

# Sparse Array Beamformer Design via ADMM (with updated Appendix D)

Huiping Huang, *Student Member, IEEE*, Hing Cheung So, *Fellow, IEEE*, and Abdelhak M. Zoubir, *Fellow, IEEE*

**Abstract**—In this paper, we devise a sparse array design algorithm for adaptive beamforming. Our strategy is based on finding a sparse beamformer weight to maximize the output signal-to-interference-plus-noise ratio (SINR). The proposed method utilizes the alternating direction method of multipliers (ADMM), and admits closed-form solutions at each ADMM iteration. The algorithm convergence properties are analyzed by showing the monotonicity and boundedness of the augmented Lagrangian function. In addition, we prove that the proposed algorithm converges to the set of Karush-Kuhn-Tucker stationary points. Numerical results exhibit its excellent performance, which is comparable to that of the exhaustive search approach, slightly better than those of the state-of-the-art solvers, including the semidefinite relaxation (SDR), its variant (SDR-V), and the successive convex approximation (SCA) approaches, and significantly outperforms several other sparse array design strategies, in terms of output SINR. Moreover, the proposed ADMM algorithm outperforms the SDR, SDR-V, and SCA methods, in terms of computational complexity.

**Index Terms**—Adaptive beamforming, ADMM, semidefinite relaxation, sparse array design, successive convex approximation, output SINR.

## I. INTRODUCTION

ADAPTIVE arrays have been widely applied in diverse practical applications, such as radar, sonar, wireless communications, to name just a few [2]. One of their important functions is beamforming, which is to extract the signal-of-interest (SOI) while suppressing interference and noise [3]. It has been reported that the performance of beamforming is affected by not only the beamformer weight, but also the array configuration [4]. In this sense, conventional uniform arrays may not be the optimal choices for adaptive beamformer design. On the other hand, since sparse arrays achieve increased array aperture and degrees of freedom while reducing the hardware complexity, as compared to conventional uniform arrays. Thus, sparse arrays could be a better option for adaptive beamformer design.

On the other hand, many research works have considered beamformer design for the communication systems of hybrid structures [5]–[8]. Hybrid architectures are one approach for providing enhanced benefits of multiple-input multiple-output (MIMO) communication at millimeter-wave (mmWave) frequencies. In such a hybrid system, there are smaller number of radio-frequency (RF) chains than antennas. Note that

H. Huang and A. M. Zoubir are with Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Germany (emails: h.huang@spg.tu-darmstadt.de, zoubir@spg.tu-darmstadt.de). H. C. So is with Department of Electrical Engineering, City University of Hong Kong, China (email: hcso@ee.cityu.edu.hk). The work of H. Huang is supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt. Parts of this paper were presented at SAM 2022 [1].

RF chains are much more expensive than antennas. Hybrid structure offers a compromise between system performance and hardware complexity. For such a configuration, it raises a question on how to adaptively switch the available RF chains to the corresponding subset of antennas, which can be interpreted as sparse array beamformer design.

## A. Related Works

In recent years, several strategies for designing sparse array beamformers have been proposed, see for example [9]–[18] and references therein. [9] investigates sparse array design problem in the presence of steering vector mismatch. [10] focuses on a specific sparse array, i.e., coprime array, for adaptive beamforming. A constrained normalized least-mean-square approach is developed in [11]. However, there is no theoretical analyses of the convergence of the algorithm. The authors in [12] consider the problem of resource allocation and interference management, and propose a group sparse beamforming method. A reconfigurable adaptive antenna array strategy is proposed to minimize the spatial correlation coefficient between the desired signal and the interference [13]. [14] studies the sparse array beamforming design problem by firstly dividing the whole array into several non-overlapped sub-groups, and then selecting one antenna per sub-group. By using the regularized switching network, [15] proposes a cognitive sparse beamformer, which is adaptive to the environmental dynamics. In [16], the authors investigate joint beamforming and antenna selection problem under imperfect channel state information, and develop an efficient branch and bound (B&B) approach and a machine learning based scheme. [17] and [18] consider sparse array design for achieving maximum signal-to-interference-plus-noise ratio (SINR) for different scenarios of single point source, multiple point sources, and wideband sources. Different from most of the existing works, our goal is to maximize the output SINR by using a smaller number of RF chains (compared to the number of antennas), and our main contributions include developing an efficient algorithm to solve the resulting problem and providing theoretical guarantee of the convergence property of the proposed algorithm.

From the perspective of algorithm, existing methods can be roughly divided into three categories: Greedy based [12], [13], [19], machine learning based [16], [20]–[23], and optimization based [16]–[18], [24]–[39] approaches. The greedy procedure in [12], [13], [19] largely reduces the combinatorial exploration space, but could result in a highly suboptimal solution. Machine learning techniques [16], [20]–[23] require prior data for their training step, which might be unavailable in some practical scenarios. On the other hand, optimization based methods include B&B [16], mixed-integer programming

(MIP) [24]–[26], semidefinite relaxation (SDR) [17], [18], [28], [29], [33], [39], and successive convex approximation (SCA) [34]–[36]. B&B and MIP are capable of finding the global optimum at the cost of a large computational burden. SDR and SCA based methods are computationally expensive when the dimension of the resulting matrix is high [40], [41]. Besides, the relaxation nature of SDR usually leads to a solution with a rank not being one, in which case extra post-processing based on randomization is needed [42]. Moreover, note that the SDR methods in [17], [18], [28], [29] utilize the  $\ell_1$ -norm square instead of the  $\ell_1$ -norm for sparsity promotion. Although the simulation results in these papers and also in Section VI-C of the present paper demonstrate the usefulness of the SDR-type methods, no theoretical support is available due to the convexity of the Pareto boundary that is not guaranteed (which has also been mentioned in [28] and [29]). On the other hand, another downside of the SCA approach lies in the fact that it requires a feasible starting point, which could be a difficult task on its own [40].

### B. Contributions

In this paper, a sparse array design algorithm based on the alternating direction method of multipliers (ADMM) is devised for adaptive beamforming. The proposed technique admits closed-form solutions at each ADMM iteration. Convergence analyses of the proposed algorithm are provided by showing the monotonicity and boundedness of the augmented Lagrangian function. Additionally, it is proved that the proposed algorithm converges to the set of Karush-Kuhn-Tucker (KKT) stationary points. Simulation results demonstrate excellent behavior of our scheme, as it outperforms several existing methods, and is comparable to the exhaustive search approach.

Note that parts of this paper were presented in its conference precursor, i.e., [1]. Compared to [1], the advancements and supplementary contributions made in the present article are listed as follows.

- In [1], we only provide an illustrative example for Proposition 1 without a proof. Differently, in the present article, besides an illustrative example we also provide a mathematical proof in a rigorous manner.
- The convergence property of the proposed ADMM has not been analysed in [1]. While in the present article, we provide such convergence analyses in theory, and extensive numerical simulations are provided to validate the theoretical development, see Sections IV and VI in the present article, respectively.
- In the present article we consider the computational complexity of the algorithms, see Section VI-B, which is not mentioned in [1].
- Moreover, more existing sparse array design strategies (such as enumeration method, random sparse array, compact uniform linear array, nested array, co-prime array, semidefinite relaxation approach, and successive convex approximation method) are taken into account in the present article.

On the other hand, our algorithms and theoretical results are developed primarily on the basis of the ideas presented in [41] and [43]. Several differences are highlighted as follows.

- Different from [41] which used ADMM to solve general quadratically constrained quadratic programming (QCQP) problems, we focus on a specific QCQP problem that arises in sparse array beamformer design. Our problem involves an  $\ell_1$ -norm regularization and thus our solution is sparse, which is not the case in [41].
- Another important difference between our work and [41] lies in the fact that the latter only provides a weaker convergence result, i.e., *if ADMM converges for their problem, then it converges to a KKT stationary point*, see Theorem 1 in [41]. On the other hand, we show stronger convergence results for our algorithm. That is, we first prove the convergence of the proposed algorithm under a mild condition, and then we prove that it converges to a KKT stationary point.
- Note that [43] needed extra assumptions on the Lipschitz gradient continuity as well as the boundedness of their objective function. In our work, we require neither such assumptions nor any other assumptions.
- Since [43] considered general non-convex problems, no explicit expressions for their parameters were derived. On the contrary, as we consider a specific non-convex problem of sparse array beamformer design, the properties of our objective function have been investigated and thus several parameters are given in an explicit manner. See for example the augmented Lagrangian parameter  $\rho$  and the strongly convex parameter  $\gamma_v$  in Lemma 1.
- The results in [43] were based on the augmented Lagrangian function, while we exploit the scaled-form augmented Lagrangian function. This results in significant differences in the following three aspects: i) the proof of the monotonicity of the augmented Lagrangian function, ii) the proof of the property of the point sequence, and iii) the proof that the algorithm converges to the set of KKT stationary points; see the proofs of Lemma 1, Theorem 1, and Theorem 2, respectively.

### C. Paper Structure

The rest of the paper is organized as follows. The signal model is established in Section II. The proposed approach is presented in Section III and the convergence analyses are given in Section IV. ADMM with re-weighted  $\ell_1$ -norm regularization is proposed in Section V. Section VI shows the simulation results, and Section VII concludes the paper and provides possible future works.

### D. Notation

The notation used throughout the paper is summarised in Table I.

## II. SIGNAL MODEL

Consider a compact uniform linear array (ULA) consisting of  $M$  antenna sensors, where the terminology *compact* means that the element-spacing of two adjacent antennas is equal to half-wavelength of the incident signals. We refer the ULA with element-spacing larger than half-wavelength to as *sparse ULA*.

TABLE I: Summary of Notation

Notation	Meaning	Notation	Meaning
regular letter	scalar	$\ \cdot\ _0$	$\ell_0$ -quasi-norm of a vector
bold-faced lower-case letter	vector	$\ \cdot\ _1$	$\ell_1$ -norm of a vector
bold-faced upper-case letter	matrix	$\ \cdot\ _2$	$\ell_2$ -norm of a vector
$\mathbf{I}$	identity matrix	$ \cdot $	absolute value
$\mathbf{1}_M$	$M \times M$ all-one matrix	$(\cdot)_+$	$(x)_+ \triangleq \max\{x, 0\}$
$\mathbf{1}$	all-one vector	$\max\{a, b\}$	maximum value of $a$ and $b$
$\mathbf{0}$	all-zero vector	$\langle \mathbf{x}, \mathbf{y} \rangle$	inner product, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y}$
$\cdot^T$	transpose	$\text{Tr}\{\cdot\}$	matrix trace
$\cdot^H$	Hermitian transpose	$E\{\cdot\}$	mathematical expectation
$\cdot^{-1}$	matrix inverse	$\mathbf{X} \succeq 0$	$\mathbf{X}$ is positive semidefinite
$\lambda_{\max}(\cdot)$	largest eigenvalue	$\mathbf{X} \succeq \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive semidefinite
$\lambda_{\min}(\cdot)$	smallest eigenvalue	$\geq$	element-wise larger than or equal to
$\mathbb{C}$	set of complex numbers	$\odot$	element-wise multiplication
$\Re\{\cdot\}$	real part	$\oslash$	element-wise division
$j$	imaginary unit, $j = \sqrt{-1}$	sign( $\cdot$ )	$\text{sign}(x) = \begin{cases} 0, & \text{if } x = 0, \\ \frac{x}{ x }, & \text{otherwise} \end{cases}$

Denote  $\mathbf{x}(t) \in \mathbb{C}^M$  as the observation vector of the compact ULA, which can be modeled as:

$$\mathbf{x}(t) = \mathbf{a}(\theta_0)s_0(t) + \sum_{k=1}^K \mathbf{a}(\theta_k)s_k(t) + \mathbf{n}(t), \quad (1)$$

where  $t = 1, 2, \dots, T$  denotes the time index, with  $T$  being the total number of available snapshots,  $\theta_0$  and  $s_0(t)$  are the direction-of-arrival (DOA) and waveform of the SOI, respectively, while  $\theta_k$  and  $s_k(t)$  denote those of the  $k$ -th interference signal. We consider one SOI and  $K$  unknown interferers in the data model, and the SOI and interferers are assumed to be narrow-band, far-field, and uncorrelated signals. In addition,  $\mathbf{n}(t) \in \mathbb{C}^M$  is a zero-mean Gaussian noise vector, and the array steering vector  $\mathbf{a}(\theta) \in \mathbb{C}^M$  takes the form as

$$\mathbf{a}(\theta) = [1, e^{-j\pi \sin(\theta)}, \dots, e^{-j\pi(M-1)\sin(\theta)}]^T. \quad (2)$$

For the simplicity of notation, we denote  $\mathbf{a}(\theta_k)$  as  $\mathbf{a}_k$ , for all  $k = 0, 1, \dots, K$ , here and subsequently.

The beamformer output is calculated as

$$y(t) = \mathbf{w}^H \mathbf{x}(t), \quad (3)$$

in which  $\mathbf{w} \in \mathbb{C}^M$  is the beamformer weight vector to be designed. The beamformer output SINR is defined as [3]

$$\text{SINR} = \frac{\sigma_s^2 |\mathbf{w}^H \mathbf{a}_0|^2}{\mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w}}, \quad (4)$$

where  $\sigma_s^2 = E\{|s_0(t)|^2\}$  is the power of the SOI, and  $\mathbf{R}_{i+n}$  is the interference-plus-noise covariance matrix, which can be written as

$$\mathbf{R}_{i+n} = \sum_{k=1}^K \sigma_k^2 \mathbf{a}_k \mathbf{a}_k^H + \sigma_n^2 \mathbf{I}, \quad (5)$$

assuming that the interference signals are uncorrelated with the noise. In (5),  $\sigma_k^2 = E\{|s_k(t)|^2\}$  is the power of the  $k$ -th interference signal, and  $\sigma_n^2$  is the noise power.

One of the most prevailing strategies for beamformer design is to maximize the SINR, which leads to the minimum variance distortionless response (MVDR) beamformer design [44]:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{i+n} \mathbf{w} \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 = 1. \quad (6)$$

The above problem can be reformulated equivalently by replacing the realistically unattainable  $\mathbf{R}_{i+n}$  with the received data covariance matrix  $\mathbf{R}_x = \sigma_s^2 \mathbf{a}_0 \mathbf{a}_0^H + \mathbf{R}_{i+n}$  [45], as:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1. \quad (7)$$

It is worth noting that the equality constraint is formulated as an inequality one, since the output power of the SOI is included as part of the objective function in (7) [17], [18].

The above problems have a closed-form solution as  $\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_{i+n}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}_{i+n}^{-1} \mathbf{a}_0}$ . Substituting  $\mathbf{w}_{\text{opt}}$  into (4) yields the corresponding optimum output SINR as

$$\text{SINR}_{\text{opt}} = \sigma_s^2 \mathbf{a}_0^H \mathbf{R}_{i+n}^{-1} \mathbf{a}_0. \quad (8)$$

### III. SPARSE ARRAY BEAMFORMER DESIGN

#### A. Sparse Beamforming Problem

Consider the situation where only  $L \leq M$  RF chains are available [28], and thus only  $L$  antennas can be simultaneously utilized for beamformer design. The problem can be formulated as [17], [18], [28], [29]:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1 \text{ and } \|\mathbf{w}\|_0 = L. \quad (9)$$

This is a combinatorial problem, and there are  $\binom{M}{L}$  possible options. It could be an extremely huge number when  $M$  is large and  $L$  is moderate. For instance, if  $M = 100$  and  $L = 20$ , there are totally  $\binom{100}{20} > 5 \times 10^{20}$  subproblems [21]. Even if a modern machine (as fast as  $10^{-10}$  seconds per subproblem) is adopted, it still needs more than 1.5 thousand years in total. This is evidently unacceptable, and thus more computationally efficient approaches are required.

One widespread method is to replace the non-convex constraint  $\|\mathbf{w}\|_0 = L$  with its convex surrogates, such as the  $\ell_1$ -norm. By doing so and writing the  $\ell_1$ -norm in the objective function as a penalty, we relax Problem (9) as [17], [28], [33]

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} + \lambda \|\mathbf{w}\|_1 \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1, \quad (10)$$

where  $\lambda > 0$  is a tuning parameter controlling the sparsity of the solution (i.e., the number of selected sensors). The above problem is QCQP [46] with  $\ell_1$ -regularization, and it is still

non-convex because of its constraint. State-of-the-art solvers include SDR, SCA, ADMM, and their variants, see [40], [41], [46]–[50] for general QCQP problems and [17], [18], [28], [29], [33], [34], [39], [51]–[55] for specific QCQP problems with applications in MIMO radar, wireless communications, and so on. In the following subsection, we develop a method based on ADMM for solving Problem (10), which shall be shown to have closed-form solutions at each iteration.

### B. ADMM for Problem (10)

To solve Problem (10) using ADMM, we first introduce an auxiliary variable  $\mathbf{v} \in \mathbb{C}^M$  and reformulate (10) into

$$\min_{\mathbf{w}, \mathbf{v}} \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \lambda \|\mathbf{w}\|_1 \quad (11a)$$

$$\text{subject to } \begin{cases} |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1 \\ \mathbf{w} = \mathbf{v}. \end{cases} \quad (11b)$$

Then we can write down the scaled-form ADMM iterations for Problem (10) as [49]

$$\mathbf{w} \leftarrow \begin{cases} \arg \min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \\ \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1 \end{cases} \quad (12a)$$

$$\mathbf{v} \leftarrow \arg \min_{\mathbf{v}} \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \quad (12b)$$

$$\mathbf{u} \leftarrow \mathbf{u} + \mathbf{w} - \mathbf{v} \quad (12c)$$

where the original variable  $\mathbf{w}$  and the auxiliary variable  $\mathbf{v}$  are separately treated in (12a) and (12b), respectively,  $\mathbf{u}$  is the scaled dual variable corresponding to the equality constraint in (11b), i.e.,  $\mathbf{w} = \mathbf{v}$ , and  $\rho > 0$  is the augmented Lagrangian parameter.

In what follows, we show that (12) has closed-form solutions at each ADMM iteration, by deducing  $\mathbf{w}$  and  $\mathbf{v}$  from (12a) and (12b), respectively. First of all, from (12b), it is simple to arrive at the closed-form solution of  $\mathbf{w}$  in a least-squares form, as

$$\mathbf{v} = \rho(2\mathbf{R}_x + \rho\mathbf{I})^{-1}(\mathbf{w} + \mathbf{u}). \quad (13)$$

Now we turn to (12a). We solve (12a) in two steps: i) We consider the unconstrained minimization problem by directly removing its constraint; ii) We check whether the solution obtained from Step i) satisfies the constraint, and update the final solution accordingly. Details are provided as follows.

**Step i):** We consider the following unconstrained minimization problem

$$\min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2. \quad (14)$$

By calculating the subgradient of the objective function in (14) with respect to (w.r.t.)  $\mathbf{w}$ , and setting the resultant expression equal to zero, we obtain its solution, denoted by  $\bar{\mathbf{w}}$ , as

$$\bar{\mathbf{w}} = \text{sign}(\mathbf{v} - \mathbf{u}) \odot \left( |\mathbf{v} - \mathbf{u}| - \frac{\lambda}{\rho} \right)_+. \quad (15)$$

The detailed derivation of (15) from Problem (14) is omitted here, and the interested readers are referred to the similar result in Lemma 1 in [56].

**Step ii):** We check whether or not  $\bar{\mathbf{w}}$  obtained from (15) satisfies  $|\bar{\mathbf{w}}^H \mathbf{a}_0|^2 \geq 1$ . If it is, then the solution to (12a),

referred to as  $\hat{\mathbf{w}}$ , is  $\hat{\mathbf{w}} = \bar{\mathbf{w}}$ . If it is not, then  $\hat{\mathbf{w}}$  can be found via the following theorem.

**Proposition 1.** Denote  $\hat{\mathbf{w}}$  and  $\bar{\mathbf{w}}$  as the solutions to Problems (12a) and (14), respectively. If  $\bar{\mathbf{w}}$  does not satisfy  $|\bar{\mathbf{w}}^H \mathbf{a}_0|^2 \geq 1$ , then as long as  $\rho \gg \lambda$ ,  $\hat{\mathbf{w}}$  equals the one in  $\{\mathbf{w} : |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1\}$ , such that it is closest (in an  $\ell_2$ -norm sense) to  $\bar{\mathbf{w}}$ .

*Proof.* Define  $f(\mathbf{w}) \triangleq \lambda \|\mathbf{w}\|_1 + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2$ , and define  $\tilde{\mathbf{w}}$  such that

$$\|\tilde{\mathbf{w}} - \bar{\mathbf{w}}\|_2 \leq \|\mathbf{w} - \bar{\mathbf{w}}\|_2 \quad (16)$$

holds for any  $\mathbf{w} \in \{\mathbf{w} : |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1\}$ . The Lagrangian parameter is chosen as  $\rho = C\lambda$ , where  $C$  is a constant. As shall be shown later, the augmented Lagrangian parameter  $\rho$  is set to be large in order to make our algorithm converge. When  $C \rightarrow \infty$  (i.e.,  $\rho \gg \lambda$ ), the objective function

$$f(\mathbf{w}) = \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 = \frac{\rho}{2} \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2, \quad (17)$$

where, in the second equality, we used  $\bar{\mathbf{w}} = \mathbf{v} - \mathbf{u}$  as  $C \rightarrow \infty$ .

Suppose that there exists a point  $\mathbf{w}' \in \{\mathbf{w} : |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1\}$ , such that  $f(\mathbf{w}') < f(\tilde{\mathbf{w}})$ . Thus, by using (17), we obtain that  $\frac{\rho}{2} \|\mathbf{w}' - \bar{\mathbf{w}}\|_2^2 < \frac{\rho}{2} \|\tilde{\mathbf{w}} - \bar{\mathbf{w}}\|_2^2$ , which contradicts (16). This implies that  $f(\tilde{\mathbf{w}}) \leq f(\mathbf{w})$  holds for all feasible  $\mathbf{w}$ , that is,  $\tilde{\mathbf{w}}$  is the solution to Problem (12a).  $\square$

**Remark 1.** Besides the above mathematical proof, we give an illustrative example for Proposition 1, by showing the near-symmetric structure of the objective function around its stationary point<sup>1</sup>. For simplicity, the variable is set to be real-valued and the dimension  $M = 1$ . The parameters are  $\lambda = 1$ ,  $\rho = 4$ ,  $\mathbf{a}_0 = 1/2$ , and  $-\mathbf{v} + \mathbf{u} = -1$ . Problem (12a) becomes

$$\min_w f(w) \triangleq |w| + 2(w - 1)^2 \quad \text{subject to } |w| \geq 2, \quad (18)$$

with its stationary point  $w_0$  falling outside its feasible region  $|w| \geq 2$ , as in Fig. 1. Thanks to the convexity and near-symmetric structure of the objective function  $f(w)$ , finding its minimum is equivalent to determining the point (inside the feasible region) closest to its stationary point  $w_0$ . In Fig. 1, it is easy to see that  $w = 2$  is such a point, and thus it is the solution to Problem (18).

Consequently, if  $\bar{\mathbf{w}}$  obtained from (15) does not satisfy  $|\bar{\mathbf{w}}^H \mathbf{a}_0|^2 \geq 1$ , then according to Proposition 1,  $\hat{\mathbf{w}}$  can be found by solving the following minimization problem

$$\hat{\mathbf{w}} \leftarrow \arg \min_{\mathbf{w}} \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2 \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1, \quad (19)$$

which is equivalent to

$$\hat{\mathbf{w}} \leftarrow \arg \min_{\mathbf{w}} \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2 \quad \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 = 1. \quad (20)$$

The equivalence between Problems (19) and (20) is shown in Appendix A, and it indicates that the solutions to Problem (19) always fall on the boundary of its feasible region. Problem (20) has the following closed-form solution as [41]

$$\hat{\mathbf{w}} = \bar{\mathbf{w}} + \frac{1 - |\bar{\mathbf{w}}^H \mathbf{a}_0|}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}^H \mathbf{a}_0|} \mathbf{a}_0 \bar{\mathbf{w}}^H \mathbf{a}_0. \quad (21)$$

<sup>1</sup>We say a function  $f(\mathbf{x})$  has a *near-symmetric* structure around a point  $\mathbf{x}_0$  if and only if  $f(\mathbf{x}_0 + \mathbf{x}) \approx f(\mathbf{x}_0 - \mathbf{x})$  holds for any  $\mathbf{x}$ .

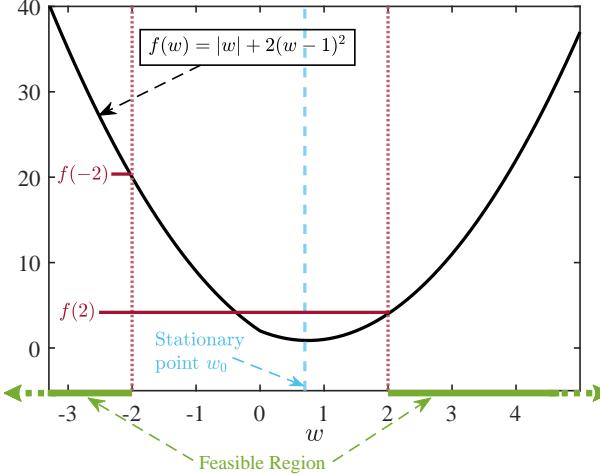


Fig. 1: Illustration of near-symmetric structure of  $f(w)$ .

Eventually, by considering Steps i) and ii) simultaneously and making use of the plus function, the solution to (12a) can be written in a single formula as

$$\hat{\mathbf{w}} = \bar{\mathbf{w}} + \frac{(1 - |\bar{\mathbf{w}}^H \mathbf{a}_0|)_+}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}^H \mathbf{a}_0|} \mathbf{a}_0 \bar{\mathbf{w}}^H \mathbf{a}_0, \quad (22)$$

where  $\bar{\mathbf{w}}$  is given in (15).

So far, we have derived closed-form solutions for  $\mathbf{w}$  and  $\mathbf{v}$  at each ADMM iteration. The complete ADMM for solving Problem (10) is summarized in Algorithm 1, where  $k_{\max}$  denotes a large scalar and  $\eta$  a small one, used to terminate the iteration, and subscript  $\cdot(k)$  denotes the variable at the  $k$ -th iteration. The convergence property of the proposed ADMM algorithm will be discussed in Section IV.

**Remark 2.** Note that we can attain any level of sparsity (i.e., any number  $L$  out of  $M$  sensors in sparse array design), by carefully tuning the value of  $\lambda$ . This shall be verified in the simulation section, see Fig. 3 in Section VI.

**Remark 3.** To ensure selection of  $L$  sensors, an appropriate value of  $\lambda$  is typically found by carrying out a binary search over a probable interval of  $\lambda$ , say  $[\lambda_L, \lambda_U]$ . To be precise, we begin by solving  $\hat{\mathbf{w}}$  using Algorithm 1, with  $\lambda = (\lambda_L + \lambda_U)/2$ . Denote  $\hat{L}$  the number of entries of  $\hat{\mathbf{w}}$  larger than (in modulus)  $\alpha$  times the maximum (in modulus) entry of  $\hat{\mathbf{w}}$ . If  $\hat{L} > L$  (resp.  $\hat{L} < L$ ), then we update  $\lambda_L = \lambda$  (resp.  $\lambda_U = \lambda$ ), and solve another  $\hat{\mathbf{w}}$  with  $\lambda = (\lambda_L + \lambda_U)/2$ . We repeat the above steps until  $\hat{L} = L$ , and the sensors corresponding to the  $L$  largest entries (in modulus) are selected. We set  $\alpha = 0.1$  in our simulations.

**Remark 4.** Note that the solution of (10) is not exactly equal to the one of (9). Therefore, after the solution of desired sparsity of (10) is obtained, one should solve a reduced-size minimization problem similar to (7) as a last step, using only the selected sensors.

#### IV. CONVERGENCE ANALYSIS

The convergence properties of Algorithm 1 are presented in this section. We start with two lemmata, which show the

---

#### Algorithm 1 ADMM for solving Problem (10)

---

**Input :**  $\mathbf{R}_x \in \mathbb{C}^{M \times M}$ ,  $\mathbf{a}_0 \in \mathbb{C}^M$ ,  $\lambda$ ,  $\rho$ ,  $k_{\max}$ ,  $\eta$

**Output :**  $\hat{\mathbf{w}} \in \mathbb{C}^M$

**Initialize:**  $\mathbf{v}_{(0)} \leftarrow \mathbf{v}_{\text{init}}$ ,  $\mathbf{u}_{(0)} \leftarrow \mathbf{u}_{\text{init}}$ ,  $k \leftarrow 0$

- 1: **while** not converged **do**
  - 2:    $\bar{\mathbf{w}}_{(k+1)} \leftarrow \text{sign}(\mathbf{v}_{(k)} - \mathbf{u}_{(k)}) \odot \left( |\mathbf{v}_{(k)} - \mathbf{u}_{(k)}| - \frac{\lambda}{\rho} \right)_+$
  - 3:    $\mathbf{w}_{(k+1)} \leftarrow \bar{\mathbf{w}}_{(k+1)} + \frac{(1 - |\bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0|)_+}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0|} \mathbf{a}_0 \bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0$
  - 4:    $\mathbf{v}_{(k+1)} \leftarrow \rho(2\mathbf{R}_x + \rho\mathbf{I})^{-1}(\mathbf{w}_{(k+1)} + \mathbf{u}_{(k)})$
  - 5:    $\mathbf{u}_{(k+1)} \leftarrow \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$
  - 6:   converged  $\leftarrow k+1 \geq k_{\max}$  or  $\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}\|_2 \leq \eta$
  - 7:    $k \leftarrow k + 1$
  - 8: **end while**
  - 9:  $\hat{\mathbf{w}} \leftarrow \mathbf{w}_{(k)}$
- 

monotonicity and boundedness of the augmented Lagrangian function of Problem (11). Two theorems are then given to show that the proposed algorithm converges and that it converges to a stationary point.

The augmented Lagrangian function regarding Problem (11) can be written as  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \triangleq \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} (\|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2)$ . As stated in Section III-B,  $\mathbf{w}$ ,  $\mathbf{v}$ , and  $\mathbf{u}$  are the original, auxiliary, and dual variables, respectively, and  $\rho > 0$  is the augmented Lagrangian parameter. In what follows, Lemma 1 shows that the function value of  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is non-increasing, and Lemma 2 shows that the function value  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is bounded from below, on the condition that  $\rho$  is larger than or equal to a certain value.

**Lemma 1.** As long as the parameter  $\rho \geq 2\sqrt{2}\lambda_{\max}(\mathbf{R}_x)$ , the point sequence produces a monotonically non-increasing objective function value sequence  $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$ . That is,  $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) \leq \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$  holds for all  $k = 0, 1, 2, \dots$ .

*Proof.* See Appendix B. □

**Lemma 2.** The function value of  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is bounded from below by 0, as long as<sup>2</sup>

$$\rho \geq \frac{2\lambda_{\max}^2(\mathbf{R}_x)}{\lambda_{\min}(\mathbf{R}_x)}. \quad (23)$$

*Proof.* See Appendix C. □

With Lemmata 1 and 2, we have the following theorem.

**Theorem 1.** As long as the augmented Lagrangian parameter

$$\rho \geq \max \left\{ 2\sqrt{2}\lambda_{\max}(\mathbf{R}_x), \frac{2\lambda_{\max}^2(\mathbf{R}_x)}{\lambda_{\min}(\mathbf{R}_x)} \right\}, \quad (24)$$

the objective function value sequence  $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$  generated by Algorithm 1 converges. Furthermore, as  $k \rightarrow \infty$ , we have  $\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)}$ ,  $\mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}$ ,  $\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)}$ , and  $\mathbf{w}_{(k)} = \mathbf{v}_{(k)}$ .

*Proof.* See Appendix D. □

<sup>2</sup>It is worth noting that the lower bound here is not tight, see Appendix C.

Moreover, the following theorem shows that the limit point of Algorithm 1 is a stationary point.

**Theorem 2.** Denote the limit point obtained by using Algorithm 1 as  $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$ . Then, it satisfies the KKT conditions of Problem (11), as

$$\mathbf{0} = 2\mathbf{R}_x \mathbf{v}_{(k+1)} - \mathbf{y}_{(k+1)}, \quad (25a)$$

$$\mathbf{w}_{(k+1)} \in \arg \min_{\mathbf{w}} \left\{ \lambda \|\mathbf{w}\|_1 + \mu^* (|\mathbf{w}^H \mathbf{a}_0|^2 - 1) \right\}, \quad (25b)$$

$$\mathbf{w}_{(k+1)} = \mathbf{v}_{(k+1)}, \quad (25c)$$

where  $\mathbf{y}_{(k+1)} = \rho \mathbf{u}_{(k+1)}$  is the dual variable corresponding to the equality constraint in (11b), and  $\mu^*$  denotes the optimal dual variable corresponding to the inequality constraint in (11b). In words, any limit point of Algorithm 1 is a stationary solution to Problem (11).

*Proof.* See Appendix E.  $\square$

**Lemma 3.** The limit point obtained by using Algorithm 1 is a stationary solution to Problem (9), provided that the tuning parameter  $\lambda$  is chosen such that the solution of Problem (11) is the same as that of Problem (9).

*Proof.* See Appendix F.  $\square$

## V. ADMM WITH RE-WEIGHTED $\ell_1$ -NORM

As has been well-documented in the literature, see e.g. [57], the iteratively re-weighted  $\ell_1$ -norm penalty has remarkable advantages over the conventional  $\ell_1$ -norm. Therefore, in this section, we propose an improved approach on the basis of Algorithm 1, by replacing the  $\ell_1$ -norm regularization in (10) with the re-weighted  $\ell_1$ -norm. That is, Problem (10) is modified as

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_x \mathbf{w} + \lambda \|\mathbf{1} \odot (|\mathbf{g}| + \epsilon) \odot \mathbf{w}\|_1 \quad (26a)$$

$$\text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1, \quad (26b)$$

where  $\mathbf{g}$  equals  $\mathbf{w}$  obtained from the previous iteration, and  $\epsilon > 0$  is a small scalar providing stability and ensuring that a zero-valued component in  $\mathbf{w}$  does not strictly prohibit a non-zero estimate at the next iteration. Note that once the non-zero entries of the solution of Problem (26) are identified, their influence is down-weighted in order to allow more sensitivity for identifying the remaining small but non-zero entries [57]. This results in a better behavior of (26) than (10), which will be corroborated in Section VI-A.

The ADMM iteration for Problem (26) is the same as (12) except for (12a) which should be replaced by

$$\mathbf{w} \leftarrow \begin{cases} \arg \min_{\mathbf{w}} \lambda \|\mathbf{1} \odot (|\mathbf{g}| + \epsilon) \odot \mathbf{w}\|_1 + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \\ \text{subject to } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1. \end{cases} \quad (27)$$

Accordingly, the result of  $\bar{\mathbf{w}}$  in (15) now becomes

$$\bar{\mathbf{w}} = \text{sign}(\mathbf{v} - \mathbf{u}) \odot (|\mathbf{v} - \mathbf{u}| - (\lambda \mathbf{1}) \odot [\rho(|\mathbf{g}| + \epsilon)])_+, \quad (28)$$

and the complete ADMM for solving Problem (26) is summarized in Algorithm 2. The convergence property of Algorithm 2, and the comparison between Algorithms 1 and 2, will be presented using simulations in Section VI-A.

---

## Algorithm 2 ADMM for solving Problem (26)

---

**Input :**  $\mathbf{R}_x \in \mathbb{C}^{M \times M}$ ,  $\mathbf{a}_0 \in \mathbb{C}^M$ ,  $\lambda$ ,  $\rho$ ,  $\epsilon$ ,  $k_{\max}$ ,  $\eta$

**Output :**  $\hat{\mathbf{w}} \in \mathbb{C}^M$

**Initialize:**  $\mathbf{v}_{(0)} \leftarrow \mathbf{v}_{\text{init}}$ ,  $\mathbf{u}_{(0)} \leftarrow \mathbf{u}_{\text{init}}$ ,  $k \leftarrow 0$

- 1: **while** not converged **do**
  - 2:    $\bar{\mathbf{w}}_{(k+1)} \leftarrow \text{sign}(\mathbf{v}_{(k)} - \mathbf{u}_{(k)}) \odot (|\mathbf{v}_{(k)} - \mathbf{u}_{(k)}| - (\lambda \mathbf{1}) \odot [\rho(|\mathbf{v}_{(k)}| + \epsilon)])_+$
  - 3:    $\mathbf{w}_{(k+1)} \leftarrow \bar{\mathbf{w}}_{(k+1)} + \frac{(1 - |\bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0|)_+}{\|\mathbf{a}_0\|_2^2 |\bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0|} \mathbf{a}_0 \bar{\mathbf{w}}_{(k+1)}^H \mathbf{a}_0$
  - 4:    $\mathbf{v}_{(k+1)} \leftarrow \rho(2\mathbf{R}_x + \rho \mathbf{I})^{-1}(\mathbf{w}_{(k+1)} + \mathbf{u}_{(k)})$
  - 5:    $\mathbf{u}_{(k+1)} \leftarrow \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$
  - 6:   converged  $\leftarrow k+1 \geq k_{\max}$  or  $\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}\|_2 \leq \eta$
  - 7:    $k \leftarrow k + 1$
  - 9: **end while**
  - 10:  $\hat{\mathbf{w}} \leftarrow \mathbf{w}_{(k)}$
- 

## VI. SIMULATION RESULTS

In this section, numerical examples are conducted to demonstrate the effectiveness of the proposed algorithms, i.e., Algorithms 1 and 2. We first examine the behaviors of the algorithms in Section VI-A, in terms of convergence property and beamformer weight sparsity control. Then, in Section VI-B, we compare the computational complexity of the proposed algorithms with those of the other state-of-the-art approaches, including SDR, an SDR variant, and SCA, presented in [28], [17], and [46], respectively. In Section VI-C, we finally test the performance of sparse array beamformers designed by using different strategies, in terms of array beampattern and output SINR.

### A. Convergence and Sparsity Control

**First example:** A compact ULA consisting of  $M = 12$  antenna sensors is utilized, while  $T = 100$  snapshots, one SOI from  $\theta_0 = 0^\circ$  and  $K = 2$  interference signals from  $\theta_1 = -10^\circ$  and  $\theta_2 = 10^\circ$ , respectively, are considered. The signal-to-noise ratio (SNR) and interference-to-noise ratio (INR) are SNR = 10 dB and INR = 20 dB, respectively. The two proposed algorithms, i.e., Algorithms 1 and 2 are examined, where the parameters are  $\epsilon = 10^{-10}$ ,  $k_{\max} = 10^3$ ,  $\eta = 10^{-12}$ ,  $\mathbf{u}_{\text{init}} = \mathbf{0}$ , and  $\mathbf{v}_{\text{init}}$  is drawn from the complex standard normal distribution. Three scenarios with different values of the tuning parameter  $\lambda$  and the augmented Lagrangian parameter  $\rho$  are considered. The results of  $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$  versus the ADMM iteration index  $k$  are given in Fig. 2. It is seen that the function value sequence  $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$  of Algorithm 1 is monotonically non-increasing and bounded from below, which is consistent with the theoretical analyses in Section IV. Additionally, we also observe that although the function value sequence  $\{\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})\}$  of Algorithm 2 is not monotonically non-increasing, it converges eventually. Moreover, Algorithm 2 converges faster than Algorithm 1.

**Second example:** A compact ULA of  $M = 12$  antenna sensors is used, and we wish to select  $L$  sensors out of them in the beamformer. We consider 8 situations with different SNR, decreasing from 20 dB to -15 dB with a stepsize of 5 dB. The interference-to-noise ratio is INR = 10 dB

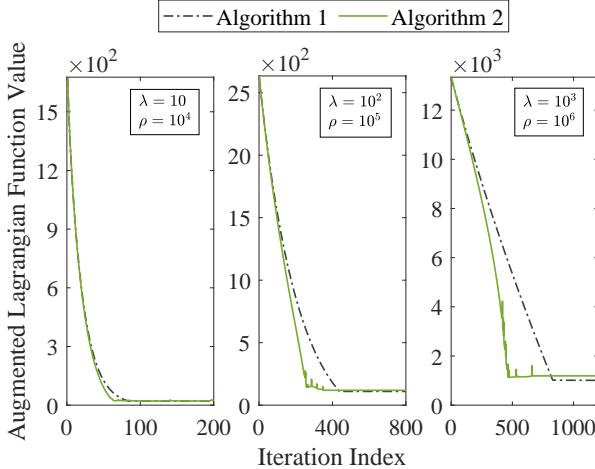


Fig. 2: Function value of  $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$  versus iteration index  $k$ . (1st example)

or INR = 20 dB, the augmented Lagrangian parameter is  $\rho = 2 \times 10^4$ , and the other parameters are the same as those in the first example. We test the sparsity of the beamformer weight obtained via Algorithms 1 and 2 w.r.t. the tuning parameter  $\lambda$ . The curves are averaged over 1000 Monte Carlo runs, and they are displayed in Fig. 3. It can be seen that, for all 8 situations, any level of sparsity (from 1 to 11) could be attained by both algorithms, and that a larger  $\lambda$  produces a smaller sparsity of  $\hat{\mathbf{w}}$ , as expected. When SNR increases, the curve becomes more gentle. In addition, the curves by using Algorithm 1 decrease far more rapidly than those by using Algorithm 2, which implies that it is much easier to tune  $\lambda$  for a specific level of sparsity when Algorithm 2 is employed.

Because of the better behavior of Algorithm 2, rather than both Algorithms 1 and 2, we solely consider Algorithm 2 in the remaining simulations, and it will be labelled as “ADMM”.

### B. Computational Complexity

We start by rewriting the state-of-the-art methods to suit our problem, i.e., Problem (26), using our notations, as [28]

$$\min_{\mathbf{W}} \quad \text{Tr}\{\mathbf{R}_x \mathbf{W}\} + \lambda \text{Tr}\{\mathbf{B}|\mathbf{W}|\} \quad (29a)$$

$$\text{subject to } \begin{cases} \text{Tr}\{\mathbf{a}_0 \mathbf{a}_0^H \mathbf{W}\} \geq 1, \\ \mathbf{W} \succeq 0, \end{cases} \quad (29b)$$

and [17]

$$\min_{\mathbf{W}, \widetilde{\mathbf{W}}} \quad \text{Tr}\{\mathbf{R}_x \mathbf{W}\} + \lambda \text{Tr}\{\mathbf{B} \widetilde{\mathbf{W}}\} \quad (30a)$$

$$\text{subject to } \begin{cases} \text{Tr}\{\mathbf{a}_0 \mathbf{a}_0^H \mathbf{W}\} \geq 1, \\ \mathbf{W} \succeq 0, \\ \widetilde{\mathbf{W}} \geq |\mathbf{W}|, \end{cases} \quad (30b)$$

where  $\mathbf{B} = \mathbf{1}_M \oslash (|\mathbf{G}| + \epsilon)$ , with  $\mathbf{G}$  being equal to  $\mathbf{W}$  obtained from the previous iteration. Problems (29) and (30) are referred to as SDR and SDR-V (short for “SDR Variant”), respectively, in the remaining simulations. If the solution  $\mathbf{W}$  to Problems (29) and (30) is of rank one, their beamformer weight can be calculated as the principal eigenvector of  $\mathbf{W}$ ; otherwise, extra post-processing based on randomization is required [42].

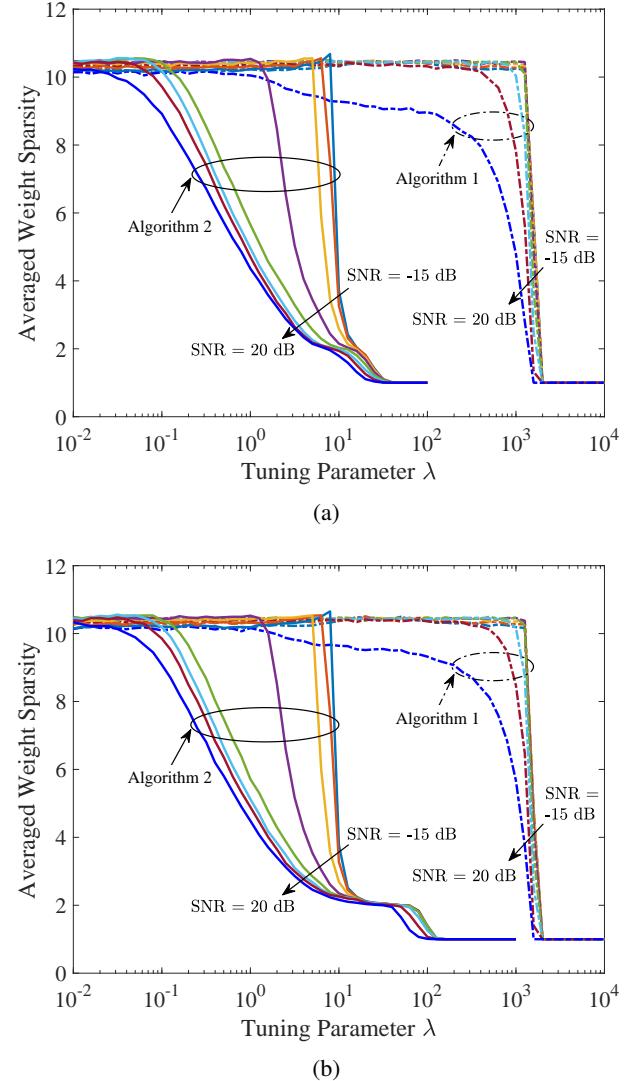


Fig. 3: Sparsity of beamformer weight versus tuning parameter  $\lambda$  with (a) INR = 10 dB and (b) INR = 20 dB. (2nd example)

On the other hand, Problem (26) can be recast as [46]

$$\min_{\mathbf{w}} \quad \mathbf{w}^H \mathbf{R}_x \mathbf{w} + \lambda \|\mathbf{b} \odot \mathbf{w}\|_1 \quad (31a)$$

$$\text{subject to } 2\Re\{\mathbf{g}^H \mathbf{a}_0 \mathbf{a}_0^H \mathbf{w}\} - |\mathbf{g}^H \mathbf{a}_0|^2 \geq 1, \quad (31b)$$

where  $\mathbf{b} = \mathbf{1} \oslash (|\mathbf{g}| + \epsilon)$  with  $\mathbf{g}$  being equal to  $\mathbf{w}$  obtained from the previous iteration, the same as what has been introduced in Section V. Problem (31) is denoted as SCA in the remaining simulations. The key of the success of SCA method lies in the fact that  $|\mathbf{w}^H \mathbf{a}_0|^2$  is a convex function w.r.t.  $\mathbf{w}$  and thus  $|\mathbf{w}^H \mathbf{a}_0|^2 \geq 2\Re\{\mathbf{g}^H \mathbf{a}_0 \mathbf{a}_0^H \mathbf{w}\} - |\mathbf{g}^H \mathbf{a}_0|^2$  holds for all  $\mathbf{w}$  and any given (known)  $\mathbf{g}$ .

If a general-purpose SDR solver, such as the interior point method, is adopted to solve Problems (29) and (30), the worst case complexity can be as high as  $\mathcal{O}(M^{6.5})$  per iteration [41]. The cost of solving Problem (31) could be smaller, if further effort is made, for instance, by taking care of the structure of the problem. However, since this is out of the scope of this paper, in our simulations, we simply utilize the CVX toolbox [58] to solve the aforementioned three problems.

The computational costs of the proposed Algorithm 1 and Algorithm 2 are the same in terms of big O notation. In what follows, we analyse the cost of Algorithm 2 in detail. The cost of computing  $\bar{\mathbf{w}}_{(k+1)}$  (Lines 2 and 3 in Algorithm 2) is  $\mathcal{O}(M)$ . The cost of computing  $\mathbf{w}_{(k+1)}$  (Line 4 in Algorithm 2) is  $\mathcal{O}(M)$ . The cost of computing  $\mathbf{v}_{(k+1)}$  (Line 5 in Algorithm 2) is  $\mathcal{O}(M^3) + \mathcal{O}(M^2)$ , where  $\mathcal{O}(M^3)$  corresponds to the inversion operation, i.e.,  $(2\mathbf{R}_x + \rho\mathbf{I})^{-1}$ , and  $\mathcal{O}(M^2)$  is from the matrix multiplication. Note that we can cache the result of  $(2\mathbf{R}_x + \rho\mathbf{I})^{-1}$ , to save computations in the subsequent iterations. Finally, the cost of computing  $\mathbf{u}_{(k+1)}$  (Line 6 in Algorithm 2) is  $\mathcal{O}(M)$ . Therefore, the total computational cost of Algorithm 2 is  $\mathcal{O}(M^3) + \mathcal{O}(K_{\text{admm}}M^2)$ , where  $K_{\text{admm}}$  denotes the number of iterations required by the proposed ADMM. Note that the above analyses are based on a fixed  $\lambda$ . We may adjust  $\lambda$  to ensure  $L$  sensors are selected, as mentioned in Remark 3. Let  $K_\lambda$  be the number of  $\lambda$  values that are considered, then the total complexity cost of the proposed ADMM is  $\mathcal{O}(K_\lambda M^3) + \mathcal{O}(K_\lambda K_{\text{admm}}M^2)$ .

It is worth noting that when  $\mathbf{B}$  in (29) and (30) and  $\mathbf{b}$  in (31) are fixed as  $\mathbf{B} = \mathbf{1}_M$  and  $\mathbf{b} = \mathbf{1}$ , Problems (29), (30), and (31) reduce to three approaches for solving Problem (10), which correspond to Algorithm 1. As has been confirmed by the numerical results in Section VI-A, algorithms with re-weighted  $\ell_1$ -norm regularization are more efficient in the sense that they converge faster and are much easier to control the sparsity of solution, compared to their counterparts with conventional  $\ell_1$ -norm regularization. Therefore, only the former group of approaches, i.e., the above-mentioned SDR (29), SDR-V (30), SCA (31), and Algorithm 2, are considered in the following simulations.

**Third example:** We wish to choose  $L = 4$  out of  $M = 12$  antenna sensors from a compact ULA. One SOI from  $\theta_0 = 0^\circ$  and  $K = 2$  interferers from  $\theta_1 = -10^\circ$  and  $\theta_2 = 10^\circ$ , respectively, are considered, while SNR = 0 dB and INR = 20 dB. The number of snapshots  $T$  varies uniformly from 10 to 150 with a stepsize of 10. The other parameters for Algorithm 2 are the same as those of the first example, except for  $\rho$  which is set to  $\rho = 10^3$  in this example. The CPU times of the examined approaches are averaged over 100 Monte Carlo runs, and they are plotted in Fig. 4. It is seen that their CPU times are almost unchanged when  $T$  varies, and that of the ADMM method is around  $10^{-1}$  seconds which is about  $10^3$  times less than those of the SDR, SDR-V, and SCA methods (which take around  $10^2$  seconds).

**Fourth example:** We wish to select  $L = 4$  out of  $M$  sensors from a compact ULA. The number of snapshots is fixed as  $T = 100$ , and the number of sensors  $M$  changes from 10 to 20. The other parameters remain unchanged as those of the third example. The CPU times of the examined methods are shown in Fig. 5, from which it is seen that the CPU times of the SDR, SDR-V, and ADMM methods increase as  $M$  increases, while the CPU time of the SCA method keeps almost unchanged when  $M$  varies. In addition, the CPU time of the proposed algorithm is much smaller than those of the other three tested approaches. Note that there is a sharp change of the ADMM curve at  $M = 15$ . This is caused by the increased number of iterations when  $M \geq 16$ .

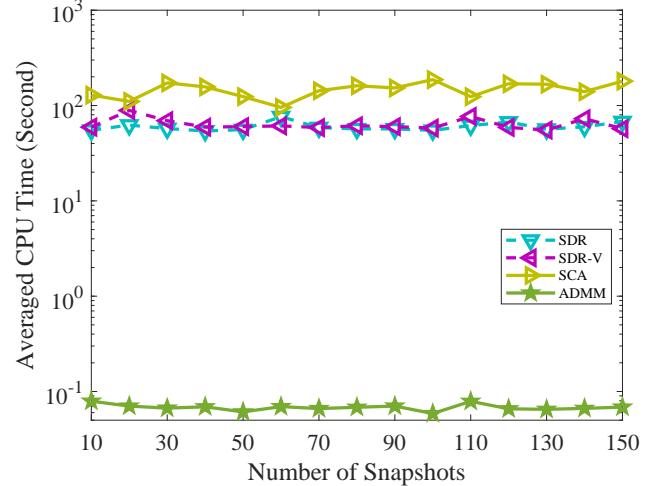


Fig. 4: Averaged CPU time versus  $T$ . (3rd example)

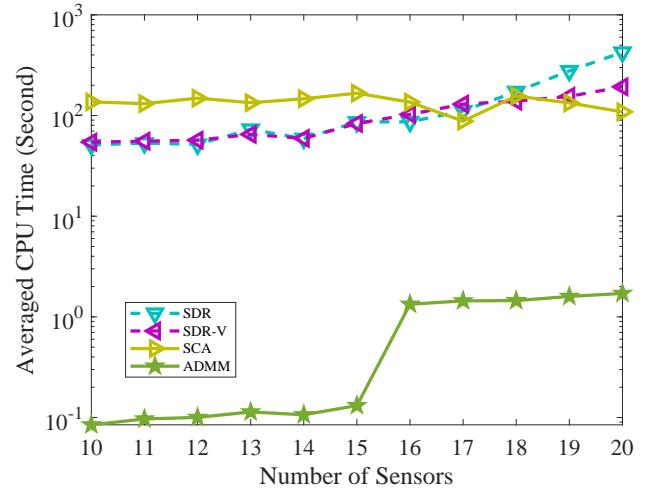


Fig. 5: Averaged CPU time versus  $M$ . (4th example)

**Fifth example:** We wish to select  $L$  out of  $M = 12$  sensors from a compact ULA. The number of snapshots is fixed as  $T = 100$ , and the number of selected sensors  $L$  changes from 3 to 12. The other parameters remain unchanged as those of the third example. The CPU times of the examined methods are drawn in Fig. 6, from which it is seen that the CPU times of all the four methods increase as  $L$  increases. Besides, the CPU time of the proposed algorithm is shown again much less than those of the other three tested methods.

### C. Beamforming Performance

In the following simulations, we examine the beamforming performance of the proposed algorithm compared with several other sparse array design strategies, in terms of beampattern and output SINR.

**Sixth example:** We choose  $L = 4$  out of  $M = 12$  sensors. One SOI from  $\theta_0 = 0^\circ$  and  $K = 2$  interferers from  $\theta_1 = -40^\circ$  and  $\theta_2 = 30^\circ$ , are considered, while SNR = 0 dB and INR = 20 dB. Different sparse array design strategies are examined, including enumeration (i.e., exhaustive search), compact ULA, sparse ULA, random array, nested array [59],

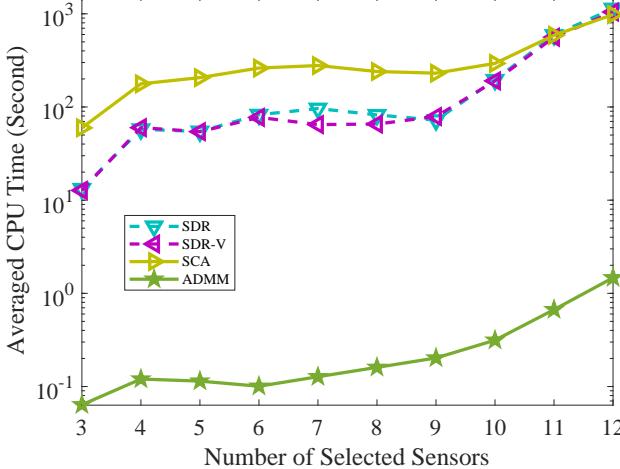


Fig. 6: Averaged CPU time versus  $L$ . (5th example)

TABLE II: Output SINR of sixth example.

Method	SINR (dB)	Method	SINR (dB)
Whole ULA	7.6160	Coprime	4.2298
Best via Enum.	5.3874	SDR	4.1792
ADMM	5.0253	Nested	4.0458
Compact ULA	4.9967	SCA	3.9229
SDR-V	4.9670	Sparse ULA	-7.4871
Random	4.9320	Worst via Enum.	-7.5373

coprime array [60], [61], SDR [28], SDR-V [17], SCA [46], and the proposed ADMM. The enumeration method tests all possible (i.e.,  $\binom{12}{4} = 495$ ) combinations, and finds out the best and worst cases. The compact ULA includes the first  $L = 4$  sensors, the sparse ULA contains the 1st, 3rd, 5th, and 7th sensors, while the random array choose  $L = 4$  out of  $M = 12$  sensors in a random manner. Besides, the nested array contains the 1st, 2nd, 3rd, and 6th sensors, and coprime array includes the 1st, 3rd, 4th and 5th sensors. SDR, SDR-V, and SCA refer to Problems (29), (30), and (31), respectively. The result of using the whole ULA is also included. Their beampatterns are depicted in Fig. 7, where we separate them into two subfigures and the ADMM is drawn in both, for a better comparison. From Fig. 7 we observe that the proposed ADMM method provides lower sidelobe and deeper nulls towards the interferences, compared to the others. Their output SINRs in this example are given in TABLE II.

**Seventh example:** We consider the scenario with one SOI whose DOA  $\theta_0$  changes from  $-60^\circ$  to  $60^\circ$ , and  $K = 2$  interferers from  $\theta_1 = \theta_0 - 10^\circ$  and  $\theta_2 = \theta_0 + 10^\circ$ , respectively. The remaining parameters are unchanged as those in the third example. The output SINR versus DOA of the SOI is plotted in Fig. 8. Note that the result of MVDR with whole ULA using the true covariance is termed as ‘‘Optimal with Whole ULA’’. It is seen that the proposed method has excellent performance, whose output SINR is less than 0.4 dB lower than that of the best case via enumeration, very close (within 0.6 dB) to the optimal SINR, slightly larger than those of the SDR, SDR-V, and SCA methods, and at least about 2 dB larger than those of the other approaches. Two exceptions occur at  $\theta_0 = -55^\circ$  and  $\theta_0 = 55^\circ$ , in which cases the output SINR of ADMM is about 2.5 dB lower than those of the best case via enumeration, SDR,

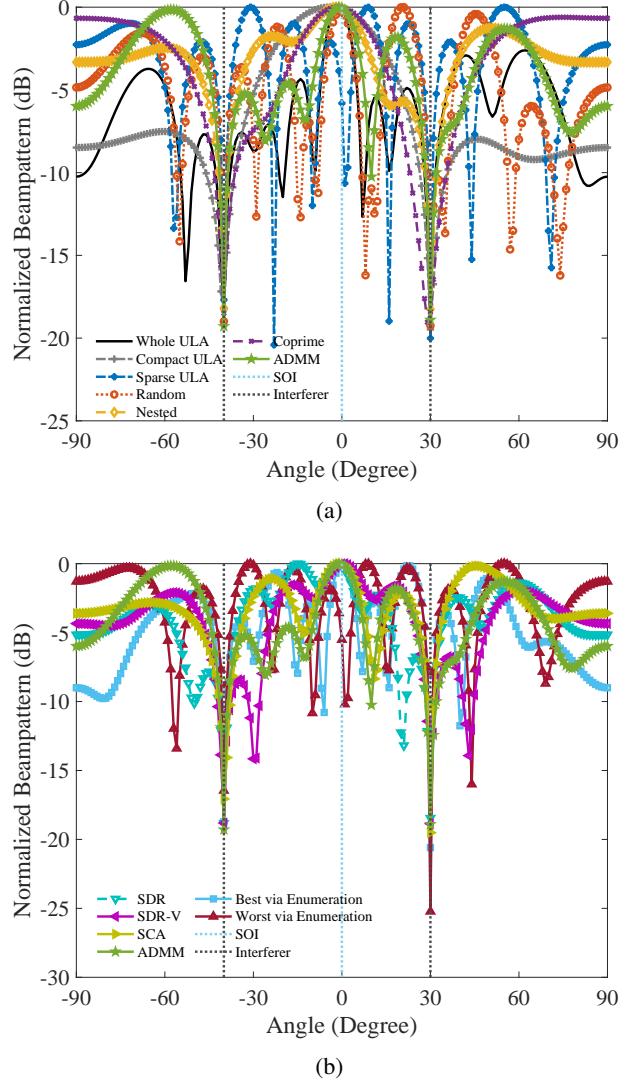


Fig. 7: Beampattern comparison with 1 SOI and 2 interferers. (6th example)

SDR-V, and SCA methods, and is still significantly higher than those of the other approaches. Another interesting result is that the performance of the nested array and the coprime array is even worse than that of the random array. This is because the goal of nested array and coprime array is to make sure more continuous virtual sensors exist in their difference coarray, such that they can estimate more sources than physical sensors. In other words, nested array and coprime array are designed to obtain better performance in DOA estimation, but not necessary to have good performance in beamforming in terms of SINR.

**Eighth example:** The SNR varies from  $-20$  dB to  $12$  dB with a stepsize of  $2$  dB. The DOA of the SOI is  $\theta_0 = 0^\circ$  and  $K = 2$  interference signals come from  $\theta_1 = -10^\circ$  and  $10^\circ$ . The other parameters are unchanged compared with those of the previous example. The output SINR versus SNR is depicted in Fig. 9(a). To provide a clearer vision of the results, we calculate the SINR departure of the corresponding methods from the optimal SINR, and draw them in Fig. 9(b). The figures demonstrate better performance of the proposed

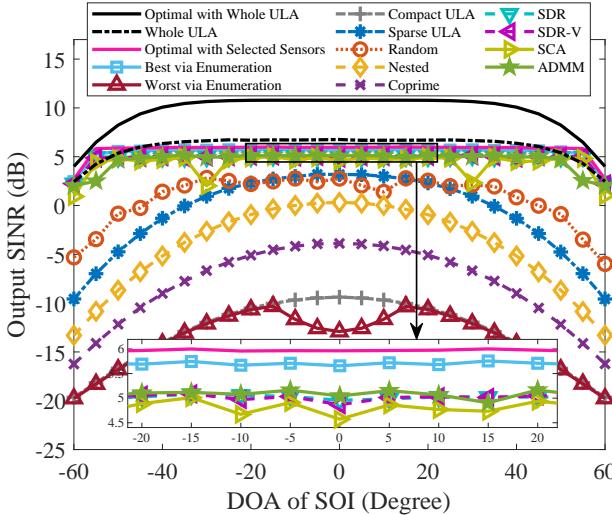


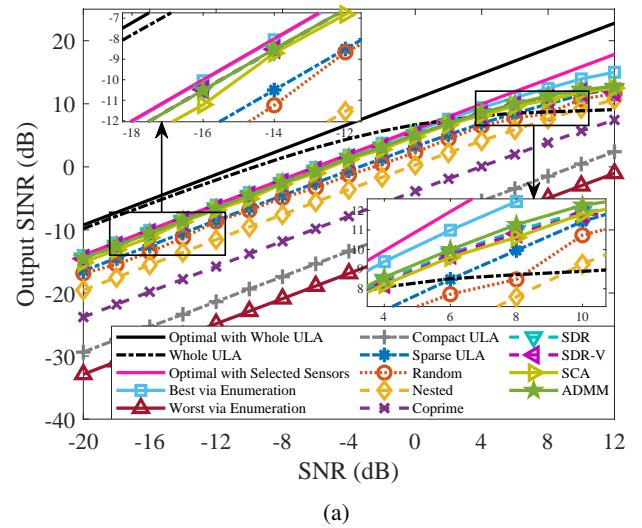
Fig. 8: Output SINR versus DOA of SOI. (7th example)

scheme than the other sparse array design techniques (except for the best case via enumeration) in terms of output SINR. It is also observed that in the large SNR region, the SINR of the whole array is even smaller than those of sparse arrays. This verifies the statement that the performance of beamforming is affected by not only the beamformer weight, but also the array configuration [4]. The reason comes from two aspects: On one hand, the performance of MVDR beamformer degrades when SNR is high, since the SOI presents in the training data [3]. On the other hand, the calculation of SINR for the whole array is different from that for the sparse arrays, that is, the length of the beamformer weight and the size of  $\mathbf{R}_{i+n}$ , used for calculating SINR, are different.

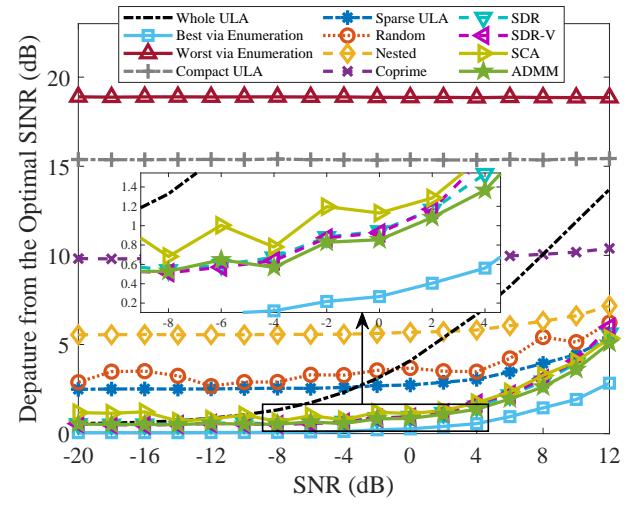
**Ninth example:** We examine the output SINR versus the number of snapshots. The simulation setup is the same as that of the third example. The results are shown in Fig. 10, from which we observe that the output SINR of the proposed ADMM is close to that of the best case via enumeration, slightly larger than those of the SDR, SDR-V, and SCA approaches, and significantly larger than those of the others.

**Tenth example:** We examine the output SINR versus the number of sensors. The simulation setup is unchanged as that of the fourth example. The results are plotted in Fig. 11, from which we again observe that the output SINR of the ADMM is close to that of the best case via enumeration, slightly larger than those of the SDR, SDR-V, and SCA methods, and significantly larger than those of the others.

**Eleventh example:** We examine the output SINR versus the number of selected sensors. The simulation setup is the same as that of the fifth example. Note that, some previous methods are not included in this example since they do not apply to the entire range of  $L$ . The results are displayed in Fig. 12. It is seen that when  $L$  approaches  $M = 12$ , the output SINR of all the tested methods converge to one point, because in this case ( $L = M$ ) all the methods select the same sensors, i.e., the whole ULA. On the other hand, when  $L < 12$ , the proposed ADMM has higher output SINR than the other tested approaches (except for the best case via enumeration).



(a)



(b)

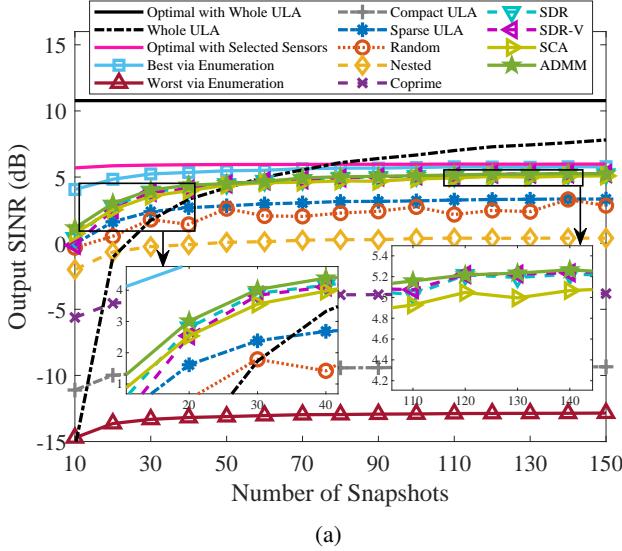
Fig. 9: Output SINR versus input SNR. (8th example)

## VII. CONCLUSION AND FUTURE WORKS

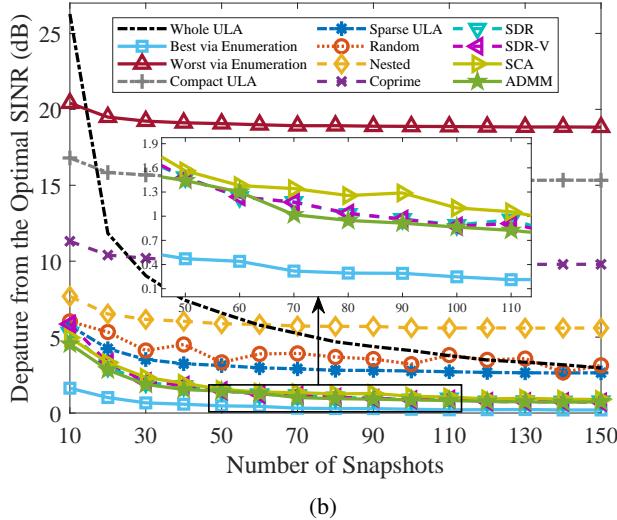
An algorithm based on alternating direction method of multipliers (ADMM) for sparse array beamformer design was proposed. Our approach provides closed-form solutions at each ADMM iteration. Theoretical analyses and numerical simulations were provided to show the convergence of the proposed algorithm. In addition, the algorithm was proved to converge to the set of stationary points. The ADMM algorithm was shown comparable to the exhaustive search method, and slightly better than the state-of-the-art solvers, including the semidefinite relaxation (SDR), an SDR variant (SDR-V), and the successive convex approximation (SCA) methods, and significantly better than several other sparse array design strategies in terms of output signal-to-interference-plus-noise ratio. Moreover, the proposed ADMM algorithm outperformed the SDR, SDR-V, and SCA approaches in terms of computational cost.

Possible future extensions of this work are given as follows.

- In this work, we consider narrow-band sources. We might extend it to wide-band sources, as investigated in [18].



(a)



(b)

Fig. 10: Output SINR versus  $T$ . (9th example)

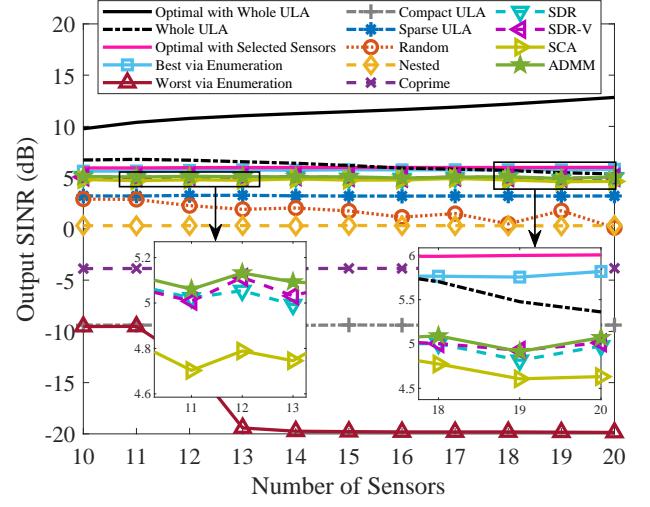
- We might also consider the situation when the steering vector of SOI is subject to uncertainties.
- We can also extend this work to multiple SOIs. In such a scenario, the problem should be formulated as one with multiple constraints. We can extend the proposed ADMM method to consensus-ADMM method. The results in our previous work [47] can be applied to this situation.

## APPENDIX A PROOF OF EQUIVALENCE OF PROBLEMS (19) AND (20)

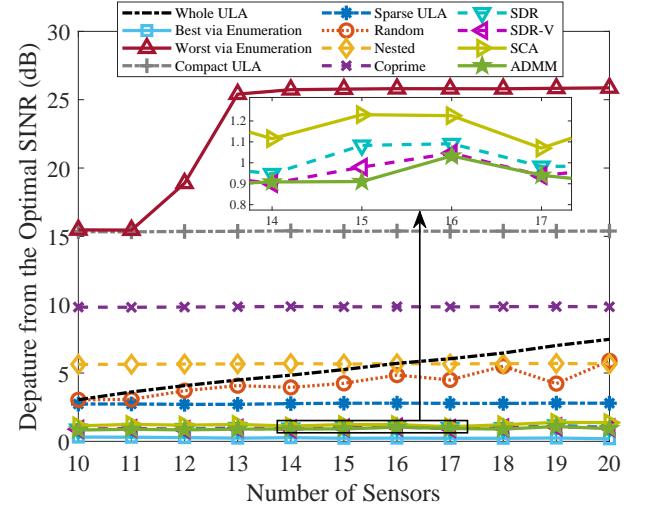
The Lagrangian function of Problem (19) is given as

$$L(\mathbf{w}, \mu) = \|\mathbf{w} - \bar{\mathbf{w}}\|_2^2 + \mu (1 - |\mathbf{w}^H \mathbf{a}_0|^2), \quad (32)$$

where  $\mu \geq 0$  is the Lagrangian dual variable. The optimal solution  $\mathbf{w}^*$  and the optimal dual variable  $\mu^*$  satisfy the



(a)



(b)

Fig. 11: Output SINR versus  $M$ . (10th example)

following identities:

$$\frac{\partial L(\mathbf{w}, \mu)}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}^*, \mu=\mu^*} = 2(\mathbf{w}^* - \bar{\mathbf{w}}) - 2\mu^* \mathbf{a}_0 \mathbf{a}_0^H \mathbf{w}^* = \mathbf{0}, \quad (33a)$$

$$\frac{\partial L(\mathbf{w}, \mu)}{\partial \mu} \Big|_{\mathbf{w}=\mathbf{w}^*, \mu=\mu^*} = 1 - |\mathbf{w}^H \mathbf{a}_0|^2 = 0. \quad (33b)$$

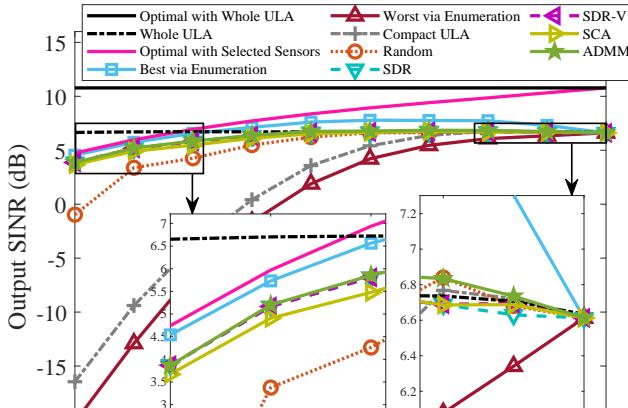
From (33a), we have  $\mathbf{w}^* = (\mathbf{I} - \mu^* \mathbf{a}_0 \mathbf{a}_0^H)^{-1} \bar{\mathbf{w}}$ . Substituting it into (33b) yields

$$1 - |\bar{\mathbf{w}}^H (\mathbf{I} - \mu^* \mathbf{a}_0 \mathbf{a}_0^H)^{-1} \mathbf{a}_0|^2 = 0. \quad (34)$$

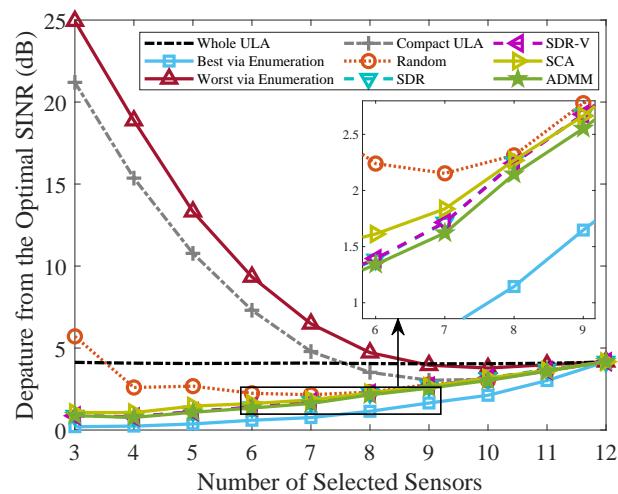
If  $\mu^* = 0$ , the above equation becomes  $1 - |\bar{\mathbf{w}}^H \mathbf{a}_0|^2 = 0$ , which contradicts the fact that  $\bar{\mathbf{w}}$  does not satisfy  $|\bar{\mathbf{w}}^H \mathbf{a}_0|^2 \geq 1$ . Therefore, the optimal dual variable can only be positive, i.e.,  $\mu^* > 0$ .

The complementary slackness condition [62] for Problem (19) is

$$\mu^* \geq 0 \quad \text{and} \quad \mu^* (1 - |\mathbf{w}^{*H} \mathbf{a}_0|^2) = 0. \quad (35)$$



(a)



(b)

Fig. 12: Output SINR versus  $L$ . (11th example)

As mentioned earlier, we have  $\mu^* > 0$ , which together with  $\mu^*(1 - |\mathbf{w}^{*H}\mathbf{a}_0|^2) = 0$ , yields  $|\mathbf{w}^{*H}\mathbf{a}_0|^2 = 1$ . This indicates that the optimal solution to Problem (19) should meet the equality constraint. In other words, Problem (19) is equivalent to Problem (20).

## APPENDIX B PROOF OF LEMMA 1

In order to show that  $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) \leq \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$  holds  $\forall k = 0, 1, 2, \dots$ , where the objective function is defined as  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \triangleq \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} (\|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2)$ , we formulate their difference as

$$\begin{aligned} & \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \\ &= [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)})] \quad (36a) \\ & \quad + [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})]. \quad (36b) \end{aligned}$$

In what follows, we separately deal with (36a) and (36b). For (36a), it is calculated as

$$\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)})$$

$$\begin{aligned} & \stackrel{(a)}{=} \frac{\rho}{2} (\|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k+1)}\|_2^2 - \|\mathbf{u}_{(k+1)}\|_2^2 \\ & \quad - \|\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k)}\|_2^2 + \|\mathbf{u}_{(k)}\|_2^2) \\ & \stackrel{(b)}{=} \frac{\rho}{2} (2\|\mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}\|_2^2 - 2\|\mathbf{u}_{(k+1)}\|_2^2 + \|\mathbf{u}_{(k)}\|_2^2) \\ & = \frac{\rho}{2} (2\|\mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}\|_2^2) \\ & \stackrel{(c)}{=} \rho \left\| \frac{2}{\rho} \mathbf{R}_x \mathbf{v}_{(k+1)} - \frac{2}{\rho} \mathbf{R}_x \mathbf{v}_{(k)} \right\|_2^2 \\ & \stackrel{(d)}{\leq} \frac{4}{\rho} \lambda_{\max}^2(\mathbf{R}_x) \|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2, \end{aligned}$$

where the definition of  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is used in (a);  $\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} = \mathbf{u}_{(k+1)} - \mathbf{u}_{(k)}$  (which is from Line 5 in Algorithm 1) has been utilized in (b);  $\mathbf{u}_{(k+1)} = \frac{2}{\rho} \mathbf{R}_x \mathbf{v}_{(k+1)}$  (which is the result by combining Lines 4 and 5 in Algorithm 1) has been employed in (c); and inequality  $\|\mathbf{R}_x \mathbf{v}\|_2^2 = \mathbf{v}^H \mathbf{R}_x^H \mathbf{R}_x \mathbf{v} \leq \mathbf{v}^H [\lambda_{\max}^2(\mathbf{R}_x) \mathbf{I}] \mathbf{v} = \lambda_{\max}^2(\mathbf{R}_x) \|\mathbf{v}\|_2^2$  has been used in (d).

Now we focus on (36b), which can be written as

$$\begin{aligned} & \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \\ &= [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})] \\ & \quad + [\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})] \\ & \stackrel{(a)}{\leq} \left[ \Re \{ \nabla_{\mathbf{v}} \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k)}) \}, \mathbf{v}_{(k+1)} - \mathbf{v}_{(k)} \right] - \frac{\gamma_{\mathbf{v}}}{2} \|\mathbf{v}_{(k+1)} \\ & \quad - \mathbf{v}_{(k)}\|_2^2 + 0 \\ & \stackrel{(b)}{=} - \frac{\gamma_{\mathbf{v}}}{2} \|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2 \\ & \stackrel{(c)}{=} - \left[ \lambda_{\min}(\mathbf{R}_x) + \frac{\rho}{2} \right] \|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2, \end{aligned}$$

where in (a) we used the following two facts: i)  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is strongly convex w.r.t.  $\mathbf{v}$  with parameter  $\gamma_{\mathbf{v}} > 0$  [63], and ii)  $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \leq 0$  since  $\mathbf{w}_{(k+1)}$  is optimal to Problem (12a); in (b) we have used the optimality condition of Problem (12b); in (c) we have used  $\gamma_{\mathbf{v}} = 2\lambda_{\min}(\mathbf{R}_x) + \rho$ , which is due to the facts that the objective function  $\mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u})$  is twice continuously differentiable w.r.t.  $\mathbf{v}$ , and thus its strong convexity parameter  $\gamma_{\mathbf{v}}$  satisfies  $\nabla_{\mathbf{v}}^2 \mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \succeq \gamma_{\mathbf{v}} \mathbf{I}$  for all  $\mathbf{v}$  [63].

By substituting the above two inequalities back to (36), and denoting  $\lambda_{\max}(\mathbf{R}_x)$  and  $\lambda_{\min}(\mathbf{R}_x)$  as  $\lambda_{\max}$  and  $\lambda_{\min}$  for brevity, respectively, we have

$$\begin{aligned} & \mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \\ & \leq \underbrace{\left[ \frac{4}{\rho} \lambda_{\max}^2 - \lambda_{\min} - \frac{\rho}{2} \right] \|\mathbf{v}_{(k+1)} - \mathbf{v}_{(k)}\|_2^2}_{(i)}. \end{aligned}$$

We observe that if  $\rho < -\sqrt{\lambda_{\min}^2 + 8\lambda_{\max}^2} - \lambda_{\min}$ , which should be deleted as  $\rho > 0$ , or  $\rho > \sqrt{\lambda_{\min}^2 + 8\lambda_{\max}^2} - \lambda_{\min}$ , the coefficient  $\frac{4}{\rho} \lambda_{\max}^2 - \lambda_{\min} - \frac{\rho}{2} < 0$  and thus (i)  $\leq 0$ . Furthermore, because of  $\sqrt{\lambda_{\min}^2 + 8\lambda_{\max}^2} - \lambda_{\min} < \lambda_{\min} + 2\sqrt{2}\lambda_{\max} - \lambda_{\min} = 2\sqrt{2}\lambda_{\max}$ , we have the conclusion that as long as  $\rho \geq 2\sqrt{2}\lambda_{\max}(\mathbf{R}_x)$ ,  $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)}) - \mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)}) \leq (i) \leq 0$ .

## APPENDIX C PROOF OF LEMMA 2

Note that the augmented Lagrangian function satisfies

$$\begin{aligned}
& \mathcal{L}(\mathbf{w}, \mathbf{v}, \mathbf{u}) \\
&= \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 - \frac{\rho}{2} \|\mathbf{u}\|_2^2 \\
&\stackrel{(a)}{=} \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 - \frac{\rho}{2} \left\| \frac{2}{\rho} \mathbf{R}_x \mathbf{v} \right\|_2^2 \\
&\stackrel{(b)}{\geq} \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 - \frac{2}{\rho} \mathbf{v}^H [\lambda_{\max}^2(\mathbf{R}_x) \mathbf{I}] \mathbf{v} \\
&= \lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \left( \mathbf{R}_x - \frac{2}{\rho} \lambda_{\max}^2(\mathbf{R}_x) \mathbf{I} \right) \mathbf{v} + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v} + \mathbf{u}\|_2^2 \\
&\stackrel{(c)}{\geq} 0,
\end{aligned}$$

where in (a) we have used  $\mathbf{u} = \frac{2}{\rho} \mathbf{R}_x \mathbf{v}$  (which is the result by combining (12c) and (13)); in (b) we have used the inequality  $\|\mathbf{R}_x \mathbf{v}\|_2^2 = \mathbf{v}^H \mathbf{R}_x^H \mathbf{R}_x \mathbf{v} \leq \mathbf{v}^H [\lambda_{\max}^2(\mathbf{R}_x) \mathbf{I}] \mathbf{v}$ ; and inequality (c) holds if  $\mathbf{R}_x - \frac{2}{\rho} \lambda_{\max}^2(\mathbf{R}_x) \mathbf{I} \succeq 0$ , which indicates (23). Note that inequality (c) is not tight, since the  $\ell_1$ -norm term and the  $\ell_2$ -norm term could be bounded from below by some large positive value. Therefore, the second term, i.e., the quadratic term w.r.t.  $\mathbf{v}$ , has much space to be tuned, which results in the fact that the lower bound for  $\rho$  in (23) is not tight.

## APPENDIX D PROOF OF THEOREM 1

Denote  $\mathcal{L}(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$  and  $\mathcal{L}(\mathbf{w}_{(k)}, \mathbf{v}_{(k)}, \mathbf{u}_{(k)})$  by  $\mathcal{L}_{(k+1)}$  and  $\mathcal{L}_{(k)}$ , respectively. According to Lemmata 1 and 2, the objective function value sequence  $\{\mathcal{L}_{(k)}\}$  produced by Algorithm 1 converges. Further, since sequence  $\{\mathcal{L}_{(k)}\}$  converges if  $\rho$  satisfies (24), we have  $\mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \rightarrow 0$  as  $k \rightarrow \infty$ . On the other hand, we know from Appendix B that  $\mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \leq (i) \leq 0$ , as long as (24) holds, where (i) is defined in Appendix B. Therefore, when (24) holds and  $k \rightarrow \infty$ , we have

$$0 = \mathcal{L}_{(k+1)} - \mathcal{L}_{(k)} \leq (i) \leq 0, \quad (37)$$

meaning that (i) = 0. This further indicates that

$$\mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}. \quad (38)$$

As already mentioned in Appendix B,  $\mathbf{u} = \frac{2}{\rho} \mathbf{R}_x \mathbf{v}$ . By jointly considering  $\mathbf{u} = \frac{2}{\rho} \mathbf{R}_x \mathbf{v}$  and  $\mathbf{v}_{(k+1)} = \mathbf{v}_{(k)}$  in (38), we obtain

$$\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)}. \quad (39)$$

Note that the update of  $\mathbf{w}_{(k+1)}$  is based on  $\mathbf{v}_{(k)}$  and  $\mathbf{u}_{(k)}$  (see Lines 2-4 in Algorithm 1). This together with (38) and (39) imply that: when  $k \rightarrow \infty$ , we have

$$\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)}. \quad (40)$$

Moreover, combining (39) and  $\mathbf{u}_{(k+1)} = \mathbf{u}_{(k)} + \mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)}$  (i.e., Line 5 in Algorithm 1) yields

$$\mathbf{w}_{(k+1)} = \mathbf{v}_{(k+1)}, \quad (41)$$

as  $k \rightarrow \infty$ . Equivalently,  $\mathbf{w}_{(k)} = \mathbf{v}_{(k)}$  as  $k \rightarrow \infty$ .

## APPENDIX E PROOF OF THEOREM 2

The Lagrangian function of (11) is given by

$$\lambda \|\mathbf{w}\|_1 + \mathbf{v}^H \mathbf{R}_x \mathbf{v} + \mu(|\mathbf{w}^H \mathbf{a}_0| - 1) + \Re\{\langle \mathbf{y}, \mathbf{w} - \mathbf{v} \rangle\}, \quad (42)$$

where  $\mu$  and  $\mathbf{y}$  are Lagrangian dual variables corresponding to the inequality and equality constraints, respectively. Note that the (Lagrangian) dual variable  $\mathbf{y}$  and the scaled dual variable  $\mathbf{u}$  are related to each other as  $\mathbf{y} = \rho \mathbf{u}$  [49]. A KKT point  $(\mathbf{w}^*, \mathbf{v}^*)$  of Problem (11), together with the corresponding dual variables  $\mu^*$  and  $\mathbf{y}^*$ , satisfies [43]

$$\mathbf{0} = 2\mathbf{R}_x \mathbf{v}^* - \mathbf{y}^*, \quad (43a)$$

$$\mathbf{w}^* \in \arg \min_{\mathbf{w}} \left\{ \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \Re\{\langle \mathbf{y}^*, \mathbf{w} - \mathbf{v}^* \rangle\} \right\}, \quad (43b)$$

$$\mathbf{w}^* = \mathbf{v}^*. \quad (43c)$$

Our aim is to show any limit point of Algorithm 1, referred to as  $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$  or  $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{y}_{(k+1)}/\rho)$ , satisfies (43). Firstly, note that Line 4 in Algorithm 1 indicates

$$2\mathbf{R}_x \mathbf{v}_{(k+1)} - \rho(\mathbf{w}_{(k+1)} - \mathbf{v}_{(k+1)} + \mathbf{u}_{(k)}) = \mathbf{0}. \quad (44)$$

Jointly considering (39), (41), (44), and  $\mathbf{y}_{(k+1)} = \rho \mathbf{u}_{(k+1)}$  yields  $2\mathbf{R}_x \mathbf{v}_{(k+1)} - \mathbf{y}_{(k+1)} = \mathbf{0}$ , which is (43a). Additionally, (41) shows that (43c) is also achieved.

Now we turn to (43b). According to (12a), we have:  $\mathbf{w}_{(k+1)}$

$$\begin{aligned}
&= \arg \min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v}_{(k)} + \mathbf{u}_{(k)}\|_2^2 \quad \text{s.t. } |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1 \\
&\stackrel{(a)}{=} \arg \min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v}_{(k)} + \mathbf{u}_{(k)}\|_2^2 \\
&= \arg \min_{\mathbf{w}} \left\{ \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \frac{\rho}{2} \|\mathbf{w} - \mathbf{v}_{(k)}\|_2^2 + \frac{\rho}{2} \|\mathbf{u}_{(k)}\|_2^2 + \rho \Re\{\langle \mathbf{u}_{(k)}, \mathbf{w} - \mathbf{v}_{(k)} \rangle\} \right\} \\
&\stackrel{(b)}{=} \arg \min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \rho \Re\{\langle \mathbf{u}_{(k)}, \mathbf{w} - \mathbf{v}_{(k)} \rangle\} \\
&\stackrel{(c)}{=} \arg \min_{\mathbf{w}} \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \Re\{\langle \mathbf{y}_{(k+1)}, \mathbf{w} - \mathbf{v}_{(k+1)} \rangle\},
\end{aligned}$$

where in (a) we have written the constraint into the objective function by involving its optimal dual variable  $\mu^*$ ; in (b) we have utilized the facts that  $\mathbf{w}_{(k+1)} = \mathbf{w}_{(k)} = \mathbf{v}_{(k)}$  at any limit point, and  $\frac{\rho}{2} \|\mathbf{u}_{(k)}\|_2^2$  is a scalar term unrelated to  $\mathbf{w}$ ; in (c) we have used  $\rho \mathbf{u}_{(k)} = \rho \mathbf{u}_{(k+1)} = \mathbf{y}_{(k+1)}$  and  $\mathbf{v}_{(k)} = \mathbf{v}_{(k+1)}$ . This completes the proof of Theorem 2.

## APPENDIX F PROOF OF LEMMA 3

First of all, by introducing an auxiliary variable into the original Problem (9), the problem is equivalently rewritten as

$$\min_{\mathbf{w}, \mathbf{v}} \mathbf{v}^H \mathbf{R}_x \mathbf{v} \quad (45a)$$

$$\text{subject to } \begin{cases} |\mathbf{w}^H \mathbf{a}_0|^2 \geq 1, \\ \|\mathbf{w}\|_0 = L, \\ \mathbf{w} = \mathbf{v}. \end{cases} \quad (45b)$$

The Lagrangian function of the above problem is given by

$$\mathbf{v}^H \mathbf{R}_x \mathbf{v} + \mu(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \nu(\|\mathbf{w}\|_0 - L) + \Re\{\langle \mathbf{y}, \mathbf{w} - \mathbf{v} \rangle\}, \quad (46)$$

where  $\mu$ ,  $\nu$ , and  $\mathbf{y}$  are Lagrangian dual variables corresponding to the three constraints, respectively. A KKT point  $(\mathbf{w}^*, \mathbf{v}^*)$  of Problem (45), together with the corresponding dual variables  $\mu^*$ ,  $\nu^*$ , and  $\mathbf{y}^*$ , satisfies

$$\mathbf{0} = 2\mathbf{R}_x\mathbf{v}^* - \mathbf{y}^*, \quad (47a)$$

$$\mathbf{w}^* \in \arg \min_{\mathbf{w}} \left\{ \begin{array}{l} \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \nu^*(\|\mathbf{w}\|_0 - L) \\ + \Re\{\langle \mathbf{y}^*, \mathbf{w} - \mathbf{v}^* \rangle\} \end{array} \right\}, \quad (47b)$$

$$\mathbf{w}^* = \mathbf{v}^*. \quad (47c)$$

Our aim is to show any limit point of Algorithm 1, referred to as  $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$ , satisfies (47). As has been proven,  $(\mathbf{w}_{(k+1)}, \mathbf{v}_{(k+1)}, \mathbf{u}_{(k+1)})$  satisfies (43). Since (47a) and (47c) are the same as (43a) and (43c), respectively. The limit point satisfies (47a) and (47c) due to the same reasons presented for (43a) and (43c), see Appendix E. Our focus is then to show the limit point satisfies (47b). We have

$$\begin{aligned} \mathbf{w}_{(k+1)} &\stackrel{(a)}{=} \arg \min_{\mathbf{w}} \left\{ \begin{array}{l} \lambda \|\mathbf{w}\|_1 + \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) \\ + \Re\{\langle \mathbf{y}_{(k+1)}, \mathbf{w} - \mathbf{v}_{(k+1)} \rangle\} \end{array} \right\} \\ &\stackrel{(b)}{=} \arg \min_{\mathbf{w}} \left\{ \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \Re\{\langle \mathbf{y}_{(k+1)}, \mathbf{w} - \mathbf{v}_{(k+1)} \rangle\} \right\} \\ &\quad \text{s.t. } \|\mathbf{w}\|_0 = L \\ &\stackrel{(c)}{=} \arg \min_{\mathbf{w}} \left\{ \begin{array}{l} \mu^*(|\mathbf{w}^H \mathbf{a}_0|^2 - 1) + \nu^*(\|\mathbf{w}\|_0 - L) \\ + \Re\{\langle \mathbf{y}^*, \mathbf{w} - \mathbf{v}^* \rangle\} \end{array} \right\}, \end{aligned}$$

where in (a) we have used the fact that the limit point satisfies (43b); in (b) we have used the fact that under some conditions, such as restricted isometry property, and by carefully tuning  $\lambda$ , the solution of (11) could be the same as that of Problem (9) [64], [65]; and in (c) we have included the constraint into the objective function by involving its optimal dual variable  $\nu^*$ . The above equation indicates that the limit point satisfies (47b). This completes the proof of Lemma 3.

## REFERENCES

- [1] H. Huang, H. C. So, and A. M. Zoubir, "Sparse array beamformer design via ADMM," in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, Norway, June 2022, pp. 336–340.
- [2] W. F. Gabriel, "Adaptive arrays - An introduction," *Proceedings of the IEEE*, vol. 64, no. 2, pp. 239–272, February 1976.
- [3] S. A. Vorobyov, "Chapter 12 - Adaptive and robust beamforming," in *Academic Press Library in Signal Processing: Volume 3*, ser. Academic Press Library in Signal Processing, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds., Elsevier, 2014, vol. 3, pp. 503–552.
- [4] H.-C. Lin, "Spatial correlations in adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 2, pp. 212–223, March 1982.
- [5] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based rf-baseband codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, November 2005.
- [6] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3501–3511, December 2006.
- [7] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, August 2010.
- [8] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.
- [9] M. B. Hawes and W. Liu, "Location optimization of robust sparse antenna arrays with physical size constraint," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 1303–1306, November 2012.
- [10] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1099–1112, February 2018.
- [11] W. Shi, Y. Li, L. Zhao, and X. Liu, "Controllable sparse antenna array for adaptive beamforming," *IEEE Access*, vol. 7, pp. 6412–6423, January 2019.
- [12] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.
- [13] X. Wang, E. Abutianios, M. Trinkle, and M. G. Amin, "Reconfigurable adaptive array beamforming by antenna selection," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2385–2396, May 2014.
- [14] X. Wang, M. S. Greco, and F. Gini, "Adaptive sparse array beamformer design by regularized complementary antenna switching," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2302–2315, March 2021.
- [15] X. Wang, W. Zhai, M. S. Greco, and F. Gini, "Cognitive sparse beamformer design in dynamic environment via regularized switching network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 2, pp. 1816–1833, April 2023.
- [16] S. Shrestha, X. Fu, and M. Hong, "Optimal solutions for joint beamforming and antenna selection: From branch and bound to graph neural imitation learning," *IEEE Transactions on Signal Processing*, vol. 71, pp. 831–846, February 2023.
- [17] S. A. Hamza and M. G. Amin, "Hybrid sparse array beamforming design for general rank signal models," *IEEE Transactions on Signal Processing*, vol. 67, no. 24, pp. 6215–6226, December 2019.
- [18] S. A. Hamza and M. G. Amin, "Sparse array beamforming design for wideband signal models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 2, pp. 1211–1226, April 2021.
- [19] G. Hegde, C. Masouros, and M. Pesavento, "Analog beamformer design for interference exploitation based hybrid beamforming," in *Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Sheffield, UK, July 2018, pp. 109–113.
- [20] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1677–1688, March 2020.
- [21] K. Diamantaras, Z. Xu, and A. Petropulu, "Sparse antenna array design for MIMO radar using softmax selection," *arXiv*, February 2021. [Online]. Available: <https://arxiv.org/abs/2102.05092>
- [22] Z. Xu, F. Liu, K. Diamantaras, C. Masouros, and A. Petropulu, "Learning to select for MIMO radar based on hybrid analog-digital beamforming," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 8228–8232.
- [23] T. X. Vu, S. Chatzinotas, V.-D. Nguyen, D. T. Hoang, D. N. Nguyen, M. D. Renzo, and B. Ottersten, "Machine learning-enabled joint antenna selection and precoding design: From offline complexity to online performance," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3710–3722, June 2021.
- [24] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3972–3987, August 2013.
- [25] T. Fischer, G. Hegde, F. Matter, M. Pesavento, M. E. Pfetsch, and A. M. Tillmann, "Joint antenna selection and phase-only beamforming using mixed-integer nonlinear programming," in *Proceedings of International ITG Workshop on Smart Antennas (WSA)*, Bochum, Germany, March 2018, pp. 1–7.
- [26] H. Li, Q. Liu, Z. Wang, and M. Li, "Transmit antenna selection and analog beamforming with low-resolution phase shifters in mmWave MISO systems," *IEEE Communications Letters*, vol. 22, no. 9, pp. 1878–1881, September 2018.
- [27] S. E. Nai, W. Ser, Z. L. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 12, pp. 3923–3930, 2010.
- [28] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Multicast beamforming with antenna selection," in *Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cesme, Turkey, September 2012, pp. 70–74.
- [29] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2660–2674, May 2013.

- [30] M. Gkizeli and G. N. Karystinos, "Maximum-SNR antenna selection among a large number of transmit antennas," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 891–901, October 2014.
- [31] M. B. Hawes and W. Liu, "Sparse array design for wideband beamforming with reduced complexity in tapped delay-lines," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 8, pp. 1236–1247, August 2014.
- [32] O. T. Demir and T. E. Tuncer, "Multicast beamforming with antenna selection using exact penalty approach," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Australia, April 2015, pp. 2489–2493.
- [33] Z. Zheng, Y. Fu, W.-Q. Wang, and H. C. So, "Sparse array design for adaptive beamforming via semidefinite relaxation," *IEEE Signal Processing Letters*, vol. 27, pp. 925–929, May 2020.
- [34] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, "Joint transceiver design with antenna selection for large-scale MU-MIMO mmWave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2085–2096, 2017.
- [35] A. Arora, C. G. Tsinos, B. Shankar Mysore R, S. Chatzinotas, and B. Ottersten, "Analog beamforming with antenna selection for large-scale antenna arrays," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 4795–4799.
- [36] Z. Zheng, Y. Fu, and W.-Q. Wang, "Sparse array beamforming design for coherently distributed sources," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 5, pp. 2628–2636, May 2021.
- [37] H. Yu, W. Qu, Y. Fu, C. Jiang, and Y. Zhao, "A novel two-stage beam selection algorithm in mmWave hybrid beamforming system," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1089–1092, June 2019.
- [38] Z. Chen, X. Zhang, S. Wang, Y. Xu, J. Xiong, and X. Wang, "Enabling practical large-scale MIMO in WLANs with hybrid beamforming," *IEEE/ACM Transactions on Networking*, vol. 29, no. 4, pp. 1605–1619, August 2021.
- [39] J. Dan, S. Geirnaert, and A. Bertrand, "Grouped variable selection for generalized eigenvalue problems," *Signal Processing*, vol. 195, p. 10847, June 2022.
- [40] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 804–808, July 2015.
- [41] K. Huang and N. D. Sidiropoulos, "Consensus-ADMM for general quadratically constrained quadratic programming," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5297–5310, October 2016.
- [42] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, June 2006.
- [43] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.
- [44] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, August 1969.
- [45] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2257–2269, September 2003.
- [46] J. Park and S. Boyd, "General heuristics for nonconvex quadratically constrained quadratic programming," *arXiv*, May 2017. [Online]. Available: <https://arxiv.org/abs/1703.07870>
- [47] H. Huang, H. C. So, and A. M. Zoubir, "Convergence analysis of consensus-ADMM for general QCQP," *Signal Processing*, vol. 208, p. 108991, July 2023.
- [48] S. Boyd and L. Vandenberghe, "Appendix B: Problems involving two quadratic functions," in *Convex Optimization*. Cambridge university press, 2004, pp. 653–659.
- [49] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundation and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, January 2011.
- [50] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.
- [51] Z. Cheng, Z. He, and B. Liao, "Hybrid beamforming design for OFDM dual-function radar-communication system," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1455–1467, November 2021.
- [52] Z. Cheng and B. Liao, "QoS-aware hybrid beamforming and DOA estimation in multi-carrier dual-function radar-communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1890–1905, March 2022.
- [53] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2755–2770, April 2018.
- [54] T. Wei, L. Wu, and B. M. R. Shankar, "Sparse array beampattern synthesis via majorization-based ADMM," in *Proceedings of IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, Norman, USA, September 2021, pp. 1–5.
- [55] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2685–2698, June 2017.
- [56] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, "Chapter 3 - Robust penalized regression in the linear model," in *Robust Statistics for Signal Processing*. Cambridge University Press, 2018, pp. 74–99.
- [57] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, December 2008.
- [58] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," March 2014. [Online]. Available: <http://cvxr.com/cvx>
- [59] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, August 2010.
- [60] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, February 2011.
- [61] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, March 2015.
- [62] S. Boyd and L. Vandenberghe, "Chapter 5 - Duality," in *Convex Optimization*. Cambridge university press, 2004, pp. 215–273.
- [63] E. K. Ryu and S. P. Boyd, "A primer on monotone operator methods," *Applied and Computational Mathematics*, vol. 15, no. 1, pp. 3–43, 2016.
- [64] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [65] S. Foucart and H. Rauhut, "Chapter 3 - Basic algorithms," in *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer New York, 2013, pp. 61–75.