obtained by Kung's algorithm [8], which is the impulse response by PCM's algorithm [9]. The absolute error of the truncated impulse response for the above realization and the realization obtained by Kung's algorithm is plotted in Fig. 1 to display the relative performance. In addition, the squared error J between the FIR realization of the original data sequence and each of the three realizations is shown in Table II along with the squared error of the truncated impulse response, i.e.,

$$\mathcal{J} = \sum_{k=1}^{22} \left| h(k) - c_r A_r^{k-1} b_r \right|^2$$

The convergence of the proposed algorithm is clearly exhibited in Fig. 2, where the squared error J as a function of t is plotted on the interval $[0, 50]$. The variation of the elements of U is displayed in Figs. 3 and 4. The variation of V is similar.

V. Conclusion

Two convergent search algorithms have been proposed for the minimization of the integral squared impulse response error between a full-order digital filter and a reduced-order model. These algorithms are applicable to the general multi-input multi-output case and have been shown to be effective by simulation studies. It is envisaged that similar algorithms can be developed to deal with the case where the reduction error is frequency weighted.

References

[1] D. Angwin and H. Kaufman, *Digital Image Restoration*. New York, NY: Springer-Verlag, 1991.

[2] L. Baratchart, M. Cardelli, and M. Olivi, "Identification and rational L_2 approximation: A gradient algorithm," *Automatica*, pp. 413–417, 1991.

[3] J. T. Spanos, M. H. Milman, and D. L. Mingori, "A new algorithm for L_2 optimal model reduction," *Automatica*, pp. 897–909, 1992.

[4] W.-Y. Yan and J. Lam, "An approximate approach to H^2 optimal model reduction," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1341–1358, July 1999.

[5] W.-Y. Yan, L. Xie, and J. Lam, "Convergent algorithms for frequency weighted L_2 model reduction," *Syst. Contr. Lett.*, vol. 31, pp. 11–20, 1997.

[6] L. Li, L. Xie, W.-Y. Yan, and Y. C. Soh, "Design of low order linear phase IIR filters via orthogonal projection," *IEEE Trans. Signal Processing*, to be published.

[7] D. S. Bernstein, L. D. Davis, and D. C. Hyland, "The optimal projection equations for reduced-order, discrete-time modeling, estimation and control," *J. Guid. Contr. Dynam.*, vol. 9, pp. 288–293, 1986.

[8] S. Kung, "A new identification and model reduction algorithm via singular value decompositions," in *Proc. 12th Asilomar Conf. Circuits, Syst., Comput.*, Pacific Grove, CA, 1978, pp. 705–714.

[9] G. M. Pitstick, J. R. Cruz, and R. J. Mulholland, "Approximate realization algorithms for truncated impulse response data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1583–1587, 1986.

A Matrix–Pencil Approach to Blind Separation of Colored Nonstationary Signals

Chunqi Chang, Zhi Ding, Sze Fong Yau, and Francis H. Y. Chan

***Abstract*—For many signal sources such as speech with distinct, nonwhite power spectral densities, second-order statistics of the received signal mixture can be exploited for signal separation. Without knowledge on noise correlation matrix, we propose a simple and yet effective signal extraction method for signal source separation under unknown temporally white noise. This new and unbiased signal extractor is derived from the matrix pencil formed between output autocorrelation matrices at different delays. Based on the matrix pencil, an ESPRIT-type algorithm is derived to get an optimal solution in least square sense. Our method is well suited for systems with colored sensor noises and for nonstationary signals.**

***Index Terms*—Beamforming, blind source separation, cancellation, interference, nonstationary signal processing.**

I. Introduction

Blind source separation has become a well-established research topic in the signal processing community. It finds useful applications in practical scenarios involving multisources and multisensors. The key objective of blind source separation is to extract source signals from sensor measurements without full knowledge of the signal propagation environment. Examples include antenna beamforming, multiple speech separation, and multichannel biomedical signal separation.

Many papers concerning this topic have been published both on separation principles and specific algorithms. The most common assumptions are the following.

1) Sensor noises are white and mutually independent.
2) Source signals are stationary.
3) At most one source is Gaussian.
4) Multiple sensor outputs are *linearly* independent.

Based on these assumptions, separation principles have been investigated [1], [2] and algorithms exploiting higher order statistics have also been proposed [3].

On the other hand, second-order statistics-based algorithms can be exploited under the assumption that source signals for separation are colored [4], [8], [10]. Second-order statistics-based methods do not require the non-Gaussian assumption. They may generate better performance than algorithms based on higher order statistics for short data length. For practical applications involving separations of speech and music signals that are typically nonwhite and nonstationary, second-order statistical methods can be more suitable and effective.

Manuscript received June 30, 1998; revised September 20, 1999. The work of C. Chang and F. H. Y. Chan was supported in part by the Research Grants Council of Hong Kong under Grants HKU553/96M and HKU7036/97E. The work of Z. Ding was supported by the United States Army Research Office and by CRASP. The work of S. F. Yau was supported in part by the Research Grants Council of Hong Kong under Grant HKUST776/96E. The associate editor coordinating the review of this paper and approving it for publication was Prof P. C. Ching.

C. Chang and F. H. Y. Chan are with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong.

Z. Ding is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA.

S. F. Yau is with the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Hong Kong.

Publisher Item Identifier S 1053-587X(00)01558-0.

However, existing second-order statistical methods often rely on statistical knowledge of sensor noise. To overcome this critical weakness, we propose a second-order statistics-based matrix pencil approach in this correspondence. The new approach yields unbiased signal estimates from signal mixtures corrupted by additive white noise. In our original work [18], we simply used a general eigendecomposition of the matrix pencil to estimate the extractor directly. In this correspodence, an optimization procedure is incorporated in least square sense to yield an ESPRIT algorithm with better performance. This paper is organized as follows. In Section II, the problem of blind source separation is described along with relevant assumptions on second-order statistics of signals for separation. Section III outlines a basic signal separability result based on second-order statistics. Section IV presents an ESPRIT algorithm based on the matrix pencil formed by two correlation matrices at different nonzero lags. Monte Carlo simulations demonstrating the performance of the proposed method are presented in Section V.

II. Problem Formulation

A memoryless mixture of multiple signals is often modeled as

$$\vec{x}(n) = A\vec{s}(n) + \vec{w}(n) \tag{1}$$

in which $\vec{s}(n)$ is a vector of source signals of dimension N, and $\vec{x}(n)$ is the received signal vector of dimension M. In (1), $\vec{w}(n)$ is the additive white noise vector, and A is an $M \times N$ memoryless mixing matrix. In order for all sources to be separable, A must have full column rank, requiring $M \geq N$. When A is not full rank, then signals can only be separated as classes [2]. In this correspondence, we only consider systems in which the sources are individually separable, i.e., A has full column rank.

Our objective is to find a signal extracting matrix $B = [\vec{b}_1\ \vec{b}_2\ \cdots\ \vec{b}_N]$ such that

$$B^H\vec{x}(n) = B^HA\vec{s}(n) + B^H\vec{w}(n) = P\vec{s}(n) + B^H\vec{w}(n) \tag{2}$$

where $P = B^HA$ is a permutation matrix that has only one nonzero element in each row and column.

We assume that the source signals

$$\vec{s}(n) = [\,s_1(n)\quad s_2(n)\quad \cdots\quad s_N(n)\,]^T$$

are uncorrelated of one another but are not (temporally) white. Hence, for stationary signals, we have

$$\begin{aligned} R_s(k) &= E\left\{\vec{s}(n)\vec{s}(n-k)^H\right\} \\ &= \text{diag}\,\{r_1(k),\, r_2(k),\, \cdots,\, r_N(k)\} \end{aligned} \tag{3}$$

where $r_i(k) = E\{s_i(n)s_i^*(n-k)\}$ and for deterministic signals

$$\begin{aligned} R_s(k) &= E\{\vec{s}(n)\vec{s}(n-k)^H\} \\ &= \lim_{N\to\infty} N^{-1} \sum_{n=1,\,N} \vec{s}(n)\vec{s}(n-k)^H \\ &= \text{diag}\{r_1(k),\, r_2(k),\, \cdots,\, r_N(k)\}. \end{aligned} \tag{4}$$

Here, the same notation E is used for both ensemble averaging and time averaging.

The additive noise $\vec{w}(n)$ is supposed to be a stationary, temporally white, zero mean complex random process uncorrelated of the sources. Therefore

$$E\{\vec{w}(n)\vec{w}(n-k)^H\} = \delta(k)R_w \tag{5}$$

where $\delta(k)$ is the Kronecker delta function, and R_w is the covariance matrix of the additive noise. However, here we do not require that the noise be spatially white as assumed in many other approaches. Allowing more freedom in array calibration, R_w can be an arbitrary unknown matrix.

III. Signal Eparability Based on Second—Order Statistics

Let $R_s(k)$ be as defined in (3) or (4). Without loss of generality, we can assume that $R_s(0) = I$. Given noise with known correlation matrix R_w, the covariance matrix of the mixtures can be shown as

$$R_x(0) = E\{\vec{x}(n)\vec{x}(n)^H\} = AR_s(0)A^H + R_w = AA^H + R_w. \tag{6}$$

Thus, we can calculate

$$R_x(0) - R_w = AA^H \tag{7}$$

which can be whitened using a whitening matrix W such that

$$I = W(R_x(0) - R_w)W^H = WAA^HW^H. \tag{8}$$

It is, hence, clear that $U = WA$ is unitary.

We note that the assumption of known noise correlation matrix is typically impractical. In the special case when the noise correlation matrix is of the form $\sigma^2 I$ and $M > N$, then σ_w^2 can be estimated as the smallest eigenvalue of $R_x(0)$.

Denoting the whitened data vector as

$$\vec{z}(n) = W\vec{x}(n) \tag{9}$$

then

$$\begin{aligned} R_z(k) &= E\{\vec{z}(n)\vec{z}(n-k)^H\} \\ &= WAR_s(k)A^HW^H \\ &= UR_s(k)U^H. \end{aligned} \tag{10}$$

Hence, columns of U are eigenvectors of $R_z(k)$.

If $R_s(k)$ has identical eigenvalues, then their eigenvectors are not unique. Hence, we cannot always use the eigenvector of $R_z(k)$ to achieve a complete separation. However, we can separate groups of signals by exploiting eigenvectors corresponding to different eigenvalues. In a totally blind situation, let $R_s(k_1)$ have only L distinct eigenvalues; then, we can have L subspaces that are orthogonal, and hence, L subclasses of signals can be extracted by

$$U_i^H\vec{z}(n), \qquad i = 1,\, 2,\, \cdots,\, L. \tag{11}$$

The dimension of U_i is equal to the multiplicity of the corresponding eigenvalue.

Once we extract a new signal vector with lower dimension $\vec{x}_i(n) = U_i^H\vec{z}(n)$, its components can be extracted based on $R_s(k_2)$ using identical steps. As we can clearly observe, to ensure that this procedure can eventually extract all the sources, it is necessary that the sources have different power spectral densities. In [10] and [11], the separation principle is discussed in detail.

The separability analysis shows that the sources are separable if they have different power spectral densities. Note that this condition is also a necessary condition if only second-order statistics is used since any unitary transform on two uncorrelated sources with identical power spectral density will maintain the spectral density and the uncorrelated property. A more rigorous proof is given in [11]. Of course, in practical situations, it is unlikely that two sources will have same spectral density unless they are both white signals; therefore, in the above procedure,

the eigendecomposition of the covariance matrix at only one time lag is enough to separate the sources, and no further iteration is necessary; this leads to the original AMUSE algorithm [8], [9]. However, when some of the spectral densities are very similar, considering more time lags will improve robustness and accuracy of the separation in practical situations. As an alternative to the iterative procedure described above, [4]–[6] proposed the joint diagonalization [7] of covariance matrices for different time lags to solve this problem and proposed the "SOBI" algorithm based on this concept.

The algorithms discussed above, including our iterative procedure, AMUSE, and SOBI, are all two-step approaches that require a whitening step under the assumption that the correlation matrix of the noise is either known or can be fully estimated. However, when the noise correlation matrix is either unknown or cannot be estimated, these methods may no longer be effective. In view of this shortcoming, we present in the next section an alternative approach that is insensitive to additive (temporally) white noise.

IV. Matrix Pencil Algorithms

A. Basic Statistical Information

Matrix pencil methods, including the well-known ESPRIT algorithms, were first proposed in the field of array signal processing to estimate the directions of arrival. For array signal processing, if the array manifold is completely unknown, it is equivalently a blind source separation problem. However, many approaches assume that the exact array manifold is known, as in MUSIC, or that the array manifold has a rotational invariance, as in the ESPRIT algorithm. These methods exploit only the spatial information of the signals. We will show that when source signals have nonwhite second-order temporal information, matrix pencil methods can be adopted to separate sources blindly.

Second-order statistics of the signals are contained in their autocorrelation matrices at all time delays. In the remaining part of this paper, we use the statistics contained in

$$R_x(k) = E\{\vec{x}(n)\vec{x}(n-k)^H\} = AR_s(k)A^H, \quad k \neq 0 \tag{12}$$

that are insensitive to additive temporally white noises. In particular, we choose the matrix pencil to be

$$\{R_1, R_2\} = \{R_x(k_1), R_x(k_2)\} \tag{13}$$

where $k_1 \neq k_2$ and propose two matrix pencil-based algorithms.

B. Generalized Eigendecomposition

Denote

$$R_1 = R_x(k_1) = AR_s(k_1)A^H = A\Lambda_1 A^H \tag{14}$$

$$R_2 = R_x(k_2) = AR_s(k_2)A^H = A\Lambda_2 A^H \tag{15}$$

where

$$\Lambda_1 = R_s(k_1) = \text{diag}\{d_{1,1}, \cdots, d_{1,N}\} \tag{16}$$

$$\Lambda_2 = R_s(k_2) = \text{diag}\{d_{2,1}, \cdots, d_{2,N}\}. \tag{17}$$

Consider the generalized eigenvalue problem

$$R_1\vec{v} = \lambda R_2\vec{v}. \tag{18}$$

Effectively, the above equation can be rewritten as

$$A\left[\Lambda_1 - \lambda\Lambda_2\right]A^H\vec{v} = 0 \tag{19}$$

or equivalently for A with full rank

$$\begin{bmatrix} d_{1,1} - \lambda d_{2,1} & 0 & \cdots & 0 \\ 0 & d_{1,2} - \lambda d_{2,2} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & d_{1,N} - \lambda d_{2,N} \end{bmatrix} \cdot A^H\vec{v} = 0. \tag{20}$$

One trivial solution to the eigenvector problem is

$$A^H\vec{v}_0 = 0. \tag{21}$$

Note that the trivial solution will exist if $M > N$.

Since A has full column rank, any other nontrivial solution requires the corresponding eigenvalue to satisfy

$$d_{1,i} - \lambda_i d_{2,i} = 0 \quad \text{or} \quad \lambda_i = \frac{d_{1,i}}{d_{2,i}} \tag{22}$$

and if the ratio λ_i is unique (i.e., the eigenvalue is unique), the corresponding eigenvector must satisfy

$$A^H\vec{v}_i = \alpha_i\mathbf{e}_i \tag{23}$$

where $\boldsymbol{e}_i$ is the ith column of the identity matrix I. In general, if the ratio of $d_{1,i}/d_{2,i}$ is not unique and

$$\frac{d_{1,i}}{d_{2,i}} = \gamma, \qquad \forall\, i \in I = \{i_1, i_2, \cdots, i_m\} \tag{24}$$

then let V_γ consist of all eigenvectors for eigenvalue γ

$$V_\gamma = [\,\vec{v}_{i_1} \quad \vec{v}_{i_2} \quad \cdots \quad \vec{v}_{i_m}\,]. \tag{25}$$

We have

$$A^H V_\gamma = \alpha_\gamma\,[\,\boldsymbol{e}_{i_1} \quad \boldsymbol{e}_{i_2} \quad \cdots \quad \boldsymbol{e}_{i_m}\,]\,U_\gamma \tag{26}$$

in which U_γ is an $i_m \times i_m$ unitary matrix. Let V be the generalized eigenvector matrix of the matrix pencil $R_1,\ R_2$; hence, we have

$$V^H A = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ 0 & U_{\gamma_1} & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & U_{\gamma_p} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} DP \tag{27}$$

where each element represents a matrix, P is a permutation, and D is a diagonal matrix.

Note that although we do not estimate the original parameter matrix A, we can get a matrix V that transforms A to a block diagonal matrix multiplied by a permutation. In many practical applications, we do not always need the parameter matrix A itself, and this block diagonal form is very useful for blind signal separation. If all ratios $\{\lambda_i\}$ are unique, then the block diagonal form of $V^H A$ becomes purely diagonal, and all signals are completely separated by V. Otherwise, V separates source signals into disjoint groups that can be further separated by using additional matrix pencil formed by correlation matrices at different lags.

Since V satisfies (27), we have

$$\vec{y}(n) = V^H \vec{x}(n) = \begin{bmatrix} \vec{y}_0(n) \\ \vec{y}_1(n) \\ \vdots \\ \vec{y}_p(n) \\ \vec{y}_{p+1}(n) \end{bmatrix} = \begin{bmatrix} D_0 \vec{s}_0(n) + V_0^H \vec{w}(n) \\ d_1 U_{\gamma_1}^H \vec{s}_1(n) + V_1^H \vec{w}(n) \\ \vdots \\ d_p U_{\gamma_p}^H \vec{s}_p(n) + V_p^H \vec{w}(n) \\ V_{p+1}^H \vec{w}(n) \end{bmatrix}. \tag{28}$$

Therefore, the signals can be partitioned into $p+2$ groups, with the first group $\vec{y}_0(n)$ as separated sources corrupted by additive noises and the last group $\vec{y}_{p+1}(n)$ purely as noises. For the groups $\vec{y}_1(n) \cdots \vec{y}_p(n)$, they are still linear combinations of the sources corrupted by additive noises. The sources contained in $\vec{y}_i(n)$, $i = 1, \cdots, p$ can further be extracted from $\vec{y}_i(n)$ by applying the same procedure on its autocorrelation matrices

$$R_{y_i}(k) = E\{\vec{y}_i(n)\vec{y}_i(n-k)^H\} = U_{\gamma_i}^H R_{s_i}(k) U_{\gamma_i} \tag{29}$$

for a pair of different delays k_3 and k_4. Alternatively, since U_{γ_i} is unitary, its estimate can be obtained directly as the eigenvectors of $R_{y_i}(k)$.

Here, we note that the groups $\vec{y}_1(n) \cdots \vec{y}_p(n)$ can be easily determined by the generalized eigendecomposition of the matrix pencil since they are associated to the nonunique eigenvalues. However, we must separate the all noise (last) output signals from other groups because for a vector $\vec{v}_0$ that satisfies $A^H \vec{v}_0 = 0$, we always have

$$A\Lambda(k_1)A^T \vec{v}_0 = \lambda A \Lambda(k_2) A^T \vec{v}_0 = 0 \quad \text{for any} \quad \lambda. \tag{30}$$

In other words, v_0 can be eigenvector of any eigenvalue. To identify these eigenvectors that generate all noise outputs, notice that signals in $\vec{y}_{p+1}(n)$ are pure noises that have the smallest power. Thus, we can always discard the $M - N$ output signals that have the smallest power. The determination of pure noise output can also be processed without explicitly computing the output signals if we utilize $R_x(0)$ since $\vec{v}^H R_x(0) \vec{v}$ equals the output signal power.

C. ESPRIT Algorithms

The ESPRIT algorithms utilize the rotation invariance structure of the data. Rewrite $\{R_1, R_2\}$ in the form

$$R_1 = AE \tag{31}$$

$$R_2 = A\Phi E \tag{32}$$

where $E = R_s(k_1)A^H$, and $\Phi = R_s(k_2)R_s^{\dagger}(k_1) = \text{diag}\{(d_{2,1}/d_{1,1}), \cdots, d_{2,N}/d_{1,N}\}$. Here, we assume that $R_s(k_1)$ has full rank; otherwise, the above relationship may not be satisfied. E has full row rank as A is of full column rank. There are three versions of ESPRIT algorithms:

1) the direct formulation [16];
2) the least square formulation [12]–[14];
3) the total least square formulation [15].

The total least squares (TLS) version of the ESPRIT algorithm solves the following minimization problem:

$$\min_{A,\Phi,E} \left\| \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} - \begin{bmatrix} A \\ A\Phi \end{bmatrix} E \right\|_F^2. \tag{33}$$

Let

$$B = AE \tag{34}$$

and

$$\Psi = E^{\dagger} \Phi E. \tag{35}$$

Then (33) can be rewritten as

$$\min_{B,\Psi} \left\| \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} - \begin{bmatrix} B \\ B\Psi \end{bmatrix} \right\|_F^2. \tag{36}$$

Solving the minimization problem in (36) results in a total least squares estimate [15] of Ψ, which is given by

$$\hat{\Psi}_{\text{TLS}} = -V_{12}V_{22}^{\dagger} \tag{37}$$

where V_{12} and V_{22} are implicitly defined by the eigendecomposition

$$\begin{bmatrix} R_1^H \\ R_2^H \end{bmatrix} [R_1 \; R_2] = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \Lambda \begin{bmatrix} V_{11}^H & V_{21}^H \\ V_{12}^H & V_{22}^H \end{bmatrix} \tag{38}$$

and $\Lambda = \text{diag}[\lambda_1, \cdots, \lambda_{2M}]$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2M}$.

Alternatively, a least squares estimate of Ψ is obtained from (36) by setting the first block to zero, i.e., letting $B = R_1$. This results in

$$\min_{\Psi} \|R_2 - R_1\Psi\|_F^2 \tag{39}$$

with the solution given by

$$\hat{\Psi}_{\text{LS}} = R_1^{\dagger} R_2 \tag{40}$$

where $R_1^{\dagger} = [R_1^H R_1]^{-1} R_1^H$ is the Moon–Penrose pseudo-inverse of R_1. This implicitly assumes that there are no measurement errors in R_1. Similarly, by letting $B = R_2\Psi^{-1}$, a solution that assumes no errors in R_2 is obtained as

$$\hat{\Psi}_{\text{LS}} = R_2^{\dagger} R_1. \tag{41}$$

Given noisy measurements, both R_1 and R_2 contain errors. The LS estimate will, thus, be biased. It is better to treat the errors in these two matrices symmetrically. This leads to the TLS solution in (37) and (38).

Equation (35) can be rewritten as

$$\Psi = (A^H)^{\dagger} R_s^{\dagger}(k_1) \Phi R_s(k_1) A^H = (A^H)^{\dagger} \Phi A^H. \tag{42}$$

Assume that Φ can be written as

$$\Phi = P_{\Phi}\Phi_0 P_{\Phi}^T = P_{\Phi} \begin{bmatrix} \phi_1 I_{\gamma_1} & 0 & \cdots & 0 \\ 0 & \phi_2 I_{\gamma_1} & \cdots & 0 \\ & & & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \phi_p I_{\gamma_p} \end{bmatrix} P_{\Phi}^T \tag{43}$$

where P_{Φ} is a permutation matrix, I_{γ_i} is an identity matrix for any $i = 1 \cdots p$, and $\phi_1 > \phi_2 > \cdots > \phi_p$. Correspondingly, we define that

$$D = P_D D_0 = P_D \begin{bmatrix} d_1 U_{\gamma_1} & 0 & \cdots & 0 \\ 0 & d_2 U_{\gamma_2} & \cdots & 0 \\ & & & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & d_p U_{\gamma_p} \end{bmatrix} \tag{44}$$

where U_{γ_i} is unitary for any $i = 1 \cdots p$, and P_D is a permutation matrix. Let $P = P_{\Phi}P_D^T$; then, it is easy to see that the eigenmatrix of Ψ is

$$V = ((APD)^H)^{\dagger}. \tag{45}$$

In fact, it can be verified by

$$\begin{aligned}\hat{\Psi}_{\mathrm{TLS}}V &= (A^{\dagger})^H\Phi A^H((APD)^H)^{\dagger}\\ &= (A^{\dagger})^H\Phi PP_D(D_0^{\dagger})^H\\ &= (A^{\dagger})^H P_\Phi P_\Phi^T\Phi P_\Phi(D_0^{\dagger})^H\\ &= (A^{\dagger})^H PP_D(D_0^{\dagger})^H\Phi_0\\ &= (A^{\dagger})^H P(D^{\dagger})^H\Phi_0\\ &= V\Phi_0. \end{aligned} \tag{46}$$

Thus, we have

$$\begin{aligned}\vec{y}(n) &= V^H\vec{x}(n)\\ &= D^{\dagger}PA^{\dagger}(A\vec{s}(n)+\vec{w}(n))\\ &= D^{\dagger}(P\vec{s}(n)+PA^{\dagger}\vec{w}(n))\\ &= \begin{bmatrix} D_0^{-1}\vec{s}_0(n)+\vec{\tilde{w}}_0(n)\\ d_1^{-1}U_{\gamma 1}^H\vec{s}_1(n)+\vec{\tilde{w}}_1(n)\\ \vdots\\ d_p^{-1}U_{\gamma p}^H\vec{s}_p(n)+\vec{\tilde{w}}_p(n)\end{bmatrix}\\ &= \begin{bmatrix}\vec{y}_0(n)\\ \vec{y}_1(n)\\ \vdots\\ \vec{y}_p(n)\end{bmatrix}. \end{aligned} \tag{47}$$

If all the diagonal entries of Φ are unique, then

$$\vec{y}(n) = D^{-1}P\vec{s}(n) + D^{-1}PA^{\dagger}\vec{w}(n) \tag{48}$$

the sources are completely separated. In general, the sources are separated into groups, as stated in (47). For those groups where the signals are still linear combination of the original sources, the sources can be further separated with a procedure, as described in Section IV-B.

In the practical case, V is estimated from $\hat{\Psi}_{\mathrm{TLS}}$ or $\hat{\Psi}_{\mathrm{LS}}$.

V. Summary and Discussion

In the previous section, we proposed two matrix pencil approaches to blind source separation using second-order statistics. We avoided using the information contained in $R_x(0)$, which is sensitive to temporally white sensor noise. Compared with other approaches such as AMUSE and SOBI, which rely on a prewhitening step, our approach is simpler and is well suited for practical systems with temporally white sensor noises that are not mutually independent.

Of course, the matrix pencil methods can also be applied to prewhitened data. It can be shown that in this case, all the versions of ESPRIT algorithms will give the same result if the matrix pencil is chosen as a pair of correlation matrices at different delays.

For all three methods, the separated signals are generated from $\vec{y}(n) = V^H\vec{x}(n)$. In short, these algorithms can be summarized as follows.

- *Generalized Eigendecomposition (GED) Approach*: Compute the generalized eigendecomposition of the matrix pencil $\{R_x(k_1),\ R_x(k_2)\}$. Denote the eigenvectors corresponding to nonzero eigenvalues as a matrix V.
- *LS-ESPRIT*: Compute the eigendecomposition of $R_1^{\dagger}R_2$. Denote the eigenvectors corresponding to nonzero eigenvalues as a matrix V.
- *TLS-ESPRIT*: Complete the eigendecomposition of

$$\begin{bmatrix}R_1^H\\ R_2^H\end{bmatrix}[R_1\ R_2] = \begin{bmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{bmatrix}\Lambda\begin{bmatrix}V_{11}^H & V_{21}^H\\ V_{12}^H & V_{22}^H\end{bmatrix}.$$

Compute the eigendecomposition of $V_{12}V_{22}^{-1}$. Denote the eigenvectors corresponding to nonzero eigenvalues as a matrix V.

Note that for LS-ESPRIT, we can compute the Nth-order pseudo inverse of $R_x(k_1)$ directly. We also note that TLS-ESPRIT has a close relation to joint diagonalization. Assume that $M_i = UD_iU^H$ for $i = 1, \cdots, L$, where U is unitary, and D_i is diagonal. Let $\hat{M}_i$ be an approximation of M_i; then, the joint diagonalization of $\{\hat{M}_i | i = 1, \cdots, L\}$, as proposed by Cardoso [6], solves the optimization problem

$$\min_U \sum_{i=1}^{L} J(U^H\hat{M}_iU) \tag{49}$$

where the matrix function $J(\cdot)$ is given by the off-diagonal elements as

$$J(M) = \|M - \mathrm{diag}(M)\|_F^2 \tag{50}$$

where $\mathrm{diag}(M)$ denotes the diagonal of matrix M. In fact, Wax *et al.* proved in [17] that (49) coincides with the least squares criterion

$$\min_{U,D_i} \sum_{i=1}^{L} \|\hat{M}_i - UD_iU^H\|_F^2. \tag{51}$$

Note that the TLS-ESPRIT (33) can be effectively rewritten as

$$\begin{aligned}&\min_{A,\,R_s(k_1),\,R_s(k_2)} \left\|\begin{bmatrix}R_1\\ R_2\end{bmatrix} - \begin{bmatrix}AR_s(k_1)A^H\\ AR_s(k_2)A^H\end{bmatrix}\right\|_F^2\\ &\quad= \sum_{i=1}^{2}\|R_i - AR_s(k_i)A^H\|_F^2. \end{aligned} \tag{52}$$

This shows that our TLS-ESPRIT algorithm is also a joint diagonalization solution of two nonwhitened correlation matrices.

We should note, however, that our solution of (36) is not the optimal solution to the TLS-ESPRIT formulation (33) since our solution is optimal if and only if the parameters B and Ψ are independent variables. The correlation between B and Ψ is obvious since they are both a combination of A, $R_s(k_1)$, and $R_s(k_2)$. Thus, our TLS solution is only a suboptimal solution of our original least square optimization problem stated in(33). Despite this, the relationship between our TLS approach and joint diagonalization is interesting and useful. In particular, we see that our TLS-ESPRIT approach is inherently an approximation of the joint diagonalization approach such as SOBI [4]. Further, the TLS approach has two advantages over the joint diagonalization approach. The first is that no prewhitening step is required, and the second is that a closed-form solution is obtained without any iterative procedure.

VI. Simulations

Before presenting our simulation results, we need to define our performance measure. The algorithms give out an extraction matrix V. We will use the signal to interference ratio (SIR) in the extracted signals as the performance measure more relevant to source separation.

Assume that the sources are extracted by

$$\hat{\vec{s}}(n) = V^H\vec{x}(n) = (V^HA)\vec{s}(n) + V^H\vec{w}(n). \tag{53}$$

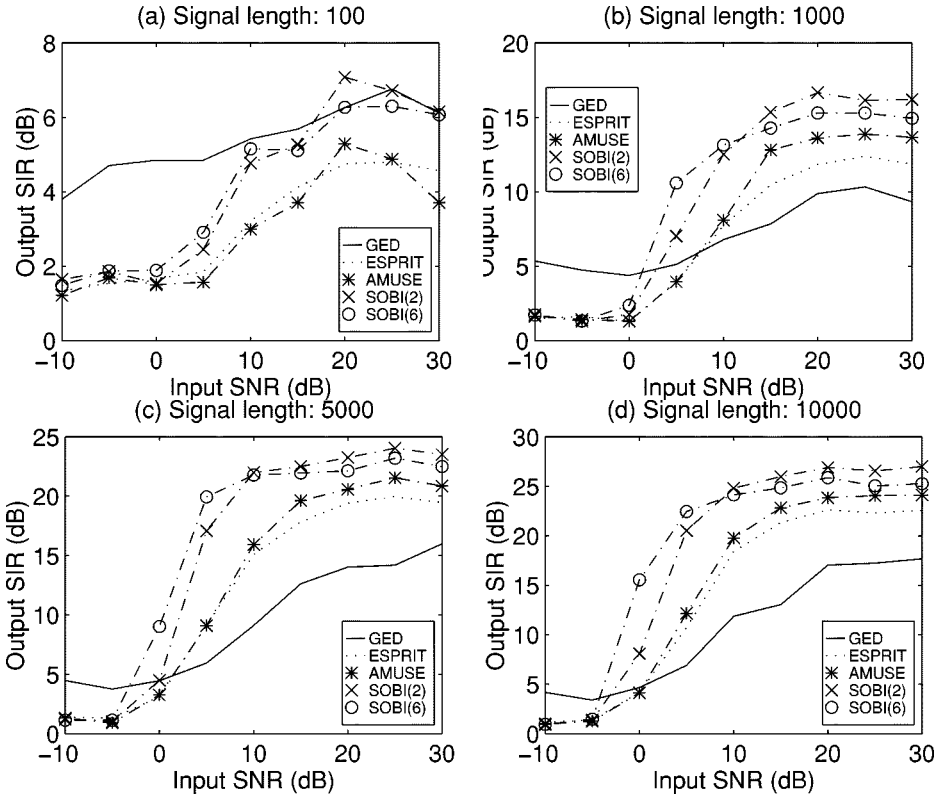

Fig. 1. Performance for the situation where sensor noises are uncorrelated.

We denote $\hat{s}_{i_0}(n)$ as the estimation of $s_i(n)$ if it has the largest SIR for $s_i(n)$ among all the estimated sources $\hat{\vec{s}}(n)$. Letting $C = V^H A$, the SIR of $s_i(n)$ can be computed as

$$\text{SIR}_i = \frac{|C_{i_0 i}|^2}{\sum_{j \neq i_0} |C_{i_0 j}|^2}. \tag{54}$$

For the overall system, the performance measure is defined as the averaged value of SIR for all the sources, which is

$$\mathcal{I}_{\text{perf}} = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^{N} \text{SIR}_i \right) \text{dB}. \tag{55}$$

In the simulation setup, we consider a five-element uniform linear array (ULA) with half wavelength sensor spacing receives two signals in the presence of stationary complex white noise. The sources are generated by filtering complex circular white Gaussian processes through first-order AR filters with poles at $0.85e^{j0.5}$ and $0.85e^{j0.55}$, respectively. The sources arrive from different angles at $\phi_1 = 10^\circ$ and $\phi_2 = 30^\circ$. The simulation is performed for different signal lengths and signal-to-noise ratio (SNR). The signal lengths of 100, 1000, 5000, and 10 000 are used, whereas the SNR varies from -10 to 30 dB at 5–dB increments.

For additive sensor noises, we tested both independent noises and correlated noise. For correlated noise, the correlation matrix is chosen as

$$R_w = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \tag{56}$$

with $\rho = 0.5$.

The algorithms used in this simulation include the generalized eigendecomposition (GED) method and the TLS-ESPRIT algorithm. For the TLS-ESPRIT algorithm, we use matrices $\{R_x(1),\ R_x(2)\}$. For comparison purposes, the AMUSE and SOBI algorithms are also applied. In AMUSE, the time delay is selected as $k = 1$. In SOBI, we consider two cases: One is to jointly diagonalize $R_z(1)$ and $R_z(2)$, which is denoted as SOBI (2), and the other is to jointly diagonalize $R_z(k)$, $k = 1 \cdots 6$, which is denoted as SOBI (6). The signal $\vec{z}(n)$ is the whitened version of original data signal $\vec{x}(n)$. Therefore, there are five algorithms for comparison.

Simulation results are plotted in Figs. 1 and 2. In the figures, the average SIR ($\mathcal{I}_{\text{perf}}$) as defined in (55) is plotted against input SNR. Note that the performance index is averaged over 100 independent Monte Carlo runs. In Fig. 1, the additive noises at the sensors are mutually independent, and in Fig. 2, these additive noises are correlated. In both figures, the result is plotted for signal sample length of 100, 1000, 5000, and, 10 000, respectively, in (a)–(d).

The following observations can be made from the simulation results.

a) For short data length and/or low input SNR, GED outperforms other methods under both independent and correlated sensor noises. In fact, as we can see, when the output SIR for AMUSE and SOBI are less than 2 dB, the GED method generates output SIR of 5 dB. The reason is that for short date length and low SNR, estimates of data correlation matrices are inaccurate. The AMUSE algorithm and the SOBI algorithm both require prewhitening of the data and are more prone to the presence of large noise and short data effects.

b) For long data length and/or high input SNR, correlation matrices become more accurate, and GED does not perform as well as other more complex algorithms.

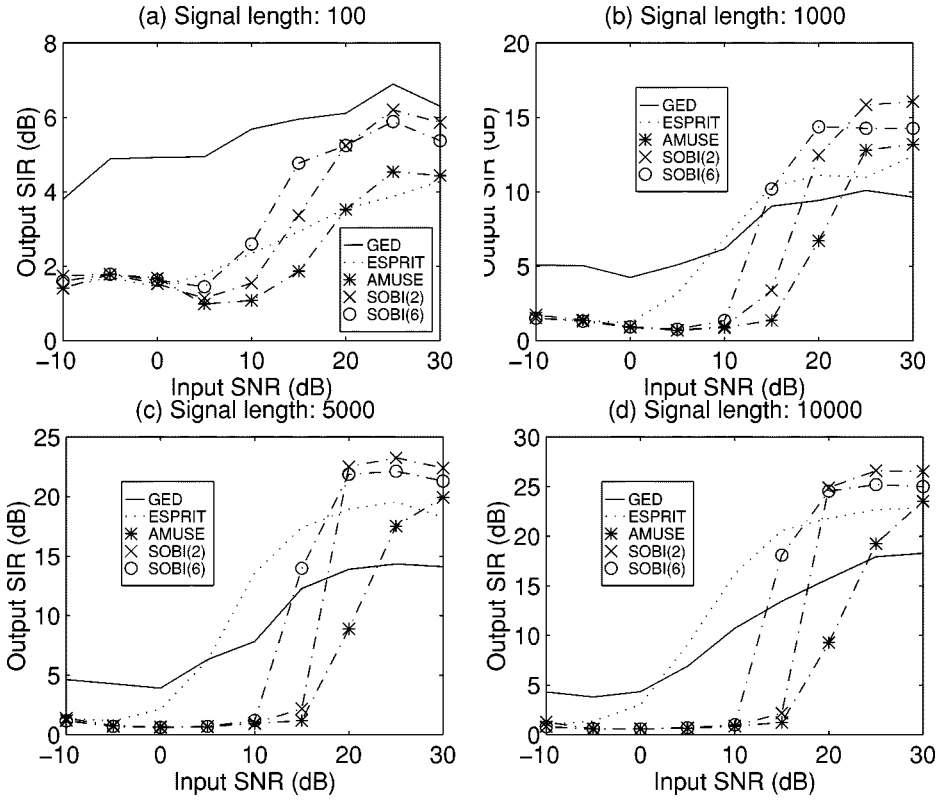

Fig. 2. Performance for the situation where sensor noises are correlated.

c) When sensor noises are uncorrelated, the performance of ESPRIT algorithm is comparable with that of the AMUSE but is inferior to SOBI (2) and SOBI (6). This is because only two correlation matrices are used in ESPRIT and AMUSE, whereas SOBI exploits the additional information in $R_x(0)$.

d) When sensor noises are correlated, performances of AMUSE and SOBI degrade greatly, especially when the input SNR is low, and the whitening procedure does not perform well. However, for the matrix pencil approach, both GED and ESPRIT perform almost the same as the case where noises are uncorrelated since they both use correlation matrices at nonzero lags that are inherently insensitive to the correlation among sensor noises.

e) For the case where sensor noises are correlated and moderate input SNR and data length, ESPRIT outperforms all the other methods.

To summarize simulation results, it can be pointed out that GED is very useful for short data length or the high noise level. When the sensor noises are correlated, ESPRIT performs better than GED under moderate level of SNR.

VII. Conclusions

In this correspondence, we present a new matrix pencil approach to blind source separation based on second-order statistics. Both the generalized eigendecomposition algorithm and the ESPRIT algorithm exploit statistical information contained in the output autocorrelation matrices at different delays that are insensitive to noise correlation. The method relies only on second-order statistics of the received signals and, thus, performs well in short data cases. It can separate colored Gaussian sources, in contrast with higher order statistical methods. It allows flexible array calibration of sensors since it does not require that sensor noises be mutually independent. It does not require signals to be stationary.

References

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, Apr. 1994.

[2] X. Cao and R. Liu, "General approach to blind source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 562–571, Mar. 1996.

[3] J. F. Cardoso, "Source separation using higher order moments," in *Proc. IEEE ICASSP*, Glasgow, U.K., May 1989, pp. 2109–2112.

[4] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[5] K. Abed-Meraim, A. Belouchrani, J. F. Cardoso, and E. Moulines, "Asymptotic performance of second order blind separation," in *Proc. IEEE ICASSP*, vol. 4, Adelaide, Australia, Apr. 1994, pp. 277–280.

[6] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1996.

[7] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 52–53, Feb. 1997.

[8] L. Tong, R. Liu, V. C. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.

[9] L. Tong, V. C. Soon, Y. F. Huang, and R. Liu, "AMUSE: A new blind identification algorithm," in *Proc. IEEE ICASSP*, Albuquerque, NM, Apr. 1990, pp. 1784–1787.

[10] C. Chang, S. F. Yau, P. Kwok, F. K. Lam, and F. H. Y. Chan, "Sequential approach to blind source separation using second order statistics," in *Proc. ICICS*, Singapore, Sept. 1997, pp. 1608–1612.

[11] ——, "Uncorrelated component analysis for blind source separation," *Circuits Syst. Signal Process.*, vol. 18, no. 3, pp. 225–239, 1999.

[12] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1340–1342, Oct. 1986.

[13] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotation invariance techniques—ESPRIT," in *Proc. 19th Asilomar Conf. Circuits, Syst., Comput.*, Pacific Grove, CA, Nov. 1985, pp. 90–93.
[14] ——, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, pp. 1044–1045, July 1986.
[15] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.
[16] F. A. Sakarya and M. H. Hayes, "A subspace rotation-based technique for estimating 2-D arrival angles using nonlinear array configurations," *IEEE Trans. Signal Processing*, vol. 42, pp. 409–411, Feb. 1994.
[17] M. Wax and J. Sheinvald, "A least-square approach to joint diagonalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 52–53, Feb. 1997.
[18] C. Chang, Z. Ding, S. Yau, and F. Chan, "A matrix-pencil approach to blind separation of nonwhite signals in white noise," in *Proc. IEEE ICASSP*, vol. 4, Seattle, WA, May 1998, pp. 2485–2488.

Transmitter Optimization for Noisy ISI Channels in the Presence of Crosstalk

Naofal Al-Dhahir

***Abstract*—Transmitter optimization techniques for maximizing the throughput of linear ISI channels impaired by additive-Gaussian noise and crosstalk are presented. Transmitter ends of both single carrier and multicarrier transceiver structures are optimized subject to a fixed average input energy constraint. The effect of transmitter optimization on channel throughput is quantified by comparison with scenarios where both the desired user and the crosstalker use a flat energy distribution across the transmission bandwidth.**

***Index Terms*—Crosstalk, intersymbol interference, throughput, transmitter.**

I. Introduction

Limited bandwidth resources in many spectrally efficient digital communications systems often result in having multiple users, who typically have the same transmission power spectral density characteristics, share the same frequency band, and, thus, interfere with each other. Among the scenarios where this interference is performance limiting is *crosstalk* (both near-end and far-end) in the emerging high-speed digital subscriber line (DSL) systems [1], [5].

Effective signal processing techniques are implemented at the *receiver* to mitigate crosstalk such as *decision-feedback equalization* (DFE) in single-carrier modulation systems [1] or *FFT processing* in multicarrier modulation systems [5]. Full optimization of a communication system entails optimizing both the receiver *and* transmitter ends, where the second task requires optimizing the transmission bandwidth and the power spectral density shape of the input signal. While transmitter optimization for multicarrier systems on noisy ISI channels with crosstalk has received considerable attention recently [4], [6], this has not been the case for single-carrier systems, where published studies either assume no crosstalk (see [3], [6], and the references therein) or an infinite-length transmit filter, as in [8]. In addition, previous transmitter optimization studies for multicarrier systems are for the discrete multitone (DMT) implementation, where channel spectrum partitioning is effected by using the IFFT/FFT modulating vectors and adding a cyclic prefix to the input block. We are not aware of any multicarrier transmitter optimization studies in the presence of crosstalk for vector coding (VC) multicarrier systems [9], where zero stuffing and optimum eigenvector-based modulating/demodulating vectors are used.

In this correspondence, we present a unified framework for optimizing the transmitter of *finite-complexity* single carrier and multicarrier modulation systems on linear ISI channels impaired by additive-Gaussian noise and crosstalk. The *performance metric* assumed for transmitter optimization is channel throughput (in bits/symbol) at a given symbol rate. The rest of this paper is organized as follows. Section II formulates the general channel throughput maximization problem and specializes it to single-carrier systems. Application to VC-based and DMT-based multicarrier systems is studied in Section III. Simulation results are given in Section IV, and the correspondence is concluded in Section V.

A. Notation

Throughout this correspondence, vectors are denoted in lowercase bold and matrices in uppercase bold. The transpose is denoted by $(\cdot)^t$, the complex-conjugate transpose by $(\cdot)^*$, the determinant by $|\cdot|$, and the inverse by $(\cdot)^{-1}$. Furthermore, $\mathbf{0}$ is a vector of all zeros, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. When the components of a vector are to be emphasized, the first and last components, separated by a colon, are given as a subscript to the vector as in $\mathbf{x}_{k+N-1:k-\nu}$.

II. Single-Carrier Modulation Transmitter Optimization

A. Input–Output Model

We adopt the following discrete-time representation of an additive-noise dispersive channel impaired by crosstalk:

$$\mathbf{y}_k = \sum_{m=0}^{\nu} \mathbf{h}_m x_{k-m} + \mathbf{n}_k + \sum_{i=0}^{\nu_x} \mathbf{g}_i \tilde{x}_{k-i} \tag{1}$$

where $\mathbf{h}_m \stackrel{\text{def}}{=} [h_{l-1,m} \ \cdots \ h_{0,m}]^t$ and $\mathbf{g}_i \stackrel{\text{def}}{=} [g_{l-1,i} \ \cdots \ g_{0,i}]^t$ are the mth main channel and the ith crosstalk channel (vector) impulse response coefficients having memories of ν and ν_x, respectively, and oversampled by a factor of l. We assume a continuous transmission bandwidth and perfect knowledge of the desired and crosstalker channel and the noise characteristics at the transmitter and receiver ends. The input sequence $\{x_k\}$, the crosstalker sequence $\{\tilde{x}_k\}$, and the noise sequence $\{\mathbf{n}_k\}$ are assumed to be stationary, zero-mean, independent of each other, and have nonsingular autocorrelation matrices denoted by $\mathbf{R}_{xx}$, $\mathbf{R}_{\tilde{x}\tilde{x}}$, and $\mathbf{R}_{nn}$, respectively.

The input and crosstalker sequences are generated by the same FIR transmit filter according to

$$x_k = \sum_{n=0}^{\nu_t} p_n \epsilon_{k-n} \quad \text{and} \quad \tilde{x}_k = \sum_{n=0}^{\nu_t} p_n \eta_{k-n} \tag{2}$$

where $\{p_i\}_{i=0}^{\nu_t}$ are the transmit filter coefficients, and $\{\epsilon_k\}$ and $\{\eta_k\}$ are zero-mean sequences with uncorrelated samples and unit variance.

Manuscript received August 18, 1998; revised August 26, 1999. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dimitrios Hatzinakos.

The author is with AT&T Shannon Labs, Florham Park, NJ 07932 USA.

Publisher Item Identifier 1053-587X(00)01541-5.