A SIGNAL PROCESSING APPLICATION OF RANDOMIZED LOW-RANK APPROXIMATIONS

Peter Parker†

MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420
USA
{pparker}@ll.mit.edu

Patrick J. Wolfe and Vahid Tarokh

Harvard University
Division of Engineering and Applied Sciences
33 Oxford Street
Cambridge, MA 02138
USA
{patrick,vahid}@deas.harvard.edu

ABSTRACT

Low-rank approximations to linear operators find wide use in signal processing. In the discrete case, assuming the desired rank is known a priori, such approximations are generally calculated using the singular value decomposition. In this vein, randomized algorithms have recently been developed in the context of theoretical computer science, with the goal of achieving approximations arbitrarily close to this optimal low-rank solution with very high probability. Such algorithms function by finding (deterministic) low-rank approximations to random submatrices chosen probabilistically—thereby providing significant reductions in computational complexity, and leading to their applicability even in the case of very large matrices. Here it is demonstrated that algorithms of this type also show promise in signal processing applications, in particular for the case of adaptive beamforming in both the narrowband and wideband scenarios. Quantitative simulation results are provided to indicate that near-optimal nulling performance, as measured in terms of signal-to-interference-plus-noise ratio, may be achieved via straightforward modifications of the randomized algorithms described above. Results indicate that a large computational savings is possible, relative to standard methods, with little corresponding loss in performance.

1. INTRODUCTION

Low-rank subspace estimation is a ubiquitous problem in signal processing [1–4]. Recently, various randomized algorithms for obtaining low-rank approximations to a given matrix have been proposed in the computer science literature (see, e.g., [5, 6]). Such algorithms have been shown to achieve approximations arbitrarily close to the optimal low-rank solution with very high probability; they also provide significant reductions in computational complexity. Applications in computer science include latent semantic indexing and web searching [7].

This paper studies the performance of two such algorithms in the context of signal processing for large adaptive arrays (such as those currently under development for radar, having hundreds of degrees of freedom [8]). Arrays of this type are employed to suppress a number of interfering signals, while enhancing gain in a desired direction. Practical considerations lead to the development of many sub-optimal algorithms [9]; one frequently used algorithm well-suited to the adaptive array signal model is that of dominant eigenmode rejection [3]. This algorithm has the desirable property of rapid convergence (with respect to the amount of training data required), but suffers from a significant computational burden as it is based on an exact eigendecomposition. This motivates the development of approximate solutions enabling appealing trade-offs between computational complexity and performance, such as the randomized algorithms considered here. Indeed, on the basis of this preliminary investigation, it is hypothesized that such an approach has the potential to significantly impact a wide variety of signal processing applications.

The outline of this paper is as follows. Section 2 reviews the two randomized algorithms under consideration, and Section 3 describes the adaptive beamforming signal model and performance metric used for simulations and performance evaluation. Section 4 presents results and analysis detailing the performance of the algorithms in the context of narrowband adaptive beamforming, and Section 5 introduces a modified algorithm to improve performance in the wideband case. A summary and concluding remarks are given in Section 6.

2. RANDOMIZED LOW-RANK APPROXIMATIONS

Linear operators play a fundamental role in many signal processing applications, and hence low-rank approximants thereof find widespread use. In the discrete case, assuming the desired rank is known a priori, such approximations are generally calculated using the singular value decomposition (SVD); indeed, relatively efficient algorithms for calculating the largest k singular pairs of a matrix $\mathbf{A}$ are well-known. In many practical applications, however, approximations to the rank-k singular value solution will suffice, and indeed are desirable if they can be obtained at a lower computational cost. To that end, this section provides a brief overview of two recently introduced randomized algorithms for low-rank approximation, referred to in the sequel as the submatrix [5] and sparse [6] algorithms. While numerous variations and refinements exist in the literature (see, e.g., [10]), for the sake of comparison only the basic forms of these algorithms will be considered here.

*This work was sponsored by the Department of the Air Force under Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

†The first author performed this work while at Harvard University.

Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ be a matrix for which an approximant of rank $k \ll \operatorname{rank}(\mathbf{A})$ is desired. Given $\epsilon > 0$, the algorithms under consideration generate matrices $\hat{\mathbf{A}}$ of rank $\leq k$ for which the bound

$$\|\mathbf{A} - \hat{\mathbf{A}}\|_F \leq \|\mathbf{A} - \mathbf{A}_k\|_F + \epsilon\|\mathbf{A}\|_F \tag{1}$$

holds with some probability $1-\delta$, where $\|\mathbf{A}\|_F = (\sum_{i,j} \mathbf{A}_{ij}^2)^{1/2}$ denotes the Frobenius norm, $\mathbf{A}_k$ is the corresponding optimal rank-k approximant obtained via SVD, and $\hat{\mathbf{A}} = \mathbf{H}\mathbf{H}^H\mathbf{A}$ for some matrix $\mathbf{H}$ produced by the algorithm.

The essence of the submatrix algorithm lies in the generation of a scaled random submatrix $\mathbf{W} \in \mathbb{C}^{w\times c}$ of $\mathbf{A} \in \mathbb{C}^{m\times n}$ (where $w \ll m$ and $c \ll n$ are chosen a priori), for which in turn an approximant of rank $\leq k$ is obtained via SVD [7]:

1. Calculate a probability distribution based on the norms of columns of $\mathbf{A} \in \mathbb{C}^{m\times n}$ as follows. Let $\mathbf{A}^{(i)}$ denote the i^{th} column of $\mathbf{A}$ and define $p_i \propto \|\mathbf{A}^{(i)}\|_2^2$ to be its selection probability, with $\sum_i p_i = 1$. Choose (with replacement) $c \ll n$ columns of $\mathbf{A}$ according to this distribution, and consider a matrix $\mathbf{C}$ having columns $\mathbf{C}^{(t)} = \mathbf{A}^{(i_t)}/\sqrt{cp_{i_t}}$.
2. Calculate a probability distribution based on the norms of the rows of $\mathbf{C} \in \mathbb{C}^{m\times c}$ as follows. Let $\mathbf{C}_{(i)}$ denote the i^{th} row of $\mathbf{C}$ and define $q_i \propto \|\mathbf{C}_{(i)}\|_2^2$ to be its selection probability, with $\sum_i q_i = 1$. Choose (with replacement) $w \ll m$ rows of $\mathbf{C}$ according to this distribution and consider a matrix $\mathbf{W}$ having rows $\mathbf{W}_{(t)} = \mathbf{C}_{(i_t)}/\sqrt{wq_{i_t}}$.
3. Based on the submatrix $\mathbf{W} \in \mathbb{C}^{w\times c}$, compute the SVD $\mathbf{Z}\mathbf{\Sigma}^2\mathbf{Z}^H$ of $\mathbf{W}^H\mathbf{W}$. Index the diagonal elements of $\mathbf{\Sigma}^2$ in decreasing order as $\{\sigma_t^2\}$ and set
$$\ell = \min\left\{k, \max\left\{t : \sigma_t^2 \geq \frac{c\epsilon}{8k}\|\mathbf{W}\|_F^2\right\}\right\}.$$
4. Form restrictions $\mathbf{Z}_\ell$ and $\mathbf{\Sigma}_\ell$, corresponding to the chosen ℓ principal singular vectors of $\mathbf{W}$, and set $\mathbf{H} = \mathbf{C}\mathbf{Z}_\ell\mathbf{\Sigma}_\ell^{-1}$. Output $\mathbf{H} \in \mathbb{C}^{m\times \ell}$ and $\hat{\mathbf{A}} = \mathbf{H}\mathbf{H}^H\mathbf{A}$.

The sparse algorithm consists simply of discarding a fraction of the entries of $\mathbf{A}$ in order to accelerate standard iterative methods for low-rank approximation such as orthogonal or Lanczos iteration [6]: Fix $s > 1$, let $\mathbf{S} \in \mathbb{C}^{m\times n}$ be defined element-wise as

$$\mathbf{S}_{ij} = \begin{cases} s\mathbf{A}_{ij} & \text{with probability } s^{-1} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

and set $\mathbf{H}$ equal to the optimal rank-k estimate of $\mathbf{S}$, obtained at a reduced computational cost according to the value of s.

3. ADAPTIVE BEAMFORMING MODEL AND METRIC

This section describes the signal model under consideration and motivates the investigation of the randomized low-rank approximation algorithms described in Section 2 via the simulations to follow. Specifically, consider the case of L wideband interference signals incident at angles $\{\theta_1, \theta_2, \ldots, \theta_L\}$ on an N-element adaptive array having an inter-element spacing of $d = \lambda/2$ with regard to the chosen operating frequency λ^{-1} (see Figure 1). The signal produced at the output of the array is modeled as

$$\mathbf{x}(t) = \mathbf{D}\mathbf{s}(t) + \mathbf{n}(t),$$

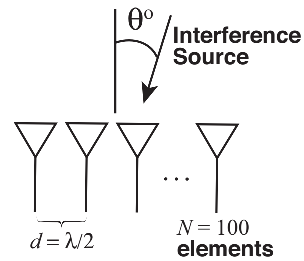

Fig. 1. The array model employed for simulations

where $\mathbf{D}(f) = \begin{bmatrix}\mathbf{d}(\theta_1, f) & \mathbf{d}(\theta_2, f) & \cdots & \mathbf{d}(\theta_L, f)\end{bmatrix}$ is an $N \times L$ matrix of interference steering vectors considered as a function of frequency, $\mathbf{s}(t)$ is an $L \times 1$ vector of interference modulations, and $\mathbf{n}(t)$ is an $N\times 1$ vector of white Gaussian noise; for the case of large arrays considered herein, it is further assumed that $L \ll N$.

The covariance matrix for this model [11] is given by

$$\mathbf{R} = \int_{\mathcal{B}} \mathbf{D}(f)\mathbf{P}(f)\mathbf{D}(f)^H df + \sigma^2\mathbf{I}, \tag{3}$$

where the region of integration $\mathcal{B}$ corresponds to the frequency band of the system, $\mathbf{P}(f)$ denotes the source covariance matrix, σ^2 is the noise power (taken here to be unity), and $(\cdot)^H$ denotes the Hermitian transpose. Further, it is assumed that there is no desirable signal component in the covariance matrix $\mathbf{R}$ (i.e., there is some "listening period" during which it is known that no signal is present; for the purposes of the current investigation it is assumed that $\mathbf{R}$ is known exactly). In this general wideband model, there are $\rho \geq L$ interference eigenvalues of $\mathbf{R}$ above the noise floor. In keeping with standard practice, the signal is termed narrowband if $\rho = L$ [11], a case which occurs when the bandwidth-time product (in the sense of propagation time across the array) is small, or when the interference powers only slightly exceed the noise power.

A standard metric for performance evaluation in adaptive beamforming is the signal-to-interference-plus-noise ratio (SINR):

$$\mathrm{SINR}(\mathbf{w}) = \frac{|\mathbf{d}(\theta_{\text{sig}})^H\mathbf{w}|^2}{\mathbf{w}^H\mathbf{R}\mathbf{w}}, \tag{4}$$

with $\mathbf{d}(\theta_{\text{sig}})$ being the array response at the angle of an assumed signal, and $\mathbf{w}$ a set of weights aimed at mitigating the interference signals while maintaining gain in the desired direction θ_{sig}. The method of dominant eigenmode rejection, which has been shown to yield near-optimal performance [3] and is frequently used in practice, is taken to form the basis for comparison in this work; in this case the weight vector $\mathbf{w}$ of (4) is given by

$$\mathbf{w} = (\mathbf{I} - \mathbf{U}_\rho\mathbf{U}_\rho^H)\mathbf{d}(\theta_{\text{sig}}), \tag{5}$$

with $\mathbf{U}_\rho$ comprising the ρ principal eigenvectors of the interference-plus-noise covariance matrix $\mathbf{R}$. Note that this method incurs a significant computational cost through the requisite eigendecomposition, particularly when repeated recalculations are required—as may be necessitated by the adaptive scenarios encountered in practice. Indeed, its complexity has inspired study into efficient means of reduced-dimension adaptive processing; here it motivates the use of randomized algorithms to approximate $\mathbf{U}_\rho$ at a reduced computational cost.

While it is possible to verify performance bounds along the lines of (1) for the randomized algorithms under consideration here, such metrics are not generally used directly in the context of adaptive beamforming. As application performance is of primary interest in this study, SINR as defined in (4) is employed to measure the quality of the resultant approximate interference subspaces as follows. Let $\mathbf{w}$ be the weight vector as defined in (5), obtained using the exact ρ dimensional subspace (computed via orthogonal iteration), and $\tilde{\mathbf{w}} = (\mathbf{I} - \mathbf{H}\mathbf{H}^H)\mathbf{d}(\theta_{\text{sig}})$ be the weight vector computed using the output $\mathbf{H}$ of either randomized approximation algorithm. To characterize the overall performance loss due to use of the approximated rather than exact subspace, define the ratio

$$\text{SINR}_{\text{rel}}(\tilde{\mathbf{w}}) = \frac{\text{SINR}(\tilde{\mathbf{w}})}{\text{SINR}(\mathbf{w})} = \frac{|\mathbf{d}(\theta_{\text{sig}})^H \tilde{\mathbf{w}}|^2}{\tilde{\mathbf{w}}^H \mathbf{R} \tilde{\mathbf{w}}} \frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{|\mathbf{d}(\theta_{\text{sig}})^H \mathbf{w}|^2}$$

to be the SINR loss relative to dominant eigenmode rejection.

4. ANALYSIS IN THE NARROWBAND CASE

As an illustrative example of the narrowband beamforming case, this section considers a uniform linear array of $N = 100$ elements. Let $L = 2$ uncorrelated interference signals have directions of arrival (DOA) given by $\theta = \{-25°, 10°\}$ with respect to array broadside, and interference-to-noise ratios (INR) of $\{40, 50\}$ dB, respectively. For the sparse algorithm, $s = 2$ in (2) was chosen such that each element of $\mathbf{R}$ was rescaled *or* zeroed independently with probability $1/2$, and orthogonal iteration was employed to obtain a rank-2 approximant thereof. For the submatrix algorithm, the submatrix dimensions (w, c) were chosen empirically to be twice the desired dimension $\rho = L = 2$, and the threshold in Step 3 was set such that $\epsilon = 10^{-4}$; while these values were seen to provide reasonable performance for the example considered, no attempt was made to optimize this choice of parameters. Note that the output of the submatrix algorithm is only approximately orthonormal; therefore, as well as using $\mathbf{H}$ in the weight calculation

$$\tilde{\mathbf{w}}_{\text{sub}} = (\mathbf{I} - \mathbf{H}\mathbf{H}^H)\mathbf{d}(\theta_{\text{sig}}), \tag{6}$$

an orthonormalized version of $\mathbf{H}$ for weight calculation is also considered:

$$\tilde{\mathbf{w}}_{\text{orth}} = (\mathbf{I} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)\mathbf{d}(\theta_{\text{sig}}). \tag{7}$$

Figure 2 shows the average value of SINR_{rel} as a function of θ_{sig}, taken over 500 different realizations of each of the three randomized algorithms (sparse, submatrix, and orthonormalized submatrix, to be considered in turn below) for a single matrix $\mathbf{R}$ corresponding to the scenario described above. From this figure it may be seen that the sparse algorithm exhibits a 12 dB loss in performance across all angles, indicating a complete breakdown in nulling capability. This phenomenon may be explained by observing (as in [6]) that the sparse matrix in (2) may be expressed as $\mathbf{S} = \mathbf{A} + \mathbf{E}$, with $s > 1$ fixed as before and

$$\mathbf{E}_{ij} = \begin{cases} (s-1)\mathbf{A}_{ij} & \text{with probability } s^{-1} \\ -\mathbf{A}_{ij} & \text{otherwise} \end{cases}.$$

Sampling $\mathbf{A}$ in this manner induces a matrix $\mathbf{E}$ having decorrelated entries, the means and variances of which are given respectively by

$$\begin{aligned} E(\mathbf{E}_{ij}) &= 0 \\ \text{Var}(\mathbf{E}_{ij}) &= \mathbf{A}_{ij}^2(s-1), \end{aligned}$$

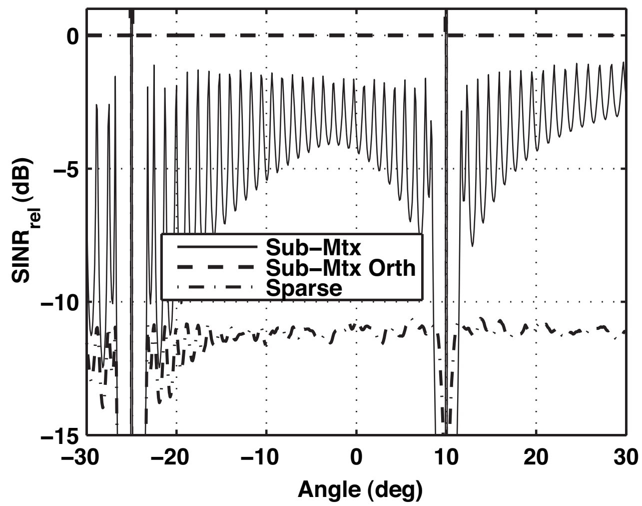

Fig. 2. Signal-to-interference-plus-noise ratio (SINR) loss in dB as a function of angle θ_{sig}, shown for each of the three randomized algorithms relative to subspace nulling by exact dominant eigenmode rejection (averaged over 500 realizations and shown for the narrowband beamforming case with 100 sensors as in Figure 1)

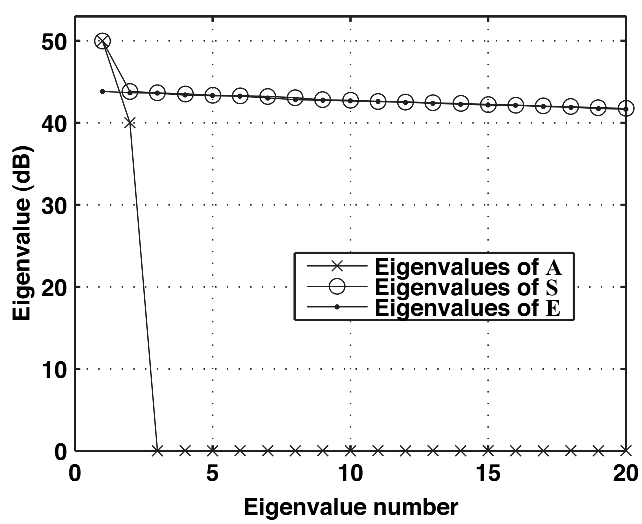

Fig. 3. The first 20 eigenvalues of the matrices associated with the sparse algorithm; $\mathbf{A}$ is the input matrix, $\mathbf{S}$ is the sparse matrix, and $\mathbf{E}$ is the error matrix (note that the second eigenvalue of $\mathbf{A}$ occurs below the approximation noise floor)

in which case it has been observed that $\mathbf{E}$ has a noise-like eigendistribution whose variance is on the order of the magnitude of the entries of $\mathbf{A}$. As a result, the matrix $\mathbf{S}$ has a much larger "noise" floor than $\mathbf{A}$, implying a corresponding degradation of the subspace approximation. To illustrate this phenomenon, Figure 3 shows a typical eigenspectrum for the narrowband example considered above. For this example, it may be seen that the second eigenvalue of $\mathbf{A}$ falls below the approximation noise, and thus the associated eigen-

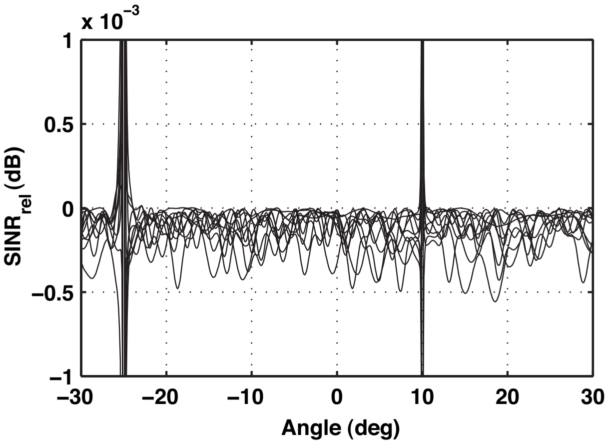

Fig. 4. A magnified view of the relative SINR shown in Figure 2 for 10 typical realizations of the orthogonalized submatrix algorithm; note the ordinate scale (10^{-3} dB)

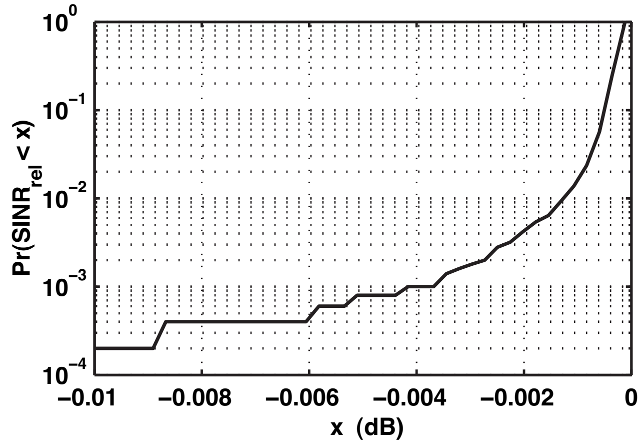

Fig. 5. Empirical cumulative distribution function of $\mathrm{SINR}_{\mathrm{rel}}$ at -10° for the narrowband beamforming example of Section 4, taken over 5000 independent realizations of the orthonormalized submatrix algorithm

vector cannot be approximated with any confidence.

It may be seen from Figure 2 that the submatrix algorithm, by contrast, exhibits a loss that is reminiscent of under-nulled interference (only slightly better than non-adaptive weights). This performance may be explained by observing that the weight vector $\tilde{\mathbf{w}}_{\mathrm{sub}}$ of (6) is computed according to a matrix $(\mathbf{I} - \mathbf{H}\mathbf{H}^H)$ having $N - k = 98$ "passband" eigenvalues equal to unity and $k = 2$ "stopband" eigenvalues equal to the eigenvalues of $\mathbf{I} - \mathbf{H}^H\mathbf{H}$. These two smaller eigenvalues impose a limit on the nulling ability; in the example considered they were observed to range from 0.05 (mediocre nulling) to 1 (no nulling).

However, when $\mathbf{H}$ is orthonormalized, the weight vector $\tilde{\mathbf{w}}_{\mathrm{orth}}$ of (7) is a function of a "true" projection matrix having eigenvalues in the set $\{0, 1\}$, and the nulling performance is consequently a direct function of the distance between the approximated subspace $\mathbf{H}$ and the exact subspace $\mathbf{U}_\rho$. Indeed, from Figure 2, it may be seen that the orthogonalized submatrix algorithm shows only a negligible loss in performance as compared with the exact solution. In this regard, Figure 4 shows a magnified view of the relative SINR for 10 typical realizations of the orthonormalized submatrix algorithm, and Figure 5 shows the empirical distribution of the resultant loss in SINR at -10° azimuth for 5000 independent realizations of the algorithm (i.e., the frequency with which $\mathrm{SINR}_{\mathrm{rel}}$ took less than a specified value). It may be seen from Figure 5 that on average, the orthonormalized submatrix algorithm exhibits a loss of only 10^{-4} dB relative to the exact subspace calculation. This result indicates that for narrowband adaptive beamforming, the submatrix algorithm produces a highly accurate (albeit non-orthonormal) approximation to the desired subspace. In combination with the additional step of orthonormalization, this algorithm appears to hold great promise for nulling applications.

The computational complexity of the submatrix algorithm is dominated by Step 3 (calculating the singular value decomposition of $\mathbf{W} \in \mathbb{C}^{w \times c}$ assuming that $w = c$; complexity $6c^3$) and Step 4 (calculating $\mathbf{H} = \mathbf{C}\mathbf{Z}_\ell \mathbf{\Sigma}_\ell^{-1}$; complexity $2Nc\ell$ where ℓ is the number of singular vector that pass the threshold of Step 3). The additional orthonormalization of $\mathbf{H}$ has complexity $N\ell^2$. By way of comparison, the complexity of a single orthogonal iteration is $2N^2k$ (for the matrix multiplication step) plus Nk^2 (for orthogonalization); for the present example, two iterations were required to calculate the two largest eigenvectors of $\mathbf{R}$. Taking into account the weight vector formation ($4N\ell$ and $4Nk$, respectively), the approximate complexity of each algorithm is given by

$$\begin{aligned} \mathcal{C}(\text{Submatrix}) &= 6c^3 + 2Nc\ell + N\ell^2 + 4N\ell \\ \mathcal{C}(\text{Orth. It.}) &= 2(2N^2k + Nk^2) + 4Nk. \end{aligned}$$

Notice that the complexity of the submatrix algorithm is linear in N whereas that of the orthogonal iteration method is quadratic in N. Indeed, for the parameters of this simulation ($N = 100$, $\ell = k = 2$, and $c = 2k = 4$), the number of operations required to generate the adaptive weight vector is reduced by a factor of approximately 25 relative to the exact solution.

5. AN IMPROVEMENT FOR THE WIDEBAND CASE

While narrowband adaptive beamforming provides a canonical example, wideband systems are of even greater interest in practice. In this case the interference covariance matrix $\mathbf{R}$ of (3) now has $\rho > L$ eigenvalues above the noise floor, and within the interference subspace the eigenvalue distribution may have a large spread. Here it is assumed that ρ is known a priori, and on the basis of the previous section's analysis only the performance of the orthonormalized submatrix algorithm is considered.

The wideband example considered here consists of an $N = 100$ element uniform linear array with $L = 2$ wideband interference signals (1% bandwidth used). The signal DOAs are given by $\{-25^\circ, 10^\circ\}$, with corresponding INRs of $\{25, 40\}$ dB, respectively. In this example, there are four eigenvalues above the noise: $\{40,\ 25,\ 15,\ 8\}$ dB; this eigenvalue spread presents a more challenging case. Recall Step 3 in the submatrix algorithm: The first k singular pairs of the submatrix $\mathbf{W}$ are calculated, but only ℓ pairs are used in the construction of $\mathbf{H}$, where

$$\ell = \min\left\{k, \max\left\{t : \sigma_t^2 \geq \frac{c\epsilon}{8k}\|\mathbf{W}\|_F^2\right\}\right\}.$$

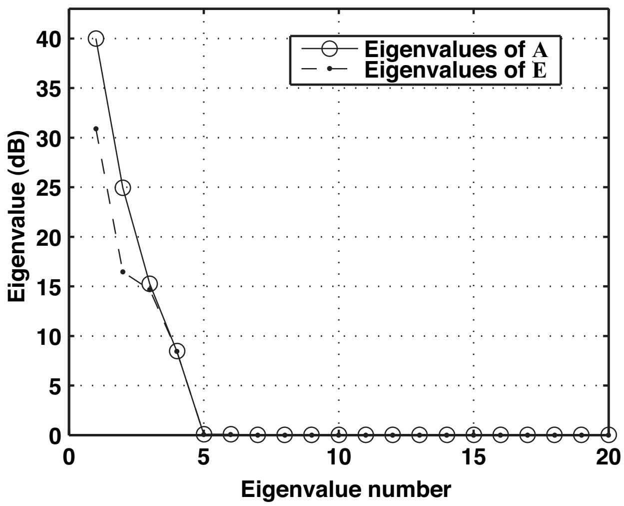

Fig. 6. First 20 eigenvalues of the wideband interference covariance matrix ($\mathbf{A}$) of Section 5 and a typical error matrix ($\mathbf{E}$); note the reduced spread of the eigenvalues of $\mathbf{E}$

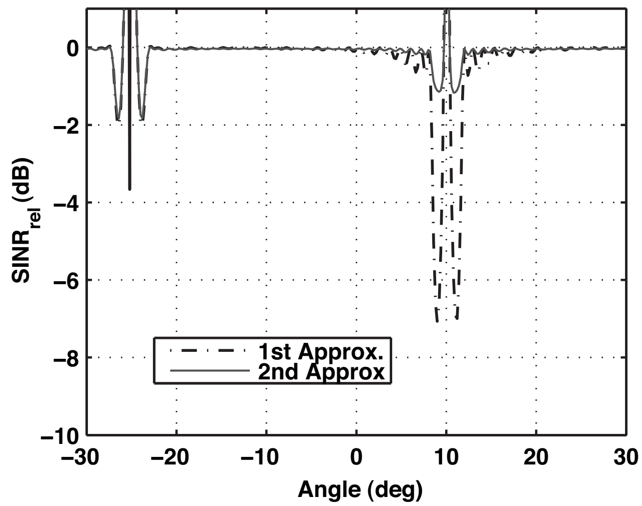

Fig. 7. SINR for a wideband array (1% bandwidth) of 100 sensors using a projection nulling technique, in which the subspace is approximated by the original and two-step submatrix algorithms and each basis is orthonormalized prior to weight vector calculation

This potential thresholding of the eigenvalues acts to control the approximation error. In practice, it has been observed that maintaining this threshold at a reasonable level—i.e., not too low—reduces the measured error, even though the resultant approximation is of a lower dimension than desired.

To approximate the desired subspace with greater accuracy, a two-step procedure is proposed as follows. Let the error matrix be defined as

$$\mathbf{E} = \mathbf{A} - \hat{\mathbf{A}} = \mathbf{A} - \mathbf{H}\mathbf{H}^H\mathbf{A}. \tag{8}$$

By noting that the subspace generated by $\mathbf{H}$ is very close to the dominant ℓ-dimensional subspace of $\mathbf{A}$ (as observed in the previous section), it may be observed that $\mathbf{E}$ differs significantly from $\mathbf{A}$ only in an ℓ-dimensional subspace. Therefore, the subspace of dimension $k - \ell$ omitted from the original basis is left essentially unperturbed in $\mathbf{E}$. However, the eigenvalue spread of the dominant subspace of $\mathbf{E}$ is much less than in $\mathbf{A}$, and hence is more likely to pass the threshold of Step 3. As a concrete illustration of this scenario, Figure 6 shows the eigenspectra for $\mathbf{A}$ and a typical realization of $\mathbf{E}$. The number of eigenvalues above the noise floor is equal in both cases, whereas the eigenvalue spread of $\mathbf{E}$ is reduced relative to that of $\mathbf{A}$.

Bearing in mind this observation, an improved approximation to $\mathbf{A}$ may be generated by executing the subspace algorithm on the error matrix $\mathbf{E}$ resulting from the first low-rank approximant. This second execution generates outputs $\mathbf{H}_e$ (the approximate dominant subspace of $\mathbf{E}$) and $\hat{\mathbf{E}} = \mathbf{H}_e\mathbf{H}_e^H\mathbf{E}$ (the low-rank approximation to $\mathbf{E}$); the outputs of the randomized approximations of $\mathbf{A}$ and $\mathbf{E}$ may be combined in the following way:

$$\begin{aligned} \mathbf{A} \approx \hat{\mathbf{A}} + \hat{\mathbf{E}} &= \mathbf{H}\mathbf{H}^H\mathbf{A} + \mathbf{H}_e\mathbf{H}_e^H\mathbf{E} \\ &= [\mathbf{H}\mathbf{H}^H + \mathbf{H}_e\mathbf{H}_e^H(\mathbf{I} - \mathbf{H}\mathbf{H}^H)]\mathbf{A}. \end{aligned}$$

Thus for the beamforming application considered here, the weight vector becomes

$$\mathbf{w} = \left\{\mathbf{I} - \left[\mathbf{H}\mathbf{H}^H + \mathbf{H}_e\mathbf{H}_e^H(\mathbf{I} - \mathbf{H}\mathbf{H}^H)\right]\right\}\mathbf{d}(\theta_{\text{sig}}). \tag{9}$$

Orthonormal bases for both $\mathbf{H}$ and $\mathbf{H}_e$ may also be used when calculating this weight vector; the complexity of the latter step may be reduced by sampling uniformly from columns of $\mathbf{E}$ (in contrast to the description of Step 1; see [5, 7]), in which case only the selected columns must be explicitly calculated.

For the wideband example described above, let $\mathbf{W} \in \mathbb{C}^{8\times 8}$ in Step 2 and $k = 4$ in Step 3 of the submatrix algorithm. Under this choice of parameters, Figure 7 shows the average value of SINR_{rel} for the orthonormalized submatrix algorithm (labeled 1$^{\text{st}}$ approx.) and the orthonormalized two-step variant (labeled 2$^{\text{nd}}$ approx.) over 500 independent realizations of the algorithms, as in the narrowband case considered in Section 4. In this example the submatrix algorithm outputs a subspace of dimension two, resulting in a 7-dB reduction in nulling ability near the interference and an average loss of about 0.25 dB away from it; as expected, this loss is commensurate with the loss of an exact dimension-two dominant eigenmode rejection filter. On the other hand, the two-step variant may be seen to exhibit a smaller loss of approximately 1 dB near the interference. It has been observed that this loss is due to a tendency of the second stage to occasionally output a subspace of dimension less than k, an occurrence which happens less frequently than in the first stage. Away from the interference, the average performance is within 0.05 dB of the corresponding exact solution. Figure 8 shows the empirical distribution of SINR_{rel} at -5° based on 5000 independent realizations (as in Section 4) of the two-step submatrix algorithm, from which it may be seen that in 99% of trials the relative SINR loss did not exceed 0.23 dB.

The computational complexity of the orthogonalized two-step variant is equal to twice that of the one-step submatrix algorithm, plus the number of operations required to form c columns of the error matrix (see (8); complexity $4Nc\ell$). The formation of the weight vector (9) also becomes significant, with complexity $16N\ell$ for the two-step algorithm vs. $4Nk$ for the one-step algorithm (note that $\ell \leq k$). For this wideband example, the orthogonal iteration method required three iterations to compute the 4 largest

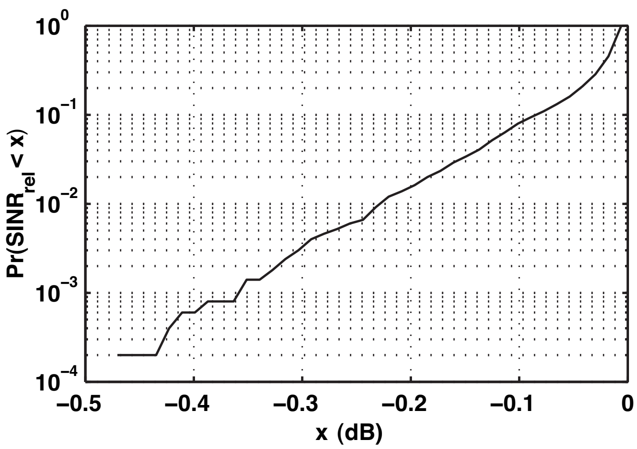

Fig. 8. Empirical cumulative distribution function of $\mathrm{SINR}_{\mathrm{rel}}$ at -5° for the wideband beamforming example of Section 5, taken over 5000 independent realizations of the orthonormalized two-step algorithm

eigenvector of $\mathbf{R}$. In summary, the approximate complexity of each algorithm is given by

$$\begin{aligned}\mathcal{C}(\mathrm{Submatrix}_2) &= 2(6c^3 + 2Nc\ell + N\ell^2) + 4Nc\ell + 16N\ell \\ \mathcal{C}(\mathrm{Orth.\ It.}) &= 3(2N^2k + Nk^2) + 4Nk.\end{aligned}$$

Notice that the complexity of the two-step algorithm remains linear in N. For a conservative estimate in the case at hand, assume that $\ell = k$; for the given simulation parameters ($N = 100$, $k = 4$, $c = 2k = 8$), the two-step submatrix algorithm requires about 6 times fewer operations (in the worst case) than the exact method of weight calculation. If the actual output dimensions ($\ell \leq k$) are used instead, then the average complexity of the submatrix algorithm over 500 realizations is calculated to be less than that of the orthogonal iteration scheme by a factor of approximately 9.

6. CONCLUSIONS AND FUTURE WORK

The above investigation has considered the use of randomized algorithms for low-rank approximations in the signal processing application of array beamforming, in both the narrowband and wideband scenarios. It has been shown that such algorithms may be effectively employed for subspace nulling in the context of adaptive beamforming, an application in which error tolerances are very tight. Quantitative simulation results have been provided, indicating that near-optimal nulling performance, as measured in terms of SINR, may be achieved via straightforward modifications of the randomized algorithms described above. These results indicate that a large computational savings is possible, relative to standard methods, with little corresponding loss in performance.

Whereas the ineffectiveness of the sparse algorithm was demonstrated and explained, a modified variant of the submatrix algorithm was shown to support nulling of large interference signals with negligible loss; a two-step procedure based on this variant was also shown to improve performance in cases where singular values of the signal subspace exhibit a large spread. Such approximations may be calculated relatively quickly, and thus hold promise in situations where computational resources are limited.

Specifically, it was shown that both procedures based on the submatrix algorithm have computational complexity linear in the size of the input matrix, in contrast to the quadratic complexity of orthogonal iteration. Based on this observation, it is hypothesized that large-scale wideband beamforming applications (e.g., space-time adaptive processing) may stand to benefit greatly from these randomized algorithms. It is intended to address this as well as adaptive subspace tracking in future work, with the eventual goal of deriving approximations tailored to application-specific performance metrics in various areas of signal processing.

7. REFERENCES

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proceedings of the RADC Spectrum Estimation Workshop*, Rome, NY, Oct. 1979, reprinted in *IEEE Transactions on Antennas and Propagation* vol. 34, no. 3, pp. 276–280, Mar 1986.

[2] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[3] A. M. Haimovich and Y. Bar-Ness, "An eigenanalysis interference canceller," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 76–84, Jan. 1991.

[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[5] A. Frieze, R. Kannan, and S. Vempala, "Fast Monte-Carlo algorithms for finding low-rank approximations," in *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS '98)*, Palo Alto, CA, 1998.

[6] D. Achlioptas and F. McSherry, "Fast computation of low rank matrix approximations," in *Proceedings of the 33rd ACM Symposium on Theory of Computing (STOC '01)*, Hersonissos, Greece, 2001, pp. 611–618.

[7] P. Drineas, R. Kannan, and M. W. Mahoney, "Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix," Tech. Rep. YALEU/DSC/TR-1270, Yale University, Feb. 2004.

[8] B. Cantrell, J. De Graaf, L. Leibowitz, F. Willwerth, G. Meurer, C. Parris, and R. Stapleton, "Development of a digital array radar (DAR)," in *Proceedings of the 2001 IEEE Radar Conference*, 2001, pp. 421–425.

[9] P. A. Zulch, J. S. Goldstein, J. R. Guerci, and I. S. Reed, "Comparison of reduced-rank signal processing techniques," in *Proceedings of the 32nd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 1998, vol. 1, pp. 421–425.

[10] Christopher K. I. Williams and Matthias Seeger, "Using the Nyström method to speed up kernel machines," in *Advances in Neural Information Processing Systems 13*, T. Leen, T. Dietterich, and V. Tresp, Eds., pp. 682–688. MIT Press, Cambridge, MA, 2001.

[11] M. Zatman, "How narrow is narrowband?," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 145, no. 2, pp. 85–91, Apr. 1998.