DATA DIMENSION REDUCTION USING KRYLOV SUBSPACES: MAKING ADAPTIVE BEAMFORMERS ROBUST TO MODEL ORDER-DETERMINATION *

Hongya Ge,[1] *Ivars P. Kirsteins,*[2] *Louis L. Scharf*[3]

[1] Dept. of ECE, New Jersey Institute of Technology, Newark, NJ 07102, USA.
[2] Naval Undersea Warfare Center, Newport, RI 02841, USA.
[3] Dept. of ECE, Colorado State University, Fort Collins, CO 80523, USA.

ABSTRACT

In this work, we present a class of low-complexity reduced-dimension *adaptive* beamformers constructed from expanding Krylov subspaces. We demonstrate how the data dimensionality reduction obtained from Krylov pre-processing decreases the sensitivity of reduced-rank adaptive beamforming techniques to incorrect model-order selection and lessens the computational complexity of systems involving large arrays with many elements. An important advantage of the proposed dimensionality reduction scheme is that it relieves reduced-rank methods from the stringent requirement on the precise model order determination.

1. INTRODUCTION

Given the correct model order, reduced-rank methods (e.g. subspace techniques) provide high performance solutions to many common signal processing problems involving interference cancellation and direction finding. Consequently, they find a diverse range of applications in adaptive beamforming, interference cancellation, parameter and spectral estimation, and signal detection. However, these methods require knowledge of the interference and signal model orders. Their performances are often sensitive to errors in model order determination [1]. To alleviate this problem, algorithm implementers commonly rely on the traditional methods for model order determination, such as the AIC [2] and the MDL [3], to estimate the model order. However, for applications such as passive sonar, the model order can be *highly variable* and signals of interest are often times *weak* compared to the interference and noise components. Such application challenges motivate us to develop a new class of *dimensionality reduction* processing techniques that are *robust* to the model order over-determination.

Motivated by previous work on conjugate gradient methods [5, 6, 7], we develop, from the perspective of expanding Krylov subspaces, a low-complexity dimensionality reduction method for adaptive beamforming. We demonstrate the capability of our proposed Krylov subspace technique in making performance robust to model order determination. Therefore, our solutions are especially applicable to space-time adaptive signal processing and adaptive beamforming in underwater environments where the model order or the subspace rank is highly variable and time-varying due to the changing dynamics.

*THIS WORK HAS BEEN SUPPORTED IN PART BY THE OFFICE OF NAVAL RESEARCH AND THE NSF(CCF-0515061, CCF-0515192).

2. NOTATION AND PROBLEM FORMULATION

We assume that data snapshots from an array of N sensors with arbitrary geometry follows the linear model,

$$\mathbf{x}[m] = \mathbf{s}_0 a_0[m] + \sum_{k=1}^{K} \mathbf{s}_k a_k[m] + \mathbf{n}[m], m = 1, 2, \ldots, M. \tag{1}$$

where vector $\mathbf{s}_0 = \mathbf{s}(\boldsymbol{\theta}_s)$ is the mode of signal of interest (SOI); vectors $\mathbf{s}_k = \mathbf{s}(\boldsymbol{\theta}_k), k = 1, \ldots, K$, are the modes of K interfering sources, and $\boldsymbol{\theta}_k$ is the vector of spatial parameters associated with the kth mode. We assume that the fading coefficients $a_k[m], k = 0, 1, \ldots, K$, are statistically independent complex Gaussian random variables with zero means and variances σ_s^2 and σ_k^2, respectively. We also assume that the $a_k[m]$ are statistically independent of the white noise vector $\mathbf{n}[m] \overset{iid}{\sim} \mathcal{C}N(\mathbf{0}, \sigma_n^2 \mathbf{I}_N)$. For the data model in (1), the data covariance matrix has the modal form $\mathbf{R_{xx}} = \sigma_s^2 \mathbf{s}_0\, \mathbf{s}_0^H + \mathbf{R_I}$, where $\mathbf{R_I} = \sum_{k=1}^{K} \sigma_k^2\, \mathbf{s}_k\, \mathbf{s}_k^H + \sigma_n^2 \mathbf{I}_N$ denotes the covariance matrix of interference plus noise.

One popular adaptive beamformer or detector used in passive sonar is Capon's method [8] (also termed as the minimum variance distortionless response (MVDR) method). The Capon test statistic is,

$$\mathbf{z}_{cap} = \frac{1}{\mathbf{s}_0^H\, \hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\, \mathbf{s}_0}, \tag{2}$$

where $\hat{\mathbf{R}}_{\mathbf{xx}} = \frac{1}{M} \mathbf{X}\, \mathbf{X}^H$ is the sample covariance matrix calculated from M independent data snapshots collected as the columns of $N \times M$ matrix $\mathbf{X}$. In principle, Capon's method is straightforward to implement. However, arrays with many elements are often utilized for detection, hence the effective data dimensionality N can be quite large. Two important problems as a result of the large data dimensionality are the performance losses due to the small sample size (data snapshots M) relative to the data dimension N in Capon's method and the computational burden associated with inverting a large-order sample covariance matrix. In this work, we proposed a computationally efficient *data-driven* Krylov subspace dimensionality reduction scheme for implementing adaptive beamformers (ABFs) such as the Capon's beamformer (CBF) and the vector conjugate gradient (V-CG) beamformer. The proposed pre-processor not only reduces the computational complexity but also *effectively* reduce the performance sensitivity to model order over-determination faced by many ABFs.

3. DATA DIMENSION REDUCTION

We propose a two stage data-driven procedure for implementing the adaptive beamformers. In applications such as passive sonar, it is often possible to obtain an *approximately* representative secondary data set of noise and signal measurements independent of the primary data for detection, say, just prior to the detection interval, that can be used to determine the spatial basis vectors or subspace $< \mathbf{T} >$ that are relevant to detecting the signal. The dimensionality of $< \mathbf{T} >$ can be much less than the array dimensionality N. Therefore we can reduce the dimensionality of the data and the associated sample covariance matrix by the projections $\mathbf{y}[m] = \mathbf{T}^H\mathbf{x}[m]$ and $\mathbf{T}^H\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}$ respectively.

Our rationale is that in many situations, the fading coefficients may change quickly from snapshot to snapshot as a result of non-stationarity of the temporal second-order statistics. However, the underlaying spatial subspace $< \mathbf{T} >$, parameterized by the spatial parameters of the SOI and the interfering sources, may not change so rapidly over the temporal interval for processing/integration. Therefore, in data-driven adaptive space-time processing, we can use the $\mathbf{T}$ matrix calculated from the previous processing interval as the dimensionality reduction pre-processor for the subsequent data sets. The estimated subspace $< \mathbf{T} >$ for the dimension reduction stage does not have to be precise since further adaptation will be done in the second stage of processing. Through the simulation examples, we demonstrate the feasibility and effectiveness of such utilization in adaptive beamforming.

We now describe the data-driven dimensionality reduction scheme. Using the subspace expansion idea, we construct a $N \times p$ matrix $\mathbf{T}$ for data dimensionality reduction, such that

$$\mathbf{T} \in \mathcal{K}_p(\hat{\mathbf{R}}_{\mathbf{xx}}, \mathbf{s}_0), \qquad p << N,$$

where vector $\mathbf{s}_0$ is the SOI, and $\mathcal{K}_p(\hat{\mathbf{R}}_{\mathbf{xx}}, \mathbf{s}_0)$ is the Krylov subspace of rank p. The choice of parameter p should be large enough to include the SOI mode and the interfering modes that are *linearly correlated to* the SOI mode, yet small enough to maintain good detection performance.

We would like to construct a $\mathbf{T}$ with *orthonormal* columns to avoid dependencies of the adaptive beamformer test statistic on the particular realization of basis of $\mathbf{T}$. Motivated by vector conjugate gradient (V-CG) ideas in combination with simple *sequential* normalization and projections, we present a computationally efficient procedure for generating an orthonormal basis for the columns of $\mathbf{T}$. Specifically, starting from the SOI mode $\mathbf{s}_0$, the *orthonormal* columns of $\mathbf{T}$ can be constructed directly using the sequential projections and normalization, similar in spirit to the Gram-Schmidt procedure, on the nature Krylov basis vectors $\boldsymbol{\kappa}_i$'s, with $\boldsymbol{\kappa}_i = \hat{\mathbf{R}}_{\mathbf{xx}}\,\boldsymbol{\kappa}_{i-1}$ and $\boldsymbol{\kappa}_1 = \mathbf{s}_0$. That is, for $i = 2, 3, \ldots, p$, we calculate:

$$\begin{aligned} \boldsymbol{\kappa}_i &= \mathbf{P}^{\perp}_{\mathbf{T}_{i-1}}\,\hat{\mathbf{R}}_{\mathbf{xx}}\,\boldsymbol{\kappa}_{i-1} \\ \mathbf{t}_i &= \frac{\boldsymbol{\kappa}_i}{\|\boldsymbol{\kappa}_i\|} \\ \mathbf{T}_i &= \left[\mathbf{T}_{i-1}\,\mathbf{t}_i\right] \end{aligned} \quad , \tag{3}$$

where the initials are chosen as $\boldsymbol{\kappa}_1 = \mathbf{s}_0$ and $\mathbf{T}_1 = \mathbf{s}_0/\|\mathbf{s}_0\|$. The projection operator contains the sequential projections due to the orthogonality among columns of matrix $\mathbf{T}_{i-1}$, i.e., $\mathbf{P}^{\perp}_{\mathbf{T}_{i-1}} = \mathbf{I}_N - \sum_{k=1}^{i-1}\mathbf{t}_k\,\mathbf{t}_k^H$. Upon the completion of iterations in eq.(3), the dimensionality reduction matrix $\mathbf{T}$ is simply the $\mathbf{T}_p$. Using the above pre-processing matrix, we can map the original high-dimension data set $\{\,\mathbf{x}[m]\,\}_{m=1}^{M}$ into the reduced-dimension data set $\{\,\mathbf{y}[m]\,\}_{m=1}^{M}$,

$$\mathbf{y}[m] = \mathbf{T}^H\mathbf{x}[m], \qquad m = 1, 2, \ldots, M. \tag{4}$$

Ideally, as long as the rank of the Krylov subspace $< \mathbf{T} >$ is above the minimum sufficient rank, i.e. $p \geq (K+1)$, for capturing the SOI mode as well as the interfering modes linearly correlated to the SOI mode, there is no information loss for detection and beamforming.

It is well known that the major advantages brought by the dimensionality reduction pre-processing are two-fold. First, it helps to reduce the computational burden associated with filtering out a few components of special interest from a large dimensional data set. Secondly, in the reduced dimension space, data-driven adaptive filtering and beamforming methods can rapidly converge to the desired solution. In this work, we demonstrate that there is another important advantage in using the above Krylov subspace operator $\mathbf{T}$ for dimensionality reduction. That is, it makes the subsequent adaptive filtering and beamforming methods, working on the reduced-dimension data, *robust to* model order over-determination. We observe that the rank-r Krylov subspace in the original full-dimensional data space, $\mathcal{K}_r(\hat{\mathbf{R}}_{\mathbf{xx}},\,\mathbf{s}_0)$, and the corresponding rank-r Krylov subspace in the *reduced dimension* data space, $\mathcal{K}_r(\hat{\mathbf{R}}_{\mathbf{yy}},\,\mathbf{s}_{\mathbf{y}})$, are related by the dimension reduction matrix $\mathbf{T}$ (with orthonormal columns) and its projection matrix $\mathbf{P}_{\mathbf{T}} = \mathbf{T}\,\mathbf{T}^H$, i.e.,

$$\mathcal{K}_r\left(\mathbf{P}_{\mathbf{T}}\,\hat{\mathbf{R}}_{\mathbf{xx}}\,\mathbf{P}_{\mathbf{T}},\,\mathbf{P}_{\mathbf{T}}\,\mathbf{s}_0,\right) \xrightarrow{\mathbf{T}^H} \mathcal{K}_r(\hat{\mathbf{R}}_{\mathbf{yy}},\,\mathbf{s}_{\mathbf{y}}),$$

where the role of the projection operator $\mathbf{P_T}$ is to *re-shape and truncate* the eigen-values of $\hat{\mathbf{R}}_{\mathbf{xx}}$ to improve the condition of $\hat{\mathbf{R}}_{\mathbf{yy}}$, so that eigen-values of the $\hat{\mathbf{R}}_{\mathbf{yy}}$ equals the p non-zero eigenvalues of $\mathbf{P_T}\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{P_T}$. In doing so, we get around the ill-condition by trimming off the less relevant contributions from the high-order powers of eigen-values of $\hat{\mathbf{R}}_{\mathbf{xx}}$ *outside* the rank-p Krylov subspace $< \mathbf{T} >$. This can be seen from the decorrelating effect of the conjugate direction vectors $\mathbf{d}_i$'s, $\mathbf{D_T} = [\,\mathbf{d}_1\,\mathbf{d}_2\ldots\mathbf{d}_p\,]$ on the matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$. That is, $\mathbf{D}_{\mathbf{T}}^H\,\hat{\mathbf{R}}_{\mathbf{xx}}\,\mathbf{D_T} = \mathrm{diag}\{\delta_0, \delta_1, \ldots, \delta_{p-1}\}$, with $\delta_i = \mathbf{d}_i^H\,\hat{\mathbf{R}}_{\mathbf{xx}}\,\mathbf{d}_i$. Note that $< \mathbf{D_T} > = < \mathbf{T} >$ and the two matrices are related by the QR decomposition or G-S orthogonalization. Applying the projection matrix $\mathbf{P_T} = \mathbf{T}\,\mathbf{T}^H$ on the the sample correlation matrix, we have

$$\mathbf{P_T}\,\hat{\mathbf{R}}_{\mathbf{xx}}\,\mathbf{P_T} = \mathbf{U_T}\,\hat{\mathbf{\Sigma}}_{\mathbf{yy}}\,\mathbf{U}_{\mathbf{T}}^H, \tag{5}$$

where the $N \times p$ matrix $\mathbf{U_T} = \mathbf{T}\,\hat{\mathbf{U}}_{\mathbf{yy}}$ has orthogonal columns (similar to $\mathbf{T}$), the $\hat{\mathbf{U}}_{\mathbf{yy}}$ is a $p \times p$ unitary eigen-matrix of $\hat{\mathbf{R}}_{\mathbf{yy}}$, and the diagonal matrix $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ contains the eigen-values of $\hat{\mathbf{R}}_{\mathbf{yy}}$. The pre-processor $\mathbf{T}$ helps to eliminate the non-relevant components present in the original data $\mathbf{x}[m]$ as well as to alleviate the ill-condition of the problem in terms of the Krylov-spectrum truncation.

4. EXPERIMENTAL RESULTS

To demonstrate the robust performance brought by the proposed dimensionality reduction scheme for implementing various two-stage adaptive beamformers, we use an experimental setup with a line array of a given geometry on the arrangement of $N = 60$ array elements. The estimate $\hat{\mathbf{R}}_{\mathbf{xx}}$ is obtained from a total of $M = 65$ snapshots. The total number of modes in $\mathbf{x}[m]$ is fixed at $K + 1 = 3$. The two interfering modes, *approximately* 2.58 DFT-bins away from each other, are *not orthogonal* to the SOI mode. To study the detection performance (P_d versus SNR in dB) of various adaptive beamformers, we fix the false alarm rate at $P_{fa} = 0.02$. The signal-to-noise ratio for the SOI is defined as, SNR $= \sigma_s^2/\sigma_n^2$, and the interference-to-noise ratio is fixed as, I$NR = \sum_{k=1}^{K} \sigma_k^2/\sigma_n^2 = 18dB$.

4.1. Full-Dimension V-CG Adaptive Beamformer

Using the results in [9], it can be shown that the full-dimension reduced-rank V-CG beamformer (using the true $\mathbf{R}_{\mathbf{xx}}$) converges to the optimum beamformer in *at most* rank$_{vcg}$ $= K+1$ steps. However, when we try to adaptively implement the full-dimensional reduced-rank V-CG beamformer (using the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$), the performance does not increase montonically with rank$_{vcg}$. In fig. 1(a) we have plotted the performance of the *full-dimension* adaptive V-CG beamformer as a function of the hypothetic model order rank$_{vcg}$. In this example, the rank (rank$_{vcg}$) of the full-dimension adaptive V-CG beamformer was varied from 2 to 10. From these plots we see that when the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{xx}}$ is used instead of the true one, detection performance improvement starts to back off from the optimum one once the rank$_{vcg}$ goes beyond its optimal value (here the optimal value is 2) and rapidly degrades. This is a result of estimation errors in $\hat{\mathbf{R}}_{\mathbf{xx}}$. It is has been noticed that the bias and variance tradeoff results in a balance point in the choice of optimal rank$_{vcg}$ for the given values of data dimension N and the snapshot number M.

4.2. Reduced-Dimension Adaptive Capon Beamformer

We now apply our new dimensionality reduction scheme to Capon's method. The adaptive reduced-dimension Capon's test statistic used here is

$$\mathbf{z}_{cap}^{(\mathbf{y})} = \frac{1}{\mathbf{s}_{\mathbf{y}}^H\,\hat{\mathbf{R}}_{\mathbf{yy}}^{-1}\,\mathbf{s}_{\mathbf{y}}}, \tag{6}$$

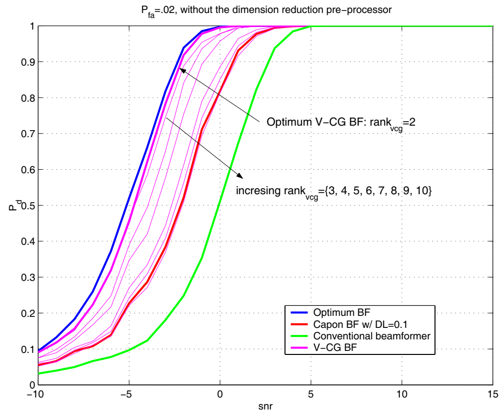
with $\hat{\mathbf{R}}_{\mathbf{yy}} = \mathbf{T}^H\hat{\mathbf{R}}_{\mathbf{xx}}\mathbf{T}$, and $\mathbf{s}_{\mathbf{y}} = \mathbf{T}^H\,\mathbf{s}_0$. In fig. 1(b) we have plotted the detection performance of the adaptive reduced dimension Capon's method as a function of dimension order p where recall, p is the column dimension of the matrix $\mathbf{T}$. One can see that the detection performance is *quite robust* to model order over-determination when we vary the dimensionality reduction parameter $p = \{\,3, \ldots, 10\}$. As expected, when the value for the dimensionality reduction parameter p is very large relative to the true model order $K + 1 = 3$, say $p = 20$, the performance of the adaptive Capon beamformer degrades significantly from the optimum one.

4.3. Reduced-Dimension V-CG (RD-VCG) Adaptive Beamformer

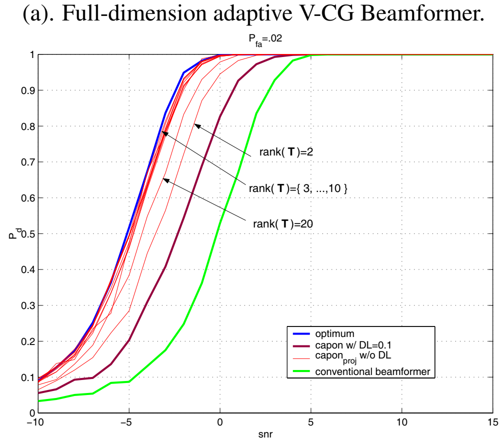
The reduced-dimension adaptive Capon beamformer still requires a $p \times p$ matrix inversion $\hat{\mathbf{R}}_{\mathbf{yy}}^{-1}$. To further reduce the computational complexity we apply the V-CG algorithm to evaluate the Capon method test statistic. Using the Krylov subspaces in the reduced-dimension data space, the RD-VCG beamformer avoids the need to invert the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{yy}}^{-1}$ yet delivers better performance than the reduced-dimension adaptive Capon's beamformer. In Fig. 2, we have plotted on the detection performance of the RD-VCG adaptive beamformer for the choice of dimensionality reduction parameter $p = \{\,4,\,10\,\}$. From Fig. 2, one can see that for a given value of dimension reduction parameter p, the performance (as well as the computational complexity) of the RD-VCG adaptive beamformer is lower-bounded (as well as upper bounded) by the reduced dimension adaptive Capon's beamformer.

5. CONCLUSIONS

We have developed a computationally efficient method for data dimensionality reduction using the expanding Krylov



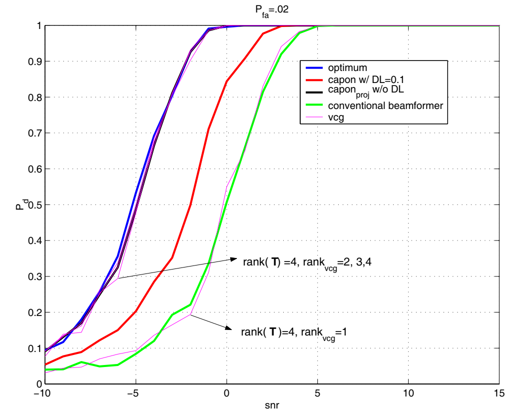
(a). Full-dimension adaptive V-CG Beamformer.



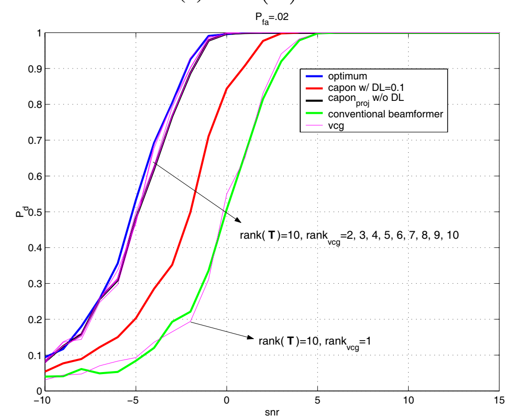
(b). Reduced-dimension adaptive Capon beamformer.

Fig. 1. Performance sensitivity of the *full-dimension* adaptive V-CG beamformer and the reduced-dimension adaptive Capon bemaformer to the over-determination on model order, $rank_{vcg}$. The order of the V-CG beamformer is varied from 2 to 10. The dimensionality reduction matrix $\mathbf{T}$ was calculated using $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\mathbf{s}_0$.

subspaces. Applied to the reduced-dimension data sets, the resultant adaptive Capon beamformer (with no need for a diagonal loading) and the vector conjugate gradient (V-CG) beamformer of various ranks show remarkable robustness to the over-determination on the model order.



(a). $\text{rank}(\mathbf{T}) = 4$



(b). $\text{rank}(\mathbf{T}) = 10$

Fig. 2. Robust performance of the *reduced-dimension* adaptive V-CG beamformer to the model order over-determination.

6. REFERENCES

[1] R. E. Zarnich, "A Fresh Look at 'Broadband' Passive Sonar Processing," *Proc. of the Adaptive Sensor Array Processing Workshop (ASAP)*, Lexington, Massachusetts, March 1999.

[2] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. on Autom. Control*, vol. AC-19, pp. 716–723, 1974.

[3] J. Rissanen, "Modelling by Shortest Data Description," *Aotomatica,* vol. 14, pp. 465–471, 1978.

[4] M. R. Hestenes, and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems," *J. Res. Nat. Bur. Stand.*, vol. 49, pp. 409–436, 1952.

[5] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A Multistage Representation of the Wiener Filter Based on Orthogonal Projections," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2943–2959, Nov. 1998.

[6] M. E. Weippert, J. D. Hiemstra, J. S. Goldstein, and M. D. Zoltowski, "Insights from the Relationship between the Multistage Wiener Filter and the Method of Conjugate Gradients," *Proceedings of the 2nd IEEE Workshop on Sensor Array and Multichannel Signal Processing*, Washington, DC, Aug. 2002.

[7] L. L. Scharf, L. T. McWhorter, E. K. P. Chong, J. S. Goldstein, and M. D. Zoltowski, "Algebraic Equivalence of Conjugate Direction and Multistage Wiener Filters," *Proceedings of the Eleventh Annual Workshop on Adaptive Sensor Array Processing (ASAP)*, Lexington, Massachusetts, March 11–13, 2003.

[8] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[9] H. Ge, M. Lundberg, and L. L. Scharf, "Warp Convergence in Conjugate Gradient Wiener Filters," *Proc. of the third IEEE Workshop on Sensor Array and Multichannel Signal Processing,* Barcelona, Spain, July, 2004.