

Dual-Domain Adaptive Beamformer Under Linearly and Quadratically Constrained Minimum Variance

Masahiro Yukawa, *Member, IEEE*, Youngchul Sung, *Senior Member, IEEE*, and Gilwon Lee, *Student Member, IEEE*

Abstract—In this paper, a novel adaptive beamforming algorithm is proposed under a linearly and quadratically constrained minimum variance (LQCMV) beamforming framework, based on a dual-domain projection approach that can efficiently implement a quadratic-inequality constraint with a possibly rank-deficient positive semi-definite matrix, and the properties of the proposed algorithm are analyzed. As an application, relaxed zero-forcing (RZF) beamforming is presented which adopts a specific quadratic constraint that bounds the power of residual interference in the beamformer output with the aid of interference-channel side-information available typically in wireless multiple-access systems. The dual-domain projection in this case plays a role in guiding the adaptive algorithm towards a better direction to minimize the interference and noise, leading to considerably faster convergence. The robustness issue against channel mismatch and ill-posedness is also addressed. Numerical examples show that the efficient use of interference side-information brings considerable gains.

Index Terms—Adaptive beamforming, dual-domain adaptive algorithm, LCMV, LQCMV, relaxed zero forcing.

I. INTRODUCTION

ADAPTIVE beamforming and smart antennas are gaining renewed interest recently for wireless communications as an effective tool to enhance the energy efficiency and data rate far beyond those achieved by current small-scale multiple-input multiple-output (MIMO) techniques [3] quite after their early consideration in 1990's [4]. In upcoming wireless communication systems, *large-scale* antenna arrays are envisioned to be deployed at basestations and beamforming based on large-scale antenna arrays is expected to be used to support high data rates and energy efficiency to the network (see Fig. 1). In typical wireless multiple-access situations, the channel (or array response) from every uplink transmitter to the basestation antenna array is known to the basestation. In order to decode the signal from one

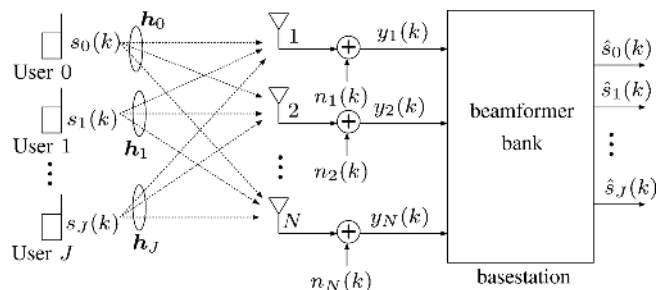


Fig. 1. System model: A single-input multiple-output (SIMO) multiple-access wireless network with a large-scale antenna array.

uplink transmitter at the basestation, one could use Capon's minimum variance distortionless response (MVDR) beamformer [5]. In this case, the channel information from the desired transmitter is used, whereas the channel information from the remaining interfering transmitters is ignored. However, it is natural to ask if there is a benefit in using the side information of the interference channels in beamforming.

In this paper, we propose a new paradigm for adaptive beamforming that can exploit the interference side-information efficiently, having the wireless multiple-access communication scenario in mind. It turns out that the use of the side information brings considerable improvements in the convergence speed of adaptive algorithms. We emphasize here that fast convergence is of great importance, particularly in the wireless multiple-access communication scenario, since it reduces the preamble size and thereby brings high data rate. As the first step to build the new paradigm, we formulate the beamforming task as a quadratic minimization problem under linear-equality and quadratic-inequality constraints. We name it *linearly and quadratically constrained minimum variance (LQCMV) beamforming*. The LQCMV beamformer can be regarded as a generalization of several existing beamformers: the MVDR beamformer, the zero-forcing (ZF) beamformer, the linearly constrained minimum variance (LCMV) [6], and the linearly and norm constrained minimum variance (LNCMV) beamformer [7]–[9]. (See [9], [10] for more extensive references on beamforming.) In a large-scale antenna array, the number of simultaneous transmitters is typically less than that of antenna elements, and it makes the function determining the quadratic constraint *not strictly convex*. Unfortunately, this important case in general LQCMV beamforming cannot be treated efficiently with the existing adaptive algorithms to implement the LNCMV beamformer (see Section II).

We here present an efficient adaptive implementation of the LQCMV beamformer based on the multi-domain adaptive learning method [11], in particular *the dual-domain adaptive*

Manuscript received September 14, 2012; revised January 12, 2013; accepted March 01, 2013. Date of publication March 22, 2013; date of current version May 09, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chong-Yung Chi. This work was partially supported by JSPS Grants-in-Aid (24760292) and Basic Science Research Program. This work was also supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0021269). Preliminary versions of this work were presented at SSP, Nice, France, June 2011 [1], and ITC-CSCC, Gyeongju, Korea, June 2011 [2].

M. Yukawa is with the Department of Electronics and Electrical Engineering, Keio University, Japan (e-mail: yukawa@elec.keio.ac.jp).

Y. Sung and G. Lee are with the Department of Electrical Engineering, KAIST, Yuseong-gu, Daejeon, 305-701, Republic of Korea (e-mail: ysung@ee.kaist.ac.kr; gwlee@kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2254481

algorithm (DDAA). The proposed DDAA is a *set-theoretic adaptive filtering approach* with a pair of closed convex sets defined in different domains. In a primary domain, a linear variety is defined with instantaneous data (which contain partial information about the data covariance matrix) as well as the linear constraint. It is the set of all vectors that make the instantaneous output power be zero as well as satisfying the linear-equality constraint. In a dual domain, on the other hand, a closed ball is defined which accommodates the quadratic constraint. The proposed DDAA updates the beamforming weight vector as follows. The current weight vector is projected onto the linear variety. (Metric projections onto convex sets have been the backbone of adaptive algorithms [12], [13].) However, the direction vector of this projection is determined by partial information about the data covariance matrix and could therefore be improved by the use of side information. To do so, the direction vector is combined convexly with an auxiliary vector which guides the adaptive algorithm towards a better direction. The auxiliary vector is obtained as follows: (i) transform the current weight vector to the dual domain, (ii) project the transformed weight vector onto the closed ball, (iii) embed the direction vector of the dual-domain projection back into the primary domain, and finally (iv) project the embedded vector in the primary domain onto the underlying subspace of the linear constraint set. These steps involve only linear operations and simple scaling, and the final step is for ensuring the linear constraint to be satisfied at each iteration. The dual-domain projection is useful because the primary-domain projection onto the quadratic constraint set has no closed form due to its ellipsoidal shape in the primary domain. Furthermore, the proposed DDAA has particular advantages for large-scale antenna array systems because its computational complexity increases only linearly as the array size increases. As will be seen in Section III, the dual-domain concept simplifies adaptive implementation of quadratic-inequality constraints, and also facilitates understanding of the overall implementation.

A convergence analysis is presented to support the proposed DDAA theoretically under the mathematically-rigorous framework of the adaptive projected subgradient method (APSM) [12]. Although an analysis of the multi-domain adaptive learning method is given in [11], there are two points to be considered for mathematical rigor here: (i) the signals under study are complex-valued and (ii) the linear constraint set has no interior. A proof is given based on (i) Wirtinger calculus [14] and (ii) a simple translation idea presented in [15]. The analysis clarifies that, under certain mild assumptions, the weight-vector sequence is convergent to a point satisfying both the linear and quadratic constraints.

As a specific application example of LQCMV, we present a minimum-variance beamformer under a distortionless constraint and a quadratic constraint that bounds the power of residual interference in the beamformer output. Compared to the zero-forcing (ZF) beamformer, the interference constraint is relaxed to prevent the noise-enhancement problem. The relaxed zero-forcing (RZF) beamformer bridges the MVDR and ZF beamformers. Simulation results show that, thanks to the efficient use of interference side-information, the RZF beamformer implemented with the DDAA significantly out-

performs the MVDR and ZF beamformers implemented with the widely-used constrained normalized LMS (CNLMS) algorithm [16].

The remainder of this paper is organized as follows. In Section II, the LQCMV framework is presented after the introduction of the data model. Section III presents the adaptive implementation of the LQCMV beamformer based on the dual-domain approach and an analysis on its properties. Section IV presents the RZF beamformer and discusses the robustness issue against channel mismatch and ill-posedness. Numerical results are provided in Section V, followed by the conclusion in Section VI.

II. LINEARLY AND QUADRATICALLY CONSTRAINED MINIMUM VARIANCE BEAMFORMER

In this paper, we will make use of standard notational conventions. \mathbb{R} , \mathbb{C} and \mathbb{N} denote the sets of all real numbers, complex numbers and nonnegative integers, respectively. Vectors and matrices are written in boldface with matrices in capitals. In particular, the identity and zero matrices of any size are denoted by \mathbf{I} and \mathbf{O} , respectively, and the zero vector of any length by $\mathbf{0}$. For a matrix \mathbf{A} , \mathbf{A}^\top and \mathbf{A}^H denote the transpose and conjugate transpose of \mathbf{A} , respectively. For any $n \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$, define the inner product $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^H \mathbf{y}$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, and its induced norm $\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}$, $\forall \mathbf{x} \in \mathbb{C}^n$. $\mathcal{R}(\mathbf{A})$ and $\mathcal{N}(\mathbf{A})$ stand for the range space and the null space of \mathbf{A} , respectively. For a random matrix \mathbf{X} , $E\{\mathbf{X}\}$ denotes the expectation of \mathbf{X} . For complex quantity c , $\Re(c)$ denotes the real part of c .

We consider a network of $J + 1$ transmitters (each equipped with a single transmit antenna) and a receiver equipped with an array of N receive antennas, as shown in Fig. 1. The received signal vector $\mathbf{y}(k) := [y_1(k), y_2(k), \dots, y_N(k)]^\top \in \mathbb{C}^N$ at time instant $k \in \mathbb{N}$ is given by [9]

$$\mathbf{y}(k) = \sum_{j=0}^J \mathbf{h}_j s_j(k) + \mathbf{n}(k), \quad k \in \mathbb{N}, \quad (1)$$

where, for each $j \in \{0, 1, \dots, J\}$, $(s_j(k))_{k \in \mathbb{N}} \subset \mathbb{C}$ is the symbol sequence of the j -th transmitter, $\mathbf{h}_j := [h_{j1}, h_{j2}, \dots, h_{jN}]^\top \in \mathbb{C}^N$ is the channel vector (referred to also as the array response vector) of the j -th signal, and $\mathbf{n}(k) := [n_1(k), n_2(k), \dots, n_N(k)]^\top \in \mathbb{C}^N$ is the noise vector at time instant k with a positive-definite covariance matrix. Here, we assume that the received signal vector $\mathbf{y}(k)$ has zero mean without loss of generality. The covariance matrix of $\mathbf{y}(k)$ is defined as $\mathbf{R} := E\{\mathbf{y}(k)\mathbf{y}^H(k)\} \in \mathbb{C}^{N \times N}$, which is positive definite due to the assumption on the noise covariance-matrix. In narrowband beamforming, the output $z(k)$ of a beamformer $\mathbf{w} := [w_1, w_2, \dots, w_N]^\top \in \mathbb{C}^N$ is given by $z(k) = \langle \mathbf{y}(k), \mathbf{w} \rangle$, $k \in \mathbb{N}$, and the output variance is given by $E\{|z(k)|^2\} = \mathbf{w}^H \mathbf{R} \mathbf{w}$.

Let $\mathbf{A} \in \mathbb{C}^{N \times L}$, $\mathbf{B} \in \mathbb{C}^{N \times M}$, $\mathbf{c} \in \mathbb{C}^L$, and $\epsilon \in [0, \infty)$, where $L \leq N$ and $M \leq N$. Without loss of generality, we assume that (i) the matrices \mathbf{A} and \mathbf{B} have full column-rank and (ii) the maximum singular value of \mathbf{B} is one. (Otherwise, the known \mathbf{B} can be normalized.) The choices of \mathbf{A} and \mathbf{B} depend on applications; one specific design of \mathbf{A} and \mathbf{B} is given in Section IV.

We now formulate the general LQCMV beamforming problem as follows:

$$\text{minimize } \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (2)$$

$$\text{subject to } \mathbf{A}^H \mathbf{w} = \mathbf{c} \quad (3)$$

$$\|\mathbf{B}^H \mathbf{w}\|^2 = \mathbf{w}^H (\mathbf{B} \mathbf{B}^H) \mathbf{w} \leq \epsilon. \quad (4)$$

Then, its solution can easily be obtained by the Karush-Kuhn-Tucker (KKT) conditions [17] and is given by [2]

$$\mathbf{w}_{\text{LQCMV}} := \mathbf{R}_\epsilon^{-1} \mathbf{A} \left(\mathbf{A}^H \mathbf{R}_\epsilon^{-1} \mathbf{A} \right)^{-1} \mathbf{c}, \quad (5)$$

where $\mathbf{R}_\epsilon := \mathbf{R} + \lambda_\epsilon \mathbf{B} \mathbf{B}^H$ and $\lambda_\epsilon \geq 0$ is the Lagrange multiplier associated with the quadratic constraint (4) depending on ϵ . If ϵ is sufficiently large so that $\|\mathbf{B}^H \mathbf{w}_{\text{LCMV}}\|^2 \leq \epsilon$ for the LCMV beamformer $\mathbf{w}_{\text{LCMV}} := \mathbf{R}^{-1} \mathbf{A} (\mathbf{A}^H \mathbf{R}^{-1} \mathbf{A})^{-1} \mathbf{c}$ [6], then it is clear that $\lambda_\epsilon = 0$ and $\mathbf{w}_{\text{LQCMV}} = \mathbf{w}_{\text{LCMV}}$. That is, when the quadratic constraint (4) is loose, the LQCMV solution is simply the LCMV solution of Frost [6]. If $\|\mathbf{B}^H \mathbf{w}_{\text{LCMV}}\|^2 > \epsilon$, on the other hand, by the KKT conditions for the convex optimization problem (2)–(4), λ_ϵ should satisfy the following equality:

$$\|\mathbf{B}^H \mathbf{w}_{\text{LQCMV}}\|^2 = \epsilon. \quad (6)$$

In this case, the linear constraint (3) is trivially satisfied by substituting (5) into (3), and the optimal λ_ϵ can be obtained by solving a nonlinear equation obtained by substituting (5) into (6) or by a semi-analytic method. Whereas the former method is not trivial, the latter semi-analytic method for obtaining λ_ϵ is simple to implement. If \mathbf{R} and a feasible range of λ_ϵ are given, $\mathbf{w}_{\text{LQCMV}}$ can be computed from (5) and λ_ϵ can be computed from (5) and (6) by a numerical method such as the bisection method. However, such a batch processing method requires estimation of \mathbf{R} and this makes the semi-analytic method less attractive from the perspective of complexity. (See Section V-A.) Fortunately, our adaptive algorithm presented in Section III does not need any explicit computation of λ_ϵ or \mathbf{R} .

The case of $\mathbf{B} = \mathbf{I}$ has been studied in [7], [8] for robust LCMV adaptive beamforming. In this case, the constraint set associated with the quadratic inequality (4) is a closed ball. Because the projection onto a closed ball is simple scaling, the constraint can easily be incorporated in adaptive algorithms. Also, the case that \mathbf{B} is not the identity matrix, but a nonsingular square-matrix, has been considered in [7]. It has been shown that the same procedure as in the case of $\mathbf{B} = \mathbf{I}$ can be applied by defining a new variable vector, say $\tilde{\mathbf{w}} := \mathbf{B}^H \mathbf{w}$, and by changing \mathbf{R} and \mathbf{A} accordingly. Through the change of variables, the ellipsoidal constraint set is changed into a closed ball and the same procedure can be applied. However, it cannot be applied to the case of $M < N$ (i.e., the case that \mathbf{B} is a tall matrix), which is of particular interest in this paper. This is because the degree of freedom is lost due to the (non-invertible) variable transformation. In the tall-matrix case, the positive semi-definite matrix $\mathbf{B} \mathbf{B}^H$ is rank deficient, and thus the constraint set associated with the quadratic inequality (4) is no longer a hyper-ellipsoid

in the entire space, although it is so in the subspace $\mathcal{R}(\mathbf{B})$. In the following section, we will present an adaptive algorithm to efficiently incorporate the quadratic constraint (4) with a tall \mathbf{B} .

Algorithm 1 (The Dual-Domain Adaptive Algorithm (DDAA) for the LQCMV Beamformer)

Requirements:

- step size $\lambda_k \in (0, 2)$
- weight $\alpha_k \in [0, 1], \forall k \in \mathbb{N}$

Algorithm: Given any initial beamformer $\mathbf{w}_0 \in K$, generate the sequence of beamformers $(\mathbf{w}_k)_{k \in \mathbb{N}} \subset K$ as follows:

$$\begin{aligned} \mathbf{w}_{k+1} &:= P_K \left[\mathbf{w}_k + \lambda_k \mu_k \left(\alpha_k \mathbf{g}_k^{(1)} + (1 - \alpha_k) \mathbf{B} \mathbf{g}_k^{(2)} \right) \right] \quad (10) \\ &= P_K (\mathbf{w}_k + \lambda_k \mu_k \mathbf{g}_k) \\ &= \mathbf{w}_k + \lambda_k \mu_k \mathbf{g}_k, \quad k \in \mathbb{N}, \quad (11) \end{aligned}$$

where

$$\mathbf{g}_k^{(1)} := P_{V_k} (\mathbf{w}_k) - \mathbf{w}_k \in \mathbb{C}^N, \quad (12)$$

$$\mathbf{g}_k^{(2)} := P_{B_\epsilon} (\mathbf{B}^H \mathbf{w}_k) - \mathbf{B}^H \mathbf{w}_k \in \mathbb{C}^M, \quad (13)$$

$$\mathbf{g}_k := \alpha_k \mathbf{g}_k^{(1)} + (1 - \alpha_k) \mathbf{Q} \mathbf{B} \mathbf{g}_k^{(2)} \in \mathbb{C}^N, \quad (14)$$

$$\mu_k := \begin{cases} \frac{\alpha_k \|\mathbf{g}_k^{(1)}\|^2 + (1 - \alpha_k) \|\mathbf{g}_k^{(2)}\|^2}{\|\mathbf{g}_k\|^2}, & \text{if } \mathbf{g}_k \neq \mathbf{0}, \\ 1, & \text{otherwise.} \end{cases} \quad (15)$$

Here, $\mathbf{Q} := \mathbf{I} - \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ and $P_C(\mathbf{x})$ is the metric projection¹ of \mathbf{x} onto C . The related projections can be computed by the following formulae:

$$P_K(\mathbf{w}) = \mathbf{Q} \mathbf{w} + \mathbf{v}, \quad \mathbf{w} \in \mathbb{C}^N, \quad (16)$$

$$P_{V_k}(\mathbf{w}) = \mathbf{w} - \frac{\langle \mathbf{y}(k), \mathbf{w} \rangle}{\mathbf{y}^H(k) \mathbf{Q} \mathbf{y}(k)} \mathbf{Q} \mathbf{y}(k), \quad \mathbf{w} \in \mathbb{C}^N, \quad (17)$$

$$P_{B_\epsilon}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \text{if } \|\mathbf{x}\|^2 \leq \epsilon, \\ \frac{\sqrt{\epsilon}}{\|\mathbf{x}\|} \mathbf{x}, & \text{otherwise,} \end{cases} \quad \mathbf{x} \in \mathbb{C}^M, \quad (18)$$

where $\mathbf{v} := (\mathbf{A}^H)^\dagger \mathbf{c}$. Here, $(\mathbf{A}^H)^\dagger$ denotes the Moore-Penrose pseudo-inverse of \mathbf{A}^H .

III. ADAPTIVE IMPLEMENTATION BY A DUAL-DOMAIN APPROACH

A. The Proposed Dual-Domain Adaptive Algorithm

We define three closed convex sets that accommodate information about the LQCMV optimization (2)–(4):

$$K := \{\mathbf{w} \in \mathbb{C}^N : \mathbf{A}^H \mathbf{w} = \mathbf{c}\}, \quad (7)$$

$$B_\epsilon := \{\mathbf{x} \in \mathbb{C}^M : \|\mathbf{x}\|^2 \leq \epsilon\}, \quad (8)$$

$$V_k := K \cap \{\mathbf{w} \in \mathbb{C}^N : \langle \mathbf{y}(k), \mathbf{w} \rangle = 0\}, \quad k \in \mathbb{N}. \quad (9)$$

Here, K is the set of weight vectors satisfying the linear constraint (3), B_ϵ is the origin-centered closed ball with radius $\sqrt{\epsilon}$

¹The metric projection of an arbitrary point $\mathbf{x} \in \mathbb{C}^N$ onto a closed convex set $C \subset \mathbb{C}^N$ is defined as $P_C(\mathbf{x}) := \arg \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$. If C is closed and convex, the existence and uniqueness of $P_C(\mathbf{x})$ for any \mathbf{x} are guaranteed. Note that all the sets K , B_ϵ and V_k are closed and convex.

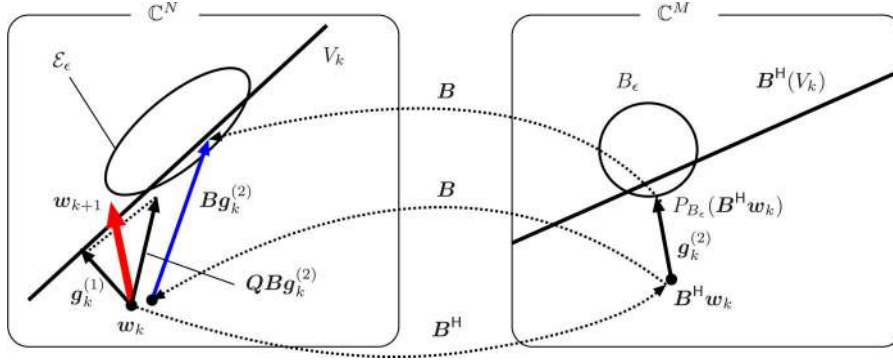


Fig. 2. A geometric interpretation of the dual-domain adaptive beamforming algorithm.

in the dual domain, and V_k is the set of weight vectors contained in K satisfying the zero instantaneous-output-power condition. Note that the set V_k , which is assumed to be nonempty, is a linear variety and accommodates only a piece of information about the true data covariance matrix \mathbf{R} . This is because in adaptive scenarios \mathbf{R} is unavailable, and therefore the beamforming vector \mathbf{w}_k at time instant k needs to be adapted only with available information, which is $\mathbf{y}(k)$ in this case. Now we present the proposed algorithm to implement the LQCMV beamformer described in (2)–(4) in Algorithm 1.

The equalities in (11) are verified by noting that $P_K(\mathbf{w}_k) = \mathbf{w}_k$ and $\mathbf{g}_k^{(1)} = \mathbf{Q}\mathbf{g}_k^{(1)}$. (This is because $\mathbf{w}_k \in K$ and $P_{V_k}(\mathbf{w}_k) \in K$.) Although the three expressions in (10) and (11) are equivalent mathematically, one could use the first or second expression to avoid the accumulation of numerical errors. In the case of $\alpha_k = 1$, the algorithm reduces to the constrained normalized LMS (CNLMS) algorithm [16], which is a normalized version of Frost's CLMS algorithm [6]. When $0 < \alpha_k < 1$, on the other hand, the proposed algorithm includes the movement in a direction for satisfying the quadratic constraint (in addition to the linear constraints), and this direction is obtained efficiently by the dual and primary transforms and simple scaling. The proposed algorithm is thus a generalization of the C(N)LMS algorithm that incorporates the general quadratic constraint in which the matrix $\mathbf{B}\mathbf{B}^H$ in the quadratic form has an arbitrary rank.

A geometric interpretation of the proposed DDAA (11) is given in Fig. 2. Note that $\mathbf{g}_k^{(1)}$ is the direction vector from the current weight vector $\mathbf{w}_k \in K$ to its projection onto V_k which is the intersection of the linear constraint set K and the hyperplane of zero instantaneous-output-power in the primary domain \mathbb{C}^N . Therefore, if $\mathbf{g}_k^{(1)}$ were added to \mathbf{w}_k , the resulting vector would minimize the instantaneous output power to zero without violating the linear constraint. On the other hand, $\mathbf{g}_k^{(2)}$ exists in the dual domain \mathbb{C}^M . It is the direction vector from $\mathbf{B}^H \mathbf{w}_k$ (the image of \mathbf{w}_k under the dual transform \mathbf{B}^H) to $P_{B_\epsilon}(\mathbf{B}^H \mathbf{w}_k)$ (its projection onto the $\sqrt{\epsilon}$ ball in the dual domain). Thus, $\mathbf{B}\mathbf{g}_k^{(2)}$ is the embedding of $\mathbf{g}_k^{(2)}$ back into the primary domain by the primary transform \mathbf{B} . Finally, $\mathbf{B}\mathbf{g}_k^{(2)}$ is multiplied by the matrix \mathbf{Q} so that the addition of $\mathbf{Q}\mathbf{B}\mathbf{g}_k^{(2)}$ to \mathbf{w}_k will not violate the linear constraint. The key property of the dual-domain projection is given in the following lemma.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE DDAA AND THE CNLMS

	initial cost	per-iteration complexity
DDAA	$O(M^2 N) + L^2 N + O(L^3)$	$(6L + 2M + 5)N + 3L^2 + M + 4$
CNLMS	0	$5N$

Lemma 1: If the maximum singular value of \mathbf{B} is less than or equal to one, we have

$$\mathbf{B}P_{B_\epsilon}(\mathbf{B}^H \mathbf{w}_k) \in \mathcal{E}_\epsilon := (\mathbf{B}^H)^{-1}(B_\epsilon) := \{\mathbf{w} \in \mathbb{C}^N : \mathbf{B}^H \mathbf{w} \in B_\epsilon\},$$

where \mathcal{E}_ϵ is the inverse image of the closed ball B_ϵ under the linear operator $\mathbf{B}^H : \mathbb{C}^N \rightarrow \mathbb{C}^M$ and is indeed the set of weight vectors satisfying the quadratic constraint (4).

Proof: We have $\mathbf{B}^H \mathbf{B}(B_\epsilon) := \{\mathbf{B}^H \mathbf{B}\mathbf{x} : \mathbf{x} \in B_\epsilon\} \subset B_\epsilon$ since $\|\mathbf{B}^H \mathbf{B}\mathbf{x}\|_2 \leq \|\mathbf{B}^H \mathbf{B}\|_2 \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_2 \leq \sqrt{\epsilon}$ for any $\mathbf{x} \in B_\epsilon$. This verifies the claim. (Here, $\|\mathbf{B}^H \mathbf{B}\|_2 \leq 1$ by the singular value condition.) \square

Lemma 1 states that the end point $\mathbf{B}P_{B_\epsilon}(\mathbf{B}^H \mathbf{w}_k)$ of $\mathbf{B}\mathbf{g}_k^{(2)}$ satisfies the quadratic constraint. Hence, if the starting point $\mathbf{B}\mathbf{B}^H \mathbf{w}_k$ of $\mathbf{B}\mathbf{g}_k^{(2)}$ is close to \mathbf{w}_k , $\mathbf{Q}\mathbf{B}\mathbf{g}_k^{(2)}$ is likely to give a rough direction towards the quadratic constraint set \mathcal{E}_ϵ . We convexly combine the two direction vectors $\mathbf{g}_k^{(1)}$ (for minimizing the overall output power based on partial information about the covariance matrix) and $\mathbf{Q}\mathbf{B}\mathbf{g}_k^{(2)}$ (for maintaining $\|\mathbf{B}^H \mathbf{w}_k\|$ below a certain level). The resulting direction \mathbf{g}_k therefore gives a better direction to minimize the interference and noise, yielding fast convergence of the proposed DDAA. The position of \mathbf{w}_{k+1} is determined by the step size λ_k after proper length normalization by μ_k (see Appendix II for an explanation of μ_k). Fast convergence of the proposed algorithm will be shown by simulations in Section V. The following subsection will present a rigorous analysis that provides important properties of the proposed algorithm.

Computational complexity. The computational complexity of the DDAA and the CNLMS is given in Table I. Here, the complexity is measured in terms of the number of complex multiplications. The initial cost for the DDAA is for computing $(\mathbf{A}^H \mathbf{A})^{-1}$ and the maximum singular value of an $N \times M$ matrix to obtain the normalized matrix \mathbf{B} . (The maximum singular value can be obtained with $O(M^2 N)$ multiplications by means of the power method for example [18].) Note that for

any vector $\mathbf{x} \in \mathbb{C}^N$, the computation of $\mathbf{Q}\mathbf{x}$ involves no more than $2LN + L^2$ multiplications since it can be computed as $\mathbf{Q}\mathbf{x} = \mathbf{x} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{x}$. Note that given L and M , the DDAA complexity only increases linearly as N increases. This is a particular advantage of the proposed algorithm in treating a non-spherical quadratic constraint. Although the per-iteration complexity of the DDAA is higher than that of CNLMS, the overall complexity of the DDAA (measured as the product of the number of iterations required for convergence and the per-iteration complexity) is comparable to that of the CNLMS due to its much faster convergence, as shown in Section V.

B. Properties of the Proposed Algorithm

We provide an analysis on the properties of the proposed DDAA implementing the LQCMV beamformer. Our analysis is based on the result from the general framework of the adaptive projected subgradient method² (APSM) [12], [19]. To present a rigorous analysis on the convergence of the proposed algorithm, we redefine the set V_k as follows:

$$V_k := K \cap \{\mathbf{w} \in \mathbb{C}^N : |\langle \mathbf{y}(k), \mathbf{w} \rangle| \leq \gamma\}, \quad k \in \mathbb{N} \quad (19)$$

for some $\gamma \geq 0$. Here, V_k is relaxed from the zero output-power set to a γ^2 output-power set contained in K . The set $\{\mathbf{w} \in \mathbb{C}^N : |\langle \mathbf{y}(k), \mathbf{w} \rangle| \leq \gamma\}$ is a ‘fattened’ hyperplane and it is called a *hyper slab*.³

Given any closed convex set $C \subset \mathbb{C}^N$, we define a metric distance function as $d(\mathbf{x}, C) := \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$, $\mathbf{x} \in \mathbb{C}^N$. Our analysis is based on the following key observation:

Observation 1: Define *time-dependent* cost functions as shown in (20) at the bottom of the page, where $\nu_k := \alpha_k d(\mathbf{w}_k, V_k) + (1 - \alpha_k)d(\mathbf{B}^H\mathbf{w}_k, B_\epsilon)$. Then, Algorithm 1 is reproduced by the following scheme:

$$\mathbf{w}_{k+1} := \begin{cases} \mathbf{w}_k - \lambda_k \frac{\psi_k(\mathbf{w}_k)}{2\|\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k)\|^2} \nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k), & \text{if } \nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) \neq \mathbf{0}, \\ \mathbf{w}_k, & \text{otherwise,} \end{cases} \quad (21)$$

where $\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) \in \mathbb{C}^N$ denotes a *conjugate Wirtinger subgradient* of ψ_k at \mathbf{w}_k [14]. (See Appendix I about the conjugate Wirtinger subgradient.) The update equation in (21) for $\lambda_k = 1$

²Basically, the APSM involves the optimization of a sequence of convex cost functions under a convex constraint by the metric projection onto a hard constraint set and a subgradient projection associated with the cost function in an adaptive manner. For more information, see [12], [19].

³The use of $\lambda_k < 1$ results in *under-projection*, and the under-projection onto the ‘thin’ set, V_k with $\gamma = 0$, can be interpreted as the exact projection onto the ‘fattened’ set, V_k with an appropriately chosen $\gamma > 0$. This implies that the two parameters γ and λ_k have a similar effect to each other in practice. We therefore let $\gamma = 0$ and $\lambda_k < 1$ in the experiments in Section V.

expresses the subgradient projection relative to ψ_k , provided that the function ψ_k satisfies $\text{lev}_{\leq 0} \psi_k := \{\mathbf{w} \in \mathbb{C}^N : \psi_k(\mathbf{w}) \leq 0\} \neq \emptyset$.

Proof: See Appendix II.

Note that the cost function (20) is a nonnegative continuous convex function of \mathbf{w} , because it is a convex combination of two metric distance functions $d(\mathbf{w}, V_k)$ and $d(\mathbf{B}^H P_K(\mathbf{w}), B_\epsilon)$. It is zero if and only if $\mathbf{w} \in V_k$ and $\mathbf{B}^H P_K(\mathbf{w}) \in B_\epsilon$, i.e., \mathbf{w} satisfies all the linear, quadratic and γ^2 instantaneous-output-power constraints, unless the current weight vector satisfies $\mathbf{w}_k \in V_k$ and $\mathbf{B}^H \mathbf{w}_k \in B_\epsilon$. Observation 1 together with Lemma 3 in Appendix I states that the weight-vector sequence generated by the DDAA can be regarded as a sequence generated by the APSM with the properly defined sequence of cost functions $(\psi_k(\mathbf{w}))_{k \in \mathbb{N}}$ in (20). The convergence result of the APSM framework can therefore be applied to our DDAA. Hereafter, we will make the following assumption:

Assumption 1: There exists a $\kappa_0 \in \mathbb{N}$ such that (s.t.) $\mathcal{E}_\epsilon \cap (\bigcap_{k \geq \kappa_0} V_k) \neq \emptyset$.

Assumption 1 basically means that the adaptive LQCMV optimization problem is feasible⁴, i.e., the linear constraint set, the quadratic constraint set, and the γ^2 instantaneous-output-power sets have a nonempty intersection for sufficiently large k . Then, it is not difficult to verify that the following lemma holds:

Lemma 2:

- 1) $\psi_k^* := \min_{\mathbf{w} \in \mathbb{C}^N} \psi_k(\mathbf{w}) = 0$, $\forall k \geq \kappa_0$.
- 2) For all $k \geq \kappa_0$, $\Omega_k := \arg \min_{\mathbf{w} \in K} \psi_k(\mathbf{w}) = \widehat{\mathcal{E}}_\epsilon \cap \widehat{V}_k \supset \mathcal{E}_\epsilon \cap V_k \neq \emptyset$, where for any set $S \subset \mathbb{C}^N$ we define $\widehat{S} := S$, if $\mathbf{w}_k \notin S$, and $\widehat{S} := \mathbb{C}^N$, if $\mathbf{w}_k \in S$. For instance, $\Omega_k = V_k$ if $\mathbf{B}^H \mathbf{w}_k \in B_\epsilon$ and $\mathbf{w}_k \notin V_k$.
- 3) $\Omega := \bigcap_{k \geq \kappa_0} \Omega_k \supset \mathcal{E}_\epsilon \cap (\bigcap_{k \geq \kappa_0} V_k) \neq \emptyset$.

Now, the properties of Algorithm 1 are summarized in Proposition 1 (Its proof is given in Appendix II).

Proof: Consequence of the analysis. Proposition 1.1 states that at each update the dual-domain adaptive algorithm brings the current weight vector closer to the set of minimizers of $\psi_k(\cdot)$, where minimizing $\psi_k(\cdot)$ implies the minimal violation of the linear and quadratic constraints and the γ^2 instantaneous-output-power condition. It therefore suggests the stability, tracking capability, and also fast convergence of DDAA since $\psi_k(\cdot)$ takes into account the quadratic constraint (see also the discussion in Section III-A). Further, Propositions 1.2 and 1.3 state that in the steady state the weight vector generated by the DDAA indeed satisfies all the linear and quadratic

⁴One may think that Assumption 1 would be a strict condition because it would hardly be true for $\gamma = 0$. As mentioned already, however, in the case of $\gamma = 0$, the use of a sufficiently small $\lambda_k < 1$ brings a similar effect to the use of a sufficiently large $\gamma > 0$, making all γ^2 instantaneous-output-power sets commonly contain a sufficiently large region around the origin. Also \mathcal{E}_ϵ is centered at the origin. Therefore, it is a reasonable condition which is satisfied practically.

$$\psi_k(\mathbf{w}) := \begin{cases} \frac{1}{\nu_k} \left(\alpha_k d(\mathbf{w}_k, V_k) d(\mathbf{w}, V_k) + (1 - \alpha_k) d(\mathbf{B}^H \mathbf{w}_k, B_\epsilon) d(\mathbf{B}^H P_K(\mathbf{w}), B_\epsilon) \right), & \text{if } \nu_k \neq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (20)$$

constraints (3) and (4) while maintaining the instantaneous output power below γ^2 , if the problem is feasible. This means that the quadratic constraint is eventually satisfied although it is not ensured to be satisfied at each iteration.

Proposition 1 (Properties of Algorithm 1)

1) Stepwise property:

- a) For all $k \geq \kappa_0$, $\|\mathbf{w}_{k+1} - \mathbf{w}_k^*\| \leq \|\mathbf{w}_k - \mathbf{w}_k^*\|$, $\forall \mathbf{w}_k^* \in \Omega_k$.
- b) If in particular $\lambda_k \in (0, 2)$ and $\mathbf{w}_k \notin \Omega_k$ for $k \geq \kappa_0$, then

$$\|\mathbf{w}_{k+1} - \mathbf{w}_k^*\| < \|\mathbf{w}_k - \mathbf{w}_k^*\|, \quad \forall \mathbf{w}_k^* \in \Omega_k.$$

Assume in the remainder of this proposition that there exist $\epsilon_1, \epsilon_2 > 0$ s.t. $\lambda_k \in [\epsilon_1, 2 - \epsilon_2] \subset (0, 2)$.

2) Asymptotic property:

- a) $\lim_{k \rightarrow \infty} \psi_k(\mathbf{w}_k) = 0$.
- b) Assume that Ω has a relative interior⁵ with respect to (w.r.t.) K . Then, the sequence $(\mathbf{w}_k)_{k \in \mathbb{N}}$ converges to some point $\hat{\mathbf{w}} \in K$ satisfying $\lim_{k \rightarrow \infty} \psi_k(\hat{\mathbf{w}}) = 0$.

3) Limit-point property: Assume the same interior point condition as in Proposition 1.2.b. Further, assume that (i) $\alpha_1^{\min} := \inf_{k \geq \kappa_0, \mathbf{w}_k \notin V_k} \alpha_k > 0$ and (ii) $\alpha_2^{\min} := \inf_{k \geq \kappa_0, \mathbf{B}^H \mathbf{w}_k \notin B_\epsilon} (1 - \alpha_k) > 0$. Then, $\hat{\mathbf{w}} := \lim_{k \rightarrow \infty} \mathbf{w}_k \in \liminf_{k \rightarrow \infty} \Omega_k \subset K$ and $\mathbf{B}^H \hat{\mathbf{w}} \in B_\epsilon$, where $\liminf_{k \rightarrow \infty} \Omega_k := \bigcup_{k=0}^{\infty} \bigcap_{n \geq k} \Omega_n$ and the overline denotes the closure.

IV. APPLICATION: THE RELAXED ZERO-FORCING BEAMFORMER

The LQCMV framework and the DDAA are general and can be used for adaptive receiver design in the areas of beamforming, direct-sequence code-division multiple-access or other systems. In this section, we present one specific application, RZF beamforming, for wireless multiple-access communication systems. The side information of the interference channels is used to form the quadratic constraint. The robustness issue against channel mismatch and ill-posedness is also addressed.

A. System Model

We consider the system, as shown in Fig. 1, with the data model (1), where all the channel information $\{\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_J\}$ is available at the basestation. The basestation needs to decode symbols from all transmitters based on the array output $(\mathbf{y}(k))_{k \in \mathbb{N}}$. Without loss of generality, consider the decoding of the 0-th transmitter. We assume that $E\{s_j(k)\} = 0$ and $E\{s_j^2(k)\} = \sigma_j^2$ for $j \in \{0, 1, \dots, J\}$; $E\{s_i(k)s_j(k)\} = 0$ for all $i \neq j$; $E\{s_j(k)\mathbf{n}(k)\} = \mathbf{0}$ for all $j \in \{0, 1, 2, \dots, J\}$; and the noise is independent and identically distributed (i.i.d.) complex Gaussian with zero mean and variance σ_n^2 . For the decoding of the 0-th transmitter, \mathbf{h}_0 is the channel for the desired signal and $\{\mathbf{h}_1, \dots, \mathbf{h}_J\}$ is the available side-information. Thus, in this case, we specifically design \mathbf{A} , \mathbf{B} and \mathbf{c} for

⁵The existence assumption of a relative interior in Proposition 1.2.b could be weakened. It is sufficient that Ω has a relative interior w.r.t. $K \cap \Pi$ for some $\Pi := \{\mathbf{w} \in \mathbb{C}^N : \Re(\langle \mathbf{a}, \mathbf{w} \rangle) = b\}$, $\mathbb{C}^N \ni \mathbf{a} \neq \mathbf{0}, b \in \mathbb{R}$.

the LQCMV formulation as follows: $\mathbf{A} := \mathbf{h}_0$, $\mathbf{B} := \frac{\mathbf{H}}{\sigma(\mathbf{H})}$, and $\mathbf{c} := 1$, where $\mathbf{H} := [\mathbf{h}_1, \dots, \mathbf{h}_J]$ and $\sigma(\mathbf{H})$ denotes the maximum singular value of \mathbf{H} ($L := 1$ and $M := J < N$).

B. RZF Beamforming

Our new formulation for adaptive beamforming is the *relaxed zero-forcing (RZF) beamformer*, given by solving the following optimization:

$$\text{minimize } \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (22)$$

$$\text{subject to } \mathbf{h}_0^H \mathbf{w} = 1 \quad (23)$$

$$\|\mathbf{B}^H \mathbf{w}\|^2 = \mathbf{w}^H (\mathbf{B} \mathbf{B}^H) \mathbf{w} \leq \epsilon, \quad (24)$$

where $\epsilon \geq 0$ is the interference relaxation parameter. Note that the second condition $\|\mathbf{B}^H \mathbf{w}\|^2 \leq \epsilon$ reduces to the classical ZF constraint for $\epsilon = 0$. On the other hand, for $\epsilon > 0$, the ZF constraint is relaxed and it becomes a quadratic one; the relaxation prevents the noise enhancement. Further, the matrix $\mathbf{B} \mathbf{B}^H$ is rank-deficient in most cases with large antenna arrays. The RZF beamformer coincides with the MVDR beamformer $\mathbf{w}_{\text{MVDR}} := \frac{\mathbf{R}^{-1} \mathbf{h}_0}{\mathbf{h}_0^H \mathbf{R}^{-1} \mathbf{h}_0}$ when $\epsilon = \infty$ (or ϵ is sufficiently large so that $\epsilon \geq \|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$). Thus, for $0 < \epsilon < \|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$, with the perfect \mathbf{R} , the proposed RZF beamformer is in-between the MVDR beamformer and the ZF beamformer. The SINR achieved by the RZF beamformer is given by

$$\text{SINR}_{\text{RZF}}(\epsilon) := \frac{\sigma_0^2 (\mathbf{h}_0^H \mathbf{R}_\epsilon^{-1} \mathbf{h}_0)^2}{\mathbf{h}_0^H \mathbf{R}_\epsilon^{-1} \mathbf{R}_{\text{IN}} \mathbf{R}_\epsilon^{-1} \mathbf{h}_0}, \quad (25)$$

where $\sigma_0^2 > 0$ is the signal variance of the desired user and

$$\mathbf{R}_{\text{IN}} := \sum_{j=1}^J \sigma_j^2 \mathbf{h}_j \mathbf{h}_j^H + \sigma_n^2 \mathbf{I} = \sigma_{\text{IN}}^2 \bar{\mathbf{R}}_{\text{IN}}. \quad (26)$$

Here, $\bar{\mathbf{R}}_{\text{IN}}$ is normalized to have its trace equal to the number N of array elements. The SINR achieved by the MVDR beamformer is obtained by letting $\lambda_\epsilon = 0$ ($\epsilon = \infty$), whereas the SINR achieved by the ZF beamformer is obtained as the limit of $\lambda_\epsilon = \infty$ ($\epsilon = 0$); λ_ϵ is defined under (5). A simple inspection of (25) suggests that $\text{SINR}_{\text{RZF}}(\epsilon)$ is a monotonically non-decreasing function of ϵ . Thus, the SINR of the RZF beamformer does not exceed that of the MVDR beamformer when the exact data covariance matrix \mathbf{R} is available, as expected. Now, consider the adaptive implementation of the RZF beamformer. Note that the RZF beamformer (22)–(24) falls into the LQCMV framework (2)–(4). It can therefore be implemented in an adaptive manner by using the DDAA in Algorithm 1. The advantages of the adaptive RZF beamformer based on the DDAA over the adaptive MVDR/CNLMS beamformer and the adaptive ZF/CNLMS beamformer are clear. The adaptive MVDR/CNLMS beamformer discards the interference side-information. On the other hand, the adaptive ZF/CNLMS beamformer overuses the side information by completely nulling out the interference to yield noise enhancement. The adaptive RZF/DDAA beamformer cleverly exploits the same side-information in adaptation in order to guide itself in the direction $\mathbf{B} \mathbf{g}_k^{(2)}$ in (10) for suppressing the interference effectively (but not completely) in the form of parallel projection combined with the

direction $\mathbf{g}_k^{(1)}$ for minimizing the interference plus noise used in the conventional MVDR/CNLMS approach. The RZF/DDAA can be viewed as a better modification of the MVDR/CNLMS approach than the ZF/CNLMS approach for incorporating the interference side-information. The superiority of the adaptive RZF/DDAA beamformer to the other two classical adaptive methods will be shown shortly in Section V.

Remark 1 (Controlling ϵ): Control of ϵ for the RZF beamforming will be discussed, especially for raising the convergence speed, in Section V.

C. Norm-Constrained RZF Beamforming—Channel-Mismatch Consideration

The RZF (or the LQCMV in general) beamforming may suffer from performance degradations due to channel mismatch and/or ill-posedness caused by presence of interference sources close to the signal transmitter. There exist many robustification techniques to handle such uncertainties [7], [10], [20]–[22]. Among available techniques to robustify the RZF beamformer, we adopt the regularization technique,⁶ which has an easy adaptive implementation. The norm-constrained RZF (NRZF) beamformer is given in the following general form:

$$\begin{aligned} & \text{minimize} && \mathbf{w}^H \mathbf{R} \mathbf{w}, \\ & \text{subject to} && \mathbf{A}^H \mathbf{w} = \mathbf{c}, \\ & && \|\mathbf{B}^H \mathbf{w}\|^2 \leq \epsilon, \\ & && \|\mathbf{w}\|^2 \leq \rho, \quad \epsilon \geq 0, \quad \rho \geq 0. \end{aligned} \quad (27)$$

We refer to the formulation in (27) as *the linearly, quadratically and norm constrained minimum variance (LQNCMV) beamformer*. Its solution is given by

$$\mathbf{w}_{\text{LQNCMV}} := \mathbf{R}_{\epsilon, \rho}^{-1} \mathbf{A} \left(\mathbf{A}^H \mathbf{R}_{\epsilon, \rho}^{-1} \mathbf{A} \right)^{-1} \mathbf{c}, \quad (28)$$

where $\mathbf{R}_{\epsilon, \rho} := \mathbf{R} + \lambda_\epsilon \mathbf{B} \mathbf{B}^H + \lambda_\rho \mathbf{I}$ with constants $\lambda_\epsilon \geq 0$ and $\lambda_\rho \geq 0$ depending on ϵ and ρ , respectively. The discussion on how to determine the values of those constants is omitted because the discussion is not much different than the one for LQCMV and, in addition, our dual-domain algorithm does not use the values explicitly. Our adaptive algorithm to implement LQNCMV is given below.

Algorithm 2 (The Robust Dual-Domain Adaptive Algorithm (RDDAA) for LQNCMV Beamformer)

Suppose that the LQNCMV beamforming problem (27) is feasible so that $K \cap B_\rho \neq \emptyset$, where $B_\rho := \{\mathbf{w} \in \mathbb{C}^N : \|\mathbf{w}\|^2 \leq \rho\}$. Given any initial beamformer $\mathbf{w}_0 \in K$, generate a sequence of beamformers $(\mathbf{w}_k)_{k \in \mathbb{N}}$ recursively as follows:

$$\mathbf{w}_{k+1} := P_{K \cap B_\rho} \left(\mathbf{w}_k + \lambda_k \mu_k \left(\alpha_k \mathbf{g}_k^{(1)} + (1 - \alpha_k) \mathbf{B} \mathbf{g}_k^{(2)} \right) \right).$$

Here, λ_k , μ_k , α_k , $\mathbf{g}_k^{(1)}$, and $\mathbf{g}_k^{(2)}$ are defined in the same way as in Algorithm 1.

⁶It is well known that the regularization technique properly handles the ill-posedness problem and the signal-channel mismatch. It can be shown that the regularization technique is also effective to handle the interference-channel mismatch; it is omitted to show due to space limitations.

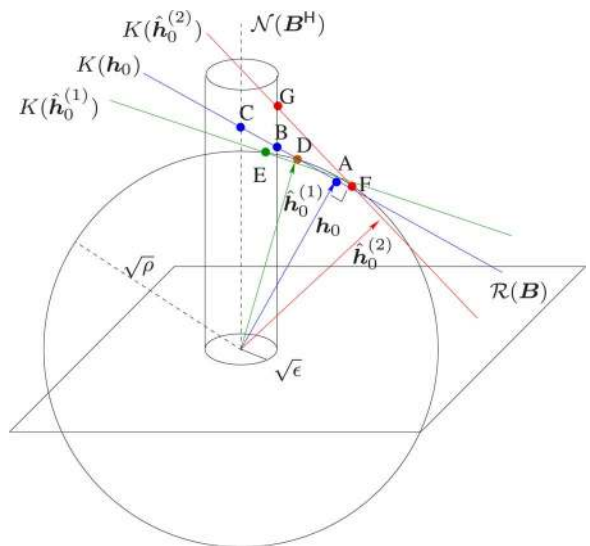


Fig. 3. Weight vector trajectories of RZF/DDAA and NRZF/RDDAA.

Algorithm 2 projects the updated weight vector onto the intersection of K and B_ρ , and the projection $P_{K \cap B_\rho}$ can be easily performed by the scaled projection method of Cox *et al.* [7]. In the case of the NRZF beamformer (i.e., the LQNCMV beamformer with the specific \mathbf{A} , \mathbf{B} , and \mathbf{c} employed in the RZF beamformer), the intersection is ensured to be nonempty if $\rho \geq \frac{1}{\|\mathbf{h}_0\|^2}$. Fig. 3 shows the trajectory of the RZF/DDAA in the weight vector space. $K(\mathbf{h}_0)$ is the set $K = \{\mathbf{w} : \mathbf{h}_0^H \mathbf{w} = 1\}$ for a given \mathbf{h}_0 . $\mathcal{N}(\mathbf{B}^H)$ is the null space of \mathbf{B}^H , and thus any point in the cylinder satisfies the quadratic constraint. When the algorithm is initialized with the matched filter $\mathbf{w}_0 = \mathbf{h}_0$, the weight vector generated by the RZF/DDAA starts from point A, moves along the hyperplane $K(\mathbf{h}_0)$, and stops at point B if the noise is ignored for simple visualization. When the regularization is applied to this case, the weight vector norm is normalized to $\sqrt{\rho}$ and the NRZF/RDDAA stops at point D since the quadratic constraint in the proposed algorithm is soft, i.e., the movement towards the cylinder of the quadratic feasible set is desired, but not a must.

V. NUMERICAL RESULTS

In this section, we present some numerical results to validate the proposed RZF beamforming implemented by the DDAA. We consider the uniform linear array with the array response $\mathbf{h}_j := \mathbf{h}(\theta_j)$ at the receiver, where θ_j is the direction of arrival (DOA) of the j -th signal, $\mathbf{h}(\theta) := \frac{1}{\sqrt{N}} [1, e^{2\pi i \frac{d}{\lambda} \cos \theta}, \dots, e^{2\pi i (N-1) \frac{d}{\lambda} \cos \theta}]^T \in \mathbb{C}^N$, $\theta \in [0, \pi]$, d is the antenna spacing and λ is the carrier wavelength [9]. We set $N = 16$, $\frac{d}{\lambda} = \frac{1}{3}$, $J = 5$ (six transmitters in the network), $[\theta_0, \dots, \theta_5] = [\frac{\pi}{6}, \frac{\pi}{12}, \frac{5\pi}{18}, \frac{7\pi}{18}, \frac{2\pi}{3}, \frac{5\pi}{6}]$, and $\alpha_k = (1 - \alpha_k) = \frac{1}{2} \forall k \in \mathbb{N}$ for the DDAA.

A. RZF/DDAA Versus MVDR/CNLMS and ZF/CNLMS

Fig. 4 shows the SINR performance of several adaptive beamforming algorithms with the matched filter $\mathbf{h}(\theta_0)$ as the initial weight vector \mathbf{w}_0 . Here, the SINR is the sample-path SINR computed as $\frac{\sigma_0^2 \mathbf{w}_k^H \mathbf{h}_0 \mathbf{h}_0^H \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{R}_{\text{IN}} \mathbf{w}_k}$ for the \mathbf{w}_k of each algorithm, which is proportional to the array gain of each algorithm.

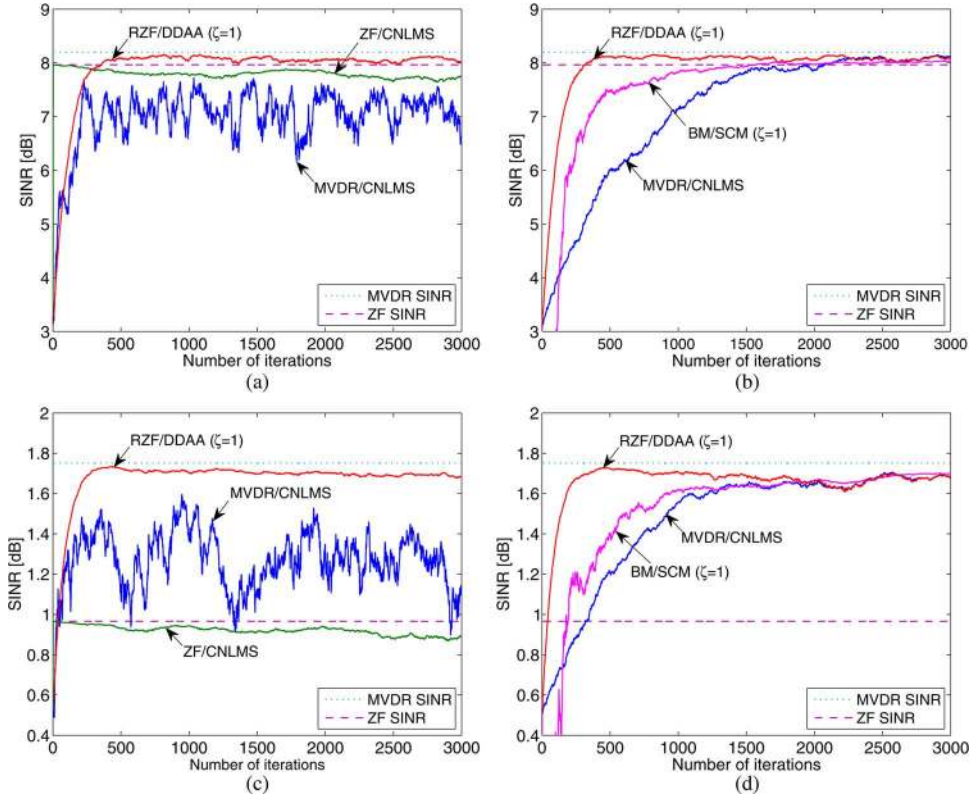


Fig. 4. SINR performance : (a) SNR = 10 dB, SIR = 0 dB, $\lambda_{\text{MVDR/CNLMS}} = 0.06$, $\lambda_{\text{ZF/CNLMS}} = \lambda_{\text{RZF/DDAA}} = 0.01$, (b) SNR = 10 dB, SIR = 0 dB, $\lambda_{\text{MVDR/CNLMS}} = \lambda_{\text{ZF/CNLMS}} = \lambda_{\text{RZF/DDAA}} = 0.01$, (c) SNR = 3 dB, SIR = 0 dB, $\lambda_{\text{MVDR/CNLMS}} = 0.08$, $\lambda_{\text{ZF/CNLMS}} = \lambda_{\text{RZF/DDAA}} = 0.01$, and (d) SNR = 3 dB, SIR = 0 dB, $\lambda_{\text{MVDR/CNLMS}} = \lambda_{\text{ZF/CNLMS}} = \lambda_{\text{RZF/DDAA}} = 0.01$.

Figs. 4(a) and (b) correspond to the case of 10 dB SNR ($\frac{\sigma_s^2}{\sigma_n^2}$) and 0 dB signal-to-interference ratio (SIR) ($\frac{\sigma_s^2}{\sigma_j^2}$) for all $j \neq 0$. In this case, the interference signals are strong compared to the noise. Thus, the theoretical performance of the ZF beamformer is close to that of the (theoretical) Capon's MVDR beamformer. (The theoretical performance is the value obtainable with the knowledge of the true covariance matrix.) In Fig. 4(a), the convergence speed of the MVDR/CNLMS is set to be the same as that of the RZF/DDAA by setting the step sizes of MVDR/CNLMS, ZF/CNLMS and RZF/DDAA to 0.06, 0.01 and 0.01, respectively. The interference relaxation parameter is set to $\epsilon = \zeta \|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$ for $\zeta = 1$. That is, the allowed amount of residual interference is set to be the same amount as the theoretically optimal MVDR beamformer \mathbf{w}_{MVDR} . It is seen that in the case of the same convergence speed, the RZF/DDAA outperforms the MVDR/CNLMS in the steady-state SINR. In Fig. 4(b), on the other hand, the step sizes for the MVDR/CNLMS and the RZF/DDAA are set to be the same as 0.01. By making the step size for the MVDR/CNLMS sufficiently small (0.01 in this case), we can average out the random perturbation in an adaptation direction and reach the optimal performance closely in the steady state. Note that the SINR of the two algorithms coincides in the steady state in this case. This can be explained as follows. With the matched-filter initialization, the interference constraint $\|\mathbf{B}^H \mathbf{w}_k\|^2 \leq \epsilon = \|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$ is not satisfied for small k . Thus, the interference constraint is activated and $\mathbf{g}_k^{(2)}$ in (10) is non-zero. This additional direction information helps

the algorithm to converge to the optimal point quickly. Once the algorithm reaches the optimal point closely (the point B in Fig. 3 with $\epsilon = \|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$), i.e., $\mathbf{w}_k \approx \mathbf{w}_{\text{MVDR}}$, the projection to the $\sqrt{\epsilon}$ ball in (18) does not change the argument, and $\mathbf{g}_k^{(2)}$ in (10) and (13) becomes zero. In this case, the length normalization factor μ_k in (15) becomes $\frac{1}{\alpha_k}$; μ_k cancels out α_k in front of $\mathbf{g}_k^{(1)}$ in (10); and the DDAA algorithm reduces to the MVDR/CNLMS. Note in the figure that the RZF/DDAA significantly outperforms the MVDR/CNLMS in the convergence speed due to the additional direction information from the interference side-information in the early stage of iteration when it has the same step size (yielding the same steady-state SINR). Fig. 4(c) and (d) correspond to the case of 3 dB SNR and 0 dB SIR for all the interference signals. In this case, the noise is relatively strong compared to the interference, and thus the ZF beamformer has a considerable performance loss compared to the MVDR beamformer. Similar trends to those in Fig. 4(a) and (b) are shown in this case, too. In all the cases, the ZF/CNLMS closely tracks the theoretical ZF SINR. To compare the RZF/DDAA with a batch processing method, we considered the batch processing semi-analytic method based on an estimated data covariance matrix mentioned in Section II. In the special case of $\zeta = 1$ as in Fig. 4, it is clear that $\lambda_\epsilon = 0$ and the RZF beamformer coincides with the MVDR beamformer. Thus, for the semi-analytic method, we used the sample covariance estimator $\hat{\mathbf{R}}(k) := \frac{1}{k} \sum_{j=1}^k \mathbf{y}(j) \mathbf{y}^H(j)$ for \mathbf{R} , computed $\mathbf{w}_{\text{LQCMV}}$ with $\hat{\mathbf{R}}(k)$ and $\lambda_\epsilon = 0$ from (5) at each k , and plotted the corresponding SINR curves (labeled by BM/SCM)

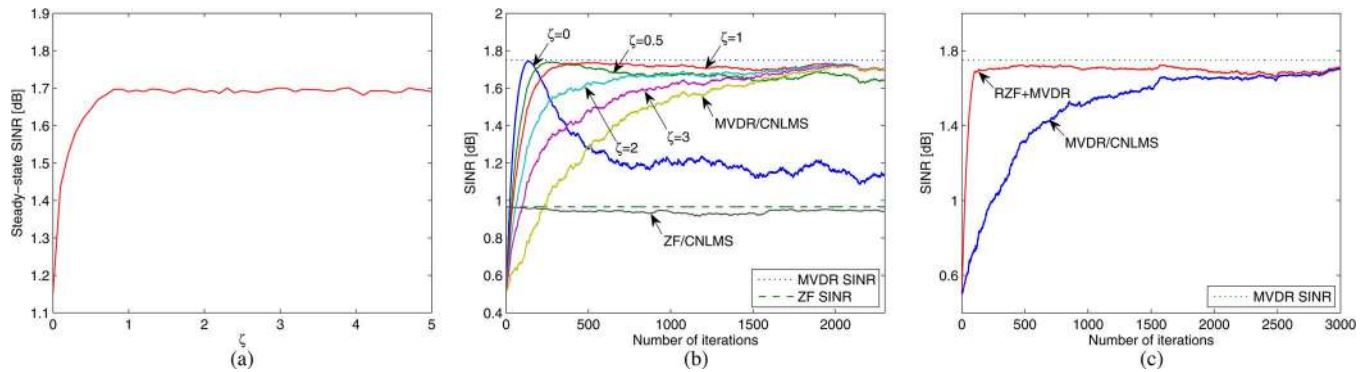


Fig. 5. Impact of the interference relaxation parameter ζ (SNR = 3 dB, SIR = 0 dB, $\lambda = 0.01$): (a) steady-state SINR as a function of ζ , (b) convergence behavior for different ζ , and (c) performance of the mixture algorithm composed of RZF/DDAA with $\epsilon = 0$ and MVDR/CNLMS switching after 100 iterations.

in Figs. 4(b) and (d). It is seen that the batch processing method requires more data, than the DDAA, to obtain a good estimate for \mathbf{R} and thus to yield a receiver beamformer with the same quality. In practice, the beamformer needs to be computed only once after an estimate of \mathbf{R} with reasonable quality is obtained. Even in this case, the batch processing method is less attractive since it requires more data (i.e. more time steps) and the per-step complexity $O(N^2)$ of the sample covariance matrix computation is heavier than the $O(N)$ per-step complexity of the DDAA.

B. Impact of the Interference Relaxation Parameter

For the results in Fig. 4, the amount of residual interference is set to be the same as that caused by the theoretically optimal MVDR beamformer. The performance of the RZF/DDAA depends on the value of the interference relaxation parameter ϵ (or ζ). To investigate the impact of ϵ , we swept ζ for the same setup as that for Fig. 4(d). The steady-state SINR as a function of ζ is shown in Fig. 5(a). As seen in the figure, the steady-state SINR increases as ζ increases and saturates around $\zeta = 1$, as expected. When $\zeta \geq 1$, the optimal MVDR beamformer lies within the feasibility set of the RZF beamforming and there is no SINR loss for the RZF beamforming. However, ζ larger than unity is not beneficial from the perspective of convergence speed. The convergence speed of the RZF/DDAA for different values of ζ is shown in Fig. 5(b). It is seen that the speed that the RZF/DDAA reaches the maximum SINR is fastest at $\zeta = 0$ and it slows down as ζ increases. When $\epsilon = \infty$ (i.e., $\zeta = \infty$), the RZF/DDAA coincides with the MVDR/CNLMS, as explained before. Thus, the optimal ϵ is $\|\mathbf{B}^H \mathbf{w}_{\text{MVDR}}\|^2$ if ϵ is fixed throughout the iteration. It is interesting to note the behavior of the RZF/DDAA with $\zeta = 0$ (or equivalently $\epsilon = 0$) and to compare it with that of the ZF/CNLMS. The SINR of the RZF/DDAA with $\zeta = 0$ reaches almost the optimal MVDR SINR first and then decreases to a steady-state value. This is because in the RZF/DDAA the interference constraint is used as a *soft* constraint. The algorithm initially moves from the matched-filter initialization (the point A in Fig. 3) in the direction for minimizing the output variance and suppressing the interference, but the interference is not fully suppressed in the early stage of iteration since the algorithm moves in that direction by small steps only. Since the optimal weight vector (the point B in Fig. 3 with $\zeta = 1$) is somewhere between the matched

filter (the point A in Fig. 3) and the ZF beamformer (the point C in Fig. 3), the algorithm reaches the optimal interference suppression level as the iteration continues from the matched-filter initialization. As the iteration continues further, the algorithm surpasses this optimal point and goes further towards the ZF solution. Thus, we have the corresponding curve for the SINR of the RZF/DDAA with $\zeta = 0$, as shown in Fig. 5(b). On the contrary, the conventional ZF/CNLMS algorithm brings the weight vector around the ZF solution at the very first iteration because the ZF interference constraint is used as a hard constraint in this case. Another thing to note is that the steady-state performance of the RZF/DDAA with $\zeta = 0$ is better than the theoretical ZF SINR. This is also because the interference constraint is used as a soft constraint for the RZF/DDAA (with the weights of $\alpha_1^{(1)} = \alpha_2^{(2)} = \frac{1}{2}$ in this example). Thus, the RZF/DDAA with $\zeta = 0$ does not fully zero-force the interference even with $\zeta = 0$ in the steady state, and it always performs better than the ZF/CNLMS for any choice of ϵ . The results in Fig. 5(a) and (b) suggest how to determine the interference relaxation parameter ϵ in practice without knowing \mathbf{w}_{MVDR} , especially for the purpose of speeding up the convergence of the conventional MVDR/CNLMS algorithm. Under the assumption of matched-filter initialization, we first set ϵ as a very small value, e.g., zero⁷, for a certain number of iterations, then set ϵ to infinity, i.e., switch to the conventional MVDR/CNLMS algorithm, and, after a sufficient number N_s of iterations guaranteeing convergence, set $\epsilon = \|\mathbf{B}^H \mathbf{w}_{N_s}\|^2$ for the remaining time to cope with possible perturbation. In this way, the interference side-information is far better exploited than in the ZF/CNLMS algorithm. Here, the initial period of $\epsilon = 0$ does not need to be exact, but it is enough that this period brings the weight vector roughly around the optimal value. (The length of this period can be determined by various simulations.) This mixture structure of RZF/DDAA and MVDR/CNLMS can be very useful in wireless digital communications in which the transmission is packetized. With such a structure, only a small preamble in front of a payload is necessary to train the beamformer adaptively before the actual data transmission begins, and the reduction in the preamble size is drastic. Fig. 5(c) shows the performance of the mixture algorithm that is composed of RZF/DDAA in the first

⁷In the extreme case of $\epsilon = 0$, the singular value normalization for \mathbf{B} is not even necessary. $\mathbf{B}\mathbf{g}_k^{(2)}$ in Algorithm 1 can be obtained by the projection onto the hyperplane determined by the zero-forcing condition.

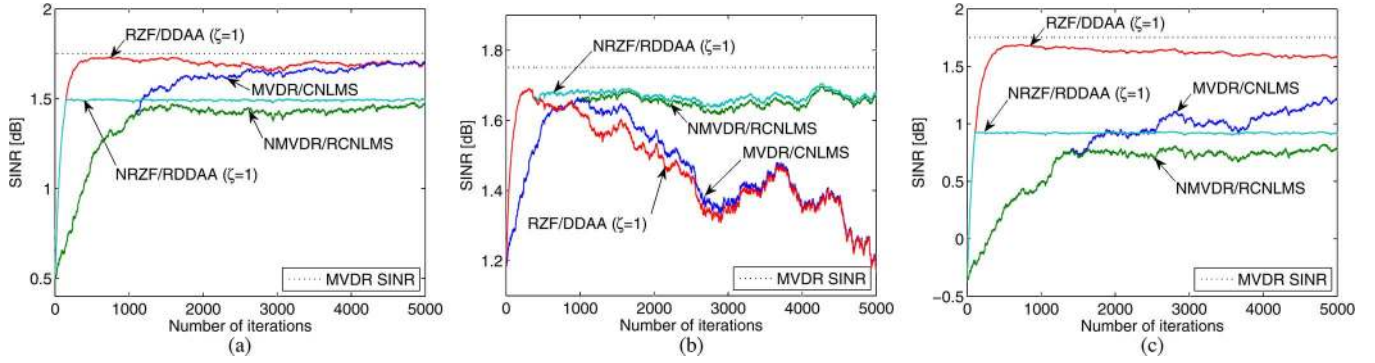


Fig. 6. NRZF with robust DDAA—signal-channel mismatch with no interference-channel mismatch ($\Delta\theta_j = 0, j = 1, \dots, J$): (a) $\Delta\theta_0 = 0$, (b) $\Delta\theta_0 = \frac{\pi}{90}$, and (c) $\Delta\theta_0 = -\frac{\pi}{90}$.

phase and MVDR/CNLMS in the second phase; the phase is switched after 100 iterations. We used the same SNR and SIR set up as that in Fig. 5(a) and (b). It is seen that the mixture algorithm performs well and significantly reduces the convergence time when compared with the MVDR/CNLMS alone. Although the performance of the mixture algorithm for different channel realizations is not shown here due to space limitations, the mixture algorithm showed similar convergence behavior for different channel realizations.

There exist some techniques [23] such as the constrained affine projection algorithm to speed up the CLMS algorithm. By redefining $V_k := K \cap \{\mathbf{w} \in \mathbb{C}^N : \mathbf{Y}^H(k)\mathbf{w} = \mathbf{0}\}$ with $\mathbf{Y}(k) := [\mathbf{y}(k), \mathbf{y}(k-1), \dots, \mathbf{y}(k-r+1)] \in \mathbb{C}^{N \times r}$, $r \in \mathbb{N} \setminus \{0\}$, we obtain an extension of the affine projection algorithm. Thus, when the interference side-information captured in the interference norm constraint is used for the affine projection algorithm, the convergence speed can be accelerated further.

C. NRZF/RDDAA (RZF/DDAA) Versus NMVDR/RCNLMS (MVDR/CNLMS) in Erroneous Cases

Fig. 6 shows the performance of the NRZF/RDDAA and that of the norm-constrained MVDR/CNLMS algorithm for the same setup as that for Fig. 4(d) (SNR = 3 dB, SIR = 0 dB, $\lambda = 0.01$) when $\theta_0 = \hat{\theta}_0 + \Delta\theta_0$ and there is no error in the interference DOAs. (Here, $\hat{\theta}_0$ denotes a signal DOA estimate available and $\Delta\theta_0$ an estimation error.) The bound of the squared norm of the weight vector is set to $\rho = 1.06$. Fig. 6(a) shows the performance in the case that the signal DOA information is also perfect. In this case, when the regularization is performed to the RZF/DDAA, the performance degrades compared to the non-regularized version. The steady-state point in this case corresponds to the point D in Fig. 3, which is different from the optimal point B. Fig. 6(b) shows the performance in the case that the available signal-DOA is in error ($\Delta\theta_0 = \frac{\pi}{90}$) and the available signal-channel is inclined towards the null space of \mathbf{B}^H (i.e., the case of $\hat{\mathbf{h}}_0^{(1)}$ in Fig. 3). In this case, when the regularization technique is not applied, the performance degrades. When it is applied, however, good performance is maintained. (The point E in Fig. 3 is the steady-state point in this case.) Note that the steady-state performance in this case is better than that of Fig. 6(a) with regularization. This is because the point E can be closer to the optimal point B than the point D, as shown in Fig. 3. Fig. 6(c) shows the performance in the

case that the available signal-DOA is in error ($\Delta\theta_0 = -\frac{\pi}{90}$) and the available signal-channel is away from the null space of \mathbf{B}^H compared to $\Delta\theta_0 = 0$ (i.e., the case of $\hat{\mathbf{h}}_0^{(2)}$ in Fig. 3). In this case, the regularization yields a worse performance than the non-regularized version. (The point G is closer to the optimal point B than the point F, as shown in Fig. 3, which is the steady-state point in this case.) Since there is movement in the direction from the point F to the point G, going out of the $\sqrt{\rho}$ ball in this case, the regularization brings the weight vector back to the ball and this enforcement is clearly seen in Fig. 6(c). A similar behavior is seen in Fig. 6(a) since there is movement from the point D to the point B, going out of the ball, in this case. In Fig. 6(b), on the other hand, the enforcement to the $\sqrt{\rho}$ ball is not clearly seen. This is because in this case the steady-state point E is already in the cylinder, as shown in Fig. 3. Therefore, $\mathbf{B}\mathbf{g}_k^{(2)}$ is not active any more and the weight vector is wandering around the point E inside the $\sqrt{\rho}$ ball. In any case, when the regularization is performed to the RZF/DDAA, the performance degradation in SINR from the theoretical maximum value is less than 1 dB and the performance degradation due to the signal-channel mismatch is kept within a reasonable level.

Although it is not shown here due to space limitations, it is observed that the interference-channel-information error is not critical to the performance of the RZF/DDAA. It is also observed that the NRZF/RDDAA performs robustly against both errors even in the case that we have errors in both the signal and interference DOA-information. Furthermore, we performed simulations when the antenna element location has an error, and observed similar behaviors to those in the DOA error case described here. The RZF/DDAA can be extended to MIMO cases. The performance of the RZF/DDAA in the MIMO case is described in [2] and the performance gain is also significant in the MIMO case.

VI. CONCLUSION

We have proposed a new paradigm for adaptive beamforming—the LQCMV beamforming framework and its efficient adaptive implementation based on the DDAA. The LQCMV beamforming is an extension of Frost's LCMV framework and it can incorporate a general quadratic-inequality constraint that may capture available side-information. We have investigated the properties of the DDAA and have presented the

convergence analysis of the algorithm under mild conditions. As a specific application of the LQCMV/DDAA, we have considered RZF beamforming which properly incorporates the interference side-information to enhance the convergence speed. The numerical results have shown that the proposed adaptive beamformer significantly outperforms the conventional adaptive beamformers based on the LCMV/C(N)LMS framework. The present study has shed new light upon the use of interference side-information, and has demonstrated its great benefit in accelerating the convergence of adaptive beamforming algorithms efficiently when it is exploited with LQCMV/DDAA.

APPENDIX I

WIRTINGER SUBGRADIENTS AND APSM

Under the mapping $\mathbf{z} := \mathbf{x} + i\mathbf{y} \rightarrow \check{\mathbf{z}} := [\mathbf{x}^T \mathbf{y}^T]^T$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, \mathbb{C}^N is isomorphic to \mathbb{R}^{2N} . We indicate the correspondence explicitly as $\mathbf{z} \rightleftharpoons \check{\mathbf{z}}$. Likewise, for any set $S \subset \mathbb{C}^N$, we define $\check{S} := \{[\mathbf{x}^T \mathbf{y}^T]^T : \mathbf{x} + i\mathbf{y} \in S\} \subset \mathbb{R}^{2N}$. The correspondence is indicated as $S \rightleftharpoons \check{S}$. For any given matrix $\mathbf{A} = \mathbf{A}_R + i\mathbf{A}_I \in \mathbb{C}^{n \times m}$ for arbitrary $n, m \in \mathbb{N}^*$, let $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \\ \mathbf{A}_I & \mathbf{A}_R \end{bmatrix} \in \mathbb{R}^{2n \times 2m}$. Also, for any given $n \in \mathbb{N}^*$, let $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^n} := \mathbf{x}^T \mathbf{y}$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and let $\|\mathbf{x}\|_{\mathbb{R}^n} := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbb{R}^n}}$, $\forall \mathbf{x} \in \mathbb{R}^n$. Then, the following is true.

Fact 1:

- 1) $\mathbf{A}^H \mathbf{z} \rightleftharpoons \bar{\mathbf{A}}^T \check{\mathbf{z}}$ for any $\mathbf{z} \in \mathbb{C}^n$.
- 2) $\mathbf{A} \mathbf{z} \rightleftharpoons \bar{\mathbf{A}} \check{\mathbf{z}}$ for any $\mathbf{z} \in \mathbb{C}^m$.
- 3) The maximum singular value of $\bar{\mathbf{A}}$ is identical to that of \mathbf{A} .
- 4) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \check{\mathbf{x}}, \check{\mathbf{y}} \rangle_{\mathbb{R}^{2N}} + i \langle \check{\mathbf{x}}, \bar{\mathbf{D}} \check{\mathbf{y}} \rangle_{\mathbb{R}^{2N}}$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, where $\bar{\mathbf{D}} := -i\mathbf{I}$.
- 5) $\|\mathbf{z}\| = \|\check{\mathbf{z}}\|_{\mathbb{R}^{2N}}$, $\forall \mathbf{z} \in \mathbb{C}^N$.
- 6) $P_C(\mathbf{z}) \rightleftharpoons P_{\check{C}}(\check{\mathbf{z}})$ for any closed convex set $C \subset \mathbb{C}^N$ and any point $\mathbf{z} \in \mathbb{C}^N$.

Definition 1 (Wirtinger subgradient): Let $\varphi : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ be a continuous convex function. Let

$$\nabla^s \varphi(\check{\mathbf{z}}) := \begin{bmatrix} \nabla_x^s \varphi(\check{\mathbf{z}}) \\ \nabla_y^s \varphi(\check{\mathbf{z}}) \end{bmatrix}$$

be a subgradient of φ at $\check{\mathbf{z}} \in \mathbb{R}^{2N}$, where $\nabla_x^s \varphi(\check{\mathbf{z}}), \nabla_y^s \varphi(\check{\mathbf{z}}) \in \mathbb{R}^N$. Due to the isomorphism $\mathbb{C}^N \simeq \mathbb{R}^{2N}$, the function φ can also be regarded as $\varphi : \mathbb{C}^N \rightarrow \mathbb{R}$, $\mathbf{z} \mapsto \varphi(\mathbf{z}) = \varphi(\check{\mathbf{z}})$, where $\mathbf{z} \rightleftharpoons \check{\mathbf{z}}$. Then, the Wirtinger subgradient of $\varphi : \mathbb{C}^N \rightarrow \mathbb{R}$ at $\mathbf{z} \in \mathbb{C}^N$ is given by [14]

$$\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}) := \frac{1}{2} (\nabla_x^s \varphi(\check{\mathbf{z}}) - i \nabla_y^s \varphi(\check{\mathbf{z}})), \quad (29)$$

and the conjugate Wirtinger subgradient (or CW-subgradient for short) of φ at \mathbf{z} is given by

$$\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}) := \frac{1}{2} (\nabla_x^s \varphi(\check{\mathbf{z}}) + i \nabla_y^s \varphi(\check{\mathbf{z}})). \quad (30)$$

The set of all conjugate Wirtinger subgradients of φ at \mathbf{z} is called the *Wirtinger subdifferential* of φ at \mathbf{z} and is denoted by $\partial_{\mathbf{f}^*} \varphi(\mathbf{z})$.

Regarding the Wirtinger and conjugate Wirtinger subgradients, we have the following.

Fact 2:

- 1) $2(\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}))^* \rightleftharpoons \nabla^s \varphi(\check{\mathbf{z}})$.

$$2) 2\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}) \rightleftharpoons \nabla^s \varphi(\check{\mathbf{z}}).$$

$$3) \frac{\varphi(\mathbf{z})}{2\|\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z})\|^2} \nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}) \rightleftharpoons \frac{\varphi(\check{\mathbf{z}})}{\|\nabla^s \varphi(\check{\mathbf{z}})\|_{\mathbb{R}^{2N}}^2} \nabla^s \varphi(\check{\mathbf{z}}).$$

4) Given any closed convex set $C \subset \mathbb{R}^N$, let $\varphi(\mathbf{z}) := d(\mathbf{z}, C)$, $\mathbf{z} \in \mathbb{C}^N$, where $d(\mathbf{z}, C)$ is the metric distance from \mathbf{z} to C . Then, a CW-subgradient is given by $\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{z}) := \frac{\mathbf{z} - P_C(\mathbf{z})}{2d(\mathbf{z}, C)} \in \partial_{\mathbf{f}^*}^s \varphi(\mathbf{z})$.

5) Let $g(\mathbf{z}) := \varphi(\mathbf{G}\mathbf{z} + \mathbf{e})$, $\mathbf{z} \in \mathbb{C}^N$, for some $\mathbf{G} \in \mathbb{C}^{M \times N}$ and $\mathbf{e} \in \mathbb{C}^M$. Then, $\nabla_{\mathbf{f}^*}^s g(\mathbf{z}) = \mathbf{G}^H \nabla_{\mathbf{f}^*}^s \varphi(\mathbf{G}\mathbf{z} + \mathbf{e})$.

6) For any continuous convex function $\varphi : \mathbb{C}^N \rightarrow \mathbb{R}$, its conjugate Wirtinger subgradient $\nabla_{\mathbf{f}^*}^s \varphi(\mathbf{w})$ at an arbitrary point $\mathbf{w} \in \mathbb{C}^N$ satisfies the following inequality:

$$\varphi(\mathbf{z}) \geq \varphi(\mathbf{w}) + 2\Re \langle \mathbf{z} - \mathbf{w}, \nabla_{\mathbf{f}^*}^s \varphi(\mathbf{w}) \rangle, \quad \forall \mathbf{z} \in \mathbb{C}^N.$$

Regarding the complex-vector sequence generated by (21) based on the conjugate Wirtinger subgradient, we have the following result which is easily verified by using the isomorphism $\mathbf{z} := \mathbf{x} + i\mathbf{y} \rightarrow \check{\mathbf{z}} := [\mathbf{x}^T \mathbf{y}^T]^T$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and Fact 2.

Lemma 3: Let $(\mathbf{w}_k)_{k \in \mathbb{N}} \subset \mathbb{C}^N$ be the (complex-vector) sequence generated by (21) for an initial vector $\mathbf{w}_0 \in \mathbb{C}^N$ and let $(\check{\mathbf{w}}_k)_{k \in \mathbb{N}} \subset \mathbb{R}^{2N}$ be the (real-vector) sequence generated as

$$\check{\mathbf{w}}_{k+1} := \begin{cases} \check{\mathbf{w}}_k - \lambda_k \frac{\psi_k(\check{\mathbf{w}}_k)}{\|\nabla^s \psi_k(\check{\mathbf{w}}_k)\|_{\mathbb{R}^{2N}}^2} \nabla^s \psi_k(\check{\mathbf{w}}_k), & \text{if } \nabla^s \psi_k(\check{\mathbf{w}}_k) \neq \mathbf{0}, \\ \check{\mathbf{w}}_k, & \text{otherwise} \end{cases}$$

for $\check{\mathbf{w}}_0 \in \mathbb{R}^{2N}$ s.t. $\mathbf{w}_0 \rightleftharpoons \check{\mathbf{w}}_0$. Then, it holds that $\mathbf{w}_k \rightleftharpoons \check{\mathbf{w}}_k$, $\forall k \in \mathbb{N}$.

Note that $(\check{\mathbf{w}}_k)_{k \in \mathbb{N}}$ is a real-vector sequence and that it is generated by the adaptive projected subgradient method (APSM) [12], [19] applied to the sequence $(\psi_k(\cdot))_{k \in \mathbb{N}}$ of convex cost functions. Thus, Lemma 3 allows us to apply the analysis of APSM (which is based on a real Hilbert space) to the complex case in this paper.

APPENDIX II PROOFS

A. Proof of Observation 1

Case 1: First, assume that $\nu_k = 0$, or equivalently that $\mathbf{w}_k \in V_k$ and $\mathbf{B}^H \mathbf{w}_k \in B_\epsilon$. In this case, $\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) = \mathbf{0}$ and hence (21) gives $\mathbf{w}_{k+1} = \mathbf{w}_k$. On the other hand, Algorithm 1 is reduced to $\mathbf{w}_{k+1} = P_K(\mathbf{w}_k) = \mathbf{w}_k$ since $P_{V_k}(\mathbf{w}_k) = \mathbf{w}_k$, $P_{B_\epsilon}(\mathbf{B}^H \mathbf{w}_k) = \mathbf{B}^H \mathbf{w}_k$, and $\mu_k = 1$. Therefore, (21) and Algorithm 1 are equivalent in this case.

Case 2: Next, assume that $\nu_k \neq 0$. In this case, by using (16) and Facts 2.4 and 2.5, it can be verified that a CW-subgradient $\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}) \in \partial_{\mathbf{f}^*}^s \psi_k(\mathbf{w})$ of ψ_k at $\mathbf{w} \in \mathbb{C}^N$ is given by

$$\begin{aligned} \nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}) = & \frac{1}{2\nu_k} \left[\alpha_k d(\mathbf{w}_k, V_k) \frac{\mathbf{w} - P_{V_k}(\mathbf{w})}{d(\mathbf{w}, V_k)} + (1 - \alpha_k) \right. \\ & \times d(\mathbf{B}^H \mathbf{w}_k, B_\epsilon) \mathbf{Q} \mathbf{B} \\ & \left. \times \frac{\mathbf{B}^H P_K(\mathbf{w}) - P_{B_\epsilon}(\mathbf{B}^H P_K(\mathbf{w}))}{d(\mathbf{B}^H P_K(\mathbf{w}), B_\epsilon)} \right]. \quad (31) \end{aligned}$$

By substituting $\mathbf{w} = \mathbf{w}_k$ into (31), we have

$$\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) = -\frac{1}{2\nu_k} \left(\alpha_k \mathbf{g}_k^{(1)} + (1 - \alpha_k) \mathbf{Q} \mathbf{B} \mathbf{g}_k^{(2)} \right). \quad (32)$$

Case 2-a: Suppose that $\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) = \mathbf{0}$. In this case, (21) gives $\mathbf{w}_{k+1} = \mathbf{w}_k$. On the other hand, Algorithm 1, i.e., (11) is reduced to $\mathbf{w}_{k+1} := \mathbf{w}_k + \lambda_k \mu_k \mathbf{g}_k = \mathbf{w}_k$ since $\mathbf{g}_k = -2\nu_k \nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) = \mathbf{0}$ by (32).

Case 2-b: Suppose now that $\nabla_{\mathbf{f}^*}^s \psi_k(\mathbf{w}_k) \neq \mathbf{0}$. In this case, by using the fact that $d(\mathbf{w}_k, V_k) = \|\mathbf{w}_k - P_{V_k}(\mathbf{w}_k)\|$ and $d(\mathbf{B}^H \mathbf{w}_k, B_\epsilon) = \|\mathbf{B}^H \mathbf{w}_k - P_{B_\epsilon}(\mathbf{B}^H \mathbf{w}_k)\|$, it can be verified that (21) reproduces Algorithm 1, i.e., (11). This completes the proof. (The length normalization parameter μ_k in (15) is designed so as to yield this equivalence.) \square

B. Proof of Proposition 1

Thanks to Lemma 3, it is sufficient to prove the proposition for its real-valued counterpart (see Appendix I for the isomorphism $\mathbb{C}^N \simeq \mathbb{R}^{2N}$). In the following, the check and bar notations defined in Appendix I are extensively used to represent the real-valued counterpart.

Proposition 1.1: By applying Lemma 2 and [19, Theorem 3.1.1] directly, the claim is readily verified.

Proposition 1.2: To apply [19, Theorems 3.1.2 and 3.1.3], there are three points to be validated here:

- (i) prove the boundedness of $(\nabla^s \psi_k(\check{\mathbf{w}}_k))_{k \in \mathbb{N}} \subset \mathbb{R}^{2N}$,
- (ii) prove the existence of a bounded sequence $(\nabla^s \psi_k(\check{\mathbf{w}}))_{k \in \mathbb{N}} \subset \mathbb{R}^{2N}$, and
- (iii) fill in the gap in the existence assumption of a relative interior.

Regarding the point (iii), in order to apply [19, Theorems 3.1.2 and 3.1.3] straightforwardly, it should be assumed that $\check{\Omega}$ has a relative interior w.r.t. a hyperplane $\check{\Pi} := \{\check{\mathbf{w}} \in \mathbb{R}^{2N} : \langle \check{\mathbf{a}}, \check{\mathbf{w}} \rangle_{\mathbb{R}^{2N}} = b\}$ for some nonzero vector $\check{\mathbf{a}} \in \mathbb{R}^{2N}$ and $b \in \mathbb{R}$. This assumption does never hold because $\check{\Omega} \subset \check{K} (\Leftrightarrow \Omega \subset K)$ and because the dimension of \check{K} is at most $2N - 2$. However we can use the same trick as in [15]. The idea is simple. As all the points of the sequence $(\check{\mathbf{w}}_k)_{k \in \mathbb{N}}$ lie on the linear variety \check{K} , we translate all those points into the underlying subspace of \check{K} , say \check{M} ; in other words we shift \check{K} to the origin. We regard the subspace \check{M} itself as a real Hilbert space, and apply the analysis of APSM to the translated vector sequence. The analysis for the original vector sequence is readily obtained by getting back to the original linear variety \check{K} through a simple translation. We mention that $\psi_k(\check{\mathbf{w}}) \leq \psi_k(\check{\mathbf{w}} + \check{\mathbf{n}})$, $\forall (\check{\mathbf{w}}, \check{\mathbf{n}}) \in \check{K} \times \check{M}^\perp$; this property is called the *constraint-embedding structure* used in the analysis in [15]. This ensures the following two points, which facilitate the analysis. First, a global minimizer of $\psi_k(\cdot)$ over \mathbb{R}^{2N} exists in \check{K} . Second, at any point $\check{\mathbf{w}} \in \check{K}$ there exists a subgradient of ψ_k in \check{M} [15, Lemma 1].

To prove the points (i) and (ii) mentioned above, we can follow the way in [11, Appendix C] with a notice to the facts that (a) $\|\check{\mathbf{Q}}\check{\mathbf{x}}\|_{\mathbb{R}^{2N}} \leq \|\check{\mathbf{x}}\|_{\mathbb{R}^{2N}}$, $\forall \check{\mathbf{x}} \in \mathbb{R}^{2N}$ and (b) $P_{\check{K}}(\check{\mathbf{w}}_k) = \check{\mathbf{w}}_k$ and $P_{\check{K}}(\check{\mathbf{w}}) = \check{\mathbf{w}}$ because $\check{\mathbf{w}}_k \in \check{K}$ and $\check{\mathbf{w}} \in \check{K} (\Leftrightarrow \mathbf{w}_k \in K$ and $\hat{\mathbf{w}} \in K)$. The fact (a) suggests the nonexpansivity of the linear operator $\check{\mathbf{Q}}$ which can be regarded as the projection operator onto the subspace \check{M} .

Proposition 1.3: The gap in the existence assumption of a relative interior in applying [19, Theorem 3.1.4] can be filled in as in the proof of Proposition 1–2. To show that the other condition to be fulfilled is satisfied, we can follow the way in [11, Appendix C]. We shortly mention that, although $\mathbf{B}^H P_K$ appearing

in the definition of ψ_k is an affine (rather than linear) mapping and $\text{lev}_{\leq 0} \psi_k = \widehat{V}_k \cap (\mathbf{B}^H P_K)^{-1}(B_\epsilon)$, it does not make any essential change in the proof. (The wide-hat notation is defined in Lemma 2.) It is therefore verified that $\hat{\mathbf{w}} \in \overline{\liminf_{k \rightarrow \infty} \Omega_k}$. In addition, since $\Omega_k \subset K$, the closedness of the linear variety K ensures that $\overline{\liminf_{k \rightarrow \infty} \Omega_k} \subset K$.

We finally show that $\mathbf{B}^H \hat{\mathbf{w}} \in B_\epsilon$. Assume that there exists $k' \in \mathbb{N}$ such that $\mathbf{B}^H \mathbf{w}_k \in B_\epsilon$ for all $k \geq k'$. In this case, the closedness of B_ϵ ensures that $\mathbf{B}^H \hat{\mathbf{w}} \in B_\epsilon$. Assume in contrast that for any $k \in \mathbb{N}$ there exists a $k' \geq k$ such that $\mathbf{B}^H \mathbf{w}_{k'} \notin B_\epsilon$. Note that $\mathbf{B}^H \mathbf{w}_{k'} \notin B_\epsilon$ implies $\Omega_{k'} \subset \mathcal{E}_\epsilon = (\mathbf{B}^H)^{-1}(B_\epsilon)$ (see Lemma 2.2). We thus obtain $\bigcap_{n \geq k} \Omega_n \subset \Omega_{k'} \subset \mathcal{E}_\epsilon$ for any $k \in \mathbb{N}$. Since $\hat{\mathbf{w}} \in \overline{\liminf_{k \rightarrow \infty} \Omega_k}$, there exists a sequence $(\mathbf{x}_i)_{i \in \mathbb{N}} \subset \liminf_{k \rightarrow \infty} \Omega_k$ satisfying $\lim_{i \rightarrow \infty} \mathbf{x}_i = \hat{\mathbf{w}}$. To prove that $\hat{\mathbf{w}} \in \mathcal{E}_\epsilon (\Leftrightarrow \mathbf{B}^H \hat{\mathbf{w}} \in B_\epsilon)$, it is sufficient to show the inclusion $\liminf_{k \rightarrow \infty} \Omega_k \subset \mathcal{E}_\epsilon$ and the closedness of \mathcal{E}_ϵ . The closedness follows from the closedness of B_ϵ and the continuity of the linear mapping \mathbf{B}^H . The closedness of \mathcal{E}_ϵ ensures that $\liminf_{k \rightarrow \infty} \Omega_k = \lim_{m \rightarrow \infty} \bigcup_{k=0}^m \bigcap_{n \geq k} \Omega_n = \lim_{m \rightarrow \infty} \bigcap_{n \geq m} \Omega_n \subset \mathcal{E}_\epsilon$, which completes the proof. \square

REFERENCES

- [1] M. Yukawa and I. Yamada, "Dual-domain adaptive beamforming based on direction-of-arrival information," in *Proc. SSP*, Nice, France, Jun. 2011.
- [2] M. Yukawa, Y. Sung, and G. Lee, "Adaptive interference suppression in MIMO multiple access channels based on dual-domain approach," in *Proc. ITC-CSCC*, Gyeongju, Korea, Jun. 2011.
- [3] GreeTouch Consortium, "GreenTouch 2010–2011 Ann. Rep.," Mobile Communications Working Group, 2011.
- [4] X. Wang and H. V. Poor, "Robust adaptive array for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1352–1366, 1998.
- [5] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [6] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.
- [7] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.
- [8] Z. Tian, K. L. Bell, and H. L. V. Trees, "A recursive least squares implementation for LCMP beamforming under quadratic constraint," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1138–1145, Jun. 2001.
- [9] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2002.
- [10] J. Li and P. Stoica, *Robust Adaptive Beamforming*. New York, NY, USA: Wiley, 2006.
- [11] M. Yukawa, K. Slavakis, and I. Yamada, "Multi-domain adaptive learning based on feasibility splitting and adaptive projected subgradient method," *IEICE Trans. Fundamentals*, vol. E93-A, no. 2, pp. 456–466, Feb. 2010.
- [12] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numer. Funct. Anal. Optim.*, vol. 25, no. 7&8, pp. 593–617, 2004.
- [13] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections: A unifying framework for linear and nonlinear classification and regression tasks," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 97–123, Jan. 2011.
- [14] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients," *IEEE Trans. Neural Netw.*, vol. 23, no. 3, pp. 425–438, Mar. 2012.
- [15] M. Yukawa and I. Yamada, "A deterministic analysis of linearly constrained adaptive filtering algorithms," in *Proc. EUSIPCO*, 2011, pp. 131–135.
- [16] J. A. Apolinário, Jr., S. Werner, P. S. R. Diniz, and T. I. Laakso, "Constrained normalized adaptive filters for CDMA mobile communications," in *Proc. EUSIPCO*, Island of Rhodes, Greece, 1998, vol. IV, pp. 2053–2056.

- [17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [18] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
- [19] K. Slavakis, I. Yamada, and N. Ogura, "Adaptive projected subgradient method over the fixed point set of strongly attracting non-expansive mappings," *Numer. Funct. Anal. Optim.*, vol. 27, no. 7&8, pp. 905–930, 2006.
- [20] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, 2003.
- [21] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, pp. 1702–1715, Jul. 2003.
- [22] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, pp. 1684–1696, 2005, 5.
- [23] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ, USA: Wiley, 2003.



Masahiro Yukawa (S'05–M'07) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology in 2002, 2004, and 2006, respectively.

After studying as a Visiting Researcher with the University of York, U.K., from October 2006 to March 2007, he worked as a Special Postdoctoral Researcher for the Next Generation Mobile Communications Laboratory at RIKEN, Saitama, Japan, from April 2007 to March 2008, and for the Brain Science Institute at RIKEN from April 2008 to March 2010. From August 2008 to November 2008,

he was a Guest Researcher at the Associate Institute for Signal Processing, the Technical University of Munich, Germany. He was an Associate Professor with the Department of Electrical and Electronic Engineering, Niigata University, Japan, from April 2010 to March 2013. He is currently an Assistant Professor with the Department of Electronics and Electrical Engineering, Keio University, Japan. His current research interests are in mathematical signal processing, nonlinear adaptive filtering, and sparse signal processing.

Dr. Yukawa serves as an Associate Editor for the IEICE TRANSACTIONS ON FUNDAMENTALS OF ELECTRONICS, COMMUNICATIONS AND COMPUTER SCIENCES, and the Springer *Journal of Multidimensional Systems and Signal Processing*. From April 2005 to March 2007, he was a recipient of the Research Fellowship of the Japan Society for the Promotion of Science (JSPS). He received the Excellent Paper Award and the Young Researcher Award from

the IEICE in 2006 and in 2010, respectively, the Yasujiro Niwa Outstanding Paper Award from Tokyo Denki University in 2007, and the Ericsson Young Scientist Award from Nippon Ericsson in 2009. He is a member of the Institute of Electrical, Information and Communication Engineers (IEICE) of Japan.



Youngchul Sung (S'92–M'93–SM'09) received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, in electronics engineering in 1993 and 1995, respectively. After working with LG Electronics Ltd., Seoul, from 1995 to 2000, he joined the Ph.D. program in 2000 and received the Ph.D. degree in 2005 in electrical and computer engineering from Cornell University, Ithaca, NY.

From 2005 to 2007, he was a Senior Engineer in the Corporate R & D Center of Qualcomm, Inc., San Diego, CA, and participated in the design of Qualcomm's 3GPP R6 WCDMA Basestation Modem. Since 2007 he has been on the faculty of the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon. His research interests include signal processing for communications, statistical signal processing, and asymptotic statistics with applications to next generation wireless communications and related areas.

Dr. Sung is an Associate Member of the IEEE Signal Processing Society Signal Processing for Communications and Networking (SPCOM) Technical Committee, an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, a member of Signal and Information Processing Theory and Methods (SIPTM) Technical Committee of Asia-Pacific Signal and Information Processing Association (APSIPA), Vice Chair of the IEEE ComSoc Asia-Pacific Board MCC, and TPC member of conferences including Globecom 2012/2011/2010/2009, ICC 2011, MILCOM 2010, DCOSS 2010, WiOpt 2009 and its sponsorship chair, APSIPA 2012/2011/2010/2009, IEEE SAM 2008.



Gilwon Lee (S'10) received the B.S. and M.S. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in electrical engineering in 2010 and 2012, respectively.

He is currently with the Wireless Information Systems Research Lab in KAIST, working towards the Ph.D. degree. His research interests include signal processing for communications and nonlinear adaptive filtering.