An Iterative Algorithm for the Computation of the MVDR Filter

Dimitris A. Pados, *Member, IEEE,* and George N. Karystinos, *Student Member, IEEE*

***Abstract*—Statistical conditional optimization criteria lead to the development of an iterative algorithm that starts from the matched filter (or constraint vector) and generates a sequence of filters that converges to the minimum-variance-distortionless-response (MVDR) solution for any positive definite input autocorrelation matrix. Computationally, the algorithm is a simple, noninvasive, recursive procedure that avoids any form of explicit autocorrelation matrix inversion, decomposition, or diagonalization. Theoretical analysis reveals basic properties of the algorithm and establishes formal convergence.**

When the input autocorrelation matrix is replaced by a conventional sample-average (positive definite) estimate, the algorithm effectively generates a sequence of MVDR filter estimators; the bias converges rapidly to zero and the covariance trace rises slowly and asymptotically to the covariance trace of the familiar sample-matrix-inversion (SMI) estimator. In fact, formal convergence of the estimator sequence to the SMI estimate is established. However, for short data records, it is the early, nonasymptotic elements of the generated sequence of estimators that offer favorable bias/covariance balance and are seen to outperform in mean-square estimation error, constraint-LMS, RLS-type, orthogonal multistage decomposition, as well as plain and diagonally loaded SMI estimates. An illustrative interference suppression example is followed throughout this presentation.

***Index Terms*—Adaptive filters, algorithms, code division multiaccess, estimation, interference suppression, iterative methods, least mean square methods.**

I. Introduction

MINIMUM-variance-distortionless-response (MVDR) filtering refers to the problem of identifying a linear filter that minimizes the variance at its output, while at the same time, the filter maintains a distortionless response toward a specific input vector direction of interest. In mathematical terms, if $\mathbf{r}$ is a random, zero mean without loss of generality, complex input vector of dimension L, $\mathbf{r} \in \mathcal{C}^L$, processed by an L-tap filter $\mathbf{w} \in \mathcal{C}^L$, then the filter output variance is $\mathbf{w}^H\mathbf{R}\mathbf{w}$, where $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$ is the input autocorrelation matrix ($E\{\cdot\}$ denotes the statistical expectation operation and $\mathbf{x}^H$ denotes the Hermitian—that is, transpose conjugate—of $\mathbf{x}$). The MVDR filter minimizes $\mathbf{w}^H\mathbf{R}\mathbf{w}$ and simultaneously satisfies $\mathbf{w}^H\mathbf{v} = 1$, or more generally, $\mathbf{w}^H\mathbf{v} = \rho \in \mathcal{C}$, where $\mathbf{v}$ is the input signal vector direction to be protected. In this setup, MVDR filtering is a standard linear constraint optimization problem and a conventional Lagrange multipliers procedure leads to the well-known solution

$$\mathbf{w}_{\mathrm{MVDR}} = \rho^* \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}} \tag{1}$$

where ρ^* denotes the conjugate of the desired response $\mathbf{w}^H\mathbf{v} = \rho$. Extensive tutorial treatments of MVDR filtering can be found in many sources, for example in [1] and [2], along with historical notes on the early work by Capon [3] and Owsley [4].

MVDR filtering has long been a workhorse for unsupervised signal processing applications where a desired scalar filter output $d \in \mathcal{C}$ cannot be identified or cannot be assumed available for each input $\mathbf{r} \in \mathcal{C}^L$. Prime examples include radar and array processing problems where the constraint vector $\mathbf{v}$ is usually referred to as the "target" or "look" direction of interest. We may also observe the close relationship between the MVDR filter and the minimum mean square error (MMSE) or Wiener [1], [2] filter. Indeed, if the constraint vector $\mathbf{v}$ is chosen to be the statistical cross-correlation vector between the desired output d and the input vector $\mathbf{r}$, that is, if $\mathbf{v} = E\{\mathbf{r}d^*\}$, then the MVDR and MMSE filters become scaled versions of each other: $c\mathbf{R}^{-1}\mathbf{v}$, $c \in \mathcal{C}$.

Conventionally, the computation of the MVDR/MMSE filter in (1) begins with the calculation of the inverse of the input autocorrelation matrix $\mathbf{R}^{-1}$ (assuming that the Hermitian $\mathbf{R}$ is strictly positive definite and, hence, invertible) based on numerical iterative diagonalization linear algebra procedures [5]. Then, the calculated matrix $\mathbf{R}^{-1}$ is used for the linear transformation of the constraint vector $\mathbf{v}$, followed by $\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$ normalization, as necessary. In this paper, we present an iterative algorithm for the *direct* calculation of the MVDR vector $\mathbf{w}_{\mathrm{MVDR}}$ in (1). The algorithm is a noninvasive procedure where no explicit matrix inversion/eigendecomposition/diagonalization is attempted. The MVDR computation algorithm creates a sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \cdots$, that begins from $\mathbf{w}_0 = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and converges to the MVDR filter ($\mathbf{w}_\infty = \mathbf{w}_{\mathrm{MVDR}}$). At each step $n = 1, 2, \cdots, \mathbf{w}_n$ is given as a simple, direct function of $\mathbf{R}$, $\mathbf{v}$, and $\mathbf{w}_{n-1}$.

Interestingly, the development of the iterative algorithm is founded solely on statistical signal processing principles. In that respect, this may be seen as another case where the mature statistical signal processing field is able to return alternative solutions and interpretations to basic numerical linear algebra problems. Understandably, the whole task may seem so far to be an academic exercise to produce yet another method for the solution of linear systems of equations of the form $\mathbf{R}\mathbf{w} = c\mathbf{v}$. However, the real motivation behind the work presented in this paper is

Manuscript received July 27, 1999; revised October 24, 2000. This work was supported by the National Science Foundation under Grant CCR-9805359 and the U.S. Air Force Office of Scientific Research under Grant F49620-99-1-0035. The associate editor coordinating the review of this paper and approving it for publication was Prof. Hideaki Sakai.

The authors are with the Department of Electrical Engineering, State University of New York at Buffalo, Amherst, NY 14260 USA (e-mail: pados@eng.buffalo.edu; cary@buffalo.edu).

Publisher Item Identifier S 1053-587X(01)00634-1.

1053–587X/01$10.00 © 2001 IEEE

adaptive signal processing, where the input autocorrelation matrix $\mathbf{R}$ is assumed unknown and it is sample-average estimated by a data record of M points $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M$:

$$\hat{\mathbf{R}}(M) = \frac{1}{M} \sum_{m=1}^{M} \mathbf{r}_m \mathbf{r}_m^H. \tag{2}$$

When $\mathbf{R}$ is substituted by $\hat{\mathbf{R}}(M)$ in the recursively generated sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \cdots$, the corresponding filter estimators $\hat{\mathbf{w}}_n(M)$ offer the means for effective control over the filter estimator bias versus (co-) variance tradeoff [6]. Starting from the *zero*-variance, high-bias (for nonwhite inputs) $\hat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ estimate, we can go all the way up to the unbiased, yet high-variance for short data record sizes M, $\hat{\mathbf{w}}_\infty(M)$ estimate and anywhere in between, $\hat{\mathbf{w}}_n(M)$, $1 \leq n < \infty$. As a result, adaptive filters from this newly developed class can be seen to outperform in expected norm-square estimation error $E\{\|\hat{\mathbf{w}}(M) - \mathbf{w}_{\mathrm{MVDR}}\|^2\}$, (constraint-)LMS [7]–[11], sample-matrix-inversion (SMI) [12] with or without diagonal loading [13], RLS-type [14], [15], and orthogonal multistage decomposition [30], [31] adaptive filter implementations. An illustrative case study drawn from the code division multiple access (CDMA) communications literature is followed throughout this paper. It may be also worth mentioning that the familiar trial-and-error tuning to problem and data-record-size specifics of the real-valued LMS gain [16] or RLS inverse matrix initialization constant [17] or SMI diagonal loading parameter [13] that plagues field practitioners is now replaced by an integer choice of one of the recursively generated filters.

The material included in this paper is organized as follows. In Section II, we present the algorithmic developments. Convergence analysis is carried out in Section III. Filter estimation issues are discussed in Section IV along with simulation examples. A few concluding remarks are given in Section V.

II. Algorithmic Developments

For a given constraint vector $\mathbf{v} \in \mathcal{C}^L$ consider the set of filters $\mathcal{D} = \{\mathbf{w} \in \mathcal{C}^L : \mathbf{w} = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} + \mathbf{u}, \mathbf{u} \in \mathcal{C}^L$, and $\mathbf{v}^H\mathbf{u} = 0\}$. $\mathcal{D}$ is the class of all filters $\mathbf{w}$ in $\mathcal{C}^L$ that are distortionless in $\mathbf{v}$, that is, $\mathbf{w}^H\mathbf{v} = \rho$. Certainly, $\mathbf{w}_{\mathrm{MVDR}}$ is in $\mathcal{D}$ for any given input autocorrelation matrix $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$. In this section, we develop an iterative algorithm for the computation of the $\mathbf{u}$ component of the MVDR filter.

Historically, algorithmic designs that focus on the MVDR filter part $\mathbf{u}$ that is orthogonal to the constraint vector, or "look," direction $\mathbf{v}$ have been widely pursued in the array processing literature and have been known as Applebaum/Howells arrays [18], [19], beam-space partially adaptive processors [20], or generalized sidelobe cancelers (GSC) [21]. Recent developments have been influenced by principal component analysis reduced-rank processing principles [22]–[24]. In general, the MVDR filter part $\mathbf{u}$ ($\mathbf{u}^H\mathbf{v} = 0$) has been approximated by $\mathbf{u}_{L\times 1} \simeq -\mathbf{B}_{L\times L}\mathbf{T}_{L\times p}\mathbf{w}^{\mathrm{GSC}}_{p\times 1}$, where

- $\mathbf{B}$ so-called "blocking matrix" (an $L \times L$ orthogonal projection operator such as $\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2$, where $\mathbf{I}$ is the identity matrix);
- $\mathbf{T}$ rank-reducing matrix with $1 \leq p < L-1$ columns to be designed;
- $\mathbf{w}^{\mathrm{GSC}}$ MS-optimum vector of weights of the p columns of $\mathbf{T}$ ($\mathbf{w}^{\mathrm{GSC}} = (\rho^*/\|\mathbf{v}\|^2)[\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{B}\mathbf{T}]^{-1}\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{v}$ [1], [2], [20]).

Owsley [25] chose the p columns of matrix $\mathbf{T}$ to be the p maximum eigenvalue eigenvectors of the *disturbance-only* autocorrelation matrix under the—not true in general—assumption that the disturbance-only eigenvectors are not rotated by the blocking matrix $\mathbf{B}$ being used. Van Veen [26], as well as Haimovich and Bar-Ness [27], addressed this concern by choosing the p maximum eigenvalue eigenvectors of the blocked data autocorrelation matrix $\mathbf{B}^H\mathbf{R}\mathbf{B}$. If, however, the columns of $\mathbf{T}$ *have to be* eigenvectors of $\mathbf{B}^H\mathbf{R}\mathbf{B}$, then the best way to choose them in the minimum output variance p-rank approximation sense was presented by Byerly and Roberts [28]: Select the p eigenvectors $\mathbf{q}_i$ of $\mathbf{B}^H\mathbf{R}\mathbf{B}$, with corresponding eigenvalues λ_i, that maximize $|\mathbf{v}^H\mathbf{R}\mathbf{q}_i|^2/\lambda_i$, $i = 1, \cdots, p$. This design algorithm has also been known as "cross-spectral metric" reduced-rank processing [29]. Non-eigenbased, l-stage, $1 \leq l \leq L-1$, orthogonal decomposition and synthesis of $\mathbf{u}$ was pursued in [30] and [31], filter decomposition using canonical correlations was considered in [32], and modular designs through factorization of the orthogonal projection operator were developed in [33]. A different approach from a different point of view is attempted in this work. A *conditional* statistical optimization procedure is shown to offer the means for *exact* computation of $\mathbf{u}$ as the convergence point of an infinite series of the form $-\sum_{n=1}^{\infty}\mu_n\mathbf{g}_n$, $\mu_n \in \mathcal{R}^+$, $\mathbf{g}_n \in \mathcal{C}^L$, and $\mathbf{g}_n^H\mathbf{v} = 0\ \forall n = 1, 2, \cdots$.

We begin the algorithmic developments from the conventional matched filter (MF) with the desired response ρ

$$\mathbf{w}_0 = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} \tag{3}$$

which is MVDR (MMSE) optimum for white $\mathcal{C}^L$ vector inputs (when $\mathbf{R} = \sigma^2\mathbf{I}$, $\sigma > 0$). We recall, w.l.o.g. and for notational simplicity, that we assume throughout this presentation that the input vectors $\mathbf{r} \in \mathcal{C}^L$ are zero mean. Next, we incorporate in $\mathbf{w}_0$ an "auxiliary" vector component that is orthogonal to $\mathbf{v}$, and we form (Fig. 1)

$$\mathbf{w}_1 = \mathbf{w}_0 - \mu_1\mathbf{g}_1 = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} - \mu_1\mathbf{g}_1 \tag{4}$$

where $\mathbf{g}_1 \in \mathcal{C}^L - \{\mathbf{0}\}$, $\mu_1 \in \mathcal{C}$, and $\mathbf{g}_1^H\mathbf{v} = 0$. We pretend for a moment that the orthogonal auxiliary vector $\mathbf{g}_1$ is arbitrary but nonzero and fixed, and we concentrate on the selection of the scalar μ_1. The value of μ_1 that minimizes the variance of the output of the filter $\mathbf{w}_1$ can be found by direct differentiation of $E\{|\mathbf{w}_1^H\mathbf{r}|^2\}$ or simply as the value that minimizes the MS error between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mu_1^*\mathbf{g}_1^H\mathbf{r}$. This is essentially a scalar version of the GSC weight determination problem and we present the solution in the form of a proposition [34].

Proposition 1: The scalar μ_1 that minimizes the variance at the output of $\mathbf{w}_1$ or equivalently minimizes the MS error between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mu_1^*\mathbf{g}_1^H\mathbf{r}$ is

$$\mu_1 = \frac{\mathbf{g}_1^H\mathbf{R}\mathbf{w}_0}{\mathbf{g}_1^H\mathbf{R}\mathbf{g}_1} \tag{5}$$

where $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$ is the input autocorrelation matrix. □

Since $\mathbf{g}_1$ is set to be orthogonal to $\mathbf{v}$, (5) shows that if the vector $\mathbf{R}\mathbf{w}_0$ happens to be "on $\mathbf{v}$" (that is $\mathbf{R}\mathbf{w}_0 = (\mathbf{v}^H\mathbf{R}\mathbf{w}_0/\|\mathbf{v}\|^2)\mathbf{v}$ or equivalently $(\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_0 = \mathbf{0}$), then $\mu_1 = 0$. Indeed, if $\mathbf{R}\mathbf{w}_0 = (\mathbf{v}^H\mathbf{R}\mathbf{w}_0/\|\mathbf{v}\|^2)\mathbf{v}$, then $\mathbf{w}_0$ is *already* the MVDR filter. To avoid this trivial case and continue with our developments, we suppose that $\mathbf{R}\mathbf{w}_0 \neq (\mathbf{v}^H\mathbf{R}\mathbf{w}_0/\|\mathbf{v}\|^2)\mathbf{v}$. By inspection, we also observe that for the MS-optimum value of μ_1, the product $\mu_1\mathbf{g}_1 = (\mathbf{g}_1^H\mathbf{R}\mathbf{w}_0/\mathbf{g}_1^H\mathbf{R}\mathbf{g}_1)\mathbf{g}_1$ is independent of the norm of $\mathbf{g}_1$. Hence, so is $\mathbf{w}_1$. At this point, we decide to choose the auxiliary vector $\mathbf{g}_1$ as the normalized vector that maximizes the magnitude of the cross-correlation between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mathbf{g}_1^H\mathbf{r}$, under the constraint that $\mathbf{g}_1^H\mathbf{v} = 0$ and $\mathbf{g}_1^H\mathbf{g}_1 = 1$:

$$\mathbf{g}_1 = \arg\max_{\mathbf{g}} |E\{\mathbf{w}_0^H\mathbf{r}(\mathbf{g}^H\mathbf{r})^*\}| = \arg\max_{\mathbf{g}} |\mathbf{w}_0^H\mathbf{R}\mathbf{g}|$$
$$\text{subject to } \mathbf{g}^H\mathbf{v} = 0 \text{ and } \mathbf{g}^H\mathbf{g} = 1. \tag{6}$$

For the sake of mathematical accuracy, we note that both the criterion function $|\mathbf{w}_0^H\mathbf{R}\mathbf{g}|$ to be maximized, as well as the orthogonality constraints, are phase invariant. In other words, if $\mathbf{g}_1$ satisfies (6), so does $\mathbf{g}_1 e^{j\phi}$ for any phase ϕ. Without loss of generality, to avoid any ambiguity in our presentation and to have a uniquely defined auxiliary vector, we seek the one and only auxiliary vector that satisfies (6) and places the cross-correlation value on the positive real line ($\mathbf{w}_0^H\mathbf{R}\mathbf{g} > 0$). This constraint optimization problem was first posed and solved in [35], where the filter $\mathbf{w}_1$ in (4) was used for multiple access interference suppression in multipath CDMA communication channels. Intuitively, the maximum magnitude cross-correlation criterion, as defined in (6), strives to identify the orthonormal to $\mathbf{v}$ auxiliary vector that can capture the most of the interference present in $\mathbf{w}_0^H\mathbf{r}$. The solution, which is derived through conventional Lagrange multipliers optimization, is given below.

Proposition 2: Suppose that $(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_0 \neq \mathbf{0}$ ($\mathbf{w}_0 \neq \mathbf{w}_{\mathrm{MVDR}}$). Then, the auxiliary vector

$$\mathbf{g}_1 = \frac{\mathbf{R}\mathbf{w}_0 - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_0}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_0 - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_0}{\|\mathbf{v}\|^2}\mathbf{v}\right\|} \tag{7}$$

maximizes the magnitude of the cross-correlation between $\mathbf{w}_0^H\mathbf{r} = (\rho/\|\mathbf{v}\|^2)\mathbf{v}^H\mathbf{r}$ and $\mathbf{g}_1^H\mathbf{r}$, $|\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1|$, subject to the constraints $\mathbf{g}_1^H\mathbf{v} = 0$ and $\mathbf{g}_1^H\mathbf{g}_1 = 1$. In addition, $\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1$ is real positive ($\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1 > 0$). □

Thus far, we have defined $\mathbf{w}_0$ in (3) and $\mathbf{w}_1$ in (4) with $\mathbf{g}_1$ and μ_1 given by (7) and (5), respectively. We are now ready to take the algorithmic developments one step further and consider a filter of the form (see Fig. 1)

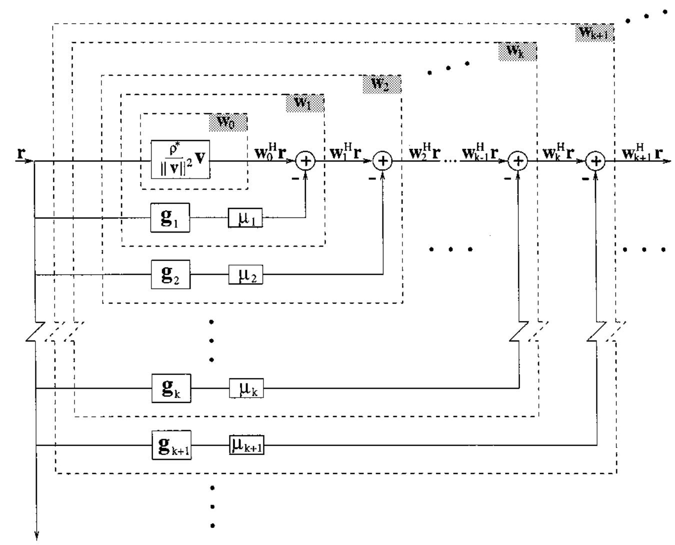

Fig. 1. Block diagram representation of the iteratively generated sequence of filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \cdots$.

$$\mathbf{w}_2 = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} - \sum_{i=1}^{2}\mu_i\mathbf{g}_i = \mathbf{w}_1 - \mu_2\mathbf{g}_2. \tag{8}$$

$\mathbf{g}_2$ and μ_2 will be *conditionally* optimized given the previously identified auxiliary vector $\mathbf{g}_1$ and scalar μ_1. $\mathbf{g}_2$ is chosen as the vector that maximizes the magnitude of the cross-correlation between $\mathbf{w}_1^H\mathbf{r}$ and $\mathbf{g}_2^H\mathbf{r}$, $|\mathbf{w}_1^H\mathbf{R}\mathbf{g}_2|$, again subject to $\mathbf{g}_2^H\mathbf{v} = 0$ and $\mathbf{g}_2^H\mathbf{g}_2 = 1$. Arguing as in Proposition 2, we suppose that $(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_1 \neq \mathbf{0}$ ($\mathbf{w}_1 \neq \mathbf{w}_{\mathrm{MVDR}}$), and we find

$$\mathbf{g}_2 = \frac{\mathbf{R}\mathbf{w}_1 - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_1}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_1 - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_1}{\|\mathbf{v}\|^2}\mathbf{v}\right\|}. \tag{9}$$

The value of μ_2 in (8) that minimizes the output variance of $\mathbf{w}_2$, $E\{|\mathbf{w}_2^H\mathbf{r}|^2\}$, given $\mathbf{w}_1$ and $\mathbf{g}_2$, is the same value that minimizes the MS error between $\mathbf{w}_1^H\mathbf{r}$ and $\mu_2^*\mathbf{g}_2^H\mathbf{r}$. Arguing as in Proposition 1, we find

$$\mu_2 = \frac{\mathbf{g}_2^H\mathbf{R}\mathbf{w}_1}{\mathbf{g}_2^H\mathbf{R}\mathbf{g}_2}. \tag{10}$$

The iterative algorithm for the generation of an infinite sequence of filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \cdots$ is already taking shape. Formally, we just need to specify the inductive step. Assuming that the filter $\mathbf{w}_k = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} - \sum_{i=1}^{k}\mu_i\mathbf{g}_i$ has been identified for some $k \geq 1$ and $\mathbf{w}_k \neq \mathbf{w}_{\mathrm{MVDR}}$, we define $\mathbf{w}_{k+1}$ as follows:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{k+1}\mathbf{g}_{k+1} \quad \text{where} \tag{11}$$

$$\mathbf{g}_{k+1} = \frac{\mathbf{R}\mathbf{w}_k - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_k}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_k - \dfrac{\mathbf{v}^H\mathbf{R}\mathbf{w}_k}{\|\mathbf{v}\|^2}\mathbf{v}\right\|} \tag{12}$$

is the orthonormal with respect to $\mathbf{v}$ auxiliary vector that, given $\mathbf{w}_k$, maximizes conditionally the cross-correlation magnitude $|E\{\mathbf{w}_k^H\mathbf{r}(\mathbf{g}_{k+1}^H\mathbf{r})^*\}| = |\mathbf{w}_k^H\mathbf{R}\mathbf{g}_{k+1}|$, and

$$\mu_{k+1} = \frac{\mathbf{g}_{k+1}^H\mathbf{R}\mathbf{w}_k}{\mathbf{g}_{k+1}^H\mathbf{R}\mathbf{g}_{k+1}} \tag{13}$$

is the scalar that minimizes the MS error between $\mathbf{w}_k^H\mathbf{r}$ and $\mu_{k+1}^*\mathbf{g}_{k+1}^H\mathbf{r}$ (minimizes $E\{|\mathbf{w}_{k+1}^H\mathbf{r}|^2\}$).

It is important to note that while the generated auxiliary vectors $\mathbf{g}_1, \mathbf{g}_2, \cdots$ are all constrained to be orthogonal to $\mathbf{v}$, orthogonality among the auxiliary vectors is *not* imposed. This is in sharp contrast to previous work that involved filtering with up to $L-1$ orthogonal to each other and to $\mathbf{v}$ auxiliary vectors (AVs) [36]–[38], where L is the data input vector dimension. We observe, however, that *successive* AVs generated by the proposed recursive conditional optimization procedure (11)–(13) *do* come up orthogonal: $\mathbf{g}_i^H\mathbf{g}_{i+1} = 0, \forall i = 1,2,3,\cdots$. For completeness purposes, we present this observation in the form of the following Lemma. A brief algebraic proof is given in the Appendix.

Lemma 1: Successive auxiliary vectors generated through (11)–(13) are orthogonal: $\mathbf{g}_i^H\mathbf{g}_{i+1} = 0, i = 1,2,3,\cdots$. □

The algorithm developed in this section is summarized in Fig. 2. The conceptual simplicity of the conditional statistical optimization process that we adopted led to a computationally simple recursion. In Fig. 2, we chose to drop the unnecessary, as previously explained, normalization of the auxiliary vectors and we also factorized their numerator to make the orthogonal projection operator apparent. Formal convergence of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \cdots$ to the MVDR filter $\rho^*(\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})$ is established in the following section.

Initialization:

$$\mathbf{w}_0 := \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}.$$

Iterative computation:

For n=1, 2, ... do
begin

$\mathbf{g}_n := \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\,\mathbf{w}_{n-1}$

if $\mathbf{g}_n = \mathbf{0}$ then EXIT

$\mu_n := \frac{\mathbf{g}_n^H\mathbf{R}\,\mathbf{w}_{n-1}}{\mathbf{g}_n^H\mathbf{R}\,\mathbf{g}_n}$

$\mathbf{w}_n := \mathbf{w}_{n-1} - \mu_n\mathbf{g}_n$

end

Output:

Filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

Fig. 2. Algorithm for the iterative generation of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \cdots$.

III. Convergence Analysis

Posing and solving an inductively defined statistical conditional optimization problem led to the design of the iterative algorithm summarized in Fig. 2. Next, we switch our attention from design to analysis. We begin with a theorem that tabulates basic properties of the generated sequence of auxiliary-vector weights $\{\mu_n\}, n = 1,2,\cdots$, and the sequence of corresponding auxiliary vectors $\{\mathbf{g}_n\}, n = 1,2,\cdots$. This theorem—the proof of which is given in the Appendix—will also serve as the foundation for establishing convergence of the filter sequence $\{\mathbf{w}_n\}$, $n = 0,1,2,\cdots$, to the MVDR filter $\rho^*(\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})$.

Theorem 1: Let $\mathbf{R}$ be a Hermitian positive definite matrix. Consider the iterative algorithm of Fig. 2.

i) The generated sequence of auxiliary-vector weights $\{\mu_n\}, n = 1,2,\cdots$, is real-valued, positive, and bounded

$$0 < \frac{1}{\lambda_{\max}} \le \mu_n \le \frac{1}{\lambda_{\min}}, \qquad n = 1,2,\cdots \tag{14}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum, correspondingly, eigenvalues of $\mathbf{R}$.

ii) The sequence of auxiliary vectors $\{\mathbf{g}_n\}, n = 1,2,\cdots$, converges to the $\mathbf{0}$ vector

$$\lim_{n\to\infty}\mathbf{g}_n = \mathbf{0}. \quad \square \tag{15}$$

We are now ready to show that $\mathbf{w}_n \to \mathbf{w}_{\text{MVDR}}$. We break the proof into two parts. First, we show that $\mathbf{w}_n$ converges. Then, we establish the actual limit. We begin from Theorem 1, Part (ii), and the definition of $\mathbf{g}_n$ in Fig. 2, which imply

$$\left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_n \underset{n\to\infty}{\longrightarrow} \mathbf{0}. \tag{16}$$

We focus on the vector $\mathbf{R}\mathbf{w}_n$ and we notice that (16) states that the component of $\mathbf{R}\mathbf{w}_n$ that is orthogonal to $\mathbf{v}$ goes to $\mathbf{0}$, as $n \to \infty$. Therefore, to prove that $\mathbf{R}\mathbf{w}_n$ converges somewhere, it suffices to show that the projection of $\mathbf{R}\mathbf{w}_n$ onto $\mathbf{v}$, $(\mathbf{v}^H\mathbf{R}\mathbf{w}_n/\|\mathbf{v}\|^2)\mathbf{v}$, converges. To achieve that, we continue from (16) and we multiply both sides by $\mathbf{R}^{-1}$ to obtain

$$\mathbf{w}_n - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{R}^{-1}\mathbf{v} \underset{n\to 0}{\longrightarrow} \mathbf{0}. \tag{17}$$

For simplicity of notation, we define $c_n \triangleq (\mathbf{w}_n^H\mathbf{R}\mathbf{v}/\|\mathbf{v}\|^2), n = 0,1,2,\cdots$. Then, we translate (17) back to the basic definition of convergence in a Euclidean space. Equation (17) implies that for every given $\epsilon > 0$, there exists some $N > 0$ such that for every $n > N$

$$\|\mathbf{w}_n - c_n^*\mathbf{R}^{-1}\mathbf{v}\| < \frac{\epsilon}{\|\mathbf{v}\|}. \tag{18}$$

Multiplying both sides of (18) by $\|\mathbf{v}\|$ and using the fact that $\|\mathbf{w}_n - c_n^*\mathbf{R}^{-1}\mathbf{v}\|\,\|\mathbf{v}\| \ge \|(\mathbf{w}_n - c_n^*\mathbf{R}^{-1}\mathbf{v})^H\mathbf{v}\|$, we obtain

$$\|\mathbf{w}_n^H\mathbf{v} - c_n\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}\| < \epsilon. \tag{19}$$

We recall that by design, $\mathbf{w}_n^H\mathbf{v} = \rho\ \forall n = 0,1,2,\cdots$. Therefore, (19) becomes $\|\rho - c_n\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}\| < \epsilon\ \forall n > N$ and we immediately conclude that c_n converges to $\rho/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$ or

$$c_n^*\mathbf{v} = \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{v} \longrightarrow \frac{\rho^*}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\mathbf{v} \quad \text{as } n\to\infty. \tag{20}$$

Equations (16) and (20) together establish that $\mathbf{R}\mathbf{w}_n$ converges; hence, $\mathbf{w}_n$ converges. Knowing this, we are now free to take limits on the left-hand side of (17) to derive

$$\lim_{n\to\infty} \mathbf{w}_n = \rho^* \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}} \tag{21}$$

which establishes formal convergence of the filter sequence $\mathbf{w}_n$ to the MVDR filter. Another interpretation of the auxiliary-vector algorithm in the context of steepest-descent theory is given in the Appendix.

We conclude this section with an illustration. We draw a signal model example from the synchronous direct-sequence CDMA (DS/CDMA) communications literature and we assume that the input signal vector $\mathbf{r} \in \mathcal{R}^L$ is given by

$$\mathbf{r} = \sum_{k=1}^{K} \sqrt{E_k} b_k \mathbf{s}_k + \mathbf{n}. \tag{22}$$

In this setup, K denotes the total number of signals ("users") present and each signal is defined through an L-dimensional, normalized, binary-antipodal vector waveform (or "user signature") $\mathbf{s}_k$, $k = 1, 2, \cdots, K$. The signature vector dimension L is usually referred to as the system "spreading gain." With respect to the kth user signal, E_k is the received signal energy and $b_k \in \{-1, +1\}$ is the information bit modeled as a random variable with equally probable values and assumed to be statistically independent from all other user bits b_j, $j \neq k$. Additive white Gaussian noise contributions are accounted for by $\mathbf{n}$ with autocorrelation matrix $E\{\mathbf{n}\mathbf{n}^T\} = \sigma^2 \mathbf{I}_{L\times L}$ ($\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$). With this notation and normalized user signatures, the signal-to-noise ratio (SNR) of the kth user signal is defined by $\text{SNR}_k \triangleq 10\log_{10}(E_k/\sigma^2)$ dB, $k = 1, 2, \cdots, K$.

MVDR (MMSE) filtering for DS/CDMA-type problems has attracted a great deal of interest [39]–[44]. If we wish to recover the information bits of, say, user *1*, then all other signals constitute multiple-access interference and the MVDR filter is built with constraint vector $\mathbf{v} = \mathbf{s}_1$, desired response $\mathbf{w}^T\mathbf{s}_1 = 1$, and autocorrelation matrix $\mathbf{R} = \sum_{k=1}^{K} E_k \mathbf{s}_k \mathbf{s}_k^T + \sigma^2\mathbf{I}$. We choose $L = 32$, $K = 13$, and we draw an arbitrary set of signatures $\mathbf{s}_1$, $\mathbf{s}_2, \cdots, \mathbf{s}_{13}$. For purposes of completeness in presentation, the exact signature assignment is given in the Appendix. We fix the SNR of the user of interest at $\text{SNR}_1 = 12$ dB, whereas the "interferers" $k = 2, \cdots, 13$ are at $\text{SNR}_{2-5} = 10$ dB, $\text{SNR}_{6-9} =$ 12 dB, and $\text{SNR}_{10-13} = 14$ dB. This signal setup is maintained without any change throughout the rest of this paper and is used for illustration purposes. Fig. 3 shows how the sequence of filters $\mathbf{w}_0$, $\mathbf{w}_1, \cdots$ generated by the algorithm in Fig. 2 converges to the MVDR solution. The convergence is captured in terms of the norm-square metric $\|\mathbf{w}_n - \mathbf{w}_{\text{MVDR}}\|^2$ as a function of the iteration step (or number of AV's used) n.

IV. Filter Estimation

Consider a constraint vector $\mathbf{v}$ and a Hermitian positive definite autocorrelation matrix $\mathbf{R}$ of an input vector $\mathbf{r} \in \mathcal{C}^L$. The algorithm developed in Section II and analyzed in Section III can be characterized as a "greedy" procedure that, at each iteration step n, identifies an orthogonal-to-$\mathbf{v}$ ("auxiliary") vector $\mathbf{g}_n$

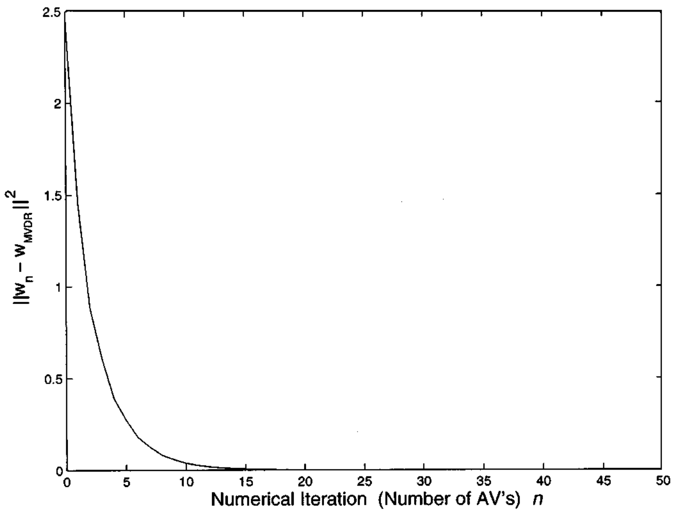

Fig. 3. Convergence of the sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \cdots$, to the MVDR solution for the signal model example in (22).

that can capture the most (in the maximum magnitude cross-correlation sense) of the interference present at the previous iteration filter output $\mathbf{w}_{n-1}^H\mathbf{r}$. Each auxiliary vector $\mathbf{g}_n$ is free to move anywhere in the $\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2$ vector space and accounts for globally present interference without any preconditions or *a priori* approximations of the interference subspace. The algorithm converges to the *exact* MVDR solution.

Assume now that $\mathbf{R}$ is in fact unknown and it is sample-average estimated by a data record of M points: $\hat{\mathbf{R}}(M) = (1/M)\sum_{m=1}^{M} \mathbf{r}_m\mathbf{r}_m^H$. For Gaussian inputs, the Hermitian matrix $\hat{\mathbf{R}}(M)$ is a maximum-likelihood (ML), consistent, unbiased estimator of $\mathbf{R}$ [6], [49]. For a large class of multivariate elliptically contoured input distributions that includes the Gaussian, if $M \geq L$, then $\hat{\mathbf{R}}(M)$ is positive definite (hence invertible) with probability 1 (w.p. 1) [45]–[47]. Then, the analysis of Section III shows that

$$\hat{\mathbf{w}}_n(M) \underset{n\to\infty}{\longrightarrow} \hat{\mathbf{w}}_\infty(M) = \rho^* \frac{[\hat{\mathbf{R}}(M)]^{-1}\mathbf{v}}{\mathbf{v}^H[\hat{\mathbf{R}}(M)]^{-1}\mathbf{v}} \tag{23}$$

where $\hat{\mathbf{w}}_\infty(M)$ is the widely used MVDR filter estimator known as the sample-matrix-inversion (SMI) filter [12].

The output sequence begins from $\hat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$, which is a *zero*-variance, fixed-valued, estimator that may be severely biased ($\hat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} \neq \mathbf{w}_{\text{MVDR}}$), unless $\mathbf{R} = \sigma^2\mathbf{I}$ for some $\sigma > 0$. In the latter trivial case, $\hat{\mathbf{w}}_0(M)$ is already the perfect MVDR filter. Otherwise, the next filter estimator in the sequence $\hat{\mathbf{w}}_1(M)$ has a significantly reduced bias due to the greedy optimization procedure employed at the expense of nonzero estimator (co-)variance. As we move up in the sequence of filter estimators $\hat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \cdots$, the bias decreases rapidly to zero,[1] while the variance rises gracefully to the SMI ($\hat{\mathbf{w}}_\infty(M)$) levels [cf. (23)]. To quantify these remarks, we plot in Fig. 4 the norm-square bias $\|E\{\hat{\mathbf{w}}_n(M)\} - \mathbf{w}_{\text{MVDR}}\|^2$ and the trace of the covariance matrix $E\{[\hat{\mathbf{w}}_n(M) - E\{\hat{\mathbf{w}}_n(M)\}][\hat{\mathbf{w}}_n(M) - E\{\hat{\mathbf{w}}_n(M)\}]^H\}$ as a function of the iteration step n for the signal model example of Fig. 3 and data

[1]The SMI estimator is unbiased for multivariate elliptically contoured input distributions [47], [48]: $E\{\hat{\mathbf{w}}_\infty(M)\} = \mathbf{w}_{\text{MVDR}} = \rho^*(\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})$.

record size $M = 256$.[2] Bias and cov-trace values are calculated from 100 000 independent filter estimator realizations for each iteration point n. Formal, theoretical statistical analysis of the generated estimators $\hat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \cdots$, is beyond the scope of this presentation and will be reported in the future.[3] From the results in Fig. 4 for $M = 256$, we see that the estimators $\hat{\mathbf{w}}_1(M)$, $\hat{\mathbf{w}}_2(M), \cdots$, up to about $\hat{\mathbf{w}}_{20}(M)$ are particularly appealing. In contrast, the estimators $\hat{\mathbf{w}}_n(M)$ for $n > 20$ do not justify their increased cov-trace cost since they have almost nothing to offer in terms of further bias reduction.

The mean-square estimation error expression $E\{\|\hat{\mathbf{w}}_n(M) - \mathbf{w}_{\mathrm{MVDR}}\|^2\}$ captures the bias/variance balance of the individual members of the estimator sequence $\hat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \cdots$. In Fig. 5, we plot the MS estimation error as a function of the iteration step n for the case study in Fig. 4 for $M = 256$ [Part (a)] and $M = 2,048$ [Part (b)]. As a reference, we also include the MS error of the constraint-LMS estimator [7]–[11]

$$\hat{\mathbf{w}}_{\mathrm{LMS}}(m) = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)[\hat{\mathbf{w}}_{\mathrm{LMS}}(m-1) - \mu \mathbf{r}_m \mathbf{r}_m^H \hat{\mathbf{w}}_{\mathrm{LMS}}(m-1)] + \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}, \quad m = 1, \cdots, M \tag{24}$$

with $\hat{\mathbf{w}}_{\mathrm{LMS}}(0) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and some $\mu > 0$ and the RLS estimator [1], [2], [14], [15] with matrix-inversion-lemma-based $\mathbf{R}^{-1}$ estimation

$$\hat{\mathbf{R}}^{-1}(m) = \hat{\mathbf{R}}^{-1}(m-1) - \frac{\hat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m\mathbf{r}_m^H\hat{\mathbf{R}}^{-1}(m-1)}{1 + \mathbf{r}_m^H\hat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m}, \quad m = 1, \cdots, M \tag{25}$$

with $\hat{\mathbf{R}}^{-1}(0) = (1/\epsilon_0)\mathbf{I}$ for some $\epsilon_0 > 0$. Theoretically, it is known that the LMS gain parameter $\mu > 0$ [16] has to be less than $1/(2\lambda_{\max}^{\mathrm{blocked}})$, where $\lambda_{\max}^{\mathrm{blocked}}$ is the maximum eigenvalue of the "blocked-data" $\mathbf{I} - (\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)$ autocorrelation matrix. While this is a theoretical upper bound, practitioners are well aware that empirical, data-dependent "optimization" or "tuning" of the LMS gain $\mu > 0$ or

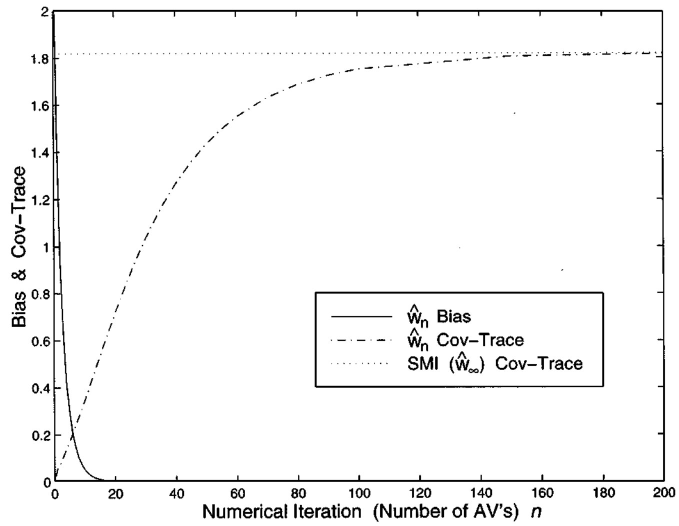

Fig. 4. Norm-square bias and covariance trace for the sequence of estimators $\hat{\mathbf{w}}_n(M)$, $n = 0, 1, \cdots$. The signal model is as in Fig. 3 and $M = 256$.

the RLS initialization parameter $\epsilon_0 > 0$ [17] is necessary to achieve acceptable performance [in our study, we set $\mu = 1/(200 \cdot \lambda_{\max}^{\mathrm{blocked}})$ and $\epsilon_0 = 20$, respectively]. This data-specific tuning frequently results in misleading, over-optimistic conclusions about the short-data-record performance of the LMS/RLS algorithms. In contrast, when the filter estimators $\hat{\mathbf{w}}_n$ generated by the algorithm of Fig. 2 are considered instead, tuning of the real-valued parameters μ and ϵ_0 is virtually replaced by an integer choice among the first several members of the $\{\hat{\mathbf{w}}_n\}$ sequence. Adaptive, data-dependent criteria for the selection of the most appropriate number of AVs for a given data record are developed and reported in [50]. In Fig. 5(a), for $M = 256$, all estimators $\hat{\mathbf{w}}_n$ from $n = 2$ up to about $n = 55$ outperform in MS error their RLS, LMS, and SMI ($\hat{\mathbf{w}}_\infty$) counterparts. $\hat{\mathbf{w}}_8$ ($n = 8$ AVs) has the least MS error of all (best bias/variance tradeoff). When the data record size is increased to $M = 2048$ [see Fig. 5(b)], we can afford more iterations (more AVs) , and $\hat{\mathbf{w}}_{13}$ offers the best bias/variance tradeoff (lowest MS error). All filter estimators $\hat{\mathbf{w}}_n$ for $n > 8$ outperform the LMS/RLS/SMI ($\hat{\mathbf{w}}_\infty$) estimators. For such large data record sets ($M = 2048$), the RLS and the SMI ($\hat{\mathbf{w}}_\infty$) MS error are almost identical.

In Fig. 6(a), we plot the MS estimation error of $\hat{\mathbf{w}}_8(M)$ ($n = 8$ AVs) and the RLS, LMS, SMI ($n = \infty$) estimators as a function of the *data record size* M. In Fig. 6(b), we repeat exactly the same study to compare various AV filter estimators only: $\hat{\mathbf{w}}_n(M)$ for $n = 0$ (matched-filter), $n = 8$, $n = 13$, and $n = 30$. Fig. 7 offers a 3-D plot of the MS estimation error as a function of the number of AVs n and the sample support M. The dark line that traces the bottom of the MS estimation error surface identifies the best number of AVs for any given data record size M.

An alternative bias/variance trading mechanism through real-valued tuning is the diagonally loaded (DL) SMI estimator [13]

$$\hat{\mathbf{w}}_{\mathrm{DL\text{-}SMI}}(\Delta) = \rho^* \frac{[\hat{\mathbf{R}}(M) + \Delta\mathbf{I}]^{-1}\mathbf{v}}{\mathbf{v}^H[\hat{\mathbf{R}}(M) + \Delta\mathbf{I}]^{-1}\mathbf{v}} \tag{26}$$

[2]The data include the signal of interest $\sqrt{E_1}b_1(m)\mathbf{s}_1$, $m = 1, \cdots, M$, as it appears in (22). Filter estimation improvements through the use of "signal-absent" data are not pursued in the context of this presentation.

[3]For multivariate elliptically contoured input distributions, an analytic expression for the covariance matrix of the SMI estimator $\hat{\mathbf{w}}_\infty(M)$ can be found in [47]:

$$E\{[\hat{\mathbf{w}}_\infty(M) - E\{\hat{\mathbf{w}}_\infty(M)\}][\hat{\mathbf{w}}_\infty(M) - E\{\hat{\mathbf{w}}_\infty(M)\}]^H\} = \frac{|\rho|^2}{(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})(M - L + 1)}\left(\mathbf{R}^{-1} - \frac{\mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}}{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right).$$

Since under these input distribution conditions $\hat{\mathbf{w}}_\infty(M)$ is unbiased, the trace of the covariance matrix is the MS filter estimation error. It is important to observe that the covariance matrix—and, therefore, the MS filter estimation error—depends on the data record size M, the filter length L, as well as the specifics of the signal processing problem at hand ($\mathbf{R}$ and $\mathbf{v}$). It is also important to note that for the signal model example in (22), the input is Gaussian-mixture distributed. Therefore, the results discussed in this footnote are not directly applicable and the analytic covariance matrix expression can only be thought of as an approximation (a rather close approximation as we concluded after our studies).

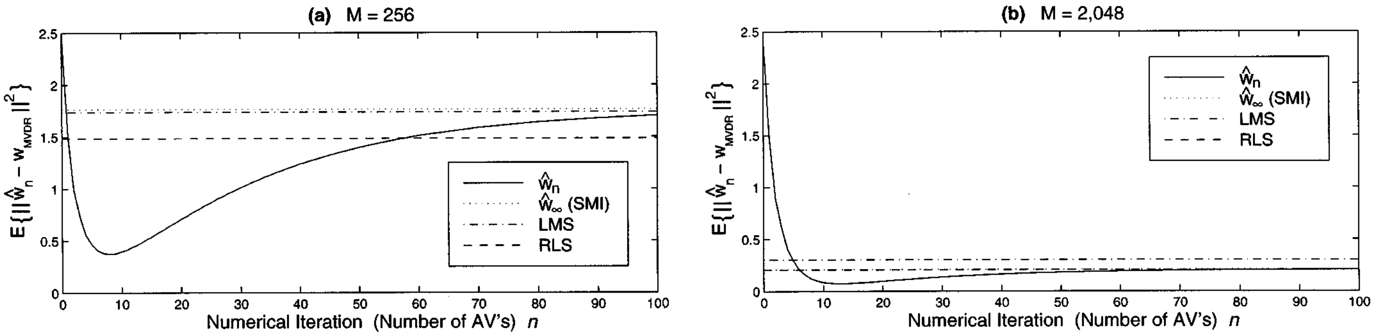

Fig. 5. MS estimation error for the sequence of estimators $\hat{\mathbf{w}}_n(M)$, $n = 0, 1, \cdots$. (a) Data record size $M = 256$. (b) $M = 2048$.

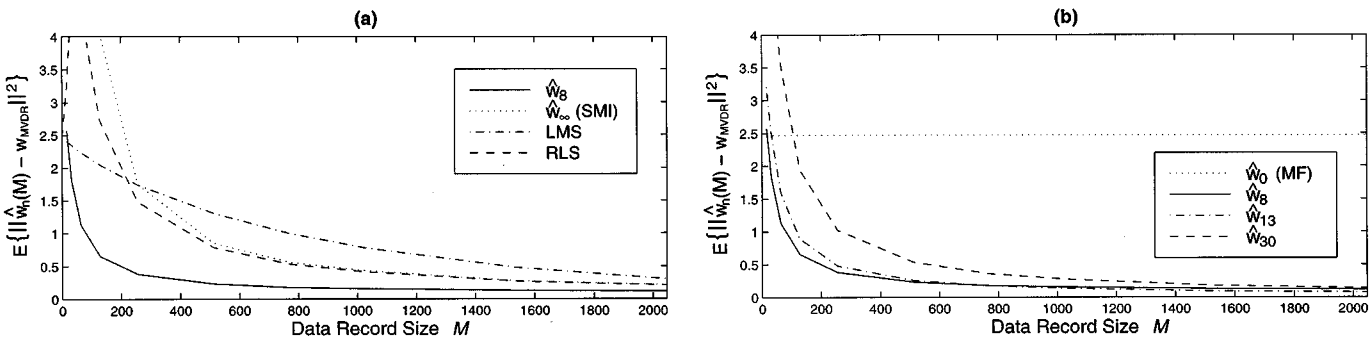

Fig. 6. MS estimation error as a function of the data record size M for various members of the $\hat{\mathbf{w}}_n(M)$ sequence: $n = 0$ (MF), $n = 8$, $n = 13$, $n = 30$, $n = \infty$ (SMI), as well as the LMS and RLS estimators.

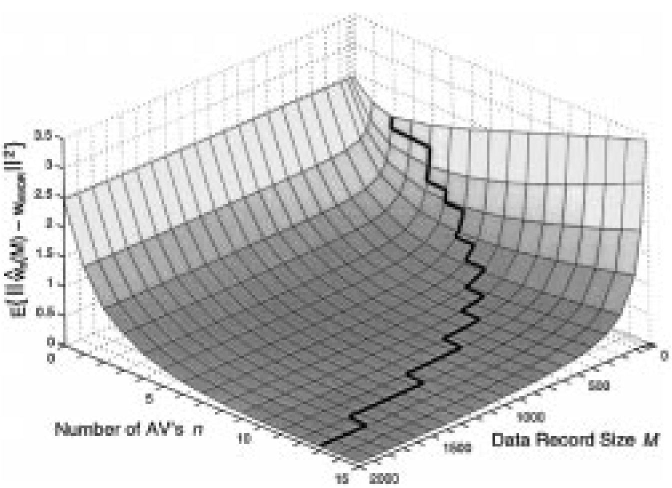

Fig. 7. MS estimation error versus number of auxiliary vectors n and sample support M.

where $\Delta \geq 0$ is the diagonal loading parameter. We observe that $\hat{\mathbf{w}}_{\text{DL-SMI}}(\Delta = 0)$ is the regular SMI estimator, whereas $\lim_{\Delta\to\infty} \hat{\mathbf{w}}_{\text{DL-SMI}}(\Delta) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$, which is the properly scaled matched filter. In Fig. 8(a), we plot the MS estimation error of the DL-SMI estimator as a function of the diagonal loading parameter Δ ($M = 60$). We identify the *best possible* diagonal loading value $\Delta \simeq 3.45$ (at significant computational cost) and in Fig. 8(c), we compare the best DL-SMI estimator against the AV estimator sequence for which *no* diagonal loading is performed. Interestingly, the AV estimators $\hat{\mathbf{w}}_n$ from $n = 4$ to 7 outperform in MS error the best possible DL-SMI estimator ($\Delta \simeq 3.45$).

Finally, a *finite set* of L filter estimators with varying bias/covariance balance can be obtained through the use of the orthogonal "multistage" filter decomposition procedure in [30] and [31]. It can be shown theoretically that the l-stage filter $\mathbf{w}_{l\text{-stage}}$, $0 \leq l \leq L-1$, is equivalent to the following structure. First, change the auxiliary-vector generation recursion in (12) or Fig. 2 to impose orthogonality, not only with respect to the constraint vector $\mathbf{v}$, but also with respect to *all previously defined* auxiliary vectors $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{n-1}$, $n \leq L-1$:

$$\mathbf{y}_n = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2} - \sum_{i=1}^{n-1} \frac{\mathbf{y}_i\mathbf{y}_i^H}{\|\mathbf{y}_i\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}. \tag{27}$$

Next, terminate the recursion at $n = l$, $0 \leq l \leq L-1$, and organize the l orthogonal to each other and to $\mathbf{v}$ vectors $\mathbf{y}_1, \cdots, \mathbf{y}_l$ in the form of a blocking matrix $\mathbf{B}_{L\times l} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_l]$. Then

$$\mathbf{w}_{l\text{-stage}} = \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v} - \mathbf{B}_{L\times l}\vec{\alpha}_{l\times 1} \quad \text{where} \tag{28}$$

$$\vec{\alpha} = \frac{\rho^*}{\|\mathbf{v}\|^2}\left[\mathbf{B}^H\mathbf{R}\mathbf{B}\right]^{-1}\mathbf{B}^H\mathbf{R}\mathbf{v} \tag{29}$$

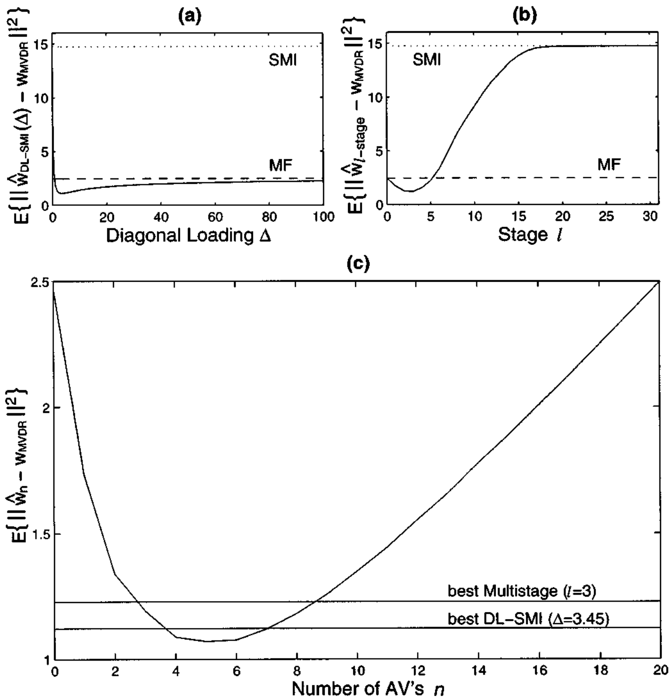

Fig. 8. MS estimation error studies for (a) diagonally loaded SMI, (b) multistage, and (c) auxiliary-vector estimators ($M = 60$).

is the MS vector-optimum (unconditionally optimum) set of weights of the vectors $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_l$.[4] In the context of filter estimation from a data record of size M, $\hat{\mathbf{w}}_{0\text{-stage}}(M)$ is the matched filter, and $\hat{\mathbf{w}}_{(L-1)\text{-stage}}(M)$ is the SMI estimator. In Fig. 8(b), we plot the MS estimation error of $\hat{\mathbf{w}}_{l\text{-stage}}(M)$ as a function of l, $0 \leq l \leq L-1 = 31$, ($M = 60$). We identify the *best* multistage estimator ($l = 3$ stages), and in Fig. 8(c), we compare it against the AV estimator sequence. We see that all AV estimators $\hat{\mathbf{w}}_n$ from $n = 3$ to 8 outperform in MS error the best multistage estimator ($l = 3$ stages). Finally, as a last study, in Fig. 9, we plot the MS error of the $\Delta = 3.45$ DL-SMI estimator together with the MS error of the *best* multistage and AV estimators over the data support range $M = L/2 = 16$ to $M = 3L = 96$.

V. Conclusions

In this paper, we relied strictly on statistical *conditional* optimization principles to derive an iterative algorithm that starts from the "white-noise matched filter" and converges to the "MVDR filter" solution for any given positive definite input autocorrelation matrix. The conceptual simplicity of the employed conditional optimization criteria led to a computationally simple iteration step. The algorithm is a greedy, yet noninvasive, procedure where no explicit autocorrelation matrix inversion/decomposition/diagonalization is attempted. We analyzed basic algorithmic properties and we established formal convergence to the MVDR filter.

When the input autocorrelation matrix is substituted by a sample-average (positive definite) estimate, the algorithm generates a sequence of filter estimators that converges to the familiar sample-matrix-inversion (SMI) unbiased estimator. The bias of the generated estimator sequence decreases rapidly to zero, whereas the estimator covariance trace rises gracefully from zero (for the initial, fixed-valued, matched-filter estimator) to the asymptotic covariance trace of SMI. Sequences of practical estimators that offer such exceptional control over favorable bias/covariance balance points are always a prime objective in the estimation theory literature. Indeed, for finite data record sets, members of the generated sequence of estimators were seen to outperform in MS estimation error, LMS/RLS-type, SMI and diagonally loaded SMI, and orthogonal multistage decomposition filter estimators. In addition, the troublesome, data-dependent tuning of the real-valued LMS learning gain parameter, the RLS initialization, or the SMI diagonal loading parameter is replaced by an integer choice among the first several members of the estimator sequence [50].

Computational simplicity and MS estimation error performance indicate that this "auxiliary-vector" adaptive processing scheme may deserve further consideration and investigation, particularly in high-dimensional adaptive signal processing applications that rely on data records of limited size.

[4]Therefore, the multistage filter in [30] and [31] is identical to the filter "$\mathbf{w}_\mathbf{B}$" as it appears in [36]–[38]. The multistage decomposition algorithm is a computationally efficient procedure for the calculation of this filter tailored to the particular structure of $\mathbf{B}^H\mathbf{R}\mathbf{B}$ (tridiagonal matrix). The same computational savings can be achieved by the general forward calculation algorithm of Liu and Van Veen [51] that returns all intermediate stage filters along the way up to the stage of interest l (total computational complexity of order $O((M+l)L^2))$. The AV algorithm in Fig. 2 has computational complexity $O((M+n)L^2)$, where n is the desired number of AVs. Again, all intermediate AV filters are returned. Estimators of practical interest have $l \ll M$ or $n \ll M$. Therefore, the complexity of all such algorithms is dominated by $O(ML^2)$, which is required for the computation of $\hat{\mathbf{R}}(M)$.

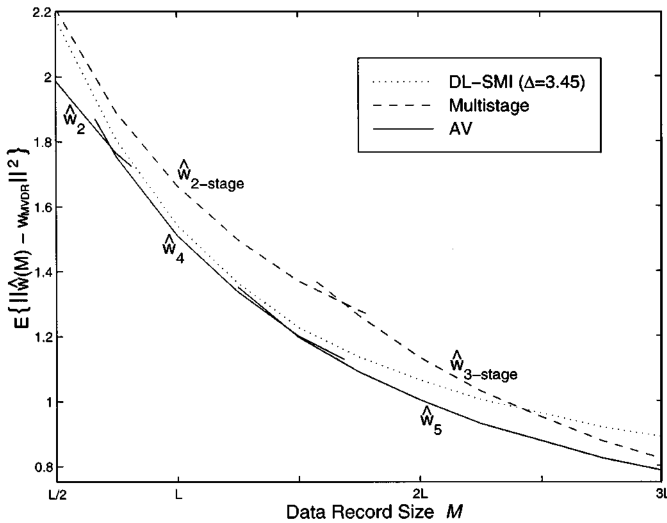

Fig. 9. MS estimation error for the *best* multistage and AV estimators over the data support range $M = L/2 = 16$ to $M = 3L = 96$. The MS estimation error of the Δ =3.45 DL-SMI estimator is also included as a reference.

Appendix

A. Proof of Lemma 1

Consider two successively generated auxiliary vectors $\mathbf{g}_i$, $\mathbf{g}_{i+1}$, $i \geq 1$, by the algorithm in Fig. 2. Since

$(\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)^2 = \mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2$ ($\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2$ is a projection operator)

$$\begin{aligned}\mathbf{g}_i &= \left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{i-1} = \left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)^2\mathbf{R}\mathbf{w}_{i-1}\\ &= \left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{g}_i. \qquad (A.1)\end{aligned}$$

Then, by the definition of $\mathbf{g}_{i+1}$ and (A.1)

$$\mathbf{g}_i^H\mathbf{g}_{i+1} = \mathbf{g}_i^H\left[\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_i\right] = \mathbf{g}_i^H\mathbf{R}\mathbf{w}_i. \qquad (A.2)$$

It is now straightforward to obtain $\mathbf{g}_i^H\mathbf{R}\mathbf{w}_i = 0$. We substitute $\mathbf{w}_i = \mathbf{w}_{i-1} - \mu_i\mathbf{g}_i$ and $\mu_i = (\mathbf{g}_i^H\mathbf{R}\mathbf{w}_{i-1}/\mathbf{g}_i^H\mathbf{R}\mathbf{g}_i)$ and calculate

$$\begin{aligned}\mathbf{g}_i^H\mathbf{R}\mathbf{w}_i &= \mathbf{g}_i^H\mathbf{R}\mathbf{w}_{i-1} - \mu_i\mathbf{g}_i^H\mathbf{R}\mathbf{g}_i\\ &= \mathbf{g}_i^H\mathbf{R}\mathbf{w}_{i-1} - \frac{\mathbf{g}_i^H\mathbf{R}\mathbf{w}_{i-1}}{\mathbf{g}_i^H\mathbf{R}\mathbf{g}_i}\mathbf{g}_i^H\mathbf{R}\mathbf{g}_i = 0. \qquad (A.3)\end{aligned}$$

From (A.2) and (A.3), we conclude $\mathbf{g}_i \perp \mathbf{g}_{i+1}$, $i = 1, 2, \cdots$.

B. Proof of Theorem 1

i) As seen in (A.1), $\mathbf{g}_n = (\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{g}_n$. Therefore

$$\mu_n = \frac{\mathbf{g}_n^H\mathbf{R}\mathbf{w}_{n-1}}{\mathbf{g}_n^H\mathbf{R}\mathbf{g}_n} = \frac{\mathbf{g}_n^H\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}}{\mathbf{g}_n^H\mathbf{R}\mathbf{g}_n} \quad n = 1, 2, \cdots. \qquad (A.4)$$

However, $(\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_{n-1} = \mathbf{g}_n$ by definition. Substituting $(\mathbf{I}-\mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}\mathbf{w}_{n-1}$ by $\mathbf{g}_n$ in the numerator of (A.4), we obtain

$$\mu_n = \frac{\mathbf{g}_n^H\mathbf{g}_n}{\mathbf{g}_n^H\mathbf{R}\mathbf{g}_n}, \qquad n = 1, 2, \cdots \qquad (A.5)$$

and we conclude [5] that

$$0 < \frac{1}{\lambda_{\max}} \le \mu_n \le \frac{1}{\lambda_{\min}}, \qquad n = 1, 2, \cdots \qquad (A.6)$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and the minimum, correspondingly, eigenvalues of $\mathbf{R}$. Equation (A.6) shows that the generated sequence of auxiliary-vector weights $\{\mu_n\}$, $n = 1, 2, \cdots$, is, in fact, positive and bounded. Equation (A.5) offers an alternative formula for the calculation of μ_n in the algorithm of Fig. 2.

ii) To prove that $\mathbf{g}_n$ in Fig. 2 converges to the $\mathbf{0}$ vector for a given Hermitian positive definite matrix $\mathbf{R}$, we find it convenient to introduce the concept of the "$\mathbf{R}$-inner-product" and the "$\mathbf{R}$-norm" [5]. If $\mathbf{x}, \mathbf{y} \in \mathcal{C}^L$, then their $\mathbf{R}$-inner-product is defined by

$$\langle\mathbf{x}, \mathbf{y}\rangle_{\mathbf{R}} \triangleq \mathbf{x}^H\mathbf{R}\mathbf{y} \qquad (A.7)$$

whereas the $\mathbf{R}$-norm of $\mathbf{x}$ is

$$\|\mathbf{x}\|_{\mathbf{R}} \triangleq \sqrt{\langle\mathbf{x}, \mathbf{x}\rangle_{\mathbf{R}}} = \sqrt{\mathbf{x}^H\mathbf{R}\mathbf{x}}. \qquad (A.8)$$

In the Proof of Lemma 1, we showed that [cf. (A.3)] $\mathbf{w}_n$ and $\mathbf{g}_n$ are $\mathbf{R}$-orthogonal. Therefore

$$\|\mathbf{w}_n + \mu_n\mathbf{g}_n\|_{\mathbf{R}}^2 = \|\mathbf{w}_n\|_{\mathbf{R}}^2 + \|\mu_n\mathbf{g}_n\|_{\mathbf{R}}^2. \qquad (A.9)$$

From the algorithmic recursion $\mathbf{w}_n = \mathbf{w}_{n-1} - \mu_n\mathbf{g}_n$, we see that

$$\|\mathbf{w}_n + \mu_n\mathbf{g}_n\|_{\mathbf{R}}^2 = \|\mathbf{w}_{n-1}\|_{\mathbf{R}}^2. \qquad (A.10)$$

Equations (A.9) and (A.10) combined show that

$$\|\mathbf{w}_n\|_{\mathbf{R}}^2 = \|\mathbf{w}_{n-1}\|_{\mathbf{R}}^2 - \|\mu_n\mathbf{g}_n\|_{\mathbf{R}}^2, \qquad n = 1, 2, \cdots. \qquad (A.11)$$

We observe that $\{\|\mathbf{w}_n\|_{\mathbf{R}}\}$, $n = 1, 2, \cdots$, is a monotonically decreasing sequence of non-negative numbers. Hence, $\|\mathbf{w}_n\|_{\mathbf{R}}$ converges. This implies that $\|\mu_n\mathbf{g}_n\|_{\mathbf{R}} \longrightarrow 0$, $\|\mu_n\mathbf{g}_n\| \longrightarrow 0$, or $|\mu_n|\ \|\mathbf{g}_n\| \longrightarrow 0$. However, in Part (i) we showed that $\{\mu_n\}$ is bounded away from zero. We conclude that $\|\mathbf{g}_n\| \longrightarrow 0$ or $\mathbf{g}_n \longrightarrow \mathbf{0}$, as $n \to \infty$.

C. Another Interpretation of the Auxiliary-Vector Algorithm

Following the notation in Sections II and III, we define the variance cost function $J(\mathbf{w}) \triangleq \mathbf{w}^H\mathbf{R}\mathbf{w}$. We seek the vector $\mathbf{w}$ that minimizes $J(\cdot)$ subject to the constraint $\mathbf{w}^H\mathbf{v} = \rho$. We modify the conventional steepest-descent algorithm [1, ch. 8] to accommodate the constraint

$$\begin{aligned}\mathbf{w}_n = &\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\left[\mathbf{w}_{n-1} - \frac{1}{2}\mu\nabla_{\mathbf{w}^*}J(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{w}_{n-1}}\right]\\ &+ \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}, \qquad n = 1, 2, \cdots \qquad (A.12)\end{aligned}$$

for an *arbitrary* $\mathbf{w}_0$ and some $\mu > 0$ "step-size" parameter.

Next, we enforce the specific initialization $\mathbf{w}_0 = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ in (3), which allows the following simplification of (A.12):

$$\begin{aligned}\mathbf{w}_n &= \mathbf{w}_{n-1} - \mu\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}\\ \mathbf{w}_0 &= \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v}, \qquad n = 1, 2, \cdots. \qquad (A.13)\end{aligned}$$

In view of (A.12), (A.13), and the algorithm in Fig. 2, the auxiliary vector $\mathbf{g}_n$, which is chosen inductively according to the maximum magnitude cross-correlation criterion in Proposition 2, can be seen as the gradient of the variance cost function evaluated at the previous iteration step filter $\mathbf{w}_{n-1}$ and projected onto the orthogonal to $\mathbf{v}$ subspace. In the same context, the auxiliary-vector weights can be seen as a variable step-size sequence that is locally optimized *at each iteration step* $n = 1, 2, \cdots$ according to the conditional MS-optimality criterion in Proposition 1:

$$\mu_n = \frac{\left[\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}\right]^H\mathbf{R}\mathbf{w}_{n-1}}{\left[\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}\right]^H\mathbf{R}\left(\mathbf{I}-\frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\mathbf{R}\mathbf{w}_{n-1}} \quad n = 1, 2, \cdots$$

[we recall that $(1/\lambda_{\max}) \leq \mu_n \leq (1/\lambda_{\min})\ \forall n = 1, 2, \cdots$ according to Theorem 1, Part (i)].

D. Signature Assignment for the DS/CDMA Example

The matrix $\mathbf{S}_{32\times 13} = [\mathbf{s}_1\ \mathbf{s}_2 \cdots \mathbf{s}_{13}]$ with columns the signature vectors $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{13}$ is given in (A.14), shown at the bottom of the page.

ACKNOWLEDGMENT

The authors would like to express their appreciation to the Associate Editor, Prof. H. Sakai of Kyoto University, and the anonymous reviewers for their stimulating comments and suggestions on the original manuscript.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[2] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[3] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
[4] N. L. Owsley, “A recent trend in adaptive spatial processing for sensor arrays: Constraint adaptation,” in *Signal Processing*, J. W. R. Griffiths *et al.*, Eds. New York: Academic, 1973, pp. 591–604.
[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1990.
[6] N. E. Nahi, *Estimation Theory and Applications*. Huntington, NY: Krieger, 1976.
[7] H. Robbins and S. Monro, “A stochastic approximation method,” *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.
[8] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, “Adaptive antenna systems,” *Proc. IEEE*, vol. 55, pp. 2143–2158, Dec. 1967.
[9] L. J. Griffiths, “A simple adaptive algorithm for real-time processing in antenna arrays,” *Proc. IEEE*, vol. 57, pp. 1696–1704, Oct. 1969.
[10] O. L. Frost, III, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.
[11] L. C. Godara and A. Cantoni, “Analysis of constrained LMS algorithm with application to adaptive beamforming using perturbation sequences,” *IEEE Trans. Antennas Propagat.*, vol. AP–34, pp. 368–379, Mar. 1986.
[12] I. S. Reed, J. D. Mallet, and L. E. Brennan, “Rapid convergence rate in adaptive arrays,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, pp. 853–863, Nov. 1974.
[13] B. D. Carlson, “Covariance matrix estimation errors and diagonal loading in adaptive arrays,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 397–401, July 1988.
[14] R. L. Plackett, “Some theorems in least squares,” *Biometrika*, vol. 37, p. 149, 1950.

$$\mathbf{S} = \frac{1}{\sqrt{32}} \begin{bmatrix}
-1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & +1 \\
-1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 \\
-1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 \\
+1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 \\
+1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 \\
-1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 \\
-1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & +1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & +1 & -1 \\
-1 & -1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 \\
-1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & +1 \\
-1 & -1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 \\
+1 & +1 & +1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 \\
-1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 \\
+1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 \\
-1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 \\
-1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 \\
+1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 \\
+1 & -1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 \\
-1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 \\
+1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\
-1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 \\
-1 & +1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1
\end{bmatrix}. \tag{A.14}$$

[15] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, vol. 70, pp. 1885–1891, 1965.

[16] V. Solo, "The limiting behavior of LMS," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1909–1922, Dec. 1989.

[17] J. M. Cioffi and T. Kailath, "Fast recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 304–337, Apr. 1984.

[18] S. P. Applebaum and D. J. Chapman, "Adaptive arrays with main beam constraints," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 650–662, Sept. 1976.

[19] P. W. Howells, "Explorations in fixed and adaptive resolution at GE and SURC," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 575–584, Sept. 1976.

[20] B. D. Van Veen and R. A. Roberts, "Partially adaptive beamformer design via output power minimization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1524–1532, Nov. 1987.

[21] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[22] D. W. Tufts, I. P. Kirsteins, and R. Kumaresan, "Data adaptive detection of a weak signal," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, pp. 313–316, Mar. 1983.

[23] D. W. Tufts and C. D. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1046–1053, Oct. 1986.

[24] L. L. Scharf and D. W. Tufts, "Reduced rank methods for modeling stationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 350–354, Mar. 1987.

[25] N. L. Owsley, *Array Signal Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[26] B. D. Van Veen, "Eigenstructure based partially adaptive array design," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 357–362, Mar. 1988.

[27] A. M. Haimovich and Y. Bar-Ness, "An eigenanalysis interference canceler," *IEEE Trans. Signal Processing*, vol. 39, pp. 76–84, Jan. 1991.

[28] K. A. Byerly and R. A. Roberts, "Output power based partially adaptive array design," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 1989, pp. 576–580.

[29] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," *IEEE Trans. Signal Processing*, vol. 45, pp. 492–496, Feb. 1997.

[30] J. S. Goldstein, I. S. Reed, P. A. Zulch, and W. L. Melvin, "A multistage STAP CFAR detection technique," in *Proc. IEEE Radar Conf.*, Dallas, TX, May 1998, pp. 111–116.

[31] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2943–2959, Nov. 1998.

[32] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.

[33] R. Sharma and B. D. Van Veen, "Large modular structures for adaptive beamforming and the Gram-Schmidt preprocessor," *IEEE Trans. Signal Processing*, vol. 42, pp. 448–451, Feb. 1994.

[34] D. A. Pados and S. N. Batalama, "Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers," *IEEE Trans. Commun.*, vol. 45, pp. 1586–1594, Dec. 1997.

[35] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1765–1773, Dec. 1998.

[36] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," in *Proc. Conf. Inform. Sci. Syst.*, vol. II. Princeton, NJ, Mar. 1998, pp. 1007–1013.

[37] D. A. Pados, J. H. Michels, T. Tsao, and M. C. Wicks, "Joint domain space-time adaptive processing with small training data sets," in *Proc. IEEE Radar Conf.*, Dallas, TX, May 1998, pp. 99–104.

[38] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406–1415, Sept. 1999.

[39] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.

[40] P. B. Rapajic and B. S. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.

[41] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.

[42] S. L. Miller, "An adaptive DS-CDMA receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, pp. 1746–1755, Feb./Mar./Apr. 1995.

[43] M. L. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.

[44] C. N. Pateros and G. J. Saulnier, "An adaptive correlator receiver for direct-sequence spread-spectrum communication," *IEEE Trans. Commun.*, vol. 44, pp. 1543–1552, Nov. 1996.

[45] R. L. Dykstra, "Establishing the positive definiteness of the sample covariance matrix," *Ann. Math. Statist.*, vol. 41, no. 6, pp. 2153–2154, 1970.

[46] A. O. Steinhardt, "Adaptive multisensor detection and estimation," in *Adaptive Radar Detection and Estimation*, S. Haykin and A. O. Steinhardt, Eds. New York: Wiley, 1992, ch. 3.

[47] C. D. Richmond, "PDF's, confidence regions, and relevant statistics for a class of sample covariance-based array processors," *IEEE Trans. Signal Processing*, vol. 44, pp. 1779–1793, July 1996.

[48] A. O. Steinhardt, "The PDF of adaptive beamforming weights," *IEEE Trans. Signal Processing*, vol. 39, pp. 1232–1235, May 1991.

[49] C. D. Richmond, "Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance," *IEEE Trans. Signal Processing*, vol. 44, pp. 305–315, Feb. 1996.

[50] H. Qian and S. N. Batalama, "Data-record-based criteria for the selection of an auxiliary-vector estimator of the MVDR filter," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2000.

[51] T.-C. Liu and B. Van Veen, "A modular structure for implementation of linearly constrained minimum variance beamformers," *IEEE Trans. Signal Processing*, vol. 39, pp. 2343–2346, Oct. 1991.

Dimitris A. Pados (M'95) was born in Athens, Greece, on October 22, 1966. He received the Diploma degree in computer engineering and science from the University of Patras, Patras, Greece in 1989 and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1994.

He worked as an Applications Manager with the Digital Systems and Telecommunications Laboratory, Computer Technology Institute, Patras, from 1989 to 1990. From 1990 to 1994, he was a Research Assistant with the Communications Systems Laboratory, Department of Electrical Engineering, University of Virginia. From 1994 to 1997, he held an Assistant Professor position with the Department of Electrical and Computer Engineering and the Center for Telecommunications Studies, University of Louisiana, Lafayette. Since August 1997, he has been an Assistant Professor with the Department of Electrical Engineering, State University of New York at Buffalo, Amherst. His research interests are in the general areas of communication theory and adaptive signal processing with emphasis on wireless multiple access communications, detection of spread-spectrum signals, adaptive antenna arrays and array radar, and neural networks.

Dr. Pados is a member of the IEEE Communications Society, the Communication Theory Technical Committee, and the IEEE Signal Processing Society.

George N. Karystinos (S'98) was born in Athens, Greece, on April 12, 1974. He received the Diploma degree in computer engineering and science from the University of Patras, Patras, Greece, in 1997. Since 1998, he has been a Research Assistant with the Communications and Signals Group, Department of Electrical Engineering, State University of New York at Buffalo, Amherst, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests are in the areas of wireless multiple access communications and statistical signal processing.

Mr. Karystinos is a student member of the IEEE Communications, Signal Processing, and Information Theory Societies and a member of Eta Kappa Nu.