Minimum Dispersion Beamforming for Non-Gaussian Signals

Xue Jiang, Student Member, IEEE, Wen-Jun Zeng, Member, IEEE, A. Yasotharan, Hing Cheung So, Senior Member, IEEE, and Thiagalingam Kirubarajan, Senior Member, IEEE

Abstract—Most of the existing beamforming methods are based on the Minimum Variance (MV) criterion. The MV approach is statistically optimal only when the signal, interferences and the noise are Gaussian-distributed. However, non-Gaussian signals arise in a variety of practical applications. In this paper, Minimum Dispersion Distortionless Response (MDDR) beamforming, which minimizes the $\ell_p$-norm of the output while constraining the desired signal response to be unity, is devised for non-Gaussian signals. It is shown that the MDDR beamformer, which implicitly exploits non-Gaussianity, can improve the performance significantly if $p > 2$ for sub-Gaussian signals or $p < 2$ for super-Gaussian signals. Three efficient algorithms, the Iteratively Reweighted Minimum Variance Distortionless Response (IR-MVDR), complex-valued full Newton’s and partial Newton’s methods, are developed to solve the resulting $\ell_p$-norm minimization with a linear constraint. Furthermore, the MDDR beamformer with a single constraint is generalized to the Linearly Constrained Minimum Dispersion (LCMD) beamformer with multiple linear constraints, which exhibits robustness against steering vector mismatch. The LCMD beamformer yields significant performance improvement over the conventional Linearly Constrained Minimum Variance (LCMV) beamformer. Simulation results are provided to demonstrate the superior performance of the proposed minimum dispersion beamforming approaches.

Index Terms—Beamforming, interference suppression, $\ell_p$-norm minimization, minimum dispersion, minimum variance distortionless response (MVDR) beamformer, non-Gaussianity.

I. INTRODUCTION

Beamforming is a spatial filtering technique [1] using a sensor array to enhance the signal of interest (SOI) and suppress interferences and noise. It is widely used in radar, sonar, wireless communications, Global Positioning System (GPS) navigation and microphone array speech processing [2], [3]. The conventional beamforming techniques, e.g., Delay-and-Sum beamforming and its variants [4], are independent of the input signal data, which results in limited capability for interference suppression. The modern beamforming techniques are data-dependent, which can adaptively achieve signal enhancement and interference mitigation [1], [4].

Many existing data-dependent beamforming methods are based on the Minimum Variance (MV) criterion [3], [5]. In particular, the Minimum Variance Distortionless Response (MVDR) beamformer constrains the response of the SOI to be unity and minimizes the variance of the output [6]. It is generally recognized as a best beamformer since minimizing the output variance with distortionless response constraint is equivalent to maximizing the output signal-to-interferences-plus-noise ratio (SINR). However, the MVDR approach can achieve optimality only when the true covariance matrix is available [3]. This requires an infinite number of snapshots, which is impractical. In fact, the performance of the MVDR beamformer degrades significantly with short data length [3], [7]. On the other hand, the eigenspace-based beamformer [8], which uses only the signal-plus-interference subspace component of the sample correlation matrix, can mitigate the adverse effects induced by the noise subspace disturbance, and hence performs better than the MVDR beamformer.

The MVDR and subspace beamformers exploit only the second-order statistics of the array output. The MV criterion is statistically optimal for Gaussian signals and noise because the first- and second-order statistics of a Gaussian distribution contain all necessary statistical information. Nevertheless, many real-world signals are non-Gaussian [9], [10]. Based on the kurtosis of a distribution, non-Gaussian distributions can be classified into two categories, namely, sub-Gaussian with kurtosis smaller than three and super-Gaussian with kurtosis larger than three [11], [12]. Many signals that arise in wireless communications, radar, sonar, and GPS navigation are sub-Gaussian [13]. On the other hand, common super-Gaussian signals include speech and biomedical data [12], [14]. Non-Gaussian noise is also frequently encountered in practice [10]. For non-Gaussian signals and noise, the higher-order and fractional lower-order statistics contain useful information and can be utilized to improve the beamformer performance. In [15], a blind beamforming method was proposed for non-Gaussian signals using fourth-order cumulants. However, it uses only the fourth-order statistics and other statistical information is discarded. The $\ell_p$-norm minimization criterion was proposed for beamforming in [16]. However, the goal of [16] was to suppress impulsive noise only with $1 \leq p < 2$. A smaller value
of $p$ was suggested in [16] although there is no theoretical justification. However, it will be shown in this paper that this recommendation is ambiguous. Moreover, there is no efficient algorithm for solving the resulting $\ell_p$-norm minimization. It uses the gradient descent method, but there is no discussion on the selection of the step size parameter. The special case with $p = 1$ was reconsidered in [17] while taking into account the steering vector error. Again, the gradient descent method was adopted to solve the optimization problem. In fact, the gradient descent scheme is quite slow and it may not even converge unless the step size is appropriate.

In this paper, we do not explicitly construct any higher- or lower-order statistics but adopt the Minimum Dispersion (MD) criterion for beamforming. The proposed Minimum Dispersion Distortionless Response (MDDR) beamformer minimizes the $\ell_p$-norm ($p \geq 1$) of the output while constraining the desired signal response to be unity. The dispersion, which is a generalization of variance, implicitly exploits the higher-order statistics for $p > 2$ or fractional lower-order statistics for $p < 2$ [18]. Compared with [16] and [17], the proposed MDDR beamformer can be tailored to Gaussian, sub-Gaussian or super-Gaussian signals and noise by choosing different values of $p$. We fully analyze the selection of $p$ for signals and noise with different statistical properties. To our best knowledge, there is no computationally simple and efficient numerical algorithm for the MD beamforming problem. Therefore, three efficient iterative algorithms, which converge fast and are computationally efficient, are also proposed.

Another drawback of the standard MVDR beamformer is that it is too sensitive to steering vector mismatch [3], [5], [19]–[21]. The SOI will be considered as an interference and hence attenuated by the MVDR beamformer if the steering vector of the SOI is imprecise. Therefore its performance dramatically degrades under these conditions. One common cause of steering vector mismatch is due to the angle-of-arrival (AOA) estimation error. Several mismatch-robust beamforming approaches have been proposed [5], [19], [20], [22]. The Linearly Constrained Minimum Variance (LCMV) beamformer [22] is a direct extension of the MVDR beamformer. It tries to cope with the AOA mismatch by imposing multiple linear constraints for a small spread of angles around the nominal AOA. Analogous to this, we derive the Linearly Constrained Minimum Dispersion (LCMD) beamformer, which uses multiple linear constraints to enhance the robustness against steering vector uncertainty.

We briefly summarize the contributions of our work on MD beamforming as follows.

i) It is pointed out that the MVDR approach is not statistically optimal in the presence of non-Gaussian signals. We show that the proposed MDDR beamformer can effectively exploit the non-Gaussianity and hence considerably improve the performance if $p > 2$ for sub-Gaussian signals or $p < 2$ for super-Gaussian signals. We also discuss the case of $p \rightarrow \infty$, i.e., the $\ell_\infty$-norm MDDR beamformer.

ii) Three efficient iterative algorithms, the Iteratively Reweighted MVDR (IR-MVDR) and the complex-valued full Newton’s and partial Newton’s methods, are developed for the efficient computation of the MDDR beamformer. In Section IV, the minimum dispersion beamformer is extended to multiple linear constraints.

We use bold capital upper case and lower case letters to represent matrices and vectors, respectively. The superscripts $\cdot^T$, $\cdot^*$ and $\cdot^H$ denote the transpose, complex conjugate and Hermitian transpose, respectively. The $\mathbf{I}$ denotes the identity matrix while $\mathbb{F}\{\cdot \}$ and $\mathbb{I}\{\cdot \}$ are the expectation operator and imaginary unit, respectively. $\mathbb{R}\{\cdot \}$ represents the real and imaginary parts of a complex-valued number, respectively, and $| \cdot |$ denotes the absolute value of a real number or the modulus of a complex number. The $\| \cdot \|$ are the Euclidean norm and the $\ell_p$-norm of a vector, respectively. Finally, $\mathbb{R}$, $\mathbb{R}^+$ and $\mathbb{C}$ are used to denote the sets of real, non-negative real, and complex numbers, respectively.

II. PROBLEM STATEMENT, MVDR AND SUBSPACE BEAMFORMERS

A. Signal Model

Consider an array of $M$ receiving sensors. The complex baseband signal received by the $m$th ($1 \leq m \leq M$) sensor at time $n$ is denoted as $x_m(n)$. The vector of the array output $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T$ is expressed as [2]

$$\mathbf{x}(n) = s(n)\mathbf{a} + \sum_{i=1}^{I} s_i(n)\mathbf{a}_i + \mathbf{n}(n),$$

where $s(n)$ is the SOI, $\{s_i(n)\}_{i=1}^{I}$ are the $I$ interferences, $\mathbf{a} \in \mathbb{C}^M$ and $\{\mathbf{a}_i\}_{i=1}^{I}$ are the steering vectors of the SOI and inter-
ferences, respectively, and $v(n)$ is the additive noise. The SOI is assumed to be uncorrelated with the interferences and noise. This assumption holds in many practical applications and it is also widely assumed in array processing, e.g., see [3], [8], [15], and [20]. Certainly, there exist cases where the signals are mutually correlated or coherent due to multipath propagation [23]. This case, which may be handled using spatial smoothing techniques [23], is beyond the scope of this paper. We collect all the $I$ interferences into a term $i(n) = \sum_i s_i(n) a_i$. Depending on the array configuration, the steering vector has different forms. For example, $a$ has the following form for a uniform linear array (ULA) [2]

$$a(\theta) = \left[ e^{j \frac{2\pi}{\lambda} d \sin \theta}, \ldots, e^{j \frac{2\pi}{\lambda} (M-1) \sin \theta} \right]^T. \quad (2)$$

where $\theta$ is the AOA, $d$ is the inter-sensor spacing, and $\lambda$ is the wavelength.

The task of data-dependent beamforming is to design a beamformer $w \in \mathbb{C}^M$ to enhance the SOI and suppress interference and noise using the observed data $X = \{x(1), \ldots, x(N)\} \in \mathbb{C}^{M \times N}$ with $N$ being the number of snapshots. The output of the beamformer is expressed as

$$y(n) = w^H x(n). \quad (3)$$

It is desired that the output $y(n)$ preserves the desired signal component and mitigates the interference and noise.

### B. MVDR Beamformer

The output SINR, which is taken as the performance measure of a beamformer, is defined as [2]

$$\text{SINR} = \frac{E\left\{ |s(n)^2|^2 \right\}}{E\left\{ |w R_{i+n} w|^2 \right\}} = \sigma_s^2 \frac{w^H a^2}{w^H R_{i+n} w} \quad (4)$$

where $\sigma_s^2 = E\{|s(n)|^2\}$ is the power of the SOI, $R_{i+n}$ is the interferences-plus-noise covariance matrix. Since the SOI is uncorrelated with the interferences and noise, the covariance matrix of $x(n)$ can be written as

$$R = E\{x(n)x^H(n)\} = \sigma_s^2 a a^H + R_{i+n}. \quad (5)$$

The MVDR beamformer [6] maximizes the output SINR by minimizing the total output variance while constraining the SOI response to be unity, i.e.,

$$\min_w \left\{ E\{|y(n)|^2\} = w^H R w \right\} \quad \text{s.t. } a^H w = 1. \quad (6)$$

Note that (6) is equivalent to minimizing $w^H R_{i+n} w$ subject to $a^H w = 1$ because

$$w^H R w = w^H R_{i+n} w + \sigma_s^2 \frac{w^H a^2}{w^H R_{i+n} w}. \quad (7)$$

The closed-form solution of (6) is known as MVDR or Capon beamformer [6] and is given by

$$w_{\text{MVDR}} = \frac{R_{i+n}^{-1} a}{a^H R_{i+n}^{-1} a}. \quad (8)$$

In practice, the true covariance matrix $R$ is unknown and it is estimated using the $N$ snapshots as

$$\hat{R} = \frac{1}{N} \sum_{n=1}^N x(n)x^H(n) - \frac{1}{N} XX^H. \quad (9)$$

By replacing $R$ in (8) with its estimate $\hat{R}$, we obtain the so-called Sample Matrix Inversion (SMI) beamformer. In other words, the SMI beamformer is a practical implementation of the MVDR beamformer. When the sample size $N$ is small, the covariance matrix $R$ cannot be estimated accurately and the performance of the SMI beamformer will degrade [3], [7], [8].

### C. Subspace Beamformer

It has been shown that the performance of the SMI beamformer is degraded mostly by the disturbed noise subspace due to finite sample effect [8]. To mitigate the adverse effects induced by the noise subspace disturbance, the eigenspace-based beamformer that uses only the signal-plus-interference subspace component of the sample correlation matrix is proposed in [8]. The eigenvalue decomposition (EVD) of $R$ is given by

$$R = UTU^H = U_s \Gamma_s U_s^H + U_n \Gamma_n U_n^H \quad (10)$$

where $U = [U_s, U_n]$, $\Gamma_s = \text{diag}\{\gamma_1, \ldots, \gamma_{I+1}\}$ is a diagonal matrix containing the $I+1$ non-increasing principal eigenvalues and $U_s$ contains the corresponding orthonormal eigenvectors. Note that $U_n$ contains the $M - (I+1)$ orthonormal eigenvectors associated with the eigenvalues listed in the diagonal elements of $\Gamma_n = \text{diag}\{\gamma_{I+2}, \ldots, \gamma_M\}$. The range space spanned by $U_s$ is the signal-plus-interference subspace and its orthogonal complement, spanned by $U_n$, is the noise subspace. The MVDR beamformer of (8) can be expressed in terms of eigenspace representation as [8]

$$w_{\text{eig}} = c \left( U_s \Gamma_s^{-1} U_s^H + U_n \Gamma_n^{-1} U_n^H \right) a \quad (11)$$

where $c$ is a constant and does not affect the performance of a beamformer. The subspace beamformer further assumes that the noise are spatially white. This requires the noise covariance matrix to be expressed as $\sigma_n^2 I$ with $\sigma_n^2$ being the noise variance. In this case, the eigenvalues of the noise subspace satisfies $\gamma_{I+2} = \cdots = \gamma_M = \sigma_n^2$. Under this assumption, one can conclude that the steering vector $a$ is orthogonal to $U_n$, i.e., $U_n^H a = 0$ [24]. Therefore (11) can be simplified as

$$w_{\text{eig}} = \frac{U_s \Gamma_s^{-1} U_s^H a}{a^H U_s \Gamma_s^{-1} U_s^H a} \quad (12)$$

Note that $w_{\text{eig}}$ in (12) is referred to as the eigenspace beamformer [8]. We also call it subspace beamformer since it can be constructed from the signal-plus-interference subspace. The idea of subspace beamforming has also been applied to multiuser detection and interference suppression in Code Division Multiple Access (CDMA) systems [25]. In practical applications, the subspace needs to be estimated by performing EVD on the sample covariance matrix $\hat{R}$. It has been shown that the
subspace beamformer is superior to the SMI beamformer since it mitigates the disturbance of the noise subspace [8].

A key issue with the subspace beamformer is the need to determine the dimension of signal-plus-interference subspace, which is equal to the value of the number of SOI plus interferences, i.e., \( I + 1 \). The Akaike Information Criterion (AIC) or the Minimum Description Length (MDL) criterion [26] can be applied to this source enumeration problem.

III. MINIMUM DISPERSION BEAMFORMER VIA \( \ell_p \)-NORM MINIMIZATION

A. Motivation by Non-Gaussianity

The MVDR and subspace beamformers utilize only the second-order statistics. For Gaussian signals and noise, the MV criterion is statistically optimal because the first- and second-order statistics of a Gaussian distribution contain all necessary and sufficient statistical information. However, many signals in practice are non-Gaussian distributed. Random signals can be classified into three classes according to the kurtosis [12]. The kurtosis of a random stationary signal \( \kappa(s(n)) \) with zero-mean is defined as

\[
\kappa(s(n)) = \frac{E[|s(n)|^4]}{(E[|s(n)|^2])^2}. \tag{13}
\]

If \( s(n) \) is Gaussian, then \( \kappa(s(n)) = 3 \). If \( \kappa(s(n)) < 3 \), \( s(n) \) is sub-Gaussian. There are a number of sub-Gaussian distributions such as uniform distribution and Bernoulli distribution. If \( \kappa(s(n)) > 3 \), \( s(n) \) is super-Gaussian. Super-Gaussian distributions, which include Laplace distribution and \( \alpha \)-stable distribution [28], are also common. For example, the Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM), radar, sonar, and GPS navigation signals are sub-Gaussian [13]. A common example of super-Gaussian signal is speech [12]. In addition to non-Gaussian signals, non-Gaussian noise is also frequently encountered [10]. For non-Gaussian signals, the higher-order (higher than 2) and lower fractional order statistics contain useful information and can be exploited to improve the performance of beamforming. In [15], a blind beamforming technique was proposed for non-Gaussian signals using the fourth-order cumulants. In the following, we will introduce the minimum dispersion beamforming, which implicitly uses the higher-order or fractional lower-order statistics.

B. Minimum Dispersion Criterion

The proposed MDDR beamformer is obtained by solving the following linearly constrained optimization problem:

\[
\min_{\mathbf{w}} \{ \mathbf{w}^H \mathbf{x}(n)^p \} \quad \text{s.t.} \quad \mathbf{a}^H \mathbf{w} = 1 \tag{14}
\]

where \( p \geq 1 \). Clearly, the MDDR beamformer is reduced to the MVDR beamformer for \( p = 2 \). In statistics, \( E\{ y(n)^p \} \) is referred to as dispersion, which is a generalization of variance [18]. Therefore, we call the solution of (14) as the minimum dispersion beamformer. It should be pointed out that the criterion of (14) has been proposed for beamforming in [16]. Furthermore, \( p = 1 \) is adopted in the beamformer of [17]. However, no insightful guideline on how to choose an appropriate \( p \) is given in [16] and it advocates to use small value of \( p \), which is not technically accurate because the choice of \( p \) is not only related to statistical characteristics of the noise but also the signal sources. Also at present, there is no efficient numerical algorithms to solve the problem in (14). In the following, we will develop three low complexity algorithms with fast convergence to calculate the MDDR beamformer and discuss how to choose an appropriate \( p \).

Replacing the expectation with the sample mean and ignoring the constant \( 1/N \), (14) can be rewritten as

\[
\min_{\mathbf{w}} \{ f_p(\mathbf{w}) = \| \mathbf{X}^H \mathbf{w} \|^p \} \quad \text{s.t.} \quad \mathbf{a}^H \mathbf{w} = 1 \tag{15}
\]

where \( \mathbf{X}^H \mathbf{w} = \mathbf{y}^* \) is the conjugate of the beamformer output \( \mathbf{y} = [y(1), \ldots, y(N)]^T \), and the \( \ell_p \)-norm is defined as

\[
\| \mathbf{y} \|^p = \left( \sum_{n=1}^{N} |y(n)|^p \right)^{1/p}. \tag{16}
\]

However, (15) has no closed-form solution except for \( p = 2 \). The optimization problem in (15) is convex for \( p \geq 1 \) and the global optimum is guaranteed using the standard interior point method for convex optimization [29]. However, we propose simpler and more efficient algorithms for solving it. Note that the \( \ell_p \)-norm minimization of (15) is a constrained optimization problem, but different from the one encountered in robust linear regression using least \( \ell_p \)-norm [30], [31], which is unconstrained. Before discussing the optimization algorithms, we give general guidelines for selecting \( p \). For Gaussian signals, the optimal \( p \) is 2. For sub-Gaussian signals, \( p > 2 \) will achieve better performance, whereas \( p < 2 \) is preferred for super-Gaussian signals. The optimal value of \( p \) depends on the probability density function (PDF) of the signals. In the simulation examples, we will further investigate the selection of \( p \).

For signals with very strong super-Gaussianity, e.g., heavy-tailed distributions, like Laplacian [27] or \( \alpha \)-stable distribution [28], one may require \( 0 < p < 1 \). However, \( 0 < p < 1 \) leads to a non-differential and nonconvex \( \ell_p \)-norm minimization problem. It has been pointed out that (15) with \( 0 < p < 1 \) is strongly NP-hard and that the global minimum is difficult to obtain [32]. Due to the mathematical difficulty, we do not consider the choice of \( 0 < p < 1 \) in this work.

C. Iteratively Reweighted MVDR Algorithm

We rewrite the objective function in (15) as

\[
f_p(\mathbf{w}) = \| \mathbf{y}^* \|^p = \sum_{n=1}^{N} y(n)^p = \sum_{n=1}^{N} |y(n)|^p 2^y(n)^2 \tag{17}
\]

where \( \mathbf{\Phi} \) is a diagonal weighting matrix

\[
\mathbf{\Phi} = \text{diag} \left\{ |y(1)|^{(p-2)/2}, \ldots, |y(N)|^{(p-2)/2} \right\}. \tag{18}
\]

with its diagonal elements being real and positive numbers. This means that the \( \ell_p \)-norm minimization problem can be converted into an \( \ell_2 \)-norm minimization one. Equation (17) can be further expressed as

\[
f_p(\mathbf{w}) = \mathbf{y}^T \mathbf{\Phi}^* \mathbf{\Phi} \mathbf{y}^* = \mathbf{y}^T \mathbf{D}(\mathbf{w}) \mathbf{y}^* = \mathbf{w}^H \mathbf{X} \mathbf{D}(\mathbf{w}) \mathbf{X}^H \mathbf{w} \tag{19}
\]
where
\[ D(w) = \text{diag}\{ |y(1)|^{-2}, \ldots, |y(N)|^{-2} \}. \] (20)

Note that \( D \) depends on the unknown \( w \) because it is related to \( y \). Therefore, it is a function of \( w \), which is written as \( D(w) \).

Now we rewrite (15) as
\[ \min_w w^H (XD(w)X^H) w \]
\[ \text{s.t. } a^H w = 1 \] (21)
whose optimal solution is given by
\[ w = \frac{(XD(w)X^H)^{-1} a}{a^H (XD(w)X^H)^{-1} a}. \] (22)

Equation (22) has a structure similar to that of the MVDR beamformer but the covariance matrix has been reweighted using the weighting matrix \( D(w) \). However, we cannot obtain a closed-form expression for the optimal \( w \) since \( D(w) \) is related to the unknown \( w \). An alternative is to use the following fixed-point iteration
\[ w^{k+1} = \frac{(XD(w^k)X^H)^{-1} a}{a^H (XD(w^k)X^H)^{-1} a}, \] (23)
to find the optimal solution, where the superscript \((\cdot)^k\) is used to denote the result at the \( k \)-th \((k = 0, 1, \ldots)\) iteration. In each iteration, the MVDR beamformer with a reweighted covariance matrix is computed. Therefore, we refer to this algorithm as IR-MVDR, which is summarized in Algorithm 1. The initial value can be taken as the data-independent beamformer
\[ w^0 = a/\|a\|^2. \] (24)

Next, we analyze the computational complexity of the IR-MVDR method. The complexity of matrix multiplication \( XD(w^k)X^H \) is \( O(NM^2) \) because \( D(w^k) \) is diagonal. The computational cost for calculating \((XD(w^k)X^H)^{-1}a\) is \( O(M^3) \). Hence the complexity of IR-MVDR is \( max(O(NM^2), O(M^3)) \) in each iteration. Since it is always assumed that the sample size is larger than the number of sensors, i.e., \( N > M \), then the complexity is \( O(NM^2) \) of each iteration.

Remark 1: The convergence behavior of the IR-MVDR is similar to that of the iteratively reweighted least-squares (IRLS) algorithm for unconstrained \( \ell_p \)-norm minimization [30], [31]. Their convergence is guaranteed only for some specific values of \( p \). We find that it does not converge when \( p \geq 3.4 \). This means that 3.4 is the critical value of the IR-MVDR scheme.

D. Complex-Valued Newton’s Methods With Equality Constraint

Despite the simplicity of the IR-MVDR algorithm, it may not converge [33]. In this subsection, we propose two Newton’s methods, whose convergence is guaranteed for \( p > 1 \), to efficiently solve the complex-valued \( \ell_p \)-norm minimization of (15). It is worth pointing out that the proposed Newton’s methods are different from those in [34]. The Newton’s methods of [34] can deal only with the unconstrained \( \ell_p \)-norm minimization problem, but the proposed schemes can handle the problem with equality constraints.

The optimization problem of (15) involves complex-valued variables. We first give the definition of the gradient with respect to (w.r.t.) complex-valued variables. The gradient of the objective \( f_p(w) \) w.r.t. the complex vector \( w \in \mathbb{C}^M \) is defined as
\[ \nabla f_p(w) = \frac{\partial f_p(w)}{\partial w} = \left[ \frac{\partial f_p}{\partial w_1^1}, \ldots, \frac{\partial f_p}{\partial w_M^1} \right]^T. \] (25)
where
\[ \frac{\partial f_p}{\partial w_i^*} = -\frac{1}{2} \left( \frac{\partial f_p}{\partial \text{Re}(w_i)} + j \frac{\partial f_p}{\partial \text{Im}(w_i)} \right), \quad i = 1, \ldots, M. \]

The Newton’s method uses the following iteration to find the minimizer of (15):
\[ w \leftarrow w + \Delta w \] (26)
where \( w \) is the point of the current iteration, and \( \Delta w \) is the update direction, which is called the Newton direction [35]. It requires that \( w \) be feasible in each iteration, which is equivalent to requiring
\[ a^H \Delta w = 0. \] (27)

The second-order Taylor expansion of \( f_p(w + \Delta w) \) around the complex vector \( w \) is given by
\[ f_p(w + \Delta w) = f_p(w) + g_w(\Delta w) + o(\|\Delta w\|^2). \] (28)
Here, $q_w(\Delta w)$ is a quadratic function w.r.t. $\Delta w$, which contains the first two order expansion terms and can be expressed as

$$q_w(\Delta w) = \nabla f_p(w)^H \Delta w + \nabla f_p(w)^T \Delta w^* + \frac{1}{2} \Delta w^H H_{ww} \Delta w + \frac{1}{2} \Delta w^H H_{ww^*} \Delta w^* + \frac{1}{2} \Delta w^H H_{ww^*} \Delta w^* + \frac{1}{2} \Delta w^H H_{ww} \Delta w^*$$

(29)

where

$$\nabla f_p(w) = \frac{\partial f_p(w)}{\partial w^*} = \frac{\nu}{2} XD(w)X^H w$$

(30)

is the gradient of $f_p(w)$, and the four $M \times M$ partial Hessian matrices are expressed as

$$H_{ww} = \frac{\partial^2 f_p(w)}{\partial \Delta w \partial \Delta w^T} = \frac{\nu(p^2)}{4} X^* E(w) X^H,$$

$$H_{ww^*} = H_{ww^*}^*, H_{ww^*}^T = H_{ww^*}^T,$$

(31)

where $E(w)$ is a diagonal matrix of the form

$$E(w) = \text{diag}\{ |y(1)|^{p-4} y^2(1), \ldots, |y(N)|^{p-4} y^2(N) \}.$$  

(32)

The partial Hessians $H_{ww^*}$ and $H_{ww^*}$ are positive definite because $D(w)$ in (20) is positive definite. Then, $q_w(\Delta w)$ can be written more compactly as

$$q_w(\Delta w) = \begin{bmatrix} \nabla f_p(w)^H \nabla f_p(w) \end{bmatrix} \Delta w + \begin{bmatrix} \Delta w^H \Delta w^* \end{bmatrix}$$

(33)

where the $2M \times 2M$ full Hessian matrix is denoted as

$$H = \begin{bmatrix} H_{ww^*} & H_{ww^*}^* \\ H_{ww^*} & H_{ww^*}^* \end{bmatrix}.$$  

(34)

The full Hessian matrix is positive definite when $p > 1$. It is noticed that the two off-diagonal block matrices $H_{ww^*}^*$ and $H_{ww^*}$ become zero if $p = 2$. In this case, these two partial Hessian matrices contain no information. When $p \neq 2$, these two matrices do not vanish and contain useful information for optimization.

1) Full Newton’s Method: For fixed $w$ in the current iteration, the Newton’s method aims to find an update direction $\Delta w$ that minimizes the quadratic function $q_w(\Delta w)$ under the linear constraint. That is,

$$\min_{\Delta w} q_w(\Delta w)$$

s.t. $a^T \Delta w = 0.$

(35)

We use the method of Lagrangian multipliers [35] to solve the constrained optimization problem. The Lagrangian function of (35) is

$$\mathcal{L}(\Delta w, \lambda) = q_w(\Delta w) + \lambda a^T \Delta w$$

(36)

where $\lambda$ is the Lagrangian multiplier. In optimization with complex-valued variables, the unknown $\Delta w$ and its conjugate $\Delta w^*$ are jointly considered. According to the optimal condition and using (29), we obtain

$$\frac{\partial \mathcal{L}(\Delta w, \lambda)}{\partial \Delta w^*} = \nabla f_p(w) + H_{ww^*} \Delta w + H_{ww^*} \Delta w^* + \lambda a = 0,$$

$$\frac{\partial \mathcal{L}(\Delta w, \lambda)}{\partial \Delta w} = \nabla f_p(w)^* + H_{ww^*} \Delta w^* + \lambda a^* = 0,$$

$$\frac{\partial \mathcal{L}(\Delta w, \lambda)}{\partial \lambda} = a^H \Delta w = 0$$

(37)

which can be written compactly as

$$\begin{bmatrix} H_{ww^*} & H_{ww^*}^* & a \\ H_{ww^*}^* & H_{ww^*} & a^* \\ a^H & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta w \\ \Delta w^* \\ \lambda \end{bmatrix} = - \begin{bmatrix} \nabla f_p(w) \\ \nabla f_p(w)^* \end{bmatrix}.$$  

(38)

By solving the linear system of (38), the Newton direction $\Delta w$ is obtained. Proposition 1 guarantees the uniqueness of the solution of (38).

Proposition 1: There is a unique solution for (37) or (38) with $p > 1$. Hence, the update direction of the full Newton’s method given by (37) is unique. In other words, the coefficient matrix of (38) must be nonsingular.

**Proof:** It follows from (37) that

$$\begin{bmatrix} H_{ww^*} & H_{ww^*}^* \\ H_{ww^*} & H_{ww^*}^* \end{bmatrix} \begin{bmatrix} \Delta w \\ \Delta w^* \end{bmatrix} = - \begin{bmatrix} \nabla f_p(w) + \lambda a \\ \nabla f_p(w)^* + \lambda a^* \end{bmatrix}.$$  

(39)

For $p > 1$, the full Hessian matrix in (34) is positive definite and hence nonsingular since the objective function $f_p(w)$ is strictly convex. Then, the linear system of equations of (39) has a unique solution:

$$\begin{bmatrix} \Delta w \\ \Delta w^* \end{bmatrix} = -H^{-1} \begin{bmatrix} \nabla f_p(w) + \lambda a \\ \nabla f_p(w)^* + \lambda a^* \end{bmatrix}.$$  

(40)

Taking the complex conjugate of both sides of the constraint $a^H \Delta w = 0$ yields $a^T \Delta w^* = 0$. Hence, we obtain the equivalent constraint

$$[a, a^*]^H \begin{bmatrix} \Delta w \\ \Delta w^* \end{bmatrix} = 0.$$  

(41)
Substituting (40) into (41) leads to a unique solution of the Lagrangian multiplier

\[
\lambda = \begin{bmatrix} a, a^* \end{bmatrix}^H H^{-1} \begin{bmatrix} \nabla f_p(w) \\ \nabla f_p(w)^* \end{bmatrix}.
\]

(42)

Note that the denominator of (42), which is a quadratic form associated with \( H^{-1} \), must be positive and hence impossible to be zero because the inverse of the full Hessian \( H^{-1} \) is positive definite and \( a \neq 0 \). This guarantees the uniqueness of \( \lambda \). Now it is clear that \((\Delta w, \Delta w^*; \lambda)\) has a unique solution and the update direction is uniquely determined. This also means that the coefficient matrix of (38) is nonsingular.

The full Newton’s method exploits all the partial Hessian matrices, which is the reason for its name. The complexity of solving (38) is \( O(M^2) \) since the size of the system is \( (2M+1)^2 \). Recalling that the cost of calculating the partial Hessian matrices is \( O(NM^2) \), the complexity of the full Newton’s method is thus \( n \times O(NM^2) \), in each iteration, which is the same as that of the IR-MVDR algorithm. After obtaining the Newton direction \( \Delta w^k \) in the \( k \)th iteration, the beamformer is updated as

\[
w^{k+1} = w^k + \mu_k \Delta w^k
\]

(43)

where \( \mu_k \geq 0 \) is the step size. The selection of \( \mu_k \) is an important issue. In the conventional Newton’s method [35], the fixed step size \( \mu_k = 1 \) is adopted, which is clearly not optimal. For a given Newton direction \( \Delta w^k \), the optimal step size is given by solving the line search

\[
\mu_k = \arg \min_{\mu \geq 0} \| X^H (w^k + \mu \Delta w^k)^* \|^p.
\]

(44)

Recalling that \( X^H w^k = y^* \) is the conjugate of the array output at the \( k \)th iteration and denoting the output increment as \( X^H \Delta w^k = \Delta y^* \), the objective function w.r.t. the step size \( \mu \) can be written as

\[
\min_{\mu \geq 0} f_p(\mu) = \| \Delta y^* \|^p.
\]

(45)

which is a simple one-dimensional optimization problem and can be easily solved by traditional line search techniques such as the golden section search or the tangential method [36]. The global optimality of \( \mu \) is guaranteed since \( f_p(\mu) \) is unimodal w.r.t. \( \mu \) if \( p \geq 1 \). The line search has a marginal computational cost of \( O(N) \).

If the initial value \( w^0 \) is feasible, then the solution of each iteration \( w^k \) is feasible due to the constraint \( a^H \Delta w = 0 \). We can initialize the beamformer using the SMI beamformer [8]

\[
w_{\text{SMI}} = \left( X X^H \right)^{-1} a
\]

(46)

or just initialize it as the data-independent beamformer of (24). These two initializations ensure that \( w^0 \) is feasible. The full Newton’s method for MDDR beamforming is summarized in Algorithm 2. The convergence of the Newton’s method has been already proved [35]–[37]. In particular, it converges to the global minimum with a quadratic convergence rate if the point is sufficiently close to the optimum since the problem of (15) is convex [35].

**Algorithm 2 Full Newton’s method for MDDR beamforming**

**Input:** Received data \( X \) and error tolerance \( \epsilon \).

**Initialize:** \( w^0 = a/\|a\|^2 \) or set \( w^0 \) using SMI beamformer of (46).

for \( k \leftarrow 0, 1, 2, \ldots \) do

Compute output \( y = (X^H w^k)^* \) and construct two diagonal matrices

\[
D(w^k) = \text{diag} \{ |y(1)|^{
u-2}, \ldots, |y(N)|^{
u-2} \}
\]

\[
E(w^k) = \text{diag} \{ |y(1)|^{\nu-4} y^2(1), \ldots, |y(N)|^{\nu-4} y^2(N) \}.
\]

Calculate gradient and Hessian matrices

\[
\nabla f_p(w^k) = \frac{p}{2} X D(w^k) X^H w^k
\]

\[
H_{w \cdot w} = \frac{p^2}{4} X D(w^k) X^H
\]

\[
H_{w w} = \frac{p(\nu-2)}{4} X^* E(w^k) X^H.
\]

Solve (38) to obtain Newton direction \( \Delta w^k \).

Determine optimal step size \( \mu_k \) by (45).

Update beamformer:

\[
w^{k+1} = w^k + \mu_k \Delta w^k.
\]

Stop if \( |\text{Re} \{ \nabla f_p(w^k)^H \Delta w^k \}| < \epsilon \).

end for

2) Partial Newton’s Method: As a simplification of the full Newton’s method, the partial Newton’s method ignores \( H_{w w} \) and \( H_{w \cdot w} \) by assuming them as \( 0 \). Then the linear system of (37) reduces to

\[
H_{w \cdot w} \Delta w + \lambda a = -\nabla f_p(w)
\]

\[
a^H \Delta w = 0
\]

(47)

(48)

From (47), we obtain

\[
\Delta w = -H_{w \cdot w}^{-1} \nabla f_p(w) + \lambda a.
\]

(49)

Substituting (49) into (48), \( \lambda \) is solved as

\[
\lambda = a^H H_{w \cdot w}^{-1} \nabla f_p(w) a
\]

(50)

Substituting (50) back into (49) leads to a closed-form solution of the update direction as

\[
\Delta w = -H_{w \cdot w}^{-1} \left( \nabla f_p(w) - \frac{a^H H_{w \cdot w}^{-1} \nabla f_p(w) a}{a^H H_{w \cdot w}^{-1} a} a \right).
\]

(51)

Once the Newton direction is determined, the beamformer is updated using (43). Again, a line search procedure in (45) can
be applied to obtain the optimal step size. We refer to this algorithm as the partial Newton’s method because it utilizes only the partial Hessian matrix \( H_{w-w} \). Since the size of the linear system of the partial Newton’s method is only half of that of the full Newton’s method, the cost for computing the Newton’s direction for the partial Newton’s method is approximately 1/8 of that of the full Newton’s method. The computational simplicity results from ignoring the off-diagonal Hessian matrices \( H_{w-w} \) and \( H_{w-w} \). These two partial Hessian matrices are not null and contain useful information for \( s \neq 2 \). Hence the performance of the partial Newton’s method is inferior to that of the full Newton’s method—its convergence rate is slower than that of the full Newton’s method.

An interesting relationship between the IR-MVDR and the partial Newton’s method for \( \ell_p \)-norm minimization is described in the following proposition.

Proposition 2: The IR-MVDR algorithm is a special case of the partial Newton’s method using a fixed step size of \( p/2 \).

Proof: The update formula of the partial Newton’s method with step size of \( p/2 \) is

\[
\mathbf{w}^{k+1} - \mathbf{w}^k + \frac{p}{2} \Delta \mathbf{w}^k.
\]  

Substituting (49) into (52) leads to

\[
\mathbf{w}^{k+1} = \mathbf{w}^k - \frac{p}{2} H_{w-w}^{-1} \left( \nabla f_p(\mathbf{w}^k) + \lambda a \right)
\]

\[
= \mathbf{w}^k - \frac{p}{2} \frac{4}{p^2} \left( XD(\mathbf{w}^k)X^H \right)^{-1}
\]

\[
\times \left( \frac{p}{2} XD(\mathbf{w}^k)X^H \mathbf{w}^k + \lambda a \right)
\]

\[
- \frac{2\lambda}{p} \left( XD(\mathbf{w}^k)X^H \right)^{-1} a.
\]  

Since \( \mathbf{w}^{k+1} \) is feasible, i.e., it satisfies the constraint \( a^H \mathbf{w}^{k+1} = 1 \), we have

\[
-\frac{2\lambda}{p} \frac{1}{a^H \left( XD(\mathbf{w}^k)X^H \right)^{-1} a}.
\]  

Plugging (54) into (53) yields (23), i.e., the update formula of the IR-MVDR algorithm.

Clearly, the fixed step size of \( p/2 \) is not optimal. Therefore the IR-MVDR is inferior to the partial Newton’s method in terms of convergence rate.

The convergence rates of the IR-MVDR, two Newton’s methods, and the gradient descent method [16] with optimal step sizes for different values of \( p \) are compared. Note that [16] does not discuss how to select a step size for the gradient descent method although it is an important issue. Here we use the optimal step size for this gradient method, which yields the best performance. Six values of \( p \), namely, \( p \in \{1,2,3,4,5,6\} \), are tried. In this numerical example, we use the experimental settings in example 1 of Section V. We are primarily interested in the behavior, as a function of the number of iterations, of the relative error \( f_p(\mathbf{w}^k) - f_p(\hat{\mathbf{w}}) / f_p(\hat{\mathbf{w}}) \), where \( \hat{\mathbf{w}} \) and \( f_p(\hat{\mathbf{w}}) \) are the minimizer and the global minimum of (15), respectively. This global minimum can be calculated exactly (in practice, up to the computer round-off precision) with a finite number of steps using the full Newton’s method or any optimization software package in advance. For fair comparison, all the methods use the same initial value of \( \mathbf{w}^0 = \mathbf{a}/|\mathbf{a}|^2 \). Fig. 1 shows the convergence rates of the three methods. We can see that the IR-MVDR algorithm does not converge for \( p \geq 3.4 \) while the two Newton’s methods converge in all cases. The gradient method [16] converges very slowly. When the IR-MVDR algorithm converges, it has a linear convergence rate. The partial Newton’s method also has a linear convergence rate but it converges faster than the IR-MVDR. The full Newton’s method has a quadratic convergence rate and converges very fast. It only needs several iterations for convergence. Ignoring the two off-diagonal Hessian matrices makes the partial Newton’s method lose the property of quadratic convergence.

Remark 2: We can use the SMI beamformer of (46) as the initial value to speed up the convergence of the IR-MVDR and the two Newton’s methods because the SMI beamformer may be closer to the true solution than the data-independent beamformer of (24). Fig. 2 compares the convergence rates using these two different initializations. It is observed that the SMI beamformer is a better initial value that can accelerate the convergence rate.

E. MDDR Beamforming Via \( \ell_\infty \)-Norm Minimization

The MDDR beamformer is applicable to the case of \( p \rightarrow \infty \), where the \( \ell_p \)-norm becomes

\[
\| \mathbf{y} \|_\infty = \max_{1 \leq n \leq N} |y(n)|.
\]  

Accordingly, the \( \ell_\infty \)-norm MMDR beamforming corresponds to

\[
\min_{\mathbf{w}} \| \mathbf{X}^H \mathbf{w} \|_\infty.
\]  

s.t. \( a^H \mathbf{w} = 1 \).

Since the function of \( \ell_\infty \)-norm is non-differentiable, the Newton’s method cannot be applied. However, we show that (56) can be converted into a second-order cone programming (SOCP). First, the complex-valued variables are split into real-valued ones, that is, \( \mathbf{w} = \mathbf{w}_R + j\mathbf{w}_I, \mathbf{a} = \mathbf{a}_R + j\mathbf{a}_I, \mathbf{y} = \mathbf{y}_R + j\mathbf{y}_I, \) and \( \mathbf{X} = \mathbf{X}_R + j\mathbf{X}_I \) with \( \mathbf{w}_R, \mathbf{w}_I, \mathbf{a}_R, \mathbf{a}_I \in \mathbb{R}^M, \mathbf{y}_R, \mathbf{y}_I \in \mathbb{R}^N, \) and \( \mathbf{X}_R, \mathbf{X}_I \in \mathbb{R}^{M \times N} \). Then we have \( y(n) = [\mathbf{y}_R(n)]^T + [\mathbf{y}_I(n)]^T \). The problem of (56) is reformulated as the following SOCP:

\[
\begin{aligned}
\min_{\mathbf{w}_R, \mathbf{w}_I, \mathbf{y}_R, \mathbf{y}_I, u} & u \\
\text{s.t.} & \sqrt{y_R^2(n) + y_I^2(n)} \leq u, \quad n = 1, \ldots, N \\
& \begin{bmatrix} \mathbf{X}_R & \mathbf{X}_I \\ \mathbf{X}_I & -\mathbf{X}_R \end{bmatrix} \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} \\
& \begin{bmatrix} \mathbf{a}_R - \mathbf{a}_I \\ \mathbf{a}_I & \mathbf{a}_R \end{bmatrix} \begin{bmatrix} \mathbf{w}_R \\ \mathbf{w}_I \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\end{aligned}
\]  

where \( u \in \mathbb{R}^+ \) is an auxiliary variable participating in the optimization. The Newton’s method cannot handle this SOCP problem because it contains an inequality constraint. A more sophisticated interior point method [29] is needed to solve it. This results in a computational complexity of \( O((N + M)^{3.5}) \) in each iteration, which is higher than those of the IR-MVDR and the Newton’s methods.
IV. EXTENSION TO MULTIPLE LINEAR CONSTRAINTS

It is known that the MVDR beamformer suffers significant performance degradation due to the uncertainty or mismatch in the steering vector [3], [5], [19]. One cause for steering vector mismatch is the AOA estimation error. When the steering vector of the SOI is imprecise, the SOI will be mistaken as interference and attenuated by the MVDR beamformer [21]. A remedy to address the AOA mismatch is to impose multiple linear constraints for a small spread of angles around the nominal AOA [22]. That is,

\[
\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \\
\text{s.t. } \mathbf{C}^H \mathbf{w} = \mathbf{g}
\]  

(58)

where \( \mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K] \in \mathbb{C}^{M \times K} \) contains \( K \) steering vectors or the derivatives of the steering vectors and \( \mathbf{g} = [g_1, \ldots, g_K]^T \) is usually taken as the vector with all elements being unit. The closed-form solution of the optimization problem in (58) is called LCMV beamformer and is given by [22]

\[
\mathbf{w}_{\text{LCMV}} = \mathbf{R}^{-1} \mathbf{C} \left( \mathbf{C}^H \mathbf{R}^{-1} \mathbf{C} \right)^{-1} \mathbf{g}. 
\]  

(59)

When the number of linear constraints is \( K = 1 \), the LCMV beamformer reduces to the MVDR beamformer. Similar to the MDDR beamformer, we can generalize the LCMV method to the following LCMD beamforming

\[
\min_{\mathbf{w}} \left\| \mathbf{X}^H \mathbf{w} \right\|_p^p \\
\text{s.t. } \mathbf{C}^H \mathbf{w} = \mathbf{g}.
\]  

(60)

Equation (60) is an extension of (15) from a single linear constraint to multiple constraints. Meanwhile, the three algorithms for solving (15) can be extended to multiple constraints in (60).
Replacing the covariance matrix \( R \) using the reweighted version \( X D(w^k)X^H \), the update rule of the Iteratively Reweighted LCMD (IR-LCMD) algorithm is given by

\[
w^{k+1} = (X D(w^k)X^H)^{-1} C (X D(w^k)X^H)^{-1} \mathbf{g}
\]

(61)

which is an extension of the IR-MVDR algorithm of (23).

The full Newton's method can also be applied to solve the LCMD problem of (60). The Newton direction \( \Delta w \) is calculated by

\[
\min_{\Delta w} q_w(\Delta w) \quad \text{subject to} \quad C^H \Delta w = 0.
\]

(62)

The Lagrangian function of (62) is

\[
L_w(\Delta w, \lambda) = q_w(\Delta w) + \lambda^T C^H \Delta w
\]

(63)

where \( \lambda = [\lambda_1, \ldots, \lambda_K]^T \) is the Lagrangian multiplier vector corresponding to the \( K \) linear constraints. Applying the optimal condition to (63) leads to

\[
\frac{\partial L_w(\Delta w, \lambda)}{\partial \Delta w^*} = \nabla f_p(w) + H_{w^*w} \Delta w^* + C \lambda = 0,
\]

\[
\frac{\partial L_w(\Delta w, \lambda)}{\partial \Delta w} = \nabla f_p(w)^* + (H_{w^*w} \cdot \Delta w^*)^* + H_{ww} \Delta w + C^* \lambda = 0,
\]

\[
\frac{\partial L_w(\Delta w, \lambda)}{\partial \lambda} = C^H \Delta w = 0
\]

which can be rewritten as

\[
\begin{bmatrix}
H_{w^*w} & H_{w^*w}^* & C
H_{ww} & H_{ww}^* & C^*
0_{K \times M} & 0_{K \times K} & C^H
\end{bmatrix}
\begin{bmatrix}
\Delta w
\Delta w^*
\lambda
\end{bmatrix} = \begin{bmatrix}
\nabla f_p(w)
\nabla f_p(w)^*
0
\end{bmatrix}.
\]

(64)

The Newton direction \( \Delta w \) is obtained by solving the linear system of (65), which requires a complexity of \( O((2M + K)^3) \).

Again, assigning the two Hessian matrices \( H_{ww} \rightarrow H_{w^*w} \rightarrow 0 \) gives the update direction of the partial Newton’s method

\[
H_{w^*w} \Delta w + C \lambda = -\nabla f_p(w)
\]

(66)

Solving (66) gives

\[
\lambda = -(C^H H_{w^*w}^{-1} C)^{-1} C^H H_{w^*w}^{-1} \nabla f_p(w)
\]

(67)

and

\[
\Delta w = -H_{w^*w}^{-1} \left( I - C (C^H H_{w^*w}^{-1} C)^{-1} C^H H_{w^*w}^{-1} \right) \nabla f_p(w).
\]

(68)

After obtaining the Newton direction, the remaining steps of the full and partial Newton’s methods are the same as the single constraint algorithms in Section III.D.

The Newton’s method exploits the second-order derivatives of the objective function, which are contained in the Hessian matrix. The gradient descent method in [16] just utilizes the first-order derivatives. Assuming that the Hessian matrix \( H_{w^*w} = I \), (68) is simplified to

\[
\Delta w = -(I - C (C^H C)^{-1} C^H) \nabla f_p(w)
\]

(69)

which is the update direction of the gradient descent method for constrained optimization problems. Note that \((I - C (C^H C)^{-1} C^H) = P_C^0 \) is the projector onto the orthogonal complementary space of range(C). Therefore the update direction \( \Delta w = -P_C^0 \nabla f_p(w) \) represents the projected gradient or gradient projection [37]. In [16], the gradient projection method is adopted to compute the LCMD beamformer. However, it does not mention how to select a step size. The step size is determined empirically. For the proposed full and partial Newton’s methods, the optimal step size is solved using line search method for a given Newton direction. Furthermore, the convergence rate of the gradient projection method is slower than that of the Newton’s method and even much slower than that of the IR-LCMV algorithm.

V. SIMULATION RESULTS

To facilitate a fair comparison, experimental parameters used in [5] and [21] are adopted in our computer simulations as well. A ULA of \( M \) omnidirectional sensors spaced half a wavelength apart is considered. The steering vector is computed using (2). Three zero-mean signals, namely, the SOI \( s_1(n) \) and two uncorrelated interferences \( s_1(n) \) and \( s_2(n) \), impinge on the array. The AOA of the SOI is \( \theta_1 = 45^\circ \) and the AOAs of the two interferences are \( \theta_1 = 30^\circ \) and \( \theta_2 = 75^\circ \). The signal-to-noise ratio (SNR) is defined as

\[
\text{SNR} = \frac{\sigma_1^2}{\sigma_n^2}.
\]

(70)

The two interferences are stronger than the SOI with variances being \( \sigma_1^2 = 4\sigma_n^2 \) and \( \sigma_2^2 = 9\sigma_n^2 \). That is, they are 6 dB and 9.5 dB above the SOI, respectively.

A. Results With Perfect Steering Vector

We first present the simulation results with a perfect steering vector. The performance of the MDJR with a variety of \( p \), MVDR (i.e., \( p = 2 \)), and subspace beamformers, as well as the optimal SINR bound, are compared. According to (4), the upper bound on the SINR is the maximum generalized eigenvalue of the matrix pair \( (\sigma_1^2 \mathbf{a} \mathbf{a}^H, R_{++}) \). Note that the subspace beamformer requires the dimension of the signal-plus-interference subspace. The MDL [26] is applied to estimate this quantity. Monte Carlo trials are conducted to evaluate the output SINR performance of the beamforming algorithms. When plotting the SINR curves, 200 Monte Carlo trials are performed to calculate the average output SINR.

Example 1: Sub-Gaussian signals

In the first example, the SOI and two interferences are QPSK signals, which are frequently encountered in communications and are sub-Gaussian distributed. The additive noise is a white
Gaussian process. Fig. 3 shows the output SINR versus SNR when the number of sensors $M = 10$ and number of snapshots $N = 100$. Fig. 4 displays the output SINR versus $N$ at $M = 10$ and $SNR = 10$ dB. Fig. 5 shows the output SINR versus $M$ at $N = 100$ and $SNR = 10$ dB.

From Figs. 3 to 5, it can be seen that the MDDR beamforming of $p > 2$ leads to an improved performance compared with the MVDR beamformer ($p = 2$) for QPSK signals. It can be seen from Fig. 3, when $SNR \geq 10$ dB, the output SINR for $p = 20$ is $10–20$ dB higher than that of the MVDR beamformer. In Figs. 4 and 5, this performance gain is still about $10$ dB when the number of snapshots is larger than $50$ or the number of sensors is more than $8$. In particular, only the $p=\infty$-MDDR beamformer ($p = \infty$) approaches the upper bound of the SINR as the SNR increases. The results demonstrate that the MDDR beamformer with larger value of $p$ has better performance. The MDDR beamformer with $p < 2$ is not recommended for sub-Gaussian signals because its performance is worse than that of the MVDR beamformer. The subspace beamformer is also superior to the MVDR beamformer and it has similar performance as the MDDR beamformer with $p = 4$. However, its performance degrades severely at low SNR.

**Example 2:** Super-Gaussian signals
In the second example, the beamforming algorithms are tested using super-Gaussian signals. The SOI, two interferences and noise are modeled as random processes satisfying a generalized Gaussian distribution (GGD) [27]. The PDF of the circular zero-mean GGD with variance $\sigma^2$ is

$$p_s(s) = \frac{\beta \Gamma(4/\beta)}{2\pi \sigma^2 \Gamma(2/\beta)} \exp\left(-\frac{|s|^4}{\sigma^2 \beta^2}\right) \quad (71)$$

where $\beta > 0$ is the shape parameter, $\Gamma(\cdot)$ is the Gamma function, and $c = (\Gamma(2/\beta)/(\Gamma(4/\beta))^{2/\beta}$ [27]. When $\beta = 2$, the GGD reduces to the circular Gaussian distribution. Note that $\beta > 2$ models sub-Gaussian signals while $\beta < 2$ models super-Gaussian ones. Especially, $\beta = 1$ corresponds to the Laplacian distribution [27], which is widely used to model speech signals [14]. The smaller the value of $\beta$, the more impulsive the signal is. We take $\beta = 0.4$ in this example.

Fig. 6 displays the output SINR versus SNR for $M = 10$ and $N = 100$. Fig. 7 illustrates the output SINR versus $N$ for $M = 10$ and $SNR = 10$ dB. Fig. 8 shows the output SINR versus $M$ for $N = 200$ and $SNR = 10$ dB.

The results of Figs. 6, 7, and 8 are opposite to those given by Figs. 3, 4, and 5, respectively. This is not surprising because the signals used in this example are super-Gaussian distributed, which are different from the sub-Gaussian signals used in the first example. These results illustrate that the MDDR beamformer with $p < 2$ will yield a SINR gain for super-Gaussian signals. The performance improvement with $p = 1$ is about $5–7$ dB when $SNR \geq 10$ dB, which is optimal in this example. In the presence of super-Gaussian signals, the MDDR beamformer with smaller $p$ has better performance. Contrary to the case of sub-Gaussian signals, $p > 2$ is not recommended for super-Gaussian signals. The subspace beamformer has a performance comparable to that of the MDDR beamformer with $p = 1$.

As SNR increases, even the optimal setting $p = 1$ cannot approach the upper bound of SINR. It still has a large gap from
Fig. 6. Output SINR versus SNR for GGD signals and noise with $\beta = 0.4$.

Fig. 7. Output SINR versus number of snapshots for GGD signals and noise with $\beta = 0.4$.

Fig. 8. Output SINR versus number of sensors for GGD signals and noise with $\beta = 0.4$.

The bound. Hence we infer that $p < 1$ may achieve better performance for GGD signals with $\beta = 0.4$. However, as mentioned above, we do not consider the choice of $0 < p < 1$ since the resulting optimization problem is difficult to solve [32].

Example 3: Gaussian signals

In the third example, all the signals and noise are Gaussian. Fig. 9 illustrates the output SINR versus SNR for $p = 2$. We can see that $p=2$ is the optimal value for Gaussian signals. It is demonstrated that all the MDDR methods have similar performance if $p$ is not far away from 2. The subspace method has the best performance for Gaussian signal model if the SNR is not very low. However, the performance of the subspace beamformer degrades at smaller SNRs.

Example 4: Impact of interference number

In the fourth example, we investigate the impact of the interference number on beamforming performance. The SOI and the interferences are QPSK modulation and noise values are Gaussian. The snapshot number is $N = 100$. The AOA of the SOI is fixed to $43^\circ$ while that of the $i$th ($1 \leq i \leq I$) interference is $\theta_i = -30^\circ + (i - 1)10^\circ$. The power of all interferers is the same and 10 dB higher than the SOI. We first consider $I = 6$ interferers. Fig. 10 plots the output SINR versus SNR. By comparing Fig. 3, which has only two interferers,
with Fig. 10, we observe that the performance of the subspace beamformer substantially degrades for larger signal number values. Fig. 11 shows the output SINR versus interference number at \( \text{SNR} = 20 \) dB. It can be seen that the subspace beamformer loses its efficiency as the interference number increases while the MVDR and MDDR beamformers maintain their performance.

### B. Results With AOA Mismatch

In this subsection, we investigate the performances of the LCMD, LCMV, and MVDR beamformers in the presence of steering vector mismatch. The steering vector mismatch is first simulated using imprecise AOA. We focus on the sub-Gaussian signal case where the QPSK signal is used. The experimental settings are the same as that of Example 1 in Section V.A. The true AOA of the SOI is 43° but the assumed AOA is 45°. For LCMV and LCMD beamformers, two linear constraints which force the responses of the signals from 42° and 48° to be unity.

Figs. 12 and 13 show the output SINR versus SNR and number of snapshots, respectively. It can be seen that the performance of the MVDR beamformer significantly degrades due to the 2° AOA error. It is not surprising that the output SINR of the MVDR beamformer decreases as the SNR increases because the signal cancellation phenomenon caused by AOA mismatch is more severe at high SNRs. The LCMV beamformer with two linear constraints enhances the robustness against AOA mismatch. The LCMD beamformer with \( p > 2 \), especially for relatively large values of \( p \), significantly improves the performance the LCMV for sub-Gaussian QPSK signals. For example, the SINR of the proposed LCMD beamformer when the number of snapshots is greater than 50.

We then investigate the robustness against the look direction of the SOI. The AOA of the SOI, i.e., \( \theta \), is varied from \(-20^\circ\) to \(20^\circ\) while the AOAs of the two interferers are fixed to \(30^\circ\) and \(75^\circ\), respectively. It is assumed that there is a 2° error in the AOA estimate. That is, the AOA estimate is \( \theta + 2^\circ \). Hence we impose two linear constraints as \( \theta - 1^\circ \) and \( \theta + 3^\circ \), which are centered around the AOA estimate \( \theta + 2^\circ \), for LCMV and LCMD beamformers. Fig. 14 shows the output SINR versus \( \delta \), from which we see that the LCMD beamformer is not sensitive to the look direction.

### C. Results With Sensor Position Mismatch

In addition to AOA estimation error, the steering vector mismatch may be caused by a variety of reasons, such as uncertainties in array response and sensor geometry position. Similar to LCMV, the LCMD beamformer was originally designed for AOA mismatch. However, it can be adapted to handle other types of mismatch. In this simulation example, the error in sensor geometry position is considered. From (2), it is observed that the steering vector is a function of the inter-sensor spacing \( \delta \) and hence denoted as \( a(d) \). In practice, the inter-sensor spacing with error can be expressed as \( d = d_0 + \Delta d \), where \( d_0 \) is the nominal value and \( \Delta d \) is the error. We investigate the effect of spacing error on beamforming performance. The nominal spacing is equal to half the wavelength, i.e., \( d_0 = 0.5\xi \). The maximal relative error \( \Delta d/d_0 \) is ±10%, which corresponds \( 0.9d_0 \leq d \leq 1.1d_0 \). Three linear
constraints, namely, \( \{ c_k^T w = 1 \}_{k=1}^3 \), are used with the LCMV and LCMD beamformers. We set \( c_1 = a(0.9d_0) \), \( c_2 = a(d_0) \), and \( c_3 = a(1.1d_0) \) in (60). The SNR is 20 dB and snapshot number \( N = 100 \). The other experimental settings are the same as those in Example 1 in Section V.A.

Fig. 15 shows the output SINR versus the relative error that varies from \(-10\%\) to \(10\%\). The performance of the MVDR beamerformer significantly degrades as the spacing error increases. The LCMV and LCMD beamformers with three linear constraints improve the robustness against sensor position. Again, the LCMD beamerformer with \( p > 2 \) yields better performance than the LCMV.

VI. CONCLUSION

By recognizing the fact that the minimum variance criterion is not statistically optimal for non-Gaussian signals, this paper investigated the minimum dispersion beamforming with either single or multiple linear constraints. The use of a single linear constraint resulted in the MDDR beamformer while the use of multiple constraints led to the LCMD beamformer. The LCMD beamformer is robust against steering vector mismatches. The MDDR and LCMD beamformers outperform their respective standard counterparts based on minimum variance, namely, the MVDR and LCMV beamformers. Three computationally efficient algorithms, i.e., the IR-MVDR, full Newton’s and partial Newton’s methods, were developed to efficiently solve the resultant \( l_2 \)-minimization problem with linear equality constraints. It was shown that the IR-MVDR is a special case of the partial Newton’s method. Simulation results demonstrated the superior performance of the minimum dispersion beamformers. An important future work is combining the minimum dispersion criterion with the nonlinear constraints, such as those in [5], [19]. These nonlinear constraints force the magnitude responses of the steering vectors in an uncertainty set to exceed unity. This will generate new beamforming techniques that are more robust against steering vector mismatch as well as enhance the SINR performance.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their useful comments.

REFERENCES


Xue Jiang (S’13) received the M.S. degree in acoustic- 
hics from Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2007. From 2007 to 2010, she was a research engineer with the Sunplus Inc., Beijing, China, where she was engaged in algorithmic research and development of mobile communication system. She is currently pursuing the Ph.D. degree and is also a research assistant at the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada. Her research interests include array processing, robust beamforming, target tracking, acoustic signal processing, and numerical optimization, with applications to radar, sonar, and wireless communications. She was one of the finalists of the Best Student Paper Award of the IEEE 2013 Radar Conference. She received the McMaster University International Excellence Award in 2012 and 2013, which is awarded to international students for a high standard of scholarly achievement.

Wen-Jun Zeng (S’10–M’11) received the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2008. He was from 2006 to 2009, he was a Research Assistant with Tsinghua University. From 2009 to 2011, he was a faculty member with the Department of Communication Engineering, Xiamen University, Xiamen, China. He is now a Senior Research Associate with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong. His research interests lie in the area of math- 
ematical signal processing, including convex optimization, array processing, sparse approximation, and inverse problem, with applications to wireless radio and underwater acoustic communications. He was a member of the Technical Program Committee of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

A. Yasotharan received the B.Sc. degree in electrical and electronic engineering with First Class Honours from the University of Peradeniya, Sri Lanka. He received the M.Eng. and Ph.D. degrees in electrical engineering from McMaster University, Canada. He is now a Defence Scientist at Defence Research and Development Canada where he works on GPS/GNSS related projects. His previous positions include: Senior Systems Engineer at Raytheon Canada Ltd., where he worked on Radar and Tracking projects; Digital Signal Processing Engineer at Unique Broad- 

Hing Cheung So (S’90–M’95–SM’07) was born in Hong Kong. He received the B.Eng. degree from the Chinese University of Hong Kong, both in electronic engineering, in 1990 and 1995, respec- 
tively. From 1990 to 1991, he was an Electronic Engineer with the Research and Development Divi- 
sion, Everett Systems Engineering Ltd., Hong Kong. During 1995–1996, he worked as a Postdoctoral Fellow with The Chinese University of Hong Kong. From 1996 to 1999, he was a Research Assistant Professor with the Department of Electronic Engineering, City University of Hong Kong, where he is currently an Associate Professor. His research interests include statistical signal processing, fast and adaptive algorithms, signal detection, parameter estimation, and source localization.

Dr. So has been on the editorial boards of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, Signal Processing, Digital Signal Processing, and ISRN Applied Mathematics, as well as a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

Thiagalingam Kirubarajan (S’95–M’98–SM’03) was born in Sri Lanka in 1969. He received the B.A. and M.A. degrees in electrical and information engineering from Cambridge University, England, in 1991 and 1993, and the M.S. and Ph.D. degrees in electrical engineering from the University of Connecticut, Storrs, in 1995 and 1998, respectively. Currently, he is a professor in the Electrical and Computer Engineering Department at McMaster University, Hamilton, Ontario. He is also serving as an Adjunct Assistant Professor and the Associate Director of the Estimation and Signal Processing Research Laboratory at the University of Connecticut. His research interests are in estimation, target tracking, multisource information fusion, sensor resource management, signal detection and fault diagnosis. His research activities at McMaster University and at the University of Connecticut are supported by U.S. Missile Defense Agency, U.S. Office of Naval Research, NASA, Qualcomm Systems, Inc., Raytheon Canada Ltd. and Defense Research Development Canada, Ottawa. In September 2001, he served in a DARPA expert panel on unattended surveillance, homeland defense and counterterrorism. He has also served as a consultant in these areas to a number of companies, including Motorola Corporation, Northrop-Grumman Corporation, Pacific-Sierra Research Cor- poration, Lockheed Martin Corporation, Inuktron Corporation and BAE systems. He has worked on the development of a number of engineering software programs, including BEARDAT for target localization from bearing and frequency measurements in clutter, FUSEDAT for fusion of multisensor data for tracking. He has also worked with Qualtech Systems, Inc., to develop an advanced fault diagnosis engine. He has published about 100 articles in areas of his research interests, in addition to one book on estimation, tracking and navigation and two edited volumes. Dr. Kirubarajan is a recipient of Ontario Premiers Research Excellence Award (2002).