Digital Signal Processing **12**, 193–222 (2002)
doi:10.1006/dspr.2002.0450

Short Data Record Adaptive Filtering: The Auxiliary-Vector Algorithm [1]

George N. Karystinos,* Haoli Qian,* Michael J. Medley,† and Stella N. Batalama*

*Department of Electrical Engineering, State University of New York at Buffalo, 332 Bonner Hall, Buffalo, New York 14260; and †Air Force Research Laboratory, IFGC, 525 Brook Road, Rome, New York 13441
E-mail: cary@eng.buffalo.edu, haoqian@eng.buffalo.edu, michael.medley@rl.af.mil, batalama@eng.buffalo.edu

Karystinos, G. N., Qian, H., Medley, M. J., and Batalama, S. N., Short Data Record Adaptive Filtering: The Auxiliary-Vector Algorithm, *Digital Signal Processing* **12** (2002) 193–222.

Based on statistical conditional optimization criteria, we developed 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 recursive procedure that avoids explicit matrix inversion, decomposition, or diagonalization operations. 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. For short data records, the early, nonasymptotic, elements of the generated sequence of estimators offer favorable bias–covariance balance and are seen to outperform in mean-square estimation error constraint-LMS, RLS-type, and orthogonal multistage decomposition estimates (also called nested Wiener filters) as well as plain and diagonally loaded SMI estimates. The problem of selecting the most successful (in some appropriate sense) filter estimator in the sequence for a given data record is addressed and two data-driven selection criteria are proposed. The first criterion minimizes the cross-validated sample average variance of the filter estimator output. The second criterion maximizes the estimated J-divergence of the filter estimator output conditional distributions. Illustrative interference suppression examples drawn from the communications literature are followed throughout this presentation.

[1] This work was supported by the National Science Foundation under Grant ECS-0073660 and the U.S. Air Force Office of Scientific Research under Grant F49620-01-1-0176.

1051-2004/02 $35.00

Key Words: adaptive filters; biased estimators; code division multiple access; cross-validation; interference suppression; iterative methods; J-divergence; least mean square methods; auxiliary-vector filters; MMSE filters; MVDR filters; filter estimation; small sample support; finite sample support; short data record estimators; Wiener filters; antenna arrays; smart antenna

1. 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. If $\mathbf{r}$ is a random, zero mean without loss of generality, complex input vector of dimension L, $\mathbf{r} \in \mathbb{C}^L$, that is processed by an L-tap filter $\mathbf{w} \in \mathbb{C}^L$, 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, the 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 $\mathbf{w}^H\mathbf{v} = \rho \in \mathbb{C}$, where $\mathbf{v} \in \mathbb{C}^L$ is the 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 [1, 2]

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

where ρ^* denotes the conjugate of the desired response $\mathbf{w}^H\mathbf{v} = \rho$. MVDR filtering has been used extensively in unsupervised signal processing applications where a desired scalar filter output $d \in \mathbb{C}$ cannot be identified or cannot be assumed available for each input $\mathbf{r} \in \mathbb{C}^L$ (for example, in 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 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, $c\mathbf{R}^{-1}\mathbf{v}$, $c \in \mathbb{C}$. For this reason, in the rest of the paper we will use the term *MMSE–MVDR filter* to refer to either filter.

In this article, first we present an iterative algorithm for the calculation of the MMSE–MVDR vector $\mathbf{w}_{\mathrm{MMSE/MVDR}}$ in (1). The algorithm is a noninvasive procedure where no explicit matrix inversion–eigendecomposition–diagonalization is attempted. The MMSE–MVDR computation algorithm creates a sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, that begins from $\mathbf{w}_0 = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and converges to the MMSE–MVDR filter ($\mathbf{w}_\infty = \mathbf{w}_{\mathrm{MMSE/MVDR}}$). At each step $n = 1, 2, \ldots$, $\mathbf{w}_n$ is given as a simple, direct function of $\mathbf{R}$, $\mathbf{v}$, and $\mathbf{w}_{n-1}$.

The development of the iterative algorithm (which we call the *auxiliary-vector algorithm* for reasons that will become apparent) is founded solely on statistical signal processing principles. The motivation behind its development is *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, $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M$:

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

When $\mathbf{R}$ is substituted by $\widehat{\mathbf{R}}(M)$ in the recursively generated sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, the corresponding filter estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, offer the means for effective control over the filter estimator bias versus (co-)variance trade-off [3]. Starting from the *0*-variance, high-bias (for nonwhite inputs) $\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ estimate, we can go all the way up to the unbiased, yet high-variance for small data record sizes M, $\widehat{\mathbf{w}}_\infty(M)$ estimate and anywhere in between, $\widehat{\mathbf{w}}_n(M)$, $1 \leq n < \infty$. As a result, adaptive filters from this newly developed class can be seen to outperform in expected norm-square estimation error, $E\{\|\widehat{\mathbf{w}}(M) - \mathbf{w}_{\text{MMSE/MVDR}}\|^2\}$, (constraint-) LMS [4], sample matrix inversion (SMI) [5] with or without diagonal loading [6], RLS-type [7, 8], and orthogonal multistage decomposition (also called nested Wiener) [9, 10] adaptive filter implementations. It is worth mentioning that the familiar trial-and-error tuning to problem and data-record-size specifics of the real-valued LMS gain or RLS inverse matrix initialization constant or SMI diagonal loading parameter that plagues field practitioners is now replaced by an integer choice of one of the recursively generated filters.

The problem of selecting the best (in some appropriate sense) filter estimator in the sequence for a given data record is addressed and two data-driven selection criteria are proposed. The first criterion is rather general and is motivated by the asymptotic minimum output variance property of the generated sequence of filter estimators. In particular, for a given data record, we select the filter estimator that has minimum cross-validated average output variance (energy). The second rule is built specifically for binary antipodal (BPSK-type) communication signals and is related to the objective of achieving maximum stochastic distance between the two conditional distributions of the filter estimator output. Under this rule, we choose the filter estimator in the sequence that exhibits maximum estimated J-divergence of the conditional output distributions. We pursue and analyze both supervised and unsupervised (blind) implementations of this criterion. Illustrative case studies drawn from the code division multiple access (CDMA) communications literature are followed throughout this article.

The rest of the article is organized as follows. In Section 2 we present the basic algorithmic development and analysis results. Filter estimation issues are discussed in Section 3. The two data-driven criteria for the selection of a filter estimator from the generated sequence are developed in Section 4. In Section 5 we examine the quality of the proposed criteria through simulations. A few concluding remarks are given in Section 6.

2. ALGORITHMIC DEVELOPMENTS AND CONVERGENCE ANALYSIS

For a given constraint vector $\mathbf{v} \in \mathbb{C}^L$ consider the set of filters $\mathcal{D} = \{\mathbf{w} \in \mathbb{C}^L$: $\mathbf{w} = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} + \mathbf{u}$, $\mathbf{u} \in \mathbb{C}^L$, and $\mathbf{v}^H\mathbf{u} = 0\}$. $\mathcal{D}$ is the class of all filters $\mathbf{w}$ in $\mathbb{C}^L$ that have a given response ρ in $\mathbf{v}$; that is, $\mathbf{w}^H\mathbf{v} = \rho$. In this section we develop an iterative algorithm for the computation of the $\mathbf{u}$ component of the MMSE–MVDR filter.

Algorithmic designs that focus on the MMSE–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 generalized sidelobe cancelers (GSC) [11] or partially adaptive beamformers [12]. Recent developments have been influenced by principal component analysis reduced-rank processing principles [13]. In general, the MMSE–MVDR filter part $\mathbf{u}$ ($\mathbf{u}^H\mathbf{v} = 0$) has been approximated by $\mathbf{u}_{L\times 1} \simeq -\mathbf{B}_{L\times(L-1)}\mathbf{T}_{(L-1)\times p}\mathbf{w}^{\mathrm{GSC}}_{p\times 1}$, where $\mathbf{B}$ is the so-called *blocking matrix* that satisfies $\mathbf{B}^H\mathbf{v} = \mathbf{0}_{L-1}$ ($\mathbf{B}$ is a full column-rank matrix that can be derived by Gram–Schmidt orthogonalization of 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), $\mathbf{T}$ is the rank reducing matrix with $1 \le p < L-1$ columns to be designed, and $\mathbf{w}^{\mathrm{GSC}}$ is the MS-optimum vector of weights of the p columns of $\mathbf{T}$ ($\mathbf{w}^{\mathrm{GSC}} = (\rho^*/\|\mathbf{v}\|^2)[\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{B}\mathbf{T}]^{-1}\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{v}$ [12]). In [14] and [15] the p columns of $\mathbf{T}$ were chosen to be 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 $\mathbf{B}^H\mathbf{R}\mathbf{B}$ (there is no documented technical optimality to this approach), then the best way to choose them in the minimum output variance p-rank approximation sense was presented in [16]: Select the p eigenvectors $\mathbf{q}_i$ of $\mathbf{B}^H\mathbf{R}\mathbf{B}$, with corresponding eigenvalues λ_i, that maximize $|\mathbf{v}^H\mathbf{R}\mathbf{B}\mathbf{q}_i|^2/\lambda_i$, $i = 1, \ldots, p$. This design algorithm was called *cross-spectral metric* reduced-rank processing in [17]. A different approach from a different point of view is described 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 $\mathbf{g}_n^H\mathbf{v} = 0$, $\forall n = 1, 2, \ldots$.

We begin the algorithmic developments from the conventional matched filter (MF) with desired response $\mathbf{w}^H\mathbf{v} = \rho$

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

which is MMSE–MVDR optimum for white $\mathbb{C}^L$ vector inputs (when $\mathbf{R} = \sigma^2\mathbf{I}$, $\sigma > 0$). We recall that, w.l.o.g. and for notational simplicity, we assume throughout this presentation that the input vectors $\mathbf{r} \in \mathbb{C}^L$ 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)

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

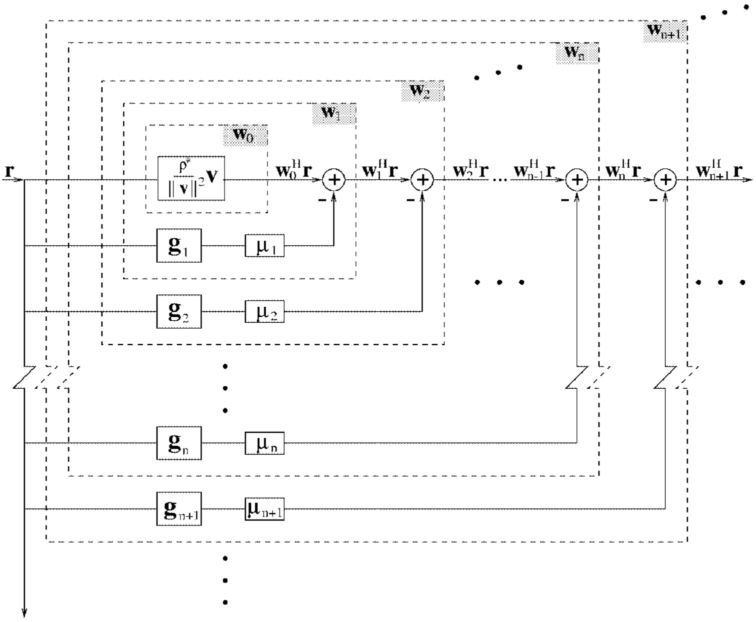

FIG. 1. Block diagram representation of the iteratively generated sequence of filters $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

where $\mathbf{g}_1 \in \mathbb{C}^L - \{\mathbf{0}\}$, $\mu_1 \in \mathbb{C}$, and $\mathbf{g}_1^H \mathbf{v} = 0$. We assume 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 [18]:

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 MMSE–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 $\mathbf{g}_1^H\mathbf{g}_1=1$:

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

For the sake of mathematical accuracy, we note that both the criterion function $|\mathbf{w}_0^H\mathbf{R}\mathbf{g}|$ to be maximized and 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 [19] 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 auxiliary vector orthonormal to $\mathbf{v}$ that can capture the most interference present in $\mathbf{w}_0^H\mathbf{r}$. The solution, derived through conventional Lagrange multipliers optimization, is given below.

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

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

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* ($\mathbf{w}_0^H\mathbf{R}\mathbf{g}_1>0$).

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

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{n+1}\mathbf{g}_{n+1}, \tag{8}$$

where

$$\mathbf{g}_{n+1} = \frac{\mathbf{R}\mathbf{w}_n - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{v}}{\left\|\mathbf{R}\mathbf{w}_n - \frac{\mathbf{v}^H\mathbf{R}\mathbf{w}_n}{\|\mathbf{v}\|^2}\mathbf{v}\right\|} \tag{9}$$

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

$$\mu_{n+1} = \frac{\mathbf{g}_{n+1}^H\mathbf{R}\mathbf{w}_n}{\mathbf{g}_{n+1}^H\mathbf{R}\mathbf{g}_{n+1}} \tag{10}$$

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

It is important to note that, while the generated auxiliary vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots$ are all constrained to be orthogonal to $\mathbf{v}$, orthogonality among the auxiliary vectors is *not* imposed [20, 21]. This is in sharp contrast to previous work that involved filtering with up to $L-1$ orthogonal to each other and to $\mathbf{v}$ vectors [22–24], where L is the data input vector dimension. We observe, however, that *successive* auxiliary vectors generated by the above recursive conditional optimization procedures (8)–(10) *do* come up orthogonal: $\mathbf{g}_n^H\mathbf{g}_{n+1}=0$, $\forall n = 1, 2, 3, \ldots$ (while $\mathbf{g}_n^H\mathbf{g}_m \neq 0$, $\forall n, m$, $|n-m| \neq 1$). For completeness purposes, we present this observation below in the form of a lemma.

LEMMA 1. *Successive auxiliary vectors generated through* (8)–(10) *are orthogonal:* $\mathbf{g}_n^H\mathbf{g}_{n+1}=0$, $n=1,2,3,\ldots$. *However,* $\mathbf{g}_n^H\mathbf{g}_m \neq 0$, $\forall n, m$, $|n-m| \neq 1$.

The algorithm is summarized in Fig. 2. The conceptual simplicity of the conditional statistical optimization process 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 $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$ to the MMSE–MVDR filter $\rho^*\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$ is established by the following theorem. The proof can be found in [21].

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

Auxiliary–Vector Algorithm

Input:

Autocovariance matrix $\mathbf{R}$, constraint vector $\mathbf{v}$, desired response $\mathbf{w}^H\mathbf{v} = \rho$.

Initialization:

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

Iterative computation:

For n=1, 2, . . . do

begin

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

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 $\mathbf{w}_0$, $\mathbf{w}_1$, $\mathbf{w}_2$, ...

FIG. 2. The algorithm for the iterative generation of the filter sequence $\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots$.

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

$$0 < \frac{1}{\lambda_{\max}} \leq \mu_n \leq \frac{1}{\lambda_{\min}}, \qquad n = 1, 2, \ldots, \tag{11}$$

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

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

$$\lim_{n\to\infty} \mathbf{g}_n = \mathbf{0}. \tag{12}$$

(iii) *The sequence of auxiliary-vector filters* $\{\mathbf{w}_n\}$, $n = 1, 2, \ldots$, *converges to the MMSE–MVDR filter:*

$$\lim_{n\to\infty} \mathbf{w}_n = \rho^* \frac{\mathbf{R}^{-1}\mathbf{v}}{\mathbf{v}^H \mathbf{R}^{-1}\mathbf{v}}. \tag{13}$$

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

$$\mathbf{r} = \sum_{k=1}^{K} \sqrt{E_k} b_k \mathbf{s}_k + \mathbf{n}. \tag{14}$$

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*) $\mathbf{s}_k$, $k = 1, 2, \ldots, K$. 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 $E\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \sigma^2 \mathbf{I}_{L\times L}$ ($\mathbf{x}^{\mathrm{T}}$ denotes the transpose of $\mathbf{x}$). With this notation and normalized user signatures, the signal-to-noise ratio of the kth user signal is defined by $\mathrm{SNR}_k \triangleq 10\log_{10} E_k/\sigma^2$ dB, $k = 1, 2, \ldots, K$.

MMSE–MVDR filtering for DS-CDMA type problems has attracted significant interest [25–29]. If we wish to recover the information bits of, say, user *1*, then all other signals constitute multiple-access interference and the MMSE–MVDR filter is built with constraint vector $\mathbf{v} = \mathbf{s}_1$, desired response $\mathbf{w}^{\mathrm{T}}\mathbf{s}_1 = 1$, and autocorrelation matrix $\mathbf{R} = \sum_{k=1}^{K} E_k \mathbf{s}_k \mathbf{s}_k^{\mathrm{T}} + \sigma^2 \mathbf{I}$. We choose $L = 32$, $K = 13$, and we draw an arbitrary set of signatures $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{13}$. 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 $\mathrm{SNR}_1 = 12$ dB while the *interferers* $k = 2, \ldots, 13$ are at $\mathrm{SNR}_{2\text{–}5} = 10$ dB, $\mathrm{SNR}_{6\text{–}9} = 12$ dB and $\mathrm{SNR}_{10\text{–}13} = 14$ dB. Figure 3 shows how the sequence of filters $\mathbf{w}_0$, $\mathbf{w}_1$, ... generated by the

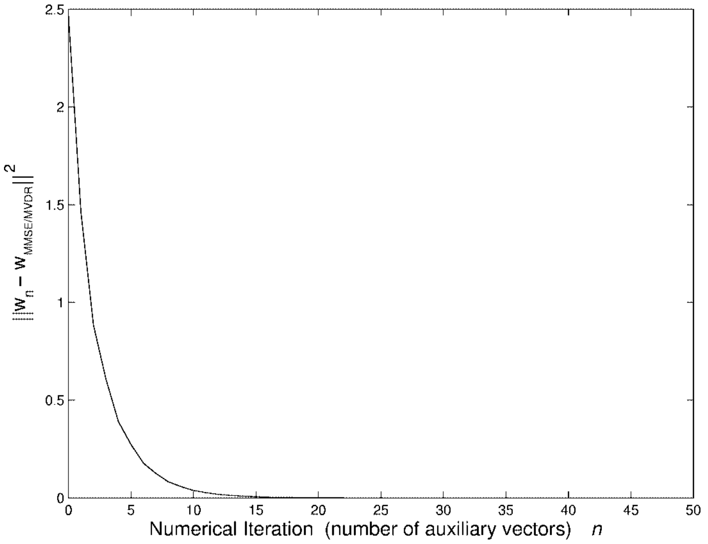

FIG. 3. Convergence of the sequence of filters $\mathbf{w}_n$, $n = 0, 1, 2, \ldots$, to the MMSE–MVDR solution for the signal model example in (14).

algorithm in Fig. 2 converges to the MMSE–MVDR solution. The convergence is captured in terms of the norm-square metric $\|\mathbf{w}_n - \mathbf{w}_{\text{MMSE/MVDR}}\|^2$ as a function of the iteration step (index of the AV filter in the sequence or number of auxiliary vectors used) n.

3. FILTER ESTIMATION

Consider a constraint vector $\mathbf{v}$ and a Hermitian positive definite autocorrelation matrix $\mathbf{R}$ of an input vector $\mathbf{r} \in \mathbb{C}^L$. Assume that $\mathbf{R}$ is in fact unknown and it is sample-average estimated from a data record of M points: $\widehat{\mathbf{R}}(M) = (1/M)\sum_{m=1}^{M} \mathbf{r}_m \mathbf{r}_m^H$. For Gaussian inputs, $\widehat{\mathbf{R}}(M)$ is a maximum-likelihood (ML), consistent, unbiased estimator of $\mathbf{R}$ [3, 30]. For a large class of multivariate elliptically contoured input distributions that includes the Gaussian, if $M \geq L$ then $\widehat{\mathbf{R}}(M)$ is positive definite (hence invertible) with probability 1 (w.p. 1) [31–33]. Then, Theorem 1 in Section 2 shows that

$$\widehat{\mathbf{w}}_n(M) \longrightarrow_{n\to\infty} \widehat{\mathbf{w}}_\infty(M) = \rho^* \frac{[\widehat{\mathbf{R}}(M)]^{-1}\mathbf{v}}{\mathbf{v}^H[\widehat{\mathbf{R}}(M)]^{-1}\mathbf{v}}, \tag{15}$$

where $\widehat{\mathbf{w}}_\infty(M)$ is the widely used MMSE–MVDR filter estimator known as the SMI filter [5].

The output sequence begins from $\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$, which is a *0*-variance, fixed-valued estimator that may be severely biased ($\widehat{\mathbf{w}}_0(M) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v} \neq \mathbf{w}_{\text{MMSE/MVDR}}$) unless $\mathbf{R} = \sigma^2\mathbf{I}$, for some $\sigma > 0$. In the latter trivial case, $\widehat{\mathbf{w}}_0(M)$ is already the perfect MMSE–MVDR filter. Otherwise, the next filter estimator in the sequence, $\widehat{\mathbf{w}}_1(M)$, has a significantly reduced bias due to the optimization procedure employed, at the expense of nonzero estimator (co-) variance. As we move up in the sequence of filter estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, the bias decreases rapidly to zero[2] while the variance increases slowly to the SMI ($\widehat{\mathbf{w}}_\infty(M)$) levels (cf. (15)). To quantify these remarks, we plot in Fig. 4 the norm-square bias $\|E\{\widehat{\mathbf{w}}_n(M)\} - \mathbf{w}_{\text{MMSE/MVDR}}\|^2$ and the trace of the covariance matrix $E\{[\widehat{\mathbf{w}}_n(M) - E\{\widehat{\mathbf{w}}_n(M)\}][\widehat{\mathbf{w}}_n(M) - E\{\widehat{\mathbf{w}}_n(M)\}]^H\}$ as a function of the iteration step n for the signal model example of Fig. 3 and data record size $M = 256$. Bias and cov-trace values are calculated from 100000 independent filter estimator realizations for each iteration point n. Formal, theoretical statistical analysis of the generated estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$, is beyond the scope of this presentation. We do note, however, that for multivariate elliptically contoured input distributions, an analytic expression for the covariance matrix of the SMI estimator $\widehat{\mathbf{w}}_\infty(M)$ can be found in [33]: $E\{[\widehat{\mathbf{w}}_\infty(M) - E\{\widehat{\mathbf{w}}_\infty(M)\}][\widehat{\mathbf{w}}_\infty(M) - E\{\widehat{\mathbf{w}}_\infty(M)\}]^H\} = [|\rho|^2 E(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})(M - L + 1)](\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})$. Since under these input distribution conditions $\widehat{\mathbf{w}}_\infty(M)$ 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 depend on the data record size M and 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 CDMA signal model example in (14) the input is Gaussian-mixture distributed. Therefore, the analytic result in [33] is not directly applicable and can only be thought of as an approximation (a rather close approximation as we concluded in our studies). In any case, from the results in Fig. 4 for $M = 256$, we see that the

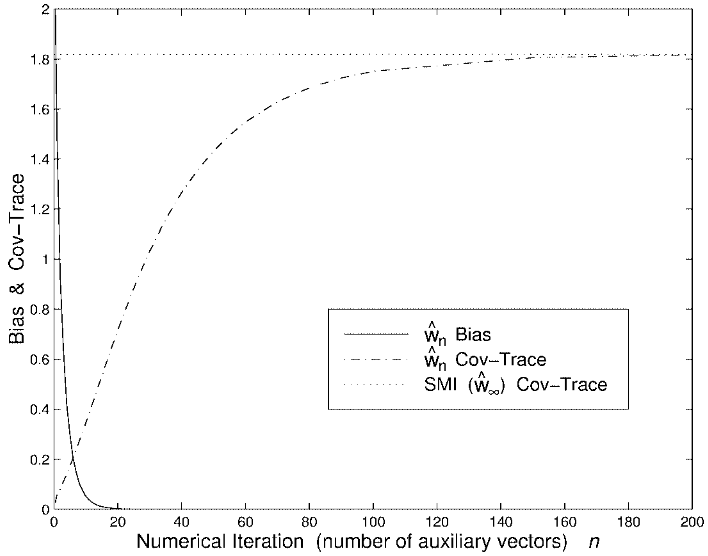

FIG. 4. Norm-square bias and covariance trace for the sequence of estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, \ldots$. The signal model is as in Fig. 3 and $M = 256$.

[2] The SMI estimator is unbiased for multivariate elliptically contoured input distributions [33, 34]: $E\{\widehat{\mathbf{w}}_\infty(M)\} = \mathbf{w}_{\text{MMSE/MVDR}} = \rho^*\mathbf{R}^{-1}\mathbf{v}/\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}$.

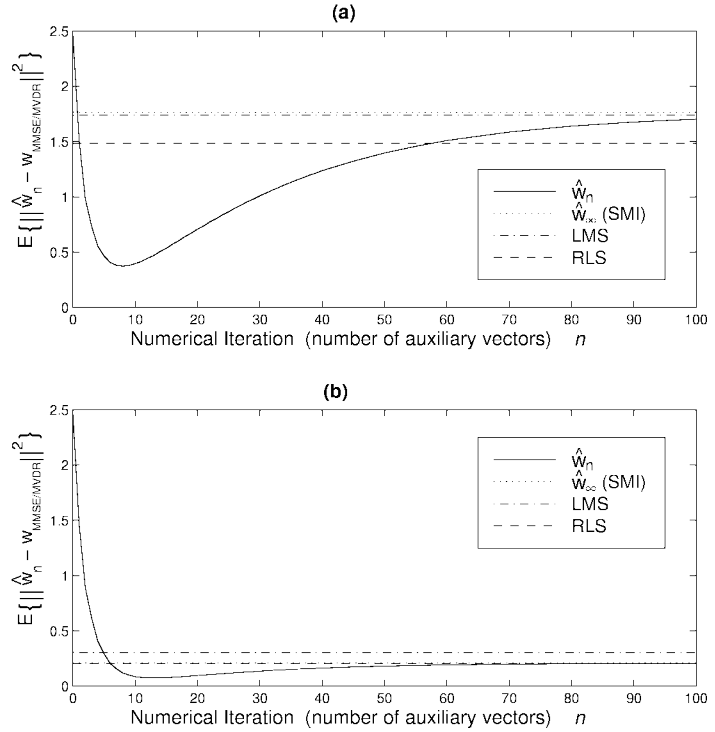

FIG. 5. MS estimation error for the sequence of estimators $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, \ldots$. Data record size (a) $M = 256$, (b) $M = 2048$.

estimators $\widehat{\mathbf{w}}_1(M), \widehat{\mathbf{w}}_2(M), \ldots$, up to about $\widehat{\mathbf{w}}_{20}(M)$ are particularly appealing. In contrast, the estimators $\widehat{\mathbf{w}}_n(M)$ 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 $E\{\|\widehat{\mathbf{w}}_n(M) - \mathbf{w}_{\text{MMSE/MVDR}}\|^2\}$ captures the bias–variance balance of the individual members of the estimator sequence $\widehat{\mathbf{w}}_n(M)$, $n = 0, 1, 2, \ldots$. In Fig. 5 we plot the MS estimation error as a function of the iteration step (index of AV filter in the sequence or number of auxiliary vectors) n for the case study in Fig. 4, for $M = 256$ (Fig. 5a) and $M = 2048$ (Fig. 5b). As a reference, we also include the MS-error of the constraint-LMS estimator [4, 35]

$$\widehat{\mathbf{w}}_{\text{LMS}}(m) = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2}\right)\left[\widehat{\mathbf{w}}_{\text{LMS}}(m-1) - \mu \mathbf{r}_m \mathbf{r}_m^H \widehat{\mathbf{w}}_{\text{LMS}}(m-1)\right] + \frac{\rho^*}{\|\mathbf{v}\|^2}\mathbf{v},$$
$$m = 1, \ldots, M, \tag{16}$$

with $\widehat{\mathbf{w}}_{\text{LMS}}(0) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ and some $\mu > 0$, and the RLS estimator [7, 8] with matrix-inversion-lemma-based $\mathbf{R}^{-1}$ estimation:

$$\widehat{\mathbf{R}}^{-1}(m) = \widehat{\mathbf{R}}^{-1}(m-1) - \frac{\widehat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m\mathbf{r}_m^H\widehat{\mathbf{R}}^{-1}(m-1)}{1 + \mathbf{r}_m^H\widehat{\mathbf{R}}^{-1}(m-1)\mathbf{r}_m}, \qquad m = 1, \ldots, M, \tag{17}$$

with $\widehat{\mathbf{R}}^{-1}(0) = (1/\epsilon_0)\mathbf{I}$ for some $\epsilon_0 > 0$. Theoretically, it is known that the LMS gain parameter $\mu > 0$ [36] has to be less than $1/(2 \cdot \lambda_{\text{max}}^{\text{blocked}})$, where

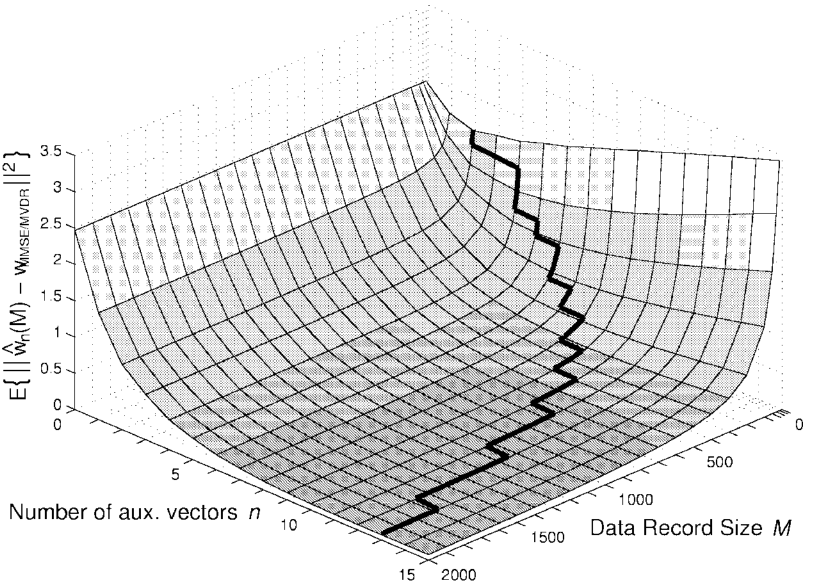

FIG. 6. MS estimation error versus number of auxiliary vectors n and sample support M.

$\lambda_{max}^{blocked}$ is the maximum eigenvalue of the blocked-data autocorrelation matrix $(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)\mathbf{R}(\mathbf{I} - \mathbf{v}\mathbf{v}^H/\|\mathbf{v}\|^2)$. 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 the RLS initialization parameter $\epsilon_0 > 0$ [37] is necessary to achieve acceptable performance (in our study we set $\mu = 1/(200 \cdot \lambda_{max}^{blocked})$ and $\epsilon_0 = 20$, respectively). This data specific tuning frequently results in misleading, overoptimistic conclusions about the short data record performance of the LMS and RLS algorithms. In contrast, when the filter estimators $\widehat{\mathbf{w}}_n$ 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 $\{\widehat{\mathbf{w}}_n\}$ sequence. Adaptive, data-dependent criteria for the selection of the most appropriate AV filter in the sequence for a given data record are developed in the next section. In Fig. 5a, for $M = 256$ all estimators $\widehat{\mathbf{w}}_n$ from $n = 2$ up to about $n = 55$ outperform in MS-error their RLS, LMS, and SMI ($\widehat{\mathbf{w}}_\infty$) counterparts. $\widehat{\mathbf{w}}_8$ ($n = 8$ auxiliary vectors) has the least MS-error of all (best bias–variance trade-off). When the data record size is increased to $M = 2048$ (Fig. 5b), we can afford more iterations (more auxiliary vectors) and $\widehat{\mathbf{w}}_{13}$ offers the best bias–variance trade-off (lowest MS-error). All filter estimators $\widehat{\mathbf{w}}_n$ for $n > 8$ outperform the LMS, RLS, and SMI ($\widehat{\mathbf{w}}_\infty$) estimators. For such large data record sets ($M = 2048$), the RLS and the SMI ($\widehat{\mathbf{w}}_\infty$) MS-errors are almost identical. Figure 6 offers a 3-dimensional plot of the MS estimation error as a function of the number of auxiliary vectors n and the sample support M. The dark line that traces the bottom of the MS estimation error surface identifies the best number of auxiliary vectors for any given data record size M.

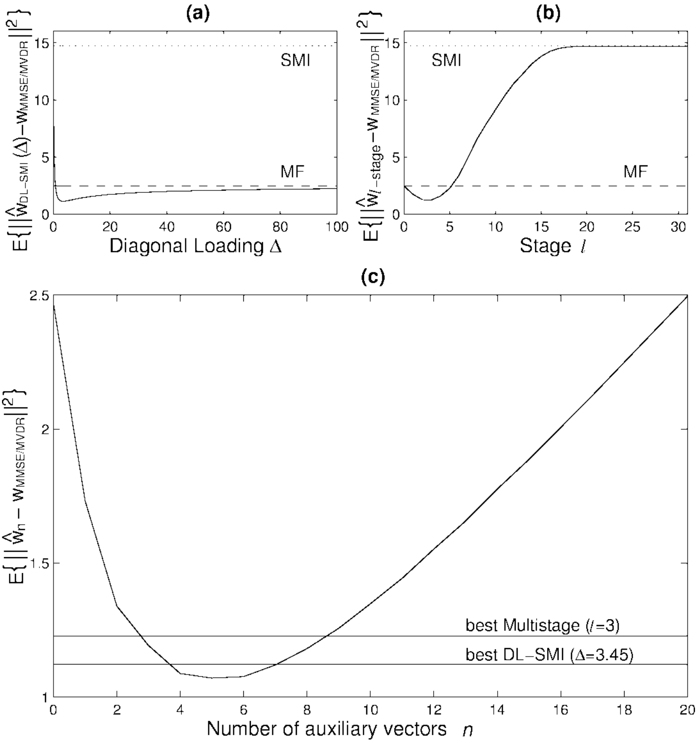

FIG. 7. MS estimation error studies for (a) diagonally loaded SMI, (b) multistage, and (c) auxiliary-vector estimators ($M = 60$).

An alternative bias–variance trading mechanism through real-valued tuning is the diagonally loaded (DL) SMI estimator [6]

$$\widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta) = \rho^* \frac{[\widehat{\mathbf{R}}(M) + \Delta \mathbf{I}]^{-1}\mathbf{v}}{\mathbf{v}^H[\widehat{\mathbf{R}}(M) + \Delta \mathbf{I}]^{-1}\mathbf{v}}, \tag{18}$$

where $\Delta \geq 0$ is the diagonal loading parameter. We observe that $\widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta = 0)$ is the regular SMI estimator, while $\lim_{\Delta \to \infty} \widehat{\mathbf{w}}_{\text{DL-SMI}}(\Delta) = (\rho^*/\|\mathbf{v}\|^2)\mathbf{v}$ which is the properly scaled matched filter. In Fig. 7a 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. 7c we compare the best DL-SMI estimator against the AV estimator sequence for which *no* diagonal loading is performed. Interestingly, the AV estimators $\widehat{\mathbf{w}}_n$ 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 [9, 10] (the resulting filters have been also referred to as nested Wiener filters). 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 (9) 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 $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n-1}$, $n \leq L-1$:

$$\mathbf{y}_n = \left(\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2} - \sum_{i=1}^{n-1} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\|\mathbf{y}_i\|^2} \right) \mathbf{R}\mathbf{w}_{n-1}. \tag{19}$$

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 $\mathbf{y}_1, \ldots, \mathbf{y}_l$ in the form of a blocking matrix $\mathbf{B}_{L \times l} = [\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots,\ \mathbf{y}_l]$. Then,

$$\mathbf{w}_{l\text{-stage}} = \frac{\rho^*}{\|\mathbf{v}\|^2} \mathbf{v} - \mathbf{B}_{L \times l} \vec{\boldsymbol{\alpha}}_{l \times 1}, \tag{20}$$

where

$$\vec{\boldsymbol{\alpha}} = \frac{\rho^*}{\|\mathbf{v}\|^2} [\mathbf{B}^H \mathbf{R} \mathbf{B}]^{-1} \mathbf{B}^H \mathbf{R} \mathbf{v} \tag{21}$$

is the MS vector-optimum (unconditionally optimum) set of weights of the vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l$.[3] In the context of MMSE–MVDR filter estimation from a data record of size M, $\widehat{\mathbf{w}}_{0\text{-stage}}(M)$ is the matched filter and $\widehat{\mathbf{w}}_{(L-1)\text{-stage}}(M)$ is the SMI estimator. In Fig. 7b we plot the MS estimation error of $\widehat{\mathbf{w}}_{l\text{-stage}}(M)$ 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. 7c we compare against the AV estimator sequence. We see that all AV estimators $\widehat{\mathbf{w}}_n$ from $n = 3$ to 8 outperform in MS-error the best multistage estimator ($l = 3$ stages). Finally, as a last study, in Fig. 8 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 $M = 3L = 96$.

4. HOW TO CHOOSE THE NUMBER OF AUXILIARY VECTORS

In this section we present two data-driven rules for the selection of the number of auxiliary vectors n [39]. The first rule selects the AV filter estimator with n auxiliary vectors that has minimum cross-validated average filter output energy. The second selection rule is specific to BPSK communications receivers that employ a sign detector at the output of the linear auxiliary-vector filter. Details are given below.

[3] Therefore, the multistage filter in [9, 10] is identical to the filter $\mathbf{w}_\mathbf{B}$ as it appears in [22–24]. 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 [38] 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 auxiliary vectors. 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 $\widehat{\mathbf{R}}(M)$.

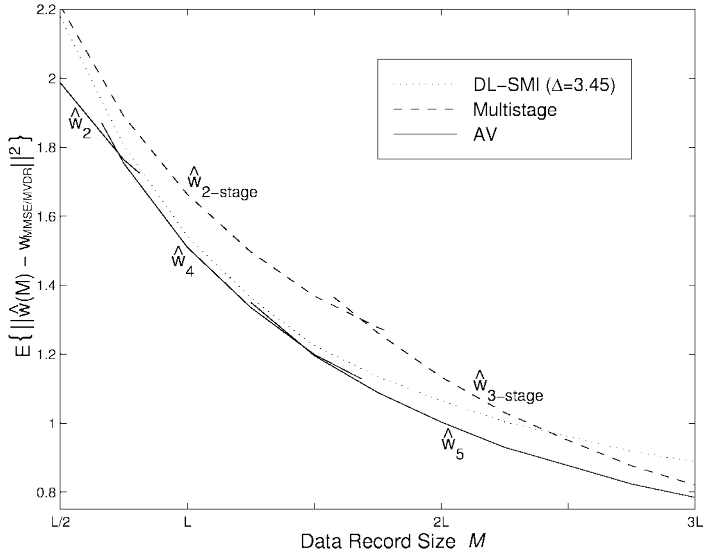

FIG. 8. MS estimation error for the *best* multistage and AV estimators over the data support range $M = L/2 = 16$ to $M = 3L = 96$. The MS estimation error of the $\Delta = 3.45$ DL-SMI estimator is also included as a reference.

4.1. Cross-Validated Minimum Output Variance Rule (CV-MOV)

Cross-validation is a well-known statistical method [40]. Here, we use cross-validation to select the filter parameter of interest (number of auxiliary vectors n) that minimizes the output variance which is estimated based on input observations that have not been used in the process of building the filter estimator itself.

A particular case of cross-validation that we use in this work is the leave-one-out method. The following criterion defines the CV-MOV AV filter estimator selection process.

CRITERION 1. *For a given data record of size M, the cross-validated minimum output variance AV filter estimator selection rule chooses the AV filter estimator $\widehat{\mathbf{w}}_{n_1}(M)$ that minimizes the cross-validated sample average output variance, i.e.,*

$$n_1 = \arg\min_n \left\{ \sum_{m=1}^{M} \widehat{\mathbf{w}}_n^H(M\backslash m)\mathbf{r}_m\mathbf{r}_m^H\widehat{\mathbf{w}}_n(M\backslash m) \right\}, \tag{22}$$

where $(M\backslash m)$ identifies the AV filter estimator that is evaluated from the available data record after removing the mth sample.

It may be important to emphasize the need for invoking the cross-validation technique for the evaluation of the sample-average output variance. If a sample-average evaluation using *all* data were attempted, then the selection rule would take the form $\min_n\{\widehat{\mathbf{w}}_n^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_n(M)\}$, where $\widehat{\mathbf{R}}(M) = (1/M)\sum_{m=1}^{M}\mathbf{r}_m\mathbf{r}_m^H$ and $\widehat{\mathbf{w}}_n(M)$ is given by the algorithm of Fig. 2 with $\widehat{\mathbf{R}}(M)$ in place of $\mathbf{R}$. Such minimization, however, would result in $n = \infty$ since we know that $\widehat{\mathbf{w}}_n^H(M)\widehat{\mathbf{R}}(M)$

$\widehat{\mathbf{w}}_n(M) \underset{n\to\infty}{\longrightarrow} \widehat{\mathbf{w}}_\infty^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_\infty(M)$ and the SMI estimator $\widehat{\mathbf{w}}_\infty(M)$ achieves minimum sample average output variance $\widehat{\mathbf{w}}_\infty^H(M)\widehat{\mathbf{R}}(M)\widehat{\mathbf{w}}_\infty(M)$ (but not minimum true output variance $\widehat{\mathbf{w}}_\infty^H(M)\mathbf{R}\widehat{\mathbf{w}}_\infty(M)$, of course).

4.2. Output J-divergence Rule

For illustration purposes we reconsider the BPSK CDMA signal model example in (14). The output J-divergence rule selects the AV filter estimator from the sequence of AV estimators that maximizes the J-divergence of the Gaussian approximated conditional filter-output distributions (conditioned on the transmitted information bit $b_1 = +1$ or $b_1 = -1$). The appropriateness of such a criterion as well as implementation details are presented below.

For a given AV filter estimator $\widehat{\mathbf{w}}_n(M)$, we denote by p_m the real part[4] of the filter output with input the mth data vector $\mathbf{r}_m$, $m = 1, 2, \ldots, M$,

$$p_m \stackrel{\triangle}{=} \mathrm{Re}\left[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m\right] = \sqrt{E_1}\,\mathrm{Re}\left[b_1(m)\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_1\right] + \sum_{k=2}^{K}\sqrt{E_k}\,\mathrm{Re}\left[b_k(m)\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k\right] + \mathrm{Re}\left[\widehat{\mathbf{w}}_n^H(M)\mathbf{n}_m\right], \tag{23}$$

where the information bits $b_k(m)$, $k = 1, 2, \ldots, K$, $m = 1, 2, \ldots, M$, are assumed to be independent identically distributed (i.i.d.) with equally probable outcomes and $\mathbf{n}_m$ is a 0-mean complex white Gaussian random vector with autocovariance matrix $\sigma^2\mathbf{I}$. Then, the scalars p_m, $m = 1, 2, \ldots, M$, are i.i.d. with common distribution $f_P(x)$ given by

$$f_P(x) = \frac{1}{2^K\sqrt{2\pi}\sigma}\sum_{i=1}^{2^K}\exp\left\{\frac{-\{x - \sum_{k=1}^{K}\sqrt{E_k}\,\mathrm{Re}[b_k^{(i)}\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k]\}^2}{2\sigma^2\|\widehat{\mathbf{w}}_n(M)\|^2}\right\}, \tag{24}$$

where $b_k^{(i)}$, $i = 1, 2, \ldots, 2^K$, is the bit of user k in the ith bit-combination.

Conditioned on the transmitted information bit of the user of interest, user *1*, the pdf of the filter output is a mixture of 2^{K-1} Gaussian distributions. However, for "effective" interference suppressive filters we can safely approximate the conditional output distribution by a Gaussian distribution as argued in [41] for MMSE–MVDR linear filtering. Under this approximation, the filter output conditional distributions given that $+1$ or -1 is transmitted are $f_{1,n} \sim \mathcal{N}[\mu(n), \sigma_{I+N}^2(n)]$ and $f_{0,n} \sim \mathcal{N}[-\mu(n), \sigma_{I+N}^2(n)]$, respectively, where $\mu(n) \stackrel{\triangle}{=} \sqrt{E_1}\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_1]$ and $\sigma_{I+N}^2(n) \stackrel{\triangle}{=} \sum_{k=2}^{K} E_k\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{s}_k]^2 + \sigma^2\|\widehat{\mathbf{w}}_n(M)\|^2$ is the conditional variance due to multiple access interference and additive white Gaussian noise (AWGN) (the index "I+N" denotes comprehensively the disturbance contribution). The effect of the above approximation on the performance of the output J-divergence selection rule will be examined in Section 5.

The J-divergence distance $J(f_{1,n}, f_{0,n})$ between the distributions $f_{1,n}(\cdot)$ and $f_{0,n}(\cdot)$ is defined as the sum of the Kullback–Leibler (K-L) distances between

[4] While the signal model in (14) is real-valued, we choose to carry out this presentation in the more general context of complex input vectors and filters.

$f_{1,n}$ and $f_{0,n}$

$$J(f_{1,n}, f_{0,n}) \stackrel{\triangle}{=} D(f_{1,n}, f_{0,n}) + D(f_{0,n}, f_{1,n}), \tag{25}$$

where the K–L distance of $f_{1,n}$ from $f_{0,n}$ is defined by $D(f_{1,n}, f_{0,n}) \stackrel{\triangle}{=} \int_{-\infty}^{\infty} f_{1,n}(x) \log(f_{1,n}(x)/f_{0,n}(x))\,dx$ [42]. Since $J(f_{1,n}, f_{0,n})$ is a function of the AV filter-estimator parameter n (number of auxiliary vectors), in the rest of this paper we will use the notation $J(n)$ to represent the J-divergence distance between $f_{1,n}(x)$ and $f_{0,n}(x)$. For the Gaussian approximated pdf's $f_{1,n}$ and $f_{0,n}$, we have $D(f_{1,n}, f_{0,n}) = D(f_{0,n}, f_{1,n}) = [2\mu(n)]^2/2\sigma_{I+N}^2(n)$ and the J-divergence simplifies to

$$J(n) = \frac{4\mu^2(n)}{\sigma_{I+N}^2(n)}. \tag{26}$$

Expression (26) justifies our choice of the output J-divergence as one of the underlying rules for the selection of the best AV filter estimator. We recall that under the same Gaussian approximation of the conditional filter-output pdf's the filter output signal-to-interference-plus-noise-ratio (SINR) can be expressed as $J(n)/4$ and, consequently, the bit-error-rate (BER) as $Q(\sqrt{J(n)}/2)$, where $Q(x) \stackrel{\triangle}{=} \int_x^{\infty} 1/\sqrt{2\pi}\exp(-u^2/2)\mathrm{d}u$. To this extent, maximization of the output J-divergence in (26) implies minimization of the BER. Therefore, we propose to select the estimator from the generated sequence of AV filter estimators that exhibits maximum *estimated* J-divergence.

(1) *Supervised output J-divergence rule*. Exploiting the symmetry of $f_{1,n}(\cdot)$ and $f_{0,n}(\cdot)$, we can show in a straightforward manner that

$$J(n) = \frac{4E^2\{b_1 \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}}{\operatorname{Var}\{b_1 \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}} \tag{27}$$

$$= \frac{4[\sum_{i=\pm 1} E\{i \operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}] | b_1 = i\} \Pr(b_1 = i)]^2}{\sum_{i=\pm 1} \operatorname{Var}\{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}] | b_1 = i\} \Pr(b_1 = i)}, \tag{28}$$

where Var$\{\cdot\}$ and Pr$\{\cdot\}$ denote variance and probability, respectively. Assuming availability of a pilot information bit sequence $\{b_1(m)\}_{m=1}^{M}$, we propose to estimate $J(n)$ by estimating statistical expectations and probabilities via sample averaging and frequencies of occurrence, respectively. We note that although (27) and (28) are ideally equivalent (when all statistical quantities are known), this is not the case in general when estimated measures are considered. So, let $\{p_1^+, p_2^+, p_3^+, \ldots, p_{M_1}^+\} \stackrel{\triangle}{=} \{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] : b_1(m) = +1,\ m = 1, 2, 3\ldots, M\}$ and $\{p_1^-, p_2^-, p_3^-, \ldots, p_{M_2}^-\} \stackrel{\triangle}{=} \{\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] : b_1(m) = -1,\ m = 1, 2, 3, \ldots, M\}$ be the sets of all filter outputs under $b_1(m) = +1$ and $b_1(m) = -1$, respectively ($M_1, M_2 \neq 0$ and $M_1 + M_2 = M$). First, it can be shown in a straightforward manner that the estimator of the numerator of (28) (that is, the weighted average of the sample average mean of the set $\{p_m^+\}_{m=1}^{M_1}$ and the sample average mean of the set $\{p_m^-\}_{m=1}^{M_2}$ weighted by the frequency of each set) is equivalent to the estimator of the numerator of (27) $4\hat{\mu}^2(n)$ where $\hat{\mu}(n) \stackrel{\triangle}{=} \sum_{m=1}^{M_1+M_2} b_1(m) p_m/(M_1 + M_2)$ (in fact, $\hat{\mu}(n)$ is the minimum variance unbiased estimator of $\mu(n)$ [43, p. 178]).

Estimators of the denominator of (27) and (28) are examined in the following proposition. The proof is included in the Appendix.

PROPOSITION 3. *Consider the estimator of $\sigma^2_{I+N}(n)$, which is the weighted average of the sample average variance of the set $\{p^+_m\}^{M_1}_{m=1}$, and the sample average variance of the set $\{p^-_m\}^{M_2}_{m=1}$, weighted by the frequency of occurrence of each set*

$$\hat{\sigma}^2_1(n) = \frac{M_1}{M_1+M_2}\hat{\sigma}^{2+}(n) + \frac{M_2}{M_1+M_2}\hat{\sigma}^{2-}(n), \tag{29}$$

where

$$\hat{\sigma}^{2+}(n) \triangleq \frac{1}{M_1}\sum_{m=1}^{M_1}(p^+_m - \hat{\mu}^+(n))^2, \qquad \hat{\mu}^+(n) \triangleq \frac{1}{M_1}\sum_{m=1}^{M_1} p^+_m,$$

$$\hat{\sigma}^{2-}(n) \triangleq \frac{1}{M_2}\sum_{m=1}^{M_2}(p^+_m - \hat{\mu}^-(n))^2, \qquad \hat{\mu}^-(n) \triangleq \frac{1}{M_2}\sum_{m=1}^{M_2} p^-_m,$$

and $\{p^+_m\}^{M_1}_{m=1}$, $\{p^-_m\}^{M_2}_{m=1}$, M_1, and M_2 are as defined previously. Consider also the direct sample average estimator of $\sigma^2_{I+N}(n)$,

$$\hat{\sigma}^2_2(n) = \frac{1}{M_1+M_2}\sum_{m=1}^{M_1+M_2}[b_1(m)p_m - \hat{\mu}(n)]^2, \tag{30}$$

where p_m is given by (23). *The estimators $\hat{\sigma}^2_1(n)$ and $\hat{\sigma}^2_2(n)$ exhibit the following properties:* (i) *They are both biased and* (ii) *$\hat{\sigma}^2_2(n)$ exhibits smaller MSE from the true value than $\hat{\sigma}^2_1(n)$.*

Utilizing Proposition 3, estimators for the filter-output J-divergence become readily available. The following theorem identifies their relative merits. The proof is included in the Appendix.

THEOREM 2. *Define the two supervised estimators of the output J-divergence $\hat{J}_{S,1}(n) \triangleq 4\hat{\mu}^2(n)/\hat{\sigma}^2_1(n)$ and $\hat{J}_{S,2}(n) \triangleq 4\hat{\mu}^2(n)/\hat{\sigma}^2_2(n)$, where the subscript "S" identifies a supervised implementation. Then, for a given information bit pilot sequence of size $M = M_1 + M_2$, where M_1 and M_2 are the cardinalities of the sets $\{b_1(m) = +1\}$ and $\{b_1(m) = -1\}$, respectively, both estimators are biased while the MSE of $\hat{J}_{S,2}(n)$ is less than the MSE of $\hat{J}_{S,1}(n)$; i.e.,*

$$E\left\{\left[\hat{J}_{S,2}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\} < E\left\{\left[\hat{J}_{S,1}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\}, \tag{31}$$

where $4\mu^2(n)/\sigma^2_{I+N}(n)$ is the true value of $J(n)$ as given by (26).

Using the preferred estimator $\hat{J}_{S,2}(n)$, the supervised implementation of the output J-divergence AV filter estimator selection rule takes the final form given by the following criterion.

CRITERION 2. *For a given information bit pilot sequence of size M, the supervised J-divergence AV filter estimator selection rule chooses the estimator $\widehat{\mathbf{w}}_{n_2}(M)$ with n_2 auxiliary vectors where*

$$n_2 = \arg\max_n \{ \hat{J}_{S,2}(n) \} = \arg\max_n \left\{ \frac{4[\frac{1}{M}\sum_{m=1}^{M} b_1(m)\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]]^2}{\frac{1}{M}\sum_{m=1}^{M}[b_1(m)\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m] - \hat{\mu}(n)]^2} \right\}. \tag{32}$$

(2) *Unsupervised (blind) output J-divergence rule.* The blind implementation of the rule is obtained by substituting the information bit b_1 in (27) by the detected bit $\hat{b}_1 = \mathrm{sgn}[\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]]$ (output of the sign detector that follows the linear filter). In particular, using $\hat{b}_1$ in place of b_1 in (27) we obtain the following J-divergence expression

$$J_B(n) = \frac{4E^2\{\hat{b}_1\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}}{\mathrm{Var}\{\hat{b}_1\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]\}} = \frac{4E^2\{|\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|\}}{\mathrm{Var}\{|\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|\}}, \tag{33}$$

where the subscript "B" identifies the blind version of the J-divergence function. The following proposition provides the conditions under which $J_B(n)$ is nearly equal to $J(n)$. The proof is included in the Appendix.

PROPOSITION 4. *If $\mu(n)/(\sigma_{I+N}(n)) \gg 1$, i.e., the filter output SINR is significantly higher than 0 dB, then $J_B(n) \approx J(n)$.*

To estimate $J_B(n)$ from a data record of finite size, we substitute the statistical expectations in (33) by sample averages. The following criterion summarizes the corresponding AV filter estimator selection rule.

CRITERION 3. *For a given data record of size M, the unsupervised (blind) J-divergence AV filter estimator selection rule chooses the estimator $\widehat{\mathbf{w}}_{n_3}(M)$ with n_3 auxiliary vectors where*

$$n_3 = \arg\max_n \{ \hat{J}_B(n) \} = \arg\max_n \left\{ \frac{4[\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|]^2}{\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|^2 - [\frac{1}{M}\sum_{m=1}^{M} |\,\mathrm{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}_m]|]^2} \right\}. \tag{34}$$

5. SIMULATION STUDIES

We examine the performance of the proposed short data record AV filter estimator selection rules for a DS-CDMA system with K users, spreading gain L, and multipath fading reception by a narrowband antenna array with N elements. All elements experience identical fading. Let J denote the number of chip interval spaced paths per baseband user signal. After conventional carrier demodulation, chip-matched filtering, and sampling at the chip rate over a multipath extended symbol interval of $L+J-1$ chips, the $L+J-1$ data samples from the ith antenna element, $i = 1, 2, \ldots, N$, are organized in the form of a vector $\mathbf{r}_m^{(i)}$ given by

$$\mathbf{r}_m^{(i)} = \sum_{k=1}^{K} \sum_{t=1}^{J} c_{k,t} \sqrt{E_k} (b_k(m) \mathbf{s}_{k,t} + b_k^-(m) \mathbf{s}_{k,t}^- + b_k^+(m) \mathbf{s}_{k,t}^+) a_{k,t}[i] + \mathbf{n}_m^{(i)},$$
$$m = 1, \ldots, M, \; i = 1, \ldots, N. \tag{35}$$

In (35), with respect to the kth user signal, E_k is the transmitted energy, $b_k(m)$, $b_k^-(m)$, and $b_k^+(m)$ are the present, the previous, and the following transmitted bits, respectively, and $c_{k,t}$ is the coefficient of the tth path of the kth user signal. The channel coefficients are modeled as independent zero-mean complex Gaussian random variables that are assumed to remain constant over the filter adaptation data record of size M. $\mathbf{s}_{k,t}$ represents the $(J-1)$-zero-padded and $(t-1)$-right-shifted version of the signature of the kth user $\mathbf{s}_k$; $\mathbf{s}_{k,t}^-$ is the *0*-filled, L-left-shifted version of $\mathbf{s}_{k,t}$; and $\mathbf{s}_{k,t}^+$ is the *0*-filled, L-right-shifted version of $\mathbf{s}_{k,t}$. Finally, $\mathbf{n}_m^{(i)}$ represents additive complex white Gaussian noise and $a_{k,t}[i]$ denotes the ith coordinate of the array response vector $\mathbf{a}_{k,t}$ that corresponds to the tth path of the kth user signal

$$a_{k,t}[i] = \exp\left\{ j2\pi(i-1) \frac{\sin\theta_{k,t} q}{\lambda} \right\}, \qquad i = 1, \ldots, N, \tag{36}$$

where $\theta_{k,t}$ is the angle of arrival, λ is the carrier wavelength, and q is the inter-element spacing (in our studies we set $q \triangleq \lambda/2$).

We vectorize the $(L + J - 1) \times N$ space–time received data matrix $[\mathbf{r}_m^{(1)}, \mathbf{r}_m^{(2)}, \ldots, \mathbf{r}_m^{(N)}]$ to form the joint space–time data vector $\mathbf{r}_m$, which is a $(L + J - 1)N$-long column vector:

$$\mathbf{r}_m = \mathrm{Vec}\left\{ \left[\mathbf{r}_m^{(1)}, \mathbf{r}_m^{(2)}, \ldots, \mathbf{r}_m^{(N)} \right]_{(L+J-1)\times N} \right\}. \tag{37}$$

The joint space–time RAKE filter for user 1 is $\mathbf{v}_1 \triangleq E_{b_1}\{\mathbf{r}_m b_1(m)\} = \mathrm{Vec}\{[\mathbf{v}_{1,1}, \mathbf{v}_{1,2}, \ldots, \mathbf{v}_{1,N}]\}$ ($E_{b_1}\{\cdot\}$ denotes statistical expectation with respect to $b_1(m)$), and $\mathbf{v}_{1,i} \triangleq \sum_{t=1}^{J} c_{1,t} \mathbf{s}_{1,t} a_{1,t}[i]$, $i = 1, 2, \ldots, N$. The MMSE–MVDR filter is built with constraint vector $\mathbf{v} = \mathbf{v}_1$, desired response $\mathbf{w}^H \mathbf{v}_1 = 1$, and autocorrelation matrix $\mathbf{R} = E\{\mathbf{r}_m \mathbf{r}_m^H\}$.

We choose $K = 20$, $N = 5$, $J = 3$ paths with independent zero-mean complex Gaussian fading coefficients of variance one (i.e., $E\{|c_{k,t}|^2\} = 1$) and Gold signatures with processing gain $L = 31$. The total SNR's (over the three paths) of the 19 interferers are set at $\mathrm{SNR}_{2\text{–}6} = 6$ dB, $\mathrm{SNR}_{7\text{–}8} = 7$ dB, $\mathrm{SNR}_{9\text{–}13} = 8$ dB, $\mathrm{SNR}_{14\text{–}15} = 9$ dB, and $\mathrm{SNR}_{16\text{–}20} = 10$ dB. The space–time product (filter length) equals $(L + J - 1)N = (31 + 2)5 = 165$. All experimental results that follow are averages over 100 different channel realizations and 10 independent data record generations per channel.

We first examine the performance of the AV filter estimator selection rules under the assumption that no info-bit pilot sequence is available. The data record size is set equal to $M = 230$ while the total SNR of the user of interest

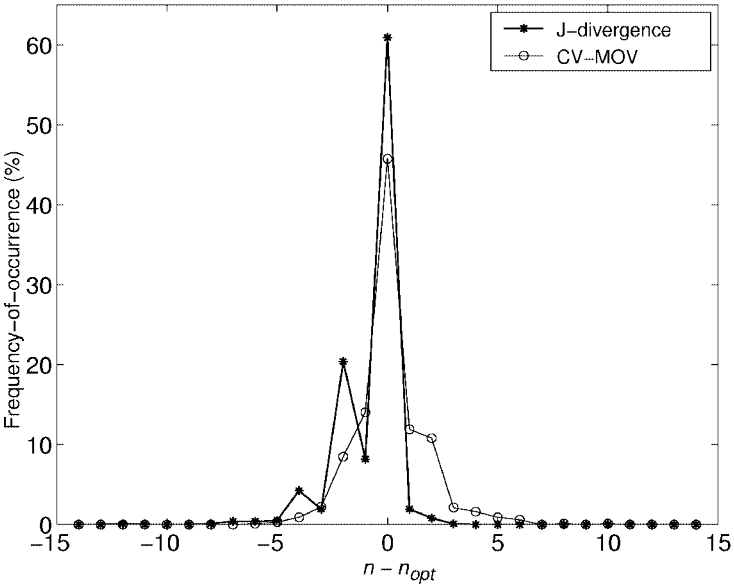

FIG. 9. Histogram of the two differences $n_1 - n_{\rm opt}$ and $n_3 - n_{\rm opt}$ where n_1 is the CV-MOV choice and n_3 is the blind J-divergence choice ($M = 230$, $\text{SNR}_1 = 8$ dB).

is set at $\text{SNR}_1 = 8$ dB. In Fig. 9, we plot the empirical pdf of the differences $(n_1 - n_{\rm opt})$ and $(n_3 - n_{\rm opt})$ where n_1 and n_3 denote selections according to Criteria 1 and 3, respectively, while $n_{\rm opt}$ denotes the "genie" maximum SINR optimum choice of the number of auxiliary vectors. We observe that both criteria provide a reliable estimate of the genie-assisted optimum number of auxiliary vectors.

The overall short data record adaptive filter performance is examined in Figs. 10 and 11. In Fig. 10, we plot the BER[5] of the AV filter estimators $\widehat{\mathbf{w}}_{n_1}(M)$ and $\widehat{\mathbf{w}}_{n_3}(M)$ as a function of the SNR of the user of interest for data records of size $M = 230$. The BER curve of the genie-assisted maximum SINR optimum filter choice $\widehat{\mathbf{w}}_{n_{\rm opt}}(M)$ as well as the corresponding curves of the ideal MMSE–MVDR filter $\mathbf{w}_{\text{MMSE/MVDR}}$, the SMI filter estimator $\widehat{\mathbf{w}}_\infty(M)$, the S-T RAKE matched filter (MF) $\widehat{\mathbf{w}}_0(M) = \mathbf{v}_1$, and the multistage filter [9, 10] with the preferred number of stages[6] $l = 7$ are also included for comparison purposes. We observe that both $\widehat{\mathbf{w}}_{n_1}(M)$ and $\widehat{\mathbf{w}}_{n_3}(M)$ are very close to the genie optimum AV filter estimator choice and outperform significantly the SMI filter estimator, the multistage filter estimator, and the matched filter. We also observe that for moderate to high SNRs of the user of interest, the J-divergence selection rule is slightly superior to the CV-MOV selection rule. The opposite is true in the low SNR range. This is explained by the fact that the J-divergence approximation

[5] The BER of each filter under consideration is approximated by $Q\left(\sqrt{\text{SINR}_{\rm out}}\right)$ [41], since the computational complexity of the BER expression for this antenna array CDMA system prohibits exact analytic evaluation.

[6] In [44], it is argued that $l = 7$ ($D = 8$ in the notation of [44]) stages are "nearly optimal over a wide range of loads and SNRs."

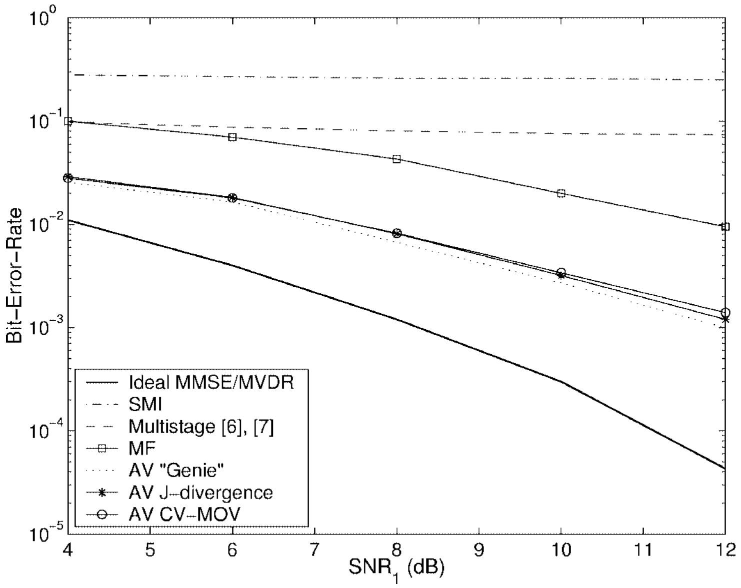

FIG. 10. BER versus SNR for the user signal of interest ($M = 230$).

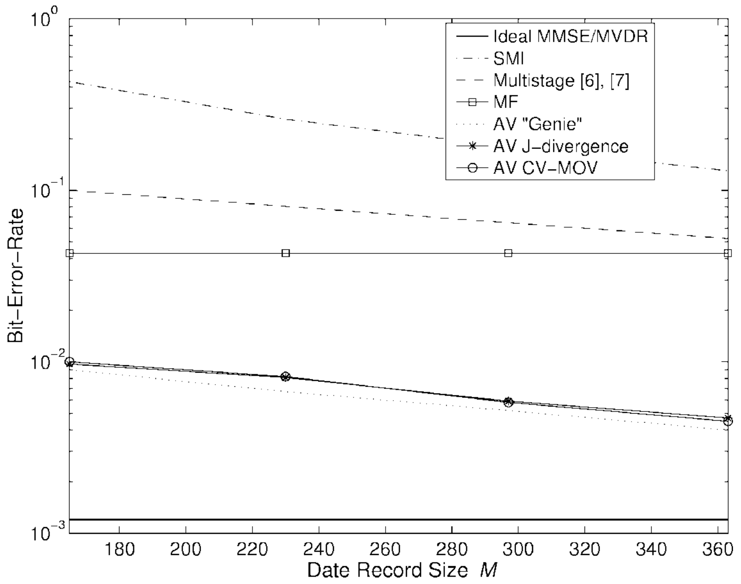

FIG. 11. BER versus data record size ($SNR_1 = 8$ dB).

$J(n) \approx J_B(n)$ used in Proposition 4 is less accurate for low filter output SINR values. On the other hand, for high filter output SINRs the discrimination capability of the CV-MOV rule is not as sharp.

Finally, Fig. 11 repeats the study of Fig. 10 as a function of the data record size. The SNR of the user of interest is fixed at 8 dB.

6. CONCLUDING REMARKS

In this article 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 MMSE–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. We analyzed basic algorithmic properties and we established formal convergence to the MMSE–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 unbiased estimator. The bias of the generated estimator sequence decreases rapidly to zero while the estimator covariance trace increases slowly 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 and RLS types, 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 parameter, or the SMI diagonal loading parameter is replaced by an integer choice among the first several members of the estimator sequence. Two data-driven criteria were proposed for the identification of the best AV filter estimator in the sequence. The first criterion calls for the minimization of the cross-validated filter-estimator output variance. The second criterion calls for the maximization of the J-divergence of the filter-estimator output-conditional distributions. Simulation studies examined and compared the operational characteristics of the proposed selection methods. With respect to the relative merits of the minimum cross-validated output variance and the maximum output J-divergence selection rules, we observed that for moderate to high output SINRs the latter method appears superior to the former (for high SINRs the cross-validated minimum output variance rule is not as sharp in discrimination ability). In contrast, in low output SINR the J-divergence method is somewhat lacking in performance (technically, the approximation in Proposition 4 is less accurate for near 0 dB or lower output SINR values). As a final general comment, the use of a sufficiently long antenna array in combination with an "effective" interference suppressive filter can result in high output SINR which favors the J-divergence selection rule, even when the *transmitted* energy of the user of interest is much lower than that of the interferers.

The auxiliary-vector algorithm in Fig. 2 together with Criteria 1, 2, and 3 form a complete toolbox for state-of-the-art estimation of MMSE–MVDR filters. The developments are of particular interest in high-dimensional adaptive signal processing applications that rely on data records of limited size.

APPENDIX

Signature Assignment for the DS-CDMA Example of Sections 3 and 4

The matrix $\mathbf{S}_{32\times 13} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_{13}]$ with columns the signature vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{13}$ is given below.

$$\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}. \quad \text{(A.1)}$$

Proof of Proposition 3. The quantities $(1/M_1)\sum_{m=1}^{M_1}(p_m^+ - \hat{\mu}^+(n))^2$ and $(1/M_2)\sum_{m=1}^{M_2}(p_m^- - \hat{\mu}^-(n))^2$ are the ML estimators for the variance of the filter output conditioned on $b_1 = +1$ and $b_1 = -1$, respectively [43, p. 179]. Both estimators are biased. (In fact, their unbiased counterparts that have multiplying factors $1/(M_1 - 1)$ and $1/(M_2 - 1)$ instead of $1/M_1$ and $1/M_2$, respectively, exhibit higher MSE.) The MSEs of the estimators of interest $\hat{\sigma}_1^2(n)$ and $\hat{\sigma}_2^2(n)$ are as follows:

$$\begin{aligned}
\mathrm{MSE}_{\hat{\sigma}_1^2(n)} &= E\left\{\left(\hat{\sigma}_1^2(n) - \sigma_{I+N}^2(n)\right)^2\right\} \\
&= E\left\{\left[\left(\frac{\sum_{m=1}^{M_1}(p_m^+ - \hat{\mu}^+(n))^2}{M_1 + M_2} - \frac{M_1\sigma_{I+N}^2(n)}{M_1 + M_2}\right)\right.\right. \\
&\qquad \left.\left. + \left(\frac{\sum_{m=1}^{M_2}(p_m^- - \hat{\mu}^-(n))^2}{M_1 + M_2} - \frac{M_2\sigma_{I+N}^2(n)}{M_1 + M_2}\right)\right]^2\right\}
\end{aligned}$$

$$= \frac{(2M_1-1)\sigma^4_{I+N}(n)}{(M_1+M_2)^2} + \frac{(2M_2-1)\sigma^4_{I+N}(n)}{(M_1+M_2)^2} + \frac{2\sigma^4_{I+N}(n)}{(M_1+M_2)^2}$$

$$= \frac{2\sigma^4_{I+N}(n)}{M_1+M_2} \tag{A.2}$$

and

$$\mathrm{MSE}_{\hat{\sigma}_2^2(n)} = E\left\{\left(\hat{\sigma}_2^2(n) - \sigma^2_{I+N}(n)\right)^2\right\} = \frac{[2(M_1+M_2)-1]\sigma^4_{I+N}(n)}{(M_1+M_2)^2}. \tag{A.3}$$

Thus,

$$\mathrm{MSE}_{\hat{\sigma}_1^2(n)} - \mathrm{MSE}_{\hat{\sigma}_2^2(n)} = \frac{\sigma^4_{I+N}(n)}{(M_1+M_2)^2} > 0. \quad \blacksquare \tag{A.4}$$

Proof of Theorem 2. $\hat{\mu}^+(n)$ and $\hat{\sigma}^{2+}(n)$ are independent random variables with distributions $\hat{\mu}^+(n) \sim \mathcal{N}(\mu(n), \sigma^2_{I+N}(n)/M_1)$ and $(M_1/\sigma^2_{I+N}(n))\hat{\sigma}^{2+}(n) \sim \chi^2_{M_1-1}$ [45] ($\chi^2_{M_1-1}$ denotes the chi-square distribution with (M_1-1) degrees of freedom). Similarly, $\hat{\mu}^-(n)$ and $\hat{\sigma}^{2-}(n)$ are independent random variables with distributions $\hat{\mu}^-(n) \sim \mathcal{N}(-\mu(n), \sigma^2_{I+N}(n)/M_2)$ and $(M_2/\sigma^2_{I+N}(n))\hat{\sigma}^{2-}(n) \sim \chi^2_{M_2-1}$. Furthermore, $(\hat{\mu}^+(n), \hat{\sigma}^{2+}(n))$ and $(\hat{\mu}^-(n), \hat{\sigma}^{2-}(n))$ are mutually independent because $(\hat{\mu}^+(n), \hat{\sigma}^{2+}(n))$ and $(\hat{\mu}^-(n), \hat{\sigma}^{2-}(n))$ are evaluated from two independent training sets.

Since $\hat{\mu}(n) = (M_1/M)\hat{\mu}^+(n) - (M_2/M)\hat{\mu}^-(n)$ and $\hat{\sigma}_1^2(n) = (M_1/M)\hat{\sigma}^{2+}(n) + (M_2/M)\hat{\sigma}^{2-}(n)$, $\hat{\mu}(n)$ and $\hat{\sigma}_1^2(n)$ are independent random variables with distributions $\hat{\mu}(n) \sim \mathcal{N}(\mu(n), \sigma^2_{I+N}(n)/M)$ and $(M/\sigma^2_{I+N}(n))\hat{\sigma}_1^2(n) \sim \chi^2_{M-2}$ where $M = M_1 + M_2$ is the size of the given information bit pilot sequence. Therefore,

$$E\{\hat{J}_{S,1}(n)\} = 4E\{\hat{\mu}^2(n)\}E\left\{\frac{1}{\hat{\sigma}_1^2(n)}\right\} = 4\left[\frac{\sigma^2_{I+N}(n)}{M} + \mu^2(n)\right]\frac{M}{\sigma^2_{I+N}(n)(M-4)}$$

$$= \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\frac{M}{M-4} + \frac{4}{M-4}, \tag{A.5}$$

$$E\left\{\hat{J}^2_{S,1}(n)\right\} = 16E\left\{\hat{\mu}^4(n)\right\}E\left\{\frac{1}{\hat{\sigma}_1^4(n)}\right\} = 16\left[\mu^4(n) + \frac{6\mu^2(n)\sigma^2_{I+N}(n)}{M} + \frac{3\sigma^4_{I+N}(n)}{M^2}\right]$$

$$\times \frac{M^2}{\sigma^4_{I+N}(n)(M-4)(M-6)}. \tag{A.6}$$

The inverse moments $E\{1/\hat{\sigma}_1^2(n)\}$ and $E\{1/\hat{\sigma}_1^4(n)\}$ exist for $M > 4$ and $M > 6$, respectively, and are given by [46]

$$E\left\{\frac{1}{\hat{\sigma}_1^2(n)}\right\} = \frac{M}{\sigma^2_{I+N}(n)(M-4)}, \tag{A.7}$$

$$E\left\{\frac{1}{\hat{\sigma}_1^4(n)}\right\} = \frac{M^2}{\sigma^4_{I+N}(n)(M-4)(M-6)}. \tag{A.8}$$

Thus, the MSE of $\hat{J}_{S,1}(n)$ is

$$\mathrm{MSE}_1 \triangleq E\left\{\left[\hat{J}_{S,1}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\} = \frac{16}{(M-4)(M-6)}a - \frac{16}{M-4}b + \frac{16\mu^4(n)}{\sigma^4_{I+N}(n)}, \tag{A.9}$$

where

$$a \triangleq \left[\mu^4(n) + \frac{6\mu^2(n)\sigma^2_{I+N}(n)}{M} + \frac{3\sigma^4_{I+N}(n)}{M^2}\right]\frac{M^2}{\sigma^4_{I+N}(n)} \quad \text{and}$$
$$b \triangleq \left[\frac{\sigma^2_{I+N}(n)}{M} + \mu^2(n)\right]\frac{2M\mu^2(n)}{\sigma^4_{I+N}(n)}. \tag{A.10}$$

On the other hand, $\hat{\mu}(n)$ and $\hat{\sigma}^2_2(n)$ are independent random variables with distributions $\hat{\mu}(n) \sim \mathcal{N}(\mu(n), \sigma^2_{I+N}(n)/M)$ and $(M/\sigma^2_{I+N}(n))\hat{\sigma}^2_2(n) \sim \chi^2_{M-1}$, respectively. Therefore,

$$E\{\hat{J}_{S,2}(n)\} = 4\left[\frac{\sigma^2_{I+N}(n)}{M} + \mu^2(n)\right]\frac{M}{\sigma^2_{I+N}(n)(M-3)}$$
$$= \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\frac{M}{M-3} + \frac{4}{M-3}, \tag{A.11}$$

$$E\{\hat{J}^2_{S,2}(n)\} = 16\left[\mu^4(n) + \frac{6\mu^2(n)\sigma^2_{I+N}(n)}{M} + \frac{3\sigma^4_{I+N}(n)}{M^2}\right]$$
$$\times \frac{M^2}{\sigma^4_{I+N}(n)(M-3)(M-5)}. \tag{A.12}$$

Thus, the MSE of $\hat{J}_{S,2}(n)$ is

$$\mathrm{MSE}_2 \triangleq E\left\{\left[\hat{J}_{S,2}(n) - \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)}\right]^2\right\}$$
$$= \frac{16}{(M-3)(M-5)}a - \frac{16}{M-3}b + \frac{16\mu^4(n)}{\sigma^4_{I+N}(n)}, \tag{A.13}$$

where the first and second inverse moments $E\{1/\hat{\sigma}^2_2(n)\}$ and $E\{1/\hat{\sigma}^4_2(n)\}$ exist for $M > 3$ and $M > 5$, respectively, and a and b are given by (A.10).

From (A.9) and (A.13), assuming $M > 6$ which is the condition for all inverse moments to exist, we obtain

$$\mathrm{MSE}_1 - \mathrm{MSE}_2 = \frac{16a}{(M-4)(M-6)} - \frac{16a}{(M-3)(M-5)} + \frac{b}{M-3} - \frac{b}{M-4}$$
$$= \frac{16}{(M-3)(M-4)}\left[\frac{2M-9}{(M-5)(M-6)}a - b\right]$$
$$> \frac{16}{(M-3)(M-4)}\left[\frac{2(M-5)}{(M-5)(M-6)}a - b\right]$$
$$= \frac{32[6M\mu^4(n) + 5M\mu^2(n)\sigma^2_{I+N}(n) + 6\mu^2(n)\sigma^2_{I+N}(n) + 3\sigma^4_{I+N}(n)]}{(M-3)(M-4)(M-6)\sigma^4_{I+N}(n)}$$
$$> 0. \tag{A.14}$$

Proof of Proposition 4. Define $Y \overset{\triangle}{=} |\operatorname{Re}[\widehat{\mathbf{w}}_n^H(M)\mathbf{r}]|$. The pdf of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_{I+N}(n)}\left[\exp\frac{-(y-\mu(n))^2}{2\sigma^2_{I+N}(n)} + \exp\frac{-(y+\mu(n))^2}{2\sigma^2_{I+N}(n)}\right]U(y), \tag{A.15}$$

where $U(y)$ is the unit step function. The mean of Y is

$$E\{Y\} = \int y f_Y(y)\,dy = \mu(n) + \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(n)}{2\sigma^2_{I+N}}\right) - 2\mu(n)Q\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right). \tag{A.16}$$

The series expansion of $Q(x)$ is [47]

$$Q(x) = \frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}\right)\left\{1 - \frac{1}{x^2} + \frac{1\cdot 3}{x^4} + \cdots + \frac{(-1)^n 1\cdot 3\cdots(2n-1)}{x^{2n}}\right\} + R_n, \tag{A.17}$$

where $R_n = (-1)^{n+1} 1\cdot 3\cdots(2n+1)\int_x^\infty \frac{1}{\sqrt{2\pi}t^{2n+2}}\exp(-t^2/2)\,dt$ is the remainder which is always less (in absolute value) than the first neglected term. If $\mu(n)/\sigma_{I+N}(n) \gg 1$, (A.16) can be written as

$$\begin{aligned} E\{Y\} &= \mu(n)\left\{1 + \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}\mu(n)}\exp\left(-\frac{\mu^2(n)}{2\sigma^2_{I+N}(n)}\right)\right. \\ &\quad \left. - \frac{2\sigma_{I+N}(n)}{\sqrt{2\pi}\mu(n)}\exp\left(-\frac{\mu^2(n)}{2\sigma^2_{I+N}(n)}\right)\left[1 - O\left[\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right)^{-2}\right]\right]\right\} \\ &= \mu(n)\left\{1 + \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2(n)}{2\sigma^2_{I+N}(n)}\right)O\left[\left(\frac{\mu(n)}{\sigma_{I+N}(n)}\right)^{-3}\right]\right\} \\ &\approx \mu(n). \end{aligned} \tag{A.18}$$

The variance of Y is

$$\operatorname{Var}\{Y\} = E\{Y^2\} - E^2\{Y\} = \mu^2(n) + \sigma^2_{I+N}(n) - E^2\{Y\} \tag{A.19}$$

and using (A.18) we may approximate

$$\operatorname{Var}\{Y\} \approx \sigma^2_{I+N}(n). \tag{A.20}$$

Finally, from (A.18) and (A.20) we obtain

$$J_B(n) = \frac{4E^2\{Y\}}{\operatorname{Var}\{Y\}} \approx \frac{4\mu^2(n)}{\sigma^2_{I+N}(n)} = J(n). \tag{A.21}$$

REFERENCES

1. Capon, J., High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57** (1969), 1408–1418.
2. Owsley, N. L., A recent trend in adaptive spatial processing for sensor arrays: Constraint adaptation. In *Signal Processing* (Griffiths, J. W. R., *et al.*, Ed.). pp. 591–604, Academic Press, New York, 1973.
3. Nahi, N. E., *Estimation Theory and Applications*. Krieger, Huntington, NY, 1976.
4. Widrow, B., Mantey, P. E., Griffiths, L. J., and Goode, B. B., Adaptive antenna systems. *Proc. IEEE* **55** (1967), 2143–2158.
5. Reed, I. S., Mallet, J. D., and Brennan, L. E., Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerospace Electron. Systems* **10** (1974), 853–863.
6. Carlson, B. D., Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Trans. Aerospace Electron. Systems* **24** (1988), 397–401.
7. Plackett, R. L., Some theorems in least squares. *Biometrika* **37** (1950), 149.
8. Wiggins, R. A. and Robinson, E. A., Recursive solution to the multichannel filtering problem. *J. Geophys. Res.* **70** (1965), 1885–1891.
9. Goldstein, J. S., Reed, I. S., Zulch, P. A., and Melvin, W. L., A multistage STAP CFAR detection technique. In *Proc. IEEE Radar Conf.* pp. 111–116, Dallas, TX, May 1998.
10. Goldstein, J. S., Reed, I. S., and Scharf, L. L., A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Inform. Theory* **44** (1998), 2943–2959.
11. Griffiths, L. J. and Jim, C. W., An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagation* **30** (1982), 27–34.
12. Van Veen, B. D. and Roberts, R. A., Partially adaptive beamformer design via output power minimization. *IEEE Trans. Acoust. Speech Signal Process.* **35** (1987), 1524–1532.
13. Scharf, L. L. and Tufts, D. W., Reduced rank methods for modeling stationary signals. *IEEE Trans. Acoust. Speech Signal Process.* **35** (1987), 350–354.
14. Van Veen, B. D., Eigenstructure based partially adaptive array design. *IEEE Trans. Antennas Propagation* **36** (1988), 357–362.
15. Haimovich, A. M. and Bar-Ness, Y., An eigenanalysis interference canceler. *IEEE Trans. Signal Process.* **39** (1991), 76–84.
16. Byerly, K. A. and Roberts, R. A., Output power based partially adaptive array design. In *Proc. Asilomar Conf. Signals, Syst., Computers*. pp. 576–580, Pacific Grove, CA, 1989.
17. Goldstein, J. S. and Reed, I. S., Reduced-rank adaptive filtering. *IEEE Trans. Signal Process.* **45** (1997), 492–496.
18. Pados, D. A. and Batalama, S. N., Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers. *IEEE Trans. Commun.* **45** (1997), 1586–1594.
19. Kansal, A., Batalama, S. N., and Pados, D. A., Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels. *IEEE J. Select. Areas Commun.* **16** (1998), 1765–1773.
20. Batalama, S. N., Medley, M. J., and Pados, D. A., Robust adaptive recovery of spread-spectrum signals with short data records. *IEEE Trans. Commun.* **48** (2000), 1725–1731.
21. Pados, D. A. and Karystinos, G. N., An iterative algorithm for the computation of the MVDR filter. *IEEE Trans. Signal Process.* **49** (2001), 290–300.
22. Pados, D. A. and Batalama, S. N., Joint space–time auxiliary-vector filtering for DS/CDMA systems with antenna arrays. In *Proc. Conf. on Inform. Sci. and Systems.* Vol. II, pp. 1007–1013, Princeton University, Princeton, NJ, March 1998.
23. Pados, D. A., Michels, J. H., Tsao, T., and Wicks, M. C., Joint domain space–time adaptive processing with small training data sets. In *Proc. IEEE Radar Conf.* pp. 99–104, Dallas, TX, May 1998.
24. Pados, D. A. and Batalama, S. N., Joint space–time auxiliary-vector filtering for DS/CDMA systems with antenna arrays. *IEEE Trans. Commun.* **47** (1999), 1406–1415.
25. Rapajic, P. B. and Vucetic, B. S., Adaptive receiver structures for asynchronous CDMA systems. *IEEE J. Select. Areas Commun.* **12** (1994), 685–697.
26. Madhow, U. and Honig, M. L., MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. Commun.* **42** (1994), 3178–3188.
27. Miller, S. L., An adaptive DS-CDMA receiver for multiuser interference rejection. *IEEE Trans. Commun.* **43** (1995), 1746–1755.

28. Honig, M. L., Madhow, U., and Verdu, S., Blind adaptive multiuser detection. *IEEE Trans. Inform. Theory* **41** (1995), 944–960.
29. Pateros, C. N. and Saulnier, G. J., An adaptive correlator receiver for direct-sequence spread-spectrum communication. *IEEE Trans. Commun.* **44** (1996), 1543–1552.
30. Richmond, C. D., Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance. *IEEE Trans. Signal Process.* **44** (1996), 305–315.
31. Dykstra, R. L., Establishing the positive definiteness of the sample covariance matrix. *Ann. Math. Statist.* **41** (1970), 2153–2154.
32. Steinhardt, A. O., Adaptive multisensor detection and estimation. In *Adaptive Radar Detection and Estimation* (Haykin, S. and Steinhardt, A. O., Eds.). Chap. 3, Wiley, New York, 1992.
33. Richmond, C. D., PDF's, confidence regions, and relevant statistics for a class of sample covariance-based array processors. *IEEE Trans. Signal Process* **44** (1996), 1779–1793.
34. Steinhardt, A. O., The PDF of adaptive beamforming weights. *IEEE Trans. Signal Process.* **39** (1991), 1232–1235.
35. Godara, L. C. and Cantoni, A., Analysis of constrained LMS algorithm with application to adaptive beamforming using perturbation sequences. *IEEE Trans. Antennas Propagation* **34** (1986), 368–379.
36. Solo, V., The limiting behavior of LMS. *IEEE Trans. Acoust., Speech, Signal Process.* **37** (1989), 1909–1922.
37. Cioffi, J. M. and Kailath, T., Fast recursive-least-squares transversal filters for adaptive filtering. *IEEE Trans. Acoust., Speech, Signal Process.* **32** (1984), 304–337.
38. Liu, T.-C. and Van Veen, B., A modular structure for implementation of linearly constrained minimum variance beamformers. *IEEE Trans. Signal Process.* **39** (1991), 2343–2346.
39. Qian, H. and Batalama, S. N., Data-record-based criteria for the selection of an auxiliary-vector estimator of the MVDR filter. In *Proc. Asilomar Conf. Signals, Systems., Computers*. pp. 802–807, Pacific Grove, CA, Oct. 2000.
40. Rao, C. R., *Handbook of Statistics 9*. Elsevier, New York, 1993.
41. Poor, H. V. and Verdu, S., Probability of error in MMSE multiuser detection. *IEEE Trans. Inform. Theory* **43** (1997), 858–871.
42. Cover, T. M. and Thomas, J. A., *Elements of Information Theory*. Wiley, New York, 1991.
43. Poor, H. V., *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1994.
44. Honig, M. L. and Xiao, W., Performance of reduced-rank linear interference suppression. *IEEE Trans. Inform. Theory* **47** (2001), 1928–1946.
45. DeGroot, M. H., *Probability and Statistics*. Addison-Wesley, Reading, MA, 1984.
46. Jones, M. C., Expressions for inverse moments of positive quadratic forms in normal variables. *Austral. J. Statist.* **28** (1986), 242–250.
47. Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*. Dover, New York, 1965.

GEORGE N. KARYSTINOS 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 in the Department of Electrical Engineering at the State University of New York at Buffalo where he is currently a Ph.D. candidate. 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 Society, Signal Processing Society, and Information Theory Society and a member of Eta Kappa Nu.

HAOLI QIAN was born in Sichuan, P.R. China, on June 2, 1977. He received the B.Sc. in electronic engineering from Tsinghua University, Beijing, P.R. China, in 1998 and the M.Sc. in electrical engineering from the State University of New York at Buffalo in 2000. Since 2000, he has been a research assistant with the Communications and Signals Group, Department of Electrical Engineering, State University of New York at Buffalo, where he is currently working toward the Ph.D. degree. His research interests are in the areas of wireless communications and

statistical signal processing. Mr. Qian is a student member of the IEEE Communications Society and Information Theory Society.

MICHAEL J. MEDLEY received his B.S., M.S., and Ph.D. in electrical engineering from Rensselaer Polytechnic Institute in 1990, 1991, and 1995, respectively. Since 1991, he has been a research engineer for the United States Air Force Research Laboratory in Rome, NY where he has been actively involved in HF channel modeling, interference suppression in spread spectrum communications, spread spectrum waveform design and multicarrier modulation. He is currently a member of the Smart Networking Radio research group developing advanced modulation and signal processing technology for next-generation military radios. His research interests include transform domain signal processing and adaptive filtering, multicarrier modulation and multiple access communication techniques. Dr. Medley is a member of the IEEE Communications Society, Eta Kappa Nu, and Tau Beta Pi.

STELLA N. BATALAMA received the Diploma degree in computer engineering and science from the University of Patras, Greece in 1989 and the Ph.D. in electrical engineering from the University of Virginia, Charlottesville, VA, in 1994. From 1989 to 1990 she was with the Computer Technology Institute, Patras, Greece. From 1990 to 1994 she was a research assistant in the Communication Systems Laboratory, Department of Electrical Engineering, University of Virginia. In 1995 she joined the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY, where she is presently an associate professor. During the summers of 1997–2002 she was Visiting Faculty in the U.S. Air Force Research Laboratory, Rome, NY. Her research interests include small sample support adaptive filtering and receiver design, adaptive multiuser detection, robust spread-spectrum communications, supervised and unsupervised optimization, and distributed detection. Dr. Batalama is currently an associate editor for the *IEEE Communications Letters* and the *IEEE Transactions on Communications*.