CASE STUDY OF PRICIPAL COMPONENT INVERSE AND CROSS SPECTRAL METRIC FOR LOW RANK INTERFERENCE ADAPTATION

B. E. Freburger and D. W. Tufts

Department of Computer and Electrical Engineering
University of Rhode Island
Kingston, RI 02881

ABSTRACT

This paper presents a review of the Principal Component Inverse method of rapidly adaptive signal detection and contrasts the use of Pricipal Components with the more recent Cross Spectral Metric method for the Generalized Sidelobe Canceller. The CSM method is optimal with known statistics and has been shown to outperfrom the PCI method in many cases of unknown covariance. This paper describes a scenario which represents a class of covariances where the PCI method can be expected to outperform the CSM method. The choice of method is therefore more subtle than previously thought.

1. INTRODUCTION

Principal Component Inverse (PCI) [10, 15, 14, 6, 7] is a method for estimating and removing an undesired interference which, over a short period of time, can be well approximated by a set of a few locally constant basis vectors. The method is an improvement on Sample Matrix Inverse (SMI) [9] in the case of low rank interference and when implemented in the context of a Generalized Sidelobe Canceller has similarities to the recently introduced Cross Spectral Metric (CSM) [3, 4] approach; However, there are performance differences which warrant investigation.

1.1. Known Statistics

When the signal is known except for a random phase embedded in noise which has a multivariate Gaussian probability distribution with zero mean vector and covariance matrix $\mathbf{R}$, then the optimum test statistic be viewed as the magnitude of the output of a noise whitening matched filter [1] By decomposing the covariance as an interference component plus a white noise component, $\mathbf{R} = \mathbf{Q} + \sigma^2\mathbf{I}$, Claus et al. [2] showed that the optimum test statistic can be re-written using a weighted projection matrix $\mathbf{P}$

$$|S^H\mathbf{R}^{-1}X| = \frac{1}{\sigma^2}|S^H(\mathbf{I}-\mathbf{P})X| = \frac{1}{\sigma^2}|S^H(X-\mathbf{P}X)| \quad (1)$$

where $\mathbf{P} = \sum_{k=1}^{N} \frac{\lambda_k}{\lambda_k+\sigma^2} Q_k Q_k^H$ and the λ_k's and Q_k's are the eigenvalues and eigenvectors associated with the matrix $\mathbf{Q}$.

By rewriting formula 1 we have an alternative formula for the test statistic [8]

$$|S^H\mathbf{R}^{-1}X| \propto |S^H X - W_{GSLC}^H(\mathbf{B}^H X)| \quad (2)$$

where $\mathbf{B}$ is a $N\times(N-1)$ matrix orthogonal to the vector S. Denoting d as the signal based coordinate value and Z as the vector

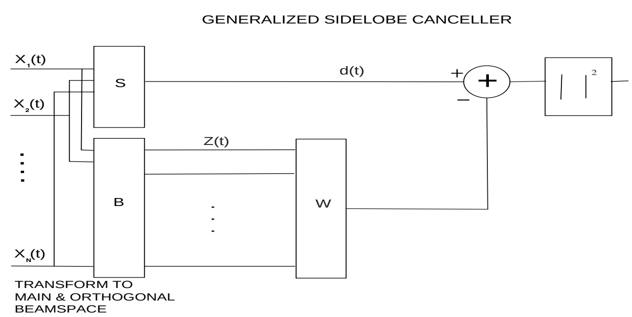

Figure 1: Generalized Sidelobe Canceller Structure

of coordinate values in the orthogonal space, the optimum weight vector is $W_{GSLC} = \mathbf{R}_Z^{-1} r_{dZ}$ [8]. This structure is commonly reffered to as a Generalized Sidelobe Canceller (GSLC) and is shown in Figure 1.

1.1.1. Estimation of Interference for a Given Rank

Suppose in the case of known covariance one wishes to design an optimal processor for which the weight vector is constrained to be in an M dimensional subspace based on the eigenvectors of the covariance matrix $\mathbf{R}_Z$. It has been shown by Goldstein and Reed [3, 4] that the optimal solution is to form the GSLC weight vector

$$\hat{W}_{GSLC} = \hat{\mathbf{U}}_M \hat{\mathbf{\Lambda}}_M^{-1} \hat{\mathbf{U}}_M^H \hat{r}_{dZ} \quad (3)$$

where $\hat{\mathbf{U}}_M$ and $\hat{\mathbf{\Lambda}}_M$ are constructed from the M eigenvectors and eigenvalues chosen from the M largest values of the set

$$\frac{|u_j^H r_{dZ}|^2}{\lambda_j} \text{ for } j = 1\cdots N-1 \quad (4)$$

which is called the Cross Spectral Metric (CSM).

1.2. Adaptive Detectors

Several different adaptive detectors can be formulated by estimating the parameters in the above versions of the optimal hypothesis test statistic. The noise covariance matrix $\mathbf{R}$ is now assumed to be unknown and multiple statistically independent observations of the data vector are available.

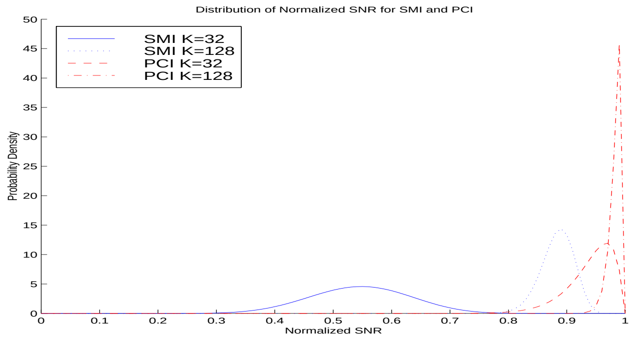

Figure 2: Distribution of Normalized SNR for SMI and PCI

1.2.1. Principal Component Inverse (PCI)

The name Principal Component Inverse (PCI) was chosen to emphasize the connection with the earlier Sample Matrix Inverse (SMI) method. PCI is a modification of SMI which can provide a more sufficient degree of adaptation with less observed data. This is achieved by using the commonly occurring a priori knowledge that during the observation interval, the noise vectors $V(k)$, over some adaptation interval of values of k, are composed of a strong component, $C(k)$ and a background component where the strong component is well represented by a set of basis vectors which do not change over the adaptation interval, $C(k) = \sum_{i=1}^{M} \alpha_i(k) Q_i$.

Starting with the form of the test statistic of formula 1, the Principal Component Inverse approach is to replace $\mathbf{P}$ in formula 1 by $\hat{\mathbf{P}} = \sum_{k=1}^{M} \hat{Q}_k \hat{Q}_k^H$ in which $\hat{Q}_k$ is the estimated k^{th} eigenvector of the data matrix of vector samples $\mathbf{Y} = \begin{bmatrix} X_1 & X_2 & \cdots & X_K \end{bmatrix}$. M is the estimated rank of the strong interference and the interference is much stronger then the background white noise, thus $\lambda_k \gg \sigma^2$ for $1 \leq k \leq M$ and $\lambda_k \approx 0$ for $k > M$ and so $\frac{\lambda_k}{\lambda_k + \sigma^2} \approx 1$ The PCI test statistic is then given by

$$\frac{1}{\sigma^2} |S^H (X - \hat{\mathbf{P}} X)| \tag{5}$$

If one defines the normalized SNR, ρ, as the ratio of the SNR with the adaptive weight vector to the SNR of the optimal known covariance weight vector, then for the SMI method, ρ was shown to have a Beta distribution dependent upon N, where N is the number of component of the data vector X and K which is the number of statistically independent vectors used in the estimate of $\hat{\mathbf{R}}$. Under the assumption that the interference is strong the PCI weight vector is approximately the SMI estimate for the problem in the reduced rank space [6] and thus the SNR has a Beta distribution where N is replaced by $M+1$. The result is that PCI can provide a higher probability than SMI for obtaining a sufficiently high value of SNR for a given sample size as shown in Figure 2.

When the interferers are strongly low rank, the rank selection for PCI can be straightforward as there will be a noticeably sharp difference between two eigenvalues. However, in many practical cases the singular values may simply fall off rapidly (only approximately low rank) making the choice of rank more difficult. The method used is based on two principles which are used to set a threshold.

- Under conditions of no strong interference, waveforms should pass through PCI processing unchanged
- Threshold should allow max expected signal and low level interference to pass through PCI processing unchanged

The rank is then determined by finding the minimum r for which $\sum_{k=0}^{r} \lambda_{N-k}^2 > T$ where λ_k is the k^{th} singular value of the sample data matrix, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. The rank is then chosen as $M = N - r_{min}$ where N is the dimension of the data vector and T is the chosen threshold setting. In cases of strong interference the threshold, T, can be set using knowledge of maximum expected signal strength and the fact that the distribution of $\frac{2}{\sigma^2} \sum_{k=0}^{r} \lambda_{N-k}^2$ without signal is approximately $\chi^2_{2(N-r)(K-r)}$ [11, 16].

The first principle for setting the threshold ensures that the processor will become a matched filter in the absence of interference. The second principle ensures that the threshold will be set high enough that the signal will not be modeled for the interference removal. This is necessary in many systems, as one can not always guarantee that training data is signal free or one may wish to train and detect on the same data interval.

In the present scenario the columns of the matrix are independent; however, in may cases such as time series signal estimation or space-time processing, the matrix formed may have a predefined structure such as Toeplitz or Hankel. Utilizing these structures in the PCI algorithm can improve the estimation performance. The reader is referred to [10, 15] for a treatment of these cases.

1.3. Adaptive CSM and Signal Based PCI

The methods of CSM and PCI offer two ways of providing an adaptive processor in the signal based coordinates of formula 2. Given a set of data $d = \begin{bmatrix} d_1 & d_2 & \cdots & d_k \end{bmatrix}$ and $Z = \begin{bmatrix} Z_1 & Z_2 & \cdots & Z_k \end{bmatrix} = \mathbf{U \Sigma V}^H$ Both can be shown to form the weight vector by using

$$\hat{\mathbf{R}}_Z = \frac{1}{K} Z Z^H \qquad \hat{r}_{dZ} = \frac{1}{K} Z d^H \tag{6}$$

and forming the weight vector as

$$\hat{W}_{GSLC} = \hat{\mathbf{U}}_M \hat{\mathbf{\Lambda}}_M^{-1} \hat{\mathbf{U}}_M^H \hat{r}_{dZ} \tag{7}$$

However, the Adaptive CSM uses the estimated cross spectral metric to choose the eigenvectors, whereas, Signal Based PCI uses only the estimated eigenvalues [6]. Also, CSM is formulated for a prescribed rank and general covariance, while PCI estimates the rank from data over the adaptation interval and assumes that the covariance is from a low rank process.

CSM is the optimal choice in case of known covariance [4] and the optimal choice with respect to mean squared error when the problem is treated as a least squares data problem (estimation and application of the weight vector is done on the same data) [3]. CSM has also been experimentally shown to provide better results for lower ranks using the Mountaintop data set [5]. These results may suggest that CSM outperforms PCI in all cases, however, the following scenario represents a large class of covariance structures for which Signal Based PCI can outperform CSM. In particular, let us confine our attention to the case where the interference is a strong low rank process and thus we can speak of a correct rank. Finite sample performance of these methods will depend upon the stability of the estimates being used for the rank selection criteria in addition to the true values of the values being estimated.

1.3.1. Subspace Swap

We now describe the concept called a subspace swap [13, 12] which is an important concept in understanding the performance of the PCI approach. In the case where we have the notion of a correct rank, divide the singular vectors of the covariance matrix into signal and orthogonal singular vectors

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_s & | & \mathbf{U}_o \end{bmatrix} \quad (8)$$

$$= \begin{bmatrix} U_1 & \cdots & U_r & | & U_{r+1} & \cdots & U_N \end{bmatrix} \quad (9)$$

where U_k corresponds to the k^{th} largest true singular value. For a given realization of the data matrix, Z, one expects the energy to be greater in the direction of the singular vectors of the signal subspace than any linear combination of the orthogonal subspace. That is,

$$|U_i U_i^H Z||_F > \sup_{|a|=1} ||U_o a a^H U_o^H Z||_F \text{ for } i = 1 \cdots r \quad (10)$$

However, it can happen that a linear combination of the singular vectors in the orthogonal space resolves more energy than a singular vector in the signal space. This is called a subspace swap and is associated with a rapid degradation in the performance of SVD based algorithms.

To facilitate the finite sample comparison of PCI and CSM, we introduce the analogous concept of a CSM swap. Let the singular vectors in equation 9 be such that U_k corresponds to the k^{th} largest true cross spectral metric. A CSM swap will occur when the output of $U_i^H Z$ for $i = 1 \cdots r$ makes a larger angle with d than for a linear combination of $i = r + 1 \cdots N$. That is, due to the finite sample size, data from a channel that should have low or no correlation with d, appears to have a higher correlation than data from a channel that shoulod be highly correlated with d.

In the case of strong low rank interference, the CSM and PCI methods with known statistics will choose the identical set of singular vectors. However, the use of different methods for adaptively choosing the singular vectors may have different probabilities of a swap. And so, although on most realizations of the training data, the chosen singular vectors and thus the performance will be identical, there will exist realizations where the two methods may differ considerably. Let us construct an example where the estimates of the CSM values will be less stable than the estimates of the eigenvalues and thus the chance of a CSM swap will be greater than a swap with the PCI method. One way to do this is to place some of the jammers in the nulls of the nominal beampattern thus giving them small true cross covariance values. One example of this type of scenario was that used by Goldstein and Reed [3] except we now consider the case of independent training and application data whereas, they considered the problem in terms of minimum mean square error where the training and application data were the same.

1.3.2. Independent Training and Test Data

The following case is constructed such that the probability of a PCI swap and rank selection error is negligible, which is to say that the jammer power is large compared to the background noise. Assume the data used to calculate the weight vector (training data) is signal free and independent of the data which will be used for signal detection. With this independence, the mean square error performance is inversely proportional to the SNR. A spatial only

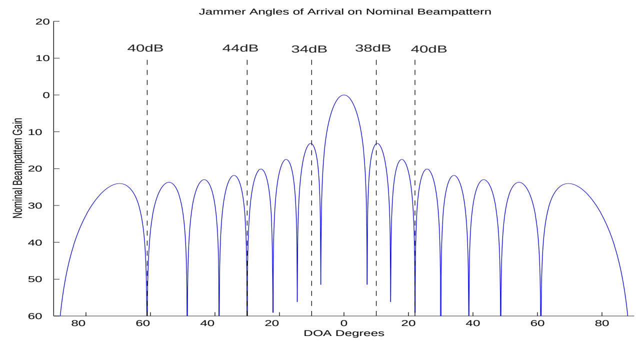

Figure 3: Jammer Angles and Powers for Simulated Scenario

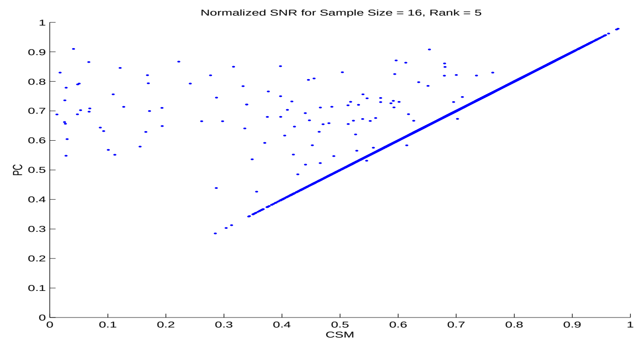

Figure 4: Normalized SNR Realizations with Rank = 5

scenario using a 16 element linear array with five interferers at angles $\begin{bmatrix} -61 & -30 & -10 & 10 & 22 \end{bmatrix}$ degrees and power levels $\begin{bmatrix} 40 & 44 & 34 & 38 & 40 \end{bmatrix}$ dB was simulated as shown in Figure 3. The desired signal was assumed to be broadside and 16 signal free samples were used for the adaptive weight calculation. Note that three of the jamming angles are near the nulls of the nominal beampattern. For each weight vector calculated, a theoretical normalized SNR was computed. This was done for 10000 trials. Figure 4 shows the results for the case when the choice of rank is 5 which corresponds to the true number of interferers. The majority of the time the two methods have identical normalized SNR and thus have chosen the same eigenvectors. However, occasionally the CSM makes a different choice which results in performance difference between the two methods.

There were a total of 109 (about 1%) realizations where CSM chose an eigenvector that was not closest to the theoretical best. These occurences were fairly evenly distributed among the three eigenvectors related to the interference in the nulls of the nominal beampattern. As shown in the Figure, the vast majority of these occurences are associated with poorer performance. There were no cases using the PCI method.

In the three samples where CSM outperformed PCI, the vectors chosen by PCI did span the interference subspace very well. With high probability this will yield good cross covariance estimates for these coordinates and thus good performance. However, with low probability there will be poor cross covariance estimates and simultaneously be a cross covariance estimate among eigenvectors that were not chosen that is good enough to compensate for the difference in the levels of interference between them. Thus

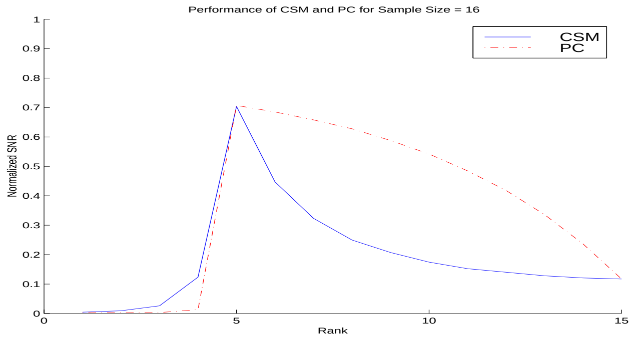

Figure 5: CSM and PCI Performance as a Function of Rank

it would have been better on this realization to choose an eigenvector that was not closest to the true interference subspace. In order for CSM to choose this eigenvector the magnitude of the poorly estimated cross covariance must be estimated low rather than high. Thus, the condition for this event to occur is that the estimated cross covariance for an eignevector with a high level of interference in the training data is estimated extremely low while a cross covariance of a low interference eigenvector is accurately estimated.

In Figure 5 we now look at the performance as a function of rank for the case of a fixed training size of 16 samples. As expected, the performance of the PCI method is poor until a rank 5 processor is used but notice that for ranks 3 and 4, the CSM method is also well below the rank 5 value due to swapping among the five largest CSM values. In the rank three case, there were 3813 times that CSM chose a singular vector that was not closest to the two best CSM singular vectors of the known covariance case. Also, from Figure5, the CSM method falls off much more quickly for ranks above the true rank. The singular vectors corresponding to singular values after the first five are essentially associated with white noise and will be very unstable within the noise subspace. Therefore using them in the processor will not improve interference cancellation. PCI simply selects the vector corresponding to the next largest eigenvalue. This selection will have the least impact on the conditioning of the inverse of the estimated covariance matrix. The CSM values above number five will have a large degree of variability since the true values are zero, and so, CSM is almost as likely to choose any of the remaining eigenvectors. When CSM chooses an eigenvector associated with a small eigenvalue, a poorly conditioned matrix will result with large weight vector errors. Note that as the rank of the processor is increased the difference between PCI and CSM becomes less since the number and magnitude of singular values on which they can possibly differ becomes smaller.

In this scenario we examined independent training and test data in which case one can often assume that the training data is signal free. However, in the same data case one must consider the effect of signal presence on the interference suppression algorithm. The effects of signal presence on SNR using CSM and Signal Based PCI is a topic for future work.

2. CONCLUSIONS

PCI and CSM give two methods of rank selection for the case of the GSLC. CSM has been shown to provide better performance than PCI in several important cases. However, PCI can provide better performance in the case of strong low rank interference due to a lower chance of a PCI swap. This perfromance difference is emphasized for small training data sets where the estimates used by the PCI and CSM methods are poor.

3. REFERENCES

[1] L. Brennan and I. Reed. Theory of adaptive radar. *IEEE Transactions on Aerospace and Electronic Systems*, AES-9(2):237–252, March 1973.

[2] A. Claus, T. Kadota, and D. Romain. Efficient approximation of a family of noises for application in adaptive spatial processing for signal detection. *IEEE Transactions on Information Theory*, IT-26(5):588–595, September 1980.

[3] J. Goldstein and I. Reed. Reduced rank adaptive filtering. *IEEE Transactions on Signal Processing*, 45(2):492, February 1997.

[4] J. Goldstein and I. Reed. Subspace selection for partially adaptive sensor array processing. *IEEE Transactions on Aerospace and Electronic Systems*, 33(2):539–544, April 1997.

[5] J. Goldstein and I. Reed. Theory of partially adaptive radar. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1309–1325, October 1997.

[6] I. Kirsteins and D. Tufts. Rapidly adaptive nulling of interference. In M. Bouvet and G. Bienvenu, editors, *High Resolution Methods in Underwater Acoustics*. Springer-Verlag, New York, N.Y., 1991.

[7] I. Kirsteins and D. Tufts. Adaptive detection using low rank approximation to a data matrix. *IEEE Transactions on Aerospace and Electronic Systems*, 30(1):55–67, January 1994.

[8] B. Picinbono. A geometric interpretation of signal detection and estimation. *IEEE Transactions on Information Theory*, IT-26:493–497, July 1980.

[9] I. Reed, J.D.Mallett, and L. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, AES-10(6):863–863, November 1974.

[10] A. A. Shah and D. W. Tufts. Estimation of the signal component of a data vector. In *IEEE Proc ICASSP-92*, pages 393–396, San Francisco, CA, March 1992.

[11] A. A. Shah and D. W. Tufts. Determination of the dimension of a signal subspace from short data records. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 42:2531–2535, September 1994.

[12] J. Thomas, L. Scharf, and D. Tufts. The probability of a subspace swap in the svd. *IEEE Transactions on Signal Processing*, 43(3):730, March 1995.

[13] D. Tufts, A. Kot, and R. Vaccaro. The analysis of threshold behavior of svd-based algorithms. In *Proceedings of the 21st Asilomar Conference on Signals, Systems and Computers*, pages 550–554, Pacific Grove, CA, November 1987.

[14] D. Tufts, R. Kumaresan, and I. Kirsteins. Data adaptive signal estimation by singular value decomposition of a data matrix. *Proceedings of the IEEE*, 70:684–685, June 1982.

[15] D. W. Tufts and A. A. Shah. Blind Weiner filtering: Estimation of a random signal in noise using little prior knowledge. In *IEEE Proc ICASSP-93*, pages IV–236–IV–239, Minneapolis, MN, April 1993.

[16] D. W. Tufts and A. A. Shah. Rank determination in time-series analysis. In *IEEE Proc ICASSP-94*, pages IV–21–IV–24, Adelaide, South Australia, April 1994.