(t)}\tilde{X}_l(t)dt\right] - \sum_{l=1}^{L}\text{Re}\left[\int_{\mathbb{T}}\overline{\mathcal{F}_M^H\mathbf{c}_l(t)}\tilde{X}_l(t)dt\right]\right). \end{aligned} \quad \text{(A6)}$$

Note that the last term in Eq. (A6) is trivial ($-\infty$), unless $\sum_{l=1}^{L}|\mathcal{F}_M^H\mathbf{c}_l(t)|^2 \le 1$ for all t. Therefore, the dual function is

$$g(\mathbf{C}) = \begin{cases} \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f, & \sup_t \sum_{l=1}^{L}|\mathcal{F}_M^H\mathbf{c}_l(t)|^2 \le 1 \\ -\infty, & \text{otherwise.} \end{cases} \quad \text{(A7)}$$

The Lagrange dual problem of Eq. (A1) is formulated by maximizing the dual function[33] $g(\mathbf{C})$,

$$\begin{aligned} &\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f \\ &\text{subject to}\ \sup_t \sum_{l=1}^{L}|\mathcal{F}_M^H\mathbf{c}_l(t)|^2 \le 1. \end{aligned} \quad \text{(A8)}$$

By Proposition 2.4 of Ref. 28, the dual problem can be recast as SDP,

$$\begin{aligned} &\underset{\mathbf{C}}{\text{maximize}}\ \langle \mathbf{Y},\mathbf{C}\rangle - \varepsilon_m\|\mathbf{C}\|_f \\ &\text{subject to}\ \begin{bmatrix} \mathbf{Q} & \mathbf{C} \\ \mathbf{C}^H & \mathbf{I} \end{bmatrix} \succcurlyeq 0, \\ &\sum_{i=1}^{M-j} Q_{i,i+j} = \begin{cases} 1, & j=0 \\ 0, & j = 1, 2, \ldots, M-1, \end{cases} \end{aligned} \quad \text{(A9)}$$

where $\mathbf{I}$ is the identity matrix of dimensions $L \times L$.

[1]H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," IEEE Signal Process. Mag. **13**(4), 67–94 (1996).
[2]H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory* (Wiley, New York, 2002), pp. 1–1433.
[3]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Simon & Schuster, New York, 1992), pp. 1–512.
[4]M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Springer, New York, 2010), pp. 1–376.
[5]E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag. **25**(2), 21–30 (2008).
[6]P. Simard and J. Antoni, "Acoustic source identification: Experimenting the l_1 minimization approach," Appl. Acoust. **74**(7), 974–986 (2013).
[7]D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," Signal Process. **86**(3), 624–638 (2006).
[8]M. Cetin, D. M. Malioutov, and A. S. Willsky, "A variational technique for source localization based on a sparse signal reconstruction perspective," in *Proceedings of the Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, FL, 13–17 May 2002, Vol. 3, pp. 2965–2968.
[9]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**(1), 120–136 (2014).
[10]G. F. Edelmann and C. F. Gaumond, "Beamforming using compressive sensing," J. Acoust. Soc. Am. **130**(4), EL232–EL237 (2011).
[11]D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005).
[12]A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," J. Acoust. Soc. Am. **136**(1), 260–271 (2014).
[13]P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, "Multiple and single snapshot compressive beamforming," J. Acoust. Soc. Am. **138**(4), 2003–2014 (2015).
[14]Y. Choo and W. Seong, "Compressive spherical beamforming for localization of incipient tip vortex cavitation," J. Acoust. Soc. Am. **140**(6), 4085–4090 (2016).
[15]J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**(8), 1408–1418 (1969).
[16]R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. thesis, Stanford University, Stanford (1982).
[17]Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," IEEE Trans. Signal Process. **59**(5), 2182–2195 (2011).
[18]D. H. Chae, P. Sadeghi, and R. A. Kennedy, "Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals," in *Proceedings of the Future Computer and Communication (ICFCC)*, Wuhan, China, 21–24 May 2010, Vol. 2, pp. 739–743.
[19]E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," Comm. Pure Appl. Math. **67**(6), 906–956 (2014).
[20]E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," J. Fourier Anal. Appl. **19**(6), 1229–1254 (2013).
[21]G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," IEEE Trans. Inf. Theory **59**(11), 7465–7490 (2013).
[22]B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," IEEE Trans. Signal Process. **61**(23), 5987–5999 (2013).
[23]J. M. Azais, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," Appl. Comput. Harmon. Anal. **38**(2), 177–195 (2015).
[24]V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," Found. Comput. Math. **15**(5), 1315–1355 (2015).
[25]R. Heckel, V. I. Morgenshtern, and M. Soltanolkotabi, "Super-resolution radar," Inf. Inference **5**(1), 22–75 (2016).
[26]A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," J. Acoust. Soc. Am. **137**(4), 1923–1935 (2015).
[27]Y. Park, W. Seong, and Y. Choo, "Compressive time delay estimation off the grid," J. Acoust. Soc. Am. **141**(6), EL585–EL591 (2017).
[28]C. Fernandez-Granda, "Super-resolution of point sources via convex programming," Inf. Inference **5**(3), 251–303 (2016).
[29]Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," IEEE Trans. Signal Process. **64**(5), 1257–1269 (2016).

[30]Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," IEEE Trans. Signal Process. **64**(19), 5145–5157 (2014).
[31]N. O. Booth, P. A. Baxley, J. A. Rice, P. W. Schey, W. S. Hodgkiss, G. L. D'Spain, and J. J. Murray, "Source localization with broad-band matched-field processing in shallow water," IEEE J. Ocean. Eng. **21**(4), 402–412 (1996).
[32]G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**(6), 3245–3265 (1999).
[33]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004), pp. 1–684.
[34]M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx (Last viewed July 17, 2017).