

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

ON THE RÔLE OF DIMENSIONALITY AND
SAMPLE SIZE FOR UNSTRUCTURED AND
STRUCTURED COVARIANCE MATRIX ESTIMATION

SALVATORE D. MORGERA⁺
DAVID B. COOPER[#]

1 APRIL 1976



(NASA-CR-147118) ON THE ROLE OF
DIMENSIONALITY AND SAMPLE SIZE FOR
UNSTRUCTURED AND STRUCTURED COVARIANCE
MATRIX ESTIMATION (Brown Univ.) 40 p HC
\$4.00

N76-22958

Unclas
CSCI 09B G3/63 26840

⁺ RAYTHEON COMPANY, SYSTEMS LABORATORY, PORTSMOUTH, R.I. 02871

[#] BROWN UNIVERSITY, DIVISION OF ENGINEERING, PROVIDENCE, R.I. 02912

D. B. Cooper's work was partially supported by NASA Grant NSG-5036

ABSTRACT

The motivation for this work emerges from a desire to explain the experimental observation that a surprisingly small sample size vis-à-vis dimension is needed to achieve good signal-to-interference ratio (SIR) performance with an adaptive predetection filter. The adaptive filter requires estimates as obtained by a recursive stochastic algorithm of the inverse of the filter input data covariance matrix. The SIR performance with sample size is compared for the situations where the covariance matrix estimates are of unstructured (generalized) form and of structured (finite Toeplitz) form; the latter case is consistent with weak stationarity of the input data stochastic process. It is argued that a recursive stochastic algorithm operating with a short update period (i.e., with statistically correlated input data vectors) naturally constrains the covariance matrix estimates to be of Toeplitz form and therefore should realize any gain achievable with a (correct) structure assumption. The expected SIR performance for a generalized covariance matrix estimate is shown to approach the optimum SIR as $\approx (1 - \frac{N}{N_S} - \frac{4}{N_S})$, where N is the filter dimension and N_S is the sample size. For a constrained Toeplitz covariance matrix estimate, the expected SIR performance is shown to approach the optimum SIR as $\approx [1 - \frac{\pi^*(N;B)}{N_S} - \frac{7}{NN_S}]$, where $\pi^*(N;B) \approx A + B \ln N + 1/2N$, where A is Euler's constant and $1 \leq B \ll N$ is the input data stochastic process correlation time. The constrained Toeplitz covariance matrix estimate therefore operates with an "effective sample size" $N_S' \approx [\frac{N}{\pi^*(N;B)}] N_S$ and offers the potential of high expected SIR at a sample size N_S for which the generalized estimator may provide exceedingly poor results. Insight is also provided into the effect of the specific form of the desired signal on expected SIR performance.

Although the work presented is cast within a specific problem framework that arises in sonar, radar, or seismic signal processing, the results also apply to the areas of pattern recognition concerned with nonparametric and parametric pattern classifier parameter estimation, e.g., biomedical image recognition and earth resource satellite multispectral data classification.

1. INTRODUCTION

A problem of considerable importance is that of detecting a known signal imbedded in an additive interference noise stochastic process. The physical implementation of the optimum predetection filter requires that certain characterizations of the additive interference noise stochastic process be a priori known. For example, if we constrain the filter structure to be causal and linear, employ an output signal-to-interference ratio (SIR) optimality criterion, and assume the interference noise covariance is a priori known, the optimum filter is the Wiener filter.

With specific regard to the detection problem as it arises in the sonar, seismic, or radar fields, it is not possible to assume that the interference noise covariance is a priori known. A reasonable approach then is to estimate the completely unknown interference noise covariance required for the implementation. This technique leads to an adaptive filter, which designs itself based on estimates formed from the input data. A rather popular selection for the adaptive filter mechanics is offered by the stochastic version of the gradient search method [1-3].

For the first order, linearly constrained gradient search algorithm, it is well known that the weight vector (filter) iterates are asymptotically unbiased, provided that the step size parameter is upper bounded by the inverse of the maximum eigenvalue of the input data covariance matrix. The convergence requirement for the variance of the weight vector iterates is considerably more demanding; it is required that the step size be upper bounded by (approximately) the inverse of the trace of the input data covariance matrix.

Therefore, the variance of the weight vector iterates and, consequently, the achievable filter SIR, are directly related to the filter dimension N . Theory establishes that in order to maintain the variance of the weight vector iterates at an acceptable level, we must have a small step size when N is large. This situation in turn suggests a possibly unbearable convergence time encompassing a large number of statistically independent input data vectors. The form of the optimum Wiener filter involves the inverse of the input data covariance matrix and appears to reinforce the belief that a long convergence time is necessary, because it would appear that N^2 elements (or, at best $\frac{N(N+1)}{2}$ elements, if symmetry is invoked) of this matrix must be estimated.

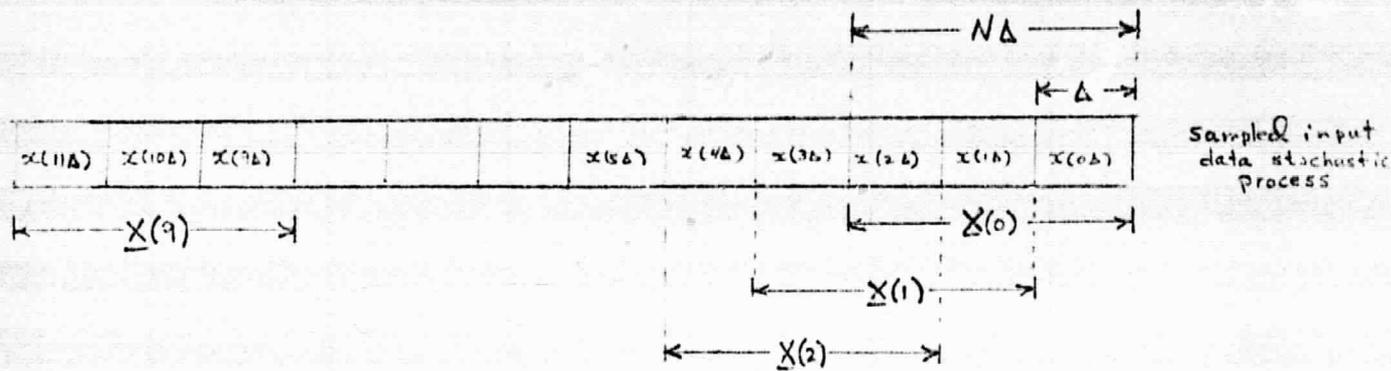
Recent work [4] provides qualitative arguments indicating that the gradient search algorithm operating with a short update period Δ based solely on the frequency content (i.e., Nyquist criteria) of the input data stochastic process does not pretend to attempt to estimate the N^2 elements of the input data covariance matrix. The sequence of input data vectors obtained at the update period Δ is generally highly correlated; the use of these vectors causes the estimate of the input data covariance matrix employed by the gradient search algorithm to be constrained to be of (finite) Toeplitz form. We denote the Toeplitz form estimate as $\hat{R}_T(\cdot)$. The situation is illustrated in Figure 1 with specific parameters utilized to aid the visualization. The stochastic sequence $\{\underline{x}(i\Delta)\}$ or, more compactly, $\{\underline{x}(i)\}$ $i=0,1,2,\dots$, represents the input data sequence with the input data vector $\underline{x}[(i+1)\Delta]$ formed from $\underline{x}(i\Delta)$ by simply shifting the temporally oldest sample out of $\underline{x}(i\Delta)$ and shifting in

the most recent sample. Also shown in Figure 1 is the generalized form input data covariance matrix estimate $\hat{R}_G(\cdot)$ that results when the update period is lengthened to $L\Delta$, where $L\Delta$ is an integer multiple of the input data stochastic process correlation time, i.e., the stochastic sequence $\{X(iL\Delta)\} i=0,1,2,\dots$, is mutually uncorrelated.

We see that operation at the short update period is implicitly consistent with the oft-invoked assumption of weak stationarity or quasi-stationarity, i.e., weak stationarity over some interval of time, of the input data stochastic process and leads to a Toeplitz form input data covariance matrix estimate which requires only N elements for complete description. Operation at the short update period is generally considered in practice to be normal operation, with operation at the long update period of academic interest primarily because it leads to a much more tractable mathematical treatment of the statistical moment properties of the weight vector iterates. The implication of the above observation is that for a fixed real time interval the transient behavior of the weight vector iterates should be better for operation at the short update period vis-à-vis operation at the long update period, assuming that the input data stochastic process is indeed weakly stationary. We do note that the reverse should be true of the steady state behaviors, simply because operation at the short update period involves large numbers of highly correlated input data vectors. This latter point is really of little practical concern, i.e., the weight vector iterates are generally always in the transient state for real sonar, seismic, or radar data.

There is experimental evidence which appears to substantiate these qualitative arguments [4]. In the referenced work, real active sonar data was processed utilizing a first order, linearly constrained gradient search algorithm. The filter dimension utilized was $N=256$ (complex, i.e., analytic signal samples) and the short update period $\Delta \cong 1/5W$, where W denotes the bandwidth of the input data stochastic process. Quite good filter SIR performance was obtained with a step size setting leading to a gradient search algorithm averaging time which encompassed only a small number $N_S \cong N$ of statistically independent input data vectors. Results of this nature are very important in signal detection problems where rapid convergence demands must be met. The surprisingly small sample size N_S required is conjectured to be a result of two factors: (1) possibility of proximity of the filter SIR to the optimum SIR although the weight vector iterates may not be near the optimum Wiener weight vector and (2) operation of the gradient search method with a short update period wherein the input data covariance matrix estimates employed are constrained to be of Toeplitz form.

We test these conjectures in this work. The starting point is the derivation of the expected filter SIR performance for input data covariance matrix estimates of both the generalized and constrained Toeplitz forms. The expected SIR is evaluated for several input data true covariance matrix examples of interest. The results are compared as a function of sample size N_S for two particular values of N with some generalization given for other values of N .



Algorithm Update Period Δ - Constrained Toeplitz Form Estimate

$$\hat{R}_T(9) = \sum_{i=0}^9 \alpha_i \underline{x}(i) \underline{x}^T(i)$$

$$= \begin{bmatrix} \sum_{j=0}^9 \alpha_j x^2(j\Delta) & \sum_{j=0}^9 \alpha_j x(j\Delta) x[(j+1)\Delta] & \sum_{j=0}^9 \alpha_j x(j\Delta) x[(j+2)\Delta] \\ \sum_{j=0}^9 \alpha_j x(j\Delta) x[(j+1)\Delta] & \sum_{j=1}^{10} \alpha_j x^2(j\Delta) & \sum_{j=1}^{10} \alpha_j x(j\Delta) x[(j+1)\Delta] \\ \sum_{j=0}^9 \alpha_j x(j\Delta) x[(j+2)\Delta] & \sum_{j=1}^{10} \alpha_j x(j\Delta) x[(j+1)\Delta] & \sum_{j=2}^{11} \alpha_j x^2(j\Delta) \end{bmatrix}$$

Algorithm Update Period $L\Delta = N\Delta$ - Generalized Form Estimate

$$\hat{R}_G(9) = \sum_{i=0}^3 \beta_i \underline{x}(iL) \underline{x}^T(iL)$$

$$= \begin{bmatrix} \sum_{j=0}^3 \beta_j x^2(jL\Delta) & \sum_{j=0}^3 \beta_j x(jL\Delta) x[(jL+1)\Delta] & \sum_{j=0}^3 \beta_j x(jL\Delta) x[(jL+2)\Delta] \\ \sum_{j=0}^3 \beta_j x(jL\Delta) x[(jL+1)\Delta] & \sum_{j=0}^3 \beta_j x^2[(jL+1)\Delta] & \sum_{j=0}^3 \beta_j x[(jL+1)\Delta] x[(jL+2)\Delta] \\ \sum_{j=0}^3 \beta_j x(jL\Delta) x[(jL+2)\Delta] & \sum_{j=0}^3 \beta_j x[(jL+1)\Delta] x[(jL+2)\Delta] & \sum_{j=0}^3 \beta_j x^2[(jL+2)\Delta] \end{bmatrix}$$

Figure 1 Relationship Between Algorithm Update Period and Input Data Correlation Estimate

2. SIGNAL-TO-INTERFERENCE NOISE RATIO PERFORMANCE CRITERION

With regard to a detection processor in which the primary goal is a decision as to whether a desired signal is present (H_1) or absent (H_0) in interference noise, the processor output signal-to-interference noise power ratio (SIR) serves as a reasonably good performance criterion. The interference noise consists of ever-present ambient noise plus noises attributed to other than the desired signal source, e.g., shipping traffic noise or reverberation noise in the passive and active sonar cases, respectively.

Using traditional binary hypothesis notation, the detection problem may be expressed as,

$$\begin{aligned} H_0: \underline{X}(i) &= \underline{N}(i) \\ H_1: \underline{X}(i) &= \underline{S} + \underline{N}(i) \end{aligned} \quad (1)$$

where $\underline{X}(i)$ and $\underline{N}(i)$ are the $N \times 1$ -dimensional stochastic (real) input data and interference noise vectors obtained at the i^{th} sampling instant and \underline{S} is the $N \times 1$ -dimensional a priori known (real) signal vector. The total signal time duration is $N\Delta$, the extent of the time interval I_d on which the detection problem is defined. The interference noise vector stochastic process is simply modeled as zero-mean and weakly stationary on a time interval several times the length of I_d and otherwise is assumed completely unknown. We call the longer time interval I_e ; it is within I_e that the interference noise vector stochastic process must be estimated.

The detection processor SIR conditioned on the $N \times 1$ -dimensional stochastic filter weight vector \underline{W} is given by,

$$[\text{SIR}|\underline{W}] = \frac{E[z(i)|H_1, \underline{W}] - E[z(i)|H_0, \underline{W}]}{E[z(i)|H_0, \underline{W}]} \quad (2)$$

where the output power statistic $z(i)$ is,

$$z(i) = \underline{W}^T \underline{X}(i) \underline{X}^T(i) \underline{W} \quad (2a)$$

From Equations (1) and (2) we obtain,

$$[\text{SIR}|\underline{W}] = \frac{(\underline{W}^T \underline{S})^2}{\underline{W}^T \underline{R}_{\text{NN}} \underline{W}} \quad (3)$$

where $\underline{R}_{\text{NN}}$ is the $(N \times N)$ -dimensional zero lag interference noise (auto) covariance matrix. The conditional SIR given by Equation

(3) is seen to be equivalent to that obtained with $\underline{R}_{\text{NN}}$ replaced by \underline{R} , the zero lag input data covariance matrix, save for a constant bias under H_1 .

We make this replacement in the ensuing calculations; this is equivalent to the assumption that H_0 is true for $i \Delta \in I_e$.

To establish the relationship pertinent to this work, we let,

$$\underline{W} = \alpha \hat{\underline{R}}^{-1} \underline{S} \quad (4)$$

where α is a scalar and \hat{R} is a nonsingular, unbiased estimate of R . Precisely put, we have $\underline{W} = \underline{W}(N; N_S)$ and $\hat{R} = \hat{R}(N; N_S)$, but we maintain the dependencies on dimension and statistically independent sample size implicit, for notational convenience. Taking the expectation of Equation (4), we obtain,

$$E[\underline{W}] \triangleq \underline{W}^* = \alpha R^{-1} \underline{S} \quad (5)$$

the optimum Wiener weight vector \underline{W}^* . Substituting Equation (4) in Equation (3),

$$[SIR|\hat{R}] = \frac{(\underline{S}^T \hat{R}^{-1} \underline{S})^2}{\underline{S}^T \hat{R}^{-1} \hat{R} \hat{R}^{-1} \underline{S}} \quad (6)$$

The estimate \hat{R} of the input data covariance matrix may be written as,

$$\hat{R} = R + E \quad (7)$$

where E is a zero-mean stochastic error matrix. From Equation (7), we have,

$$\begin{aligned} \hat{R}^{-1} &= [I + R^{-1}E]^{-1} R^{-1} \\ &\cong [I - R^{-1}E + (R^{-1}E)^2] R^{-1} \end{aligned} \quad (8)$$

taking up to the quadratic term in $R^{-1}E$ in the Neumann series expansion of \hat{R}^{-1} . Substituting Equation (8) in Equation (6) and maintaining consistency with the quadratic approximation in Equation (8), we have,

$$\begin{aligned}
 [SIR|\hat{R}] &\cong 8\underline{W}^*T\hat{R}\underline{W}^* - 3\underline{W}^*T\underline{R}\underline{W}^* \\
 &- \underline{W}^*T\hat{R}R^{-1}\hat{R}\underline{W}^* \\
 &- 3 \frac{(\underline{W}^*T\hat{R}\underline{W}^*)^2}{\underline{W}^*T\underline{R}\underline{W}^*}
 \end{aligned} \tag{9}$$

utilizing the optimum weight vector of Equation (5) with α (arbitrarily) set equal to unity. The result of Equation (9) provides a most useful means whereby the expected SIR may be evaluated conditioned on various elements of the set of nonsingular unbiased covariance estimators $\{\hat{R}\}$. We mention that the approximation of Equation (8) holds well for moderate values of SIR and is biased slightly high at low SIR. Furthermore, we deem it unnecessary to examine the conditional variance of the SIR, because the largest stochastic contributions will be to fourth order in $R^{-1}E$.

3. RELIEVING THE CONDITION FOR SPECIFIC COVARIANCE MATRIX STRUCTURES

In this section, we relieve the condition of Equation (9) for two covariance matrix estimator structures of interest: (1) a generalized estimator $\hat{R}_G(N_S)$ and a constrained Toeplitz estimator $\hat{R}_T(N_S)$ of R . The manipulations assume that the available input data sequence $\{\underline{X}(i)\}_{i \in I_e}$, say, $i=0,1,2,\dots,N_S-1$, is (N -variate) Gaussian, zero-mean, mutually uncorrelated, and weakly stationary. The Gaussian assumption facilitates the evaluation of the fourth order moments that arise in the calculations.

Case 1: The generalized estimator assumes no particular structure for the input data covariance matrix, save for symmetry; it is given by,

$$\hat{R}_G(N_S) = \frac{1}{N_S} \sum_{n=0}^{N_S-1} \underline{X}(n) \underline{X}^T(n) \quad N_S \geq N \quad (10)$$

It is well known [5] that this estimator is the maximum likelihood estimator (MLE) of the input data covariance matrix R over the domain of positive definite $N \times N$ - dimensional matrices. Inserting Equation (10) in Equation (9) and taking the expectation, we have,

$$\begin{aligned} E[SIR | \hat{R}_G] &\cong \underline{S} \underline{W}^* \underline{R} \underline{W}^* \\ &= \frac{1}{N_S^2} \sum_{n,m}^{N_S-1} \sum_{i,j}^N \sum_{k,\ell}^N \underline{W}_i^* \sigma_{jk} \underline{W}_\ell^* E[\hat{\rho}_{ij}(n) \hat{\rho}_{k\ell}(m)] \\ &= \frac{3}{N_S^2 (\underline{W}^* \underline{R} \underline{W}^*)} \sum_{n,m}^{N_S-1} \sum_{i,j}^N \sum_{k,\ell}^N \underline{W}_i^* \underline{W}_j^* \underline{W}_k^* \underline{W}_\ell^* E[\hat{\rho}_{ij}(n) \hat{\rho}_{k\ell}(m)] \end{aligned} \quad (11)$$

where w_i^* denotes the i^{th} element of \underline{W}^* and $\hat{p}_{ij}(n) = [x_i(n)x_j(n)]$ with $x_i(n)$ the i^{th} element of $\underline{X}(n)$. The scalar σ_{ij} is the $(i,j)^{\text{th}}$ element of R^{-1} ; the matrix R has the $(i,j)^{\text{th}}$ element given by ρ_{ij} . To evaluate Equation (11) we need the fourth order moment expression for the (weakly stationary) Gaussian stochastic variates, i.e.,

$$E[\hat{p}_{ij}(n)\hat{p}_{kl}(m)] = \begin{cases} \rho_{ij} \rho_{kl} & n \neq m \\ \rho_{ij}\rho_{kl} + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk} & n = m \end{cases} \quad (12)$$

Utilizing Equation (12), Equation (11) reduces to the very simple and revealing form,

$$E[\text{SIR}|\hat{R}_G] \cong \underline{W}^{*T} \underline{R} \underline{W}^* \left(\frac{N_S - H - 4}{N_S} \right) \quad (13)$$

irrespective of R and S .

We note that the expected SIR for a covariance matrix estimate of generalized form is 0.8 times the optimum SIR (i.e., within $\cong 1$ dB) given by $\underline{W}^{*T} \underline{R} \underline{W}^*$ [cf. Equation (3) evaluated at $\underline{W} = \underline{W}^*$] when $N_S = 5N + 20 \cong 5N$, for large N . The indication is that a considerably smaller sample size is required to achieve most acceptable expected SIR vis-á-vis what would be required to estimate the matrix R (or R^{-1}) to provide a refined estimate of the optimum weight vector. Thus, the first conjecture is dispatched. The behavior of Equation (13) is similar to the result obtained by Allais as discussed by Kanal, et. al. [11] for the minimum mean squared error (MMSE) associated with the MLE of the ideal predictor assuming Gaussian statistics.

Case 2: The constrained Toeplitz form estimator assumes a structure consistent with the weak stationarity of the input data stochastic process; the elements of $\hat{R}_T(N_S)$ are given by,

$$\hat{r}_{jk} = \hat{r}_{|j-k|} = \frac{1}{N_S(N-|j-k|)} \sum_{n=1}^{N_S} \sum_{q=1}^{N-|j-k|} x_q^{(n)} x_{q+|j-k|}^{(n)} \quad j, k=1, 2, \dots, N \quad (14)$$

We utilize the same notation for the elements of the covariance matrix estimator as in Case 1 with no fear of confusion. Inserting Equation (14) in Equation (9) and taking the expectation, we obtain,

$$\begin{aligned} E[SIR|\hat{R}_T] &\approx \underline{W}^* T \underline{R} \underline{W}^* \\ &= \frac{1}{N_S^2} \sum_{n=1}^{N_S-1} \sum_{m=1}^N \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^n \sum_{\ell=1}^m w_i^* \sigma_{jk} w_{\ell}^* \left\{ \frac{1}{(N-|i-j|)(N-|k-\ell|)} \cdot \right. \\ &\quad \left. \sum_{p=1}^{N-|i-j|} \sum_{q=1}^{N-|k-\ell|} E[x_p^{(n)} x_{p+|i-j|}^{(n)} x_q^{(m)} x_{q+|k-\ell|}^{(m)}] \right\} \\ &= \frac{3}{N_S^2 (\underline{W}^* T \underline{R} \underline{W}^*)} \sum_{n=1}^{N_S-1} \sum_{m=1}^N \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^n \sum_{\ell=1}^m w_i^* w_k^* w_j^* w_{\ell}^* \left\{ \frac{1}{(N-|i-j|)(N-|k-\ell|)} \cdot \right. \\ &\quad \left. \sum_{p=1}^{N-|i-j|} \sum_{q=1}^{N-|k-\ell|} E[x_p^{(n)} x_{p+|i-j|}^{(n)} x_q^{(m)} x_{q+|k-\ell|}^{(m)}] \right\} \quad (15) \end{aligned}$$

Appealing to Equation (12), we may cast Equation (15) in the form,

$$\begin{aligned}
 E[SIR\hat{R}_T] &\cong \underline{W}^* T R \underline{W}^* \\
 &- \frac{1}{N_S} \sum_i \sum_j \sum_k \sum_\ell^N w_i^* \sigma_{jk} w_\ell^* [a(|i-j|, |k-\ell|) + b(|i-j|, |k-\ell|)] \\
 &- \frac{3}{N_S (\underline{W}^* T R \underline{W}^*)} \sum_i \sum_j \sum_k \sum_\ell^N w_i^* w_k^* w_j^* w_\ell^* [a(|i-j|, |k-\ell|) + b(|i-j|, |k-\ell|)]
 \end{aligned}
 \tag{16}$$

where

$$a(|i-j|, |k-\ell|) \triangleq \frac{1}{(N-|i-j|)(N-|k-\ell|)} \sum_{p=1}^{N-|i-j|} \sum_{q=1}^{N-|k-\ell|} p^{|p-q|} p^{|p-q+|i-j|-|k-\ell||}
 \tag{16a}$$

$$b(|i-j|, |k-\ell|) \triangleq \frac{1}{(N-|i-j|)(N-|k-\ell|)} \sum_{p=1}^{N-|i-j|} \sum_{q=1}^{N-|k-\ell|} p^{|p-q-|k-\ell||} p^{|p-q+|i-j||}
 \tag{16b}$$

The quantities $a(\cdot)$, $b(\cdot)$ serve to account for the fact that the distinct elements of $\hat{R}_T(N_S)$ are obtained by summing along the diagonal and sub-diagonals of $\hat{R}_G(N_S)$ for each input data vector. There are no more than $(N-|i-j|)(N-|k-\ell|)/2$ distinct terms in $a(\cdot)$ and $b(\cdot)$, because the symmetry of R gives rise to the relations,

$$a(|i-j|, |k-l|) = a(|k-l|, |i-j|) \quad (17a)$$

$$b(|i-j|, |k-l|) = b(|k-l|, |i-j|) \quad (17b)$$

It is difficult to reduce Equation (16) further; however, bounding arguments and the aid of Equation (17) permits the following to be established,

$$E[SIR|\hat{R}_G] \leq E[SIR|\hat{R}_T] \quad (18)$$

when the approximation of Equation (8) holds for Case 1

It is important to note that with the optimum SIR held constant, Equation (13) depends simply on N_S , N ; whereas, Equation (16) depends also on the Toeplitz form input data covariance matrix R and the signal vector \underline{S} . The bound of Equation (18) is thus over all R and (non-trivial) \underline{S} . We desire a more concrete idea of how much the Case 2 results upper bound those of Case 1 for situations of practical interest; the next sections assist in this regard.

4. SEVERAL EXAMPLES OF FINITE TOEPLITZ FORM

The approach we take is to select five examples of input data covariance matrices of finite Toeplitz form and evaluate Equation (16) for each, comparing the results obtained with Equation (13). The initial results presented are for $N=5$; this dimension avoids the non-inductive results that sometimes obtain with two dimensional examples, but is not so large as to require exhaustive computational time. All covariance matrix examples are normalized in the sense that the diagonal elements are unity.

Example 1: White Process

In this case, we have $R=I$, the $N \times N$ - dimensional identity matrix. This situation corresponds to an input data white noise stochastic process (correlation time Δ). We use this example for comparative purposes, realizing that rarely does this situation obtain, because the detection processor will generally be preceded by a filter which will increase the correlation time.

Example 2: Tridiagonal Correlation

This example provides a model for R which permits first sub-diagonal correlation only, viz.,

$$R = \begin{bmatrix} 1 & \rho & & & \\ \rho & 1 & \rho & & \\ & \rho & 1 & \rho & \\ & & \rho & 1 & \rho \\ \rho & & & \rho & 1 \end{bmatrix} \quad (19)$$

To insure that R is of positive definite form, care must be taken in selecting the value of ρ . The eigenspectrum $\{\lambda\}$ of R is readily computed to be,

$$\lambda_n = 1 - 2|\rho| \cos \left(\frac{n\pi}{N+1} \right) \quad n=1,2,\dots,N \quad (20)$$

From Equation (20) we see that we must have

$$|\rho| \leq \frac{1}{2 \cos \left(\frac{\pi}{N+1} \right)} \quad (21)$$

in order for R to be of positive definite form; furthermore, $|\rho| \leq 1/2$ if R is to be of positive definite form for all N. For this example, we choose the elements of R as,

$$\rho_{ij} = \rho_{|i-j|} = \begin{cases} \rho = 0.3679 & |i-j| = 1 \\ 1.0 & |i-j| = 0 \\ 0 & \text{otherwise} \end{cases} \quad i,j=1,2,\dots,N=5 \quad (22)$$

Example 3: Markov Process

The Markov process is one of extreme practical interest; the (continuous) covariance of the input data stochastic process is of exponential form, i.e.,

$$\rho(\tau) = e^{-\gamma|\tau|} \quad \gamma > 0 \quad (23)$$

We consider two situations, which differ in terms of stochastic process correlation time: $\gamma=2.5$ and $\gamma=5.0$; the first has a correlation time

twice that of the second. It can be shown [6] that the eigenspectrum $\{\lambda\}$ of the corresponding input data covariance matrix is given by,

$$\lambda_n = \frac{1 - \rho^2}{1 - 2\rho \cos \phi_n + \rho^2} \quad n=1,2,\dots,N \quad (24a)$$

where the $\phi_n, n=1,2,\dots,N; 0 < \phi_1 < \phi_2 \dots < \phi_N < \pi$, are the solutions of,

$$\frac{\sin[(N+1)\phi_n]}{\sin\phi_n} - \frac{2\rho \sin(N\phi_n)}{\sin\phi_n} + \frac{\rho^2 \sin[(N-1)\phi_n]}{\sin\phi_n} = 0 \quad n=1,2,\dots,N \quad (24b)$$

and

$$\rho = e^{-\gamma\Delta} \quad (24c)$$

For this example, we choose the elements of R as,

$$\rho_{ij} = \rho^{|i-j|} = e^{-\gamma\Delta|i-j|} = \begin{cases} 1.0 & |i-j| = 0 \\ 0.6065 & |i-j| = 1 \\ 0.3679 & |i-j| = 2 \\ 0.2231 & |i-j| = 3 \\ 0.1353 & |i-j| = 4 \end{cases} \quad \gamma = 2.5 \quad (25a)$$

$$= \begin{cases} 1.0 & |i-j| = 0 \\ 0.3679 & |i-j| = 1 \\ 0.1353 & |i-j| = 2 \\ 0.0498 & |i-j| = 3 \\ 0.0183 & |i-j| = 4 \end{cases} \quad \gamma = 5.0 \quad (25b)$$

Example 4: Periodic Process

When the input data stochastic process is periodic with period $N\Delta$, the resulting covariance matrix is a circulant, i.e.,

$$R = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \rho_{N-1} \\ \rho_{N-1} & \rho_0 & \rho_1 & \rho_{N-2} \\ \rho_1 & \rho_2 & \rho_3 & \rho_0 \end{bmatrix} \quad (26)$$

Such matrices are a special type of Toeplitz matrix, arise in the modeling of certain spatial-temporal interference noise field models [4], and are utilized to explain the asymptotic ($N \rightarrow +\infty$) behavior of Toeplitz matrices. We may write R in the following manner [7],

$$R = \sum_{n=0}^{N-1} \rho_n J^n \quad (27a)$$

where

$$J = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad (27b)$$

From the eigenvectors of J , we directly obtain the (normalized) eigenvectors \underline{E}_n , $n=1,2,\dots,N$, of R ,

$$\underline{E}_n = N^{-1/2} e^{-2\pi j(n-1)(k-1)/N} \quad k = 1, 2, \dots, N \quad (28)$$

The eigenspectrum $\{\lambda\}$ is given by,

$$\begin{aligned} \lambda_n &= \sum_{k=1}^N \rho_{k-1} e^{-2\pi j(n-1)(k-1)/N} \\ &= \rho_0 + 2\rho_1 \cos \left[\frac{2\pi(n-1)}{N} \right] + 2\rho_2 \cos \left[\frac{4\pi(n-1)}{N} \right] + \dots \quad n=1, 2, \dots, N \end{aligned} \quad (29)$$

invoking the symmetry of R . The last term in the continued series of Equation (29) depends on whether N is even or odd. We see from Equation (29) that the eigenvalues of a circulant matrix are simply given by the discrete Fourier transform (DFT) of the first row of elements. For this example, we choose the elements of R as,

$$\rho_{ij} = \rho_{|i-j|} = \rho_{|N-|i-j||} = \begin{cases} 1.0 & |i-j| = 0 \\ 0.3679 & |i-j| = 1 \\ 0.1353 & |i-j| = 2 \end{cases} \quad (30)$$

5. PERFORMANCE EVALUATION

Figure 2 presents the results of evaluating Equation (13) and Equation (16) for the four covariance matrix examples of Section 4 as a function of sample size N_S . In the calculations, the optimum SIR is fixed at unity, viz., $\underline{W}^{*T} \underline{R} \underline{W}^* = 1$. The (Case 2) curves of Figure 2 are dependent on the choice of signal vector \underline{S} , as mentioned in Section 3. The discussion below provides some insight into the detection performance behavior as a function of \underline{S} and the rationale for the choice of \underline{S} used for the curves of Figure 2.

We find that the quantity $E[SIR|\hat{R}_T]$ appears to be close to its minimum when the N elements of \underline{S} are chosen to be equal, e.g., $\underline{S} = \underline{1}$, each element equal to unity. The quantity $E[SIR|\hat{R}_T]$ is close to its maximum if the first element of \underline{S} is chosen as unit with the remaining $N-1$ elements zero; we denote this signal vector as $\underline{S} = \underline{S1}$. A heuristic argument for these extrema may be given in terms of the asymptotic ($N \rightarrow \infty$) properties of the input data stochastic process. We note that the covariance matrix examples of Section 4, with the exception of Example 1, correspond to lowpass input data stochastic processes, which have continuous power spectral density (psd) functions. A periodogram estimate of the psd (under the zero mean, Gaussian assumption) is not consistent; specifically, the asymptotic variance at a frequency is proportional to the square of the psd at that frequency [8]. A choice $\underline{S} = \underline{1}$ accentuates the zero frequency component of the input data stochastic process, for which the spectral estimation variance is greatest;

thus, the detection performance with such a signal is expected to be poor. A selection of $\underline{S} = \underline{S1}$ causes all frequency components of the lowpass stochastic processes to be utilized, thereby reducing the statistical variability of $E[SIR|\hat{R}_T]$ through an averaging. This situation is strongly exhibited in the case of Example 1, and less so for Example 2, wherein a choice $\underline{S} = \underline{S1}$ gives rise to good detection performance by permitting the constrained Toeplitz form estimator of Equation (14) to utilize the statistical independence of the elements within each input data vector $\underline{X}(n)$ to further reduce estimate variance. Such a choice is therefore expected to provide nearly optimum detection performance for finite N and no signal bandwidth constraint, if the eigenvalue dispersion (condition number) of the input data covariance matrix is small. The situation is not unlike the use of diversity to improve communications system performance.

In view of the above, the signal vector \underline{S} admits a useful representation in terms of the members of the complete orthonormal (CON) set of (real) eigenvectors $\{\underline{E}\}$ associated with the input data covariance matrix R , i.e.,

$$\underline{S} = \sum_{n=1}^N s_n \underline{E}_n \quad (31a)$$

where the expansion coefficients are given by,

$$s_n = \underline{E}_n^T \underline{S} \quad n = 1, 2, \dots, N \quad (31b)$$

When $s_n=1$ $n=1,2,\dots,N$, \underline{S} possesses unit length along each principal component associated with the input data stochastic process. For finite N , this choice for \underline{S} parallels the asymptotic arguments that obtain when $\underline{S}=\underline{S}_1$, except for the case of Example 1, for which \underline{S} degenerates to $\underline{1}$, causing poor detection performance. The exclusion of Example 1 is explained simply by noting that the white stochastic process may be expanded in terms of any CON set. The poor detection performance we refer to obtains when \underline{E}_n has unity in the n^{th} position and zeroes elsewhere; however, we can choose the CON set $\{\underline{E}\}$ differently so that the white stochastic process, and \underline{S} as in Equation (31a) with the expansion coefficients equal to unity, does not give rise to anomalous detection performance. For \underline{S} as in Equation (31a) with the expansion coefficients equal to unity, the optimum weight vector \underline{W}^* is calculated from Equation (5) (with $\alpha=1$) to be,

$$\underline{W}^* = \sum_{n=1}^N \lambda_n^{-1} \underline{E}_n \quad (32)$$

where $\{\lambda\}$ is the eigenspectrum of R . Since \underline{S} shows no preference for one principal component vis-à-vis another, the optimum weighting for a principal component is simply inversely proportional to the respective eigenvalue.

The Case 2 solid curves of Figure 2 are for \underline{S} as computed from Equation (31) with the expansion coefficients equal to unity. In arriving at \underline{S} for the covariance matrix examples shown, the eigenvector set $\{\underline{E}\}$ was computed for each of the examples. The eigenspectrum $\{\lambda\}$ was also computed for each of

the examples and checked against Equations (20), (24), and (29). The eigenvalues are arranged in descending order (over $\left[\frac{N}{2}\right]$ in the case of Example 4) inverted, scaled to a maximum of unity, and displayed in Figure 3 as the optimum weighting applied to each principal component. The Case 2 dashed curves of Figure 2 are for Examples 1 and 2 when $\underline{S}=\underline{S}_1$.

The data presented in Figure 2 may be coalesced in a most meaningful form. Examination of Equation (13) indicates that the expected SIR for the generalized covariance matrix estimator varies as $\left(\frac{N_S - N - 4}{N_S}\right)$; the trend in the curves of Figure 2 suggests that the expected SIR for the constrained Toeplitz covariance matrix estimator varies with sample size in the following manner,

$$E[SIR|\hat{R}_T] \cong \underline{W}^* T_{RW}^* \left(\frac{N'_S - N - 4}{N'_S} \right) \quad (33a)$$

where the "effective sample size" N'_S is given by,

$$N'_S = \beta N_S \quad \beta \geq 1 \quad (33b)$$

and $\beta = \beta(\underline{S})$, i.e., β carries implicit dependence on \underline{S} . Evaluating the multiplicative factor β from Equation (33) and the data of Figure 2, we obtain the results shown in Table 1.

Covariance Matrix Example	β
1 (WP)	1.06
1 (WP, $S=S1$)	2.38
2 (TD)	2.29
2 (TD, $S=S1$)	2.30
3 (MP, $\gamma=5.0$)	2.41
3 (MP, $\gamma=2.5$)	1.87
4 (PP)	2.53

TABLE 1

EFFECTIVE SAMPLE SIZE INCREASE OF TOEPLITZ ESTIMATOR OVER GENERALIZED ESTIMATOR, N=5

We infer from the values of Table 1 that, on the average, a sample size reduction of 2.12 is possible with the constrained Toeplitz estimator vis-à-vis the generalized estimator for N=5. This result is highly significant when the input data stochastic process is quasi-stationary, because the (real) time necessary to achieve good detection performance is reduced to approximately 0.47 times that normally required.

Since the above results are tendered on a specific dimensionality, we make an attempt to establish the dependence of β on N. To do this, we repeat the calculations for N=10. The larger Example 1 and 2 matrices are natural extensions of those in Section 4; the elements of the larger Example 3 matrices are readily computed from Equation (25) with $\Delta=0.1$, $\gamma=2.5$ and 5.0; and the Example 4 matrix elements are given by,

$$\rho_{ij} = \rho_{|i-j|} = \rho_{|N-|i-j||} \left\{ \begin{array}{ll} 1.0 & |i-j| = 0 \\ 0.6065 & |i-j| = 1 \\ 0.3679 & |i-j| = 2 \\ 0.2231 & |i-j| = 3 \\ 0.1353 & |i-j| = 4 \\ 0.0821 & |i-j| = 5 \end{array} \right. \quad (34)$$

Figure 4 displays the expected SIR curves for $N=10$, and Figure 5 illustrates the optimum principal component weighting. Generally, we observe the same trends in Figure 4 as exhibited by the curves of Figure 2, except for the results pertaining to the Example 2 (tridiagonal) covariance. In this case, the difference in detection performance obtained with \underline{S} as in Equation (31) and $\underline{S}=\underline{S}_1$ is noticeably less at the higher dimension. The inference is that asymptotic arguments appear to hold for the Example 2 covariance at the moderate dimension $N=10$, i.e., the eigenvectors of the Example 2 covariance matrix are apparently near those of Equation (28). Evaluating the multiplicative factor β from Equation (33) and the data of Figure 4, we obtain the results shown in Table 2.

Covariance Matrix Example	β
1 (WP)	1.77
1 (WP, $\underline{S}=\underline{S1}$)	3.86
2 (TD)	3.66
2 (TD, $\underline{S}=\underline{S1}$)	3.66
3 (MP, $\gamma=5.0$)	3.17
3 (MP, $\gamma=2.5$)	2.16
4 (PP)	3.57

TABLE 2
EFFECTIVE SAMPLE SIZE INCREASE
OF TOEPLITZ ESTIMATOR OVER
GENERALIZED ESTIMATOR, N=10

We infer from the values of Table 2 that, on the average, a sample size reduction of 3.12 is possible with the constrained Toeplitz estimator for N=10.

We now attempt to generalize detection performance behavior for Case 2 and $N > 10$; in doing this, we obtain a rather interesting result. The expected SIR of Equation (16) may be solved in closed form when the input data stochastic process is white (Example 1) and $\underline{S}=\underline{S1}$. Performing the calculations, we obtain,

$$E[SIR|\hat{R}_T] \Big|_{WP, \underline{S1}} \cong \underline{W}^* T R \underline{W}^* \left[\frac{\frac{N}{\pi(N)} N_S - N - \frac{7}{\pi(N)}}{\frac{N}{\pi(N)} N_S} \right] \quad (35a)$$

where

$$\pi(N) = \sum_{n=1}^N \frac{1}{n} \cong A + \ln N + 1/2N \quad (35b)$$

and $A = 0.577$ is Euler's constant. The harmonic series $\pi(N)$ exhibits a \sqrt{N} behavior for small N and is divergent with an asymptotic approach consistent with the natural logarithm. The result of Equation (35) is felt to hold for a wide variety of signal vectors whose psd's are relatively flat over the unconstrained bandwidth. Examination of the behavior of Equation (16) appears to indicate that the approximation to the series $\pi(N)$ is modified to $\pi'(N)$ in the following manner when the input data stochastic process is not white, but the input data vectors still contain a large number of degrees of freedom,

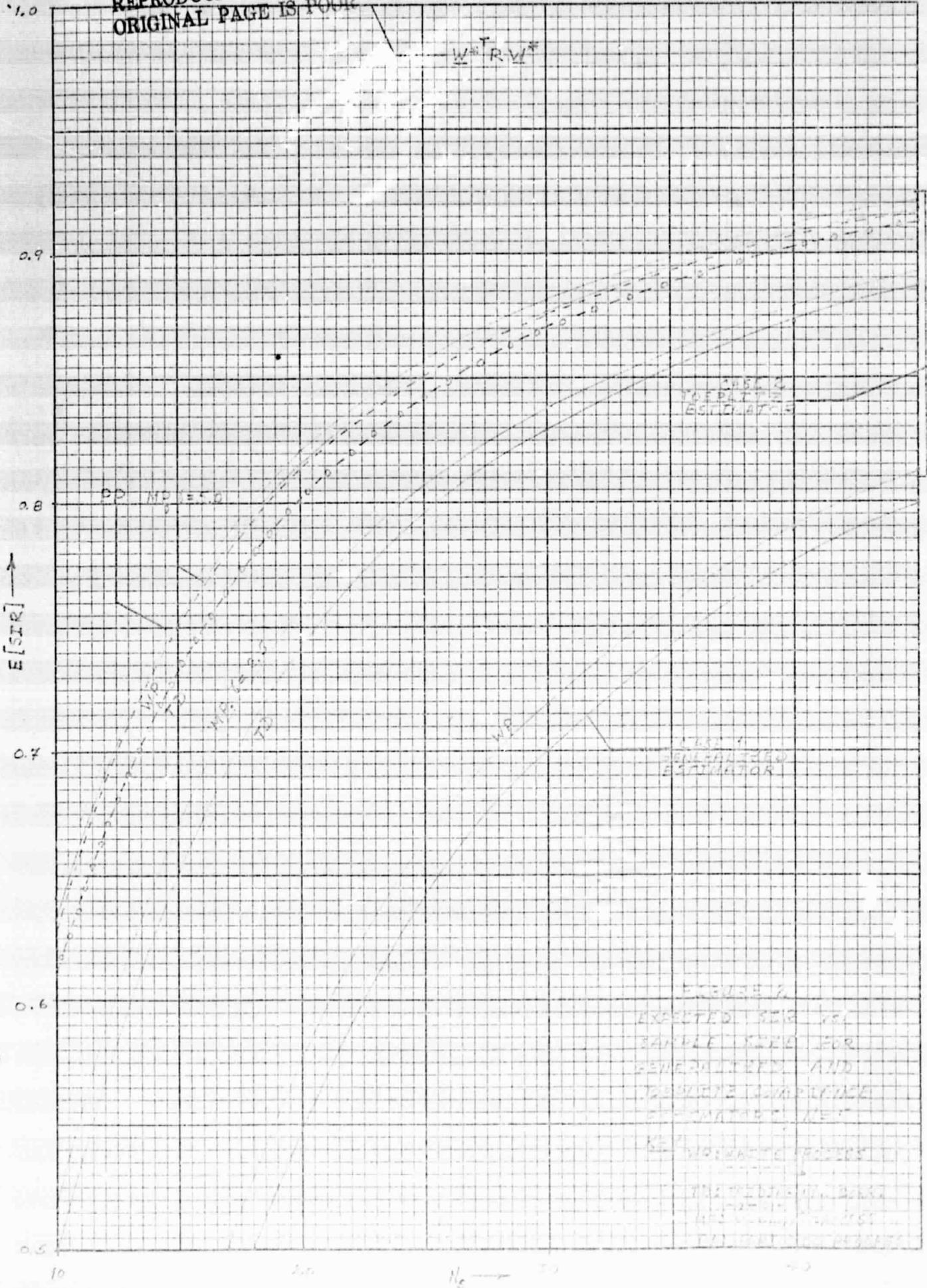
$$\pi'(N;B) \cong A + B \ln N + 1/2N \quad B \geq 1 \quad (36)$$

where $B\Delta$ is the input data stochastic process correlation time, e.g., $B=2$ for the Example 2 (triadiagonal) case. Now, comparing Equation (33) to Equations (35) and (36) we obtain a generalization for the multiplicative factor β , i.e.,

$$\begin{aligned} \beta &\cong \frac{N}{\pi'(N;B)} \\ &= \frac{N}{A + B \ln N + 1/2N} \end{aligned} \quad (37)$$

Several observations concerning Equation (37) are worth mentioning: (1) use of Equation (37) in Equation (33) with B selected appropriately provides results which compare favorably with the data of Figures 2 and 4 and Tables 1 and 2, and (2) the multiplicative factor β is divergent. The second point is very significant, indicating that a sample size $N_S < N$ for which the generalized estimator \hat{R}_G of Equation (10) is not even of full rank may be equivalent to an effective sample size N_S' which provides good detection performance when used with the constrained Toeplitz estimator \hat{R}_T of Equation (14). An example will aid in understanding these remarks and also serve to dispatch the second conjecture put forth in Section 1. For the short update experimental results mentioned in Section 1, the filter dimension $N=256$. The correlation time of the input data stochastic process is approximately $B\Delta = 8/5W$, where W denotes the bandwidth of the stochastic process; thus, there are approximately 32 uncorrelated time segments within each input data vector. The value of β as computed from Equation (37) is 5.70. A sample size $N_S = N = 256$ then provides us with an expected SIR of .827 times the optimum SIR (performance equivalent to 0.82 dB from the optimum SIR) from Equation (33). The real time required to gather the input data necessary to achieve this performance is 4 s, the gradient search algorithm averaging time utilized. We note that for $N_S = N = 256$, Equation (13) for the generalized estimator is way out of the range for which the approximation of Equation (8) is valid.

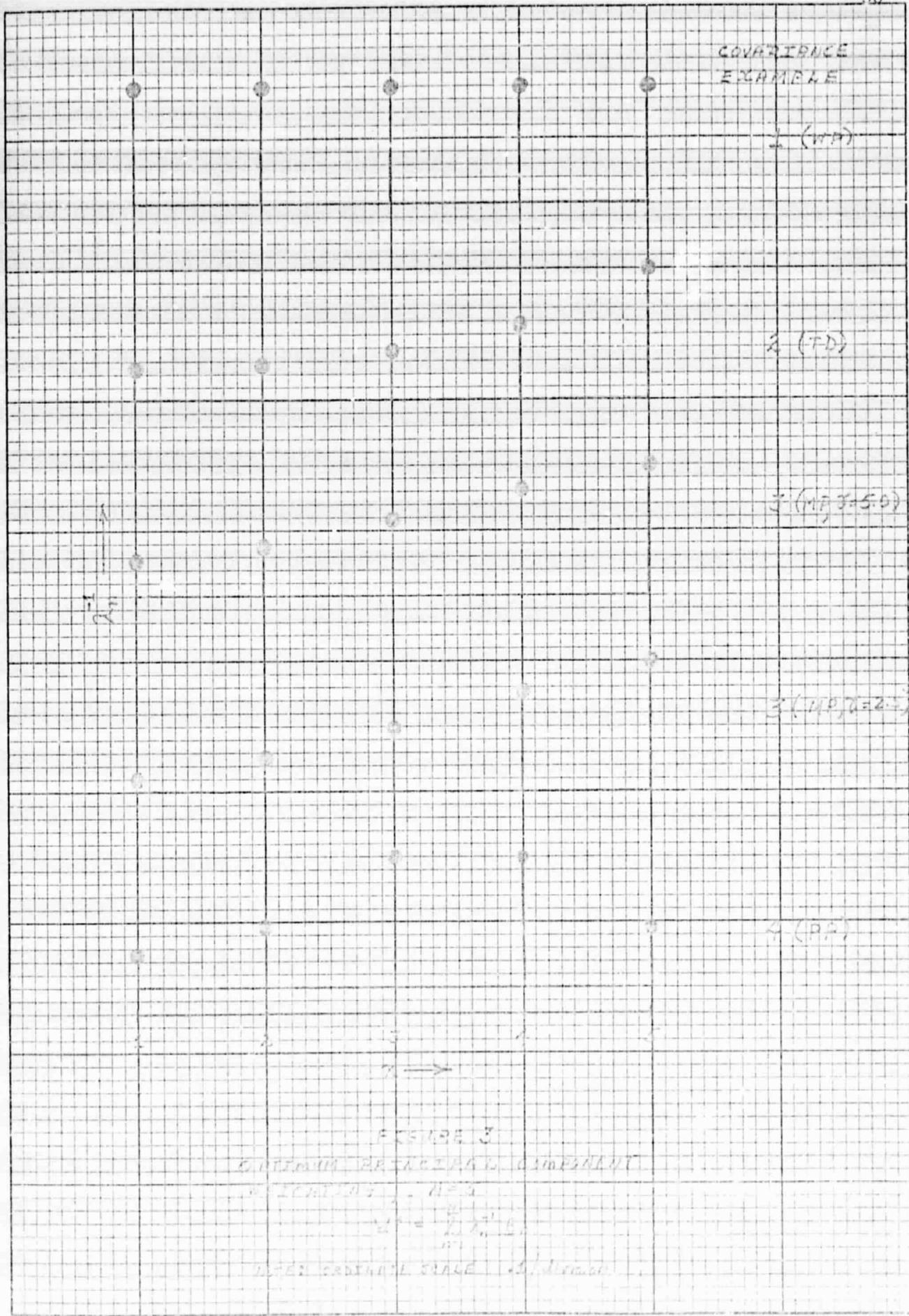
REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR



NO. 10 X 10 TO THE INCH 46 0703
7 X 10 INCHES
MADE IN U.S.A.
KEUFFEL & ESSER CO.

FIGURE 7
EXPECTED SIR FOR
SAMPLE SIZE FOR
GENERALIZED AND
ROBUST COMPLIANCE
ESTIMATES, $ME = 5.0$
KEY: $WTRW^*$ (TOP)
 $WTRW$ (MIDDLE)
 $ESTIMATES$ (BOTTOM)

COVARIANCE
EXAMPLE



λ_1

1 (WD)

2 (FD)

3 (MF 3-5.0)

3 (MF 3-5.0)

4 (PP)

λ_2

FIGURE 3

CONTINUUM PRINCIPAL COMPONENT

IDENTITY $H=3$

$$\lambda_i = \frac{1}{n} \sum_{j=1}^n \lambda_{ij}^2$$

WITH COORDINATE SCALE λ_i distance

K&E 10 X 10 TO THE INCH 46 0782
7 X 10 INCHES
MADE IN U.S.A.
KEUFFEL & ESSER CO.

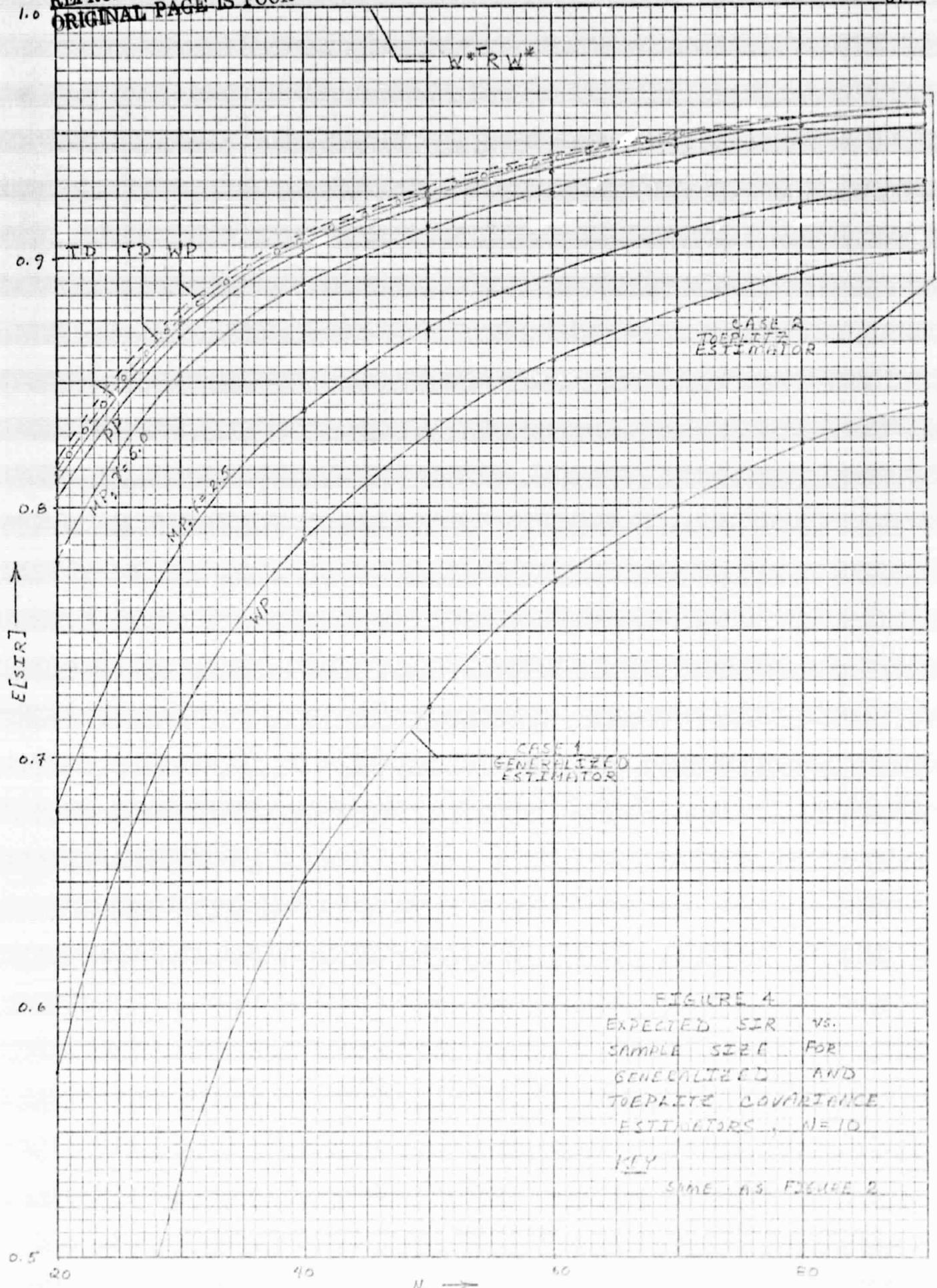
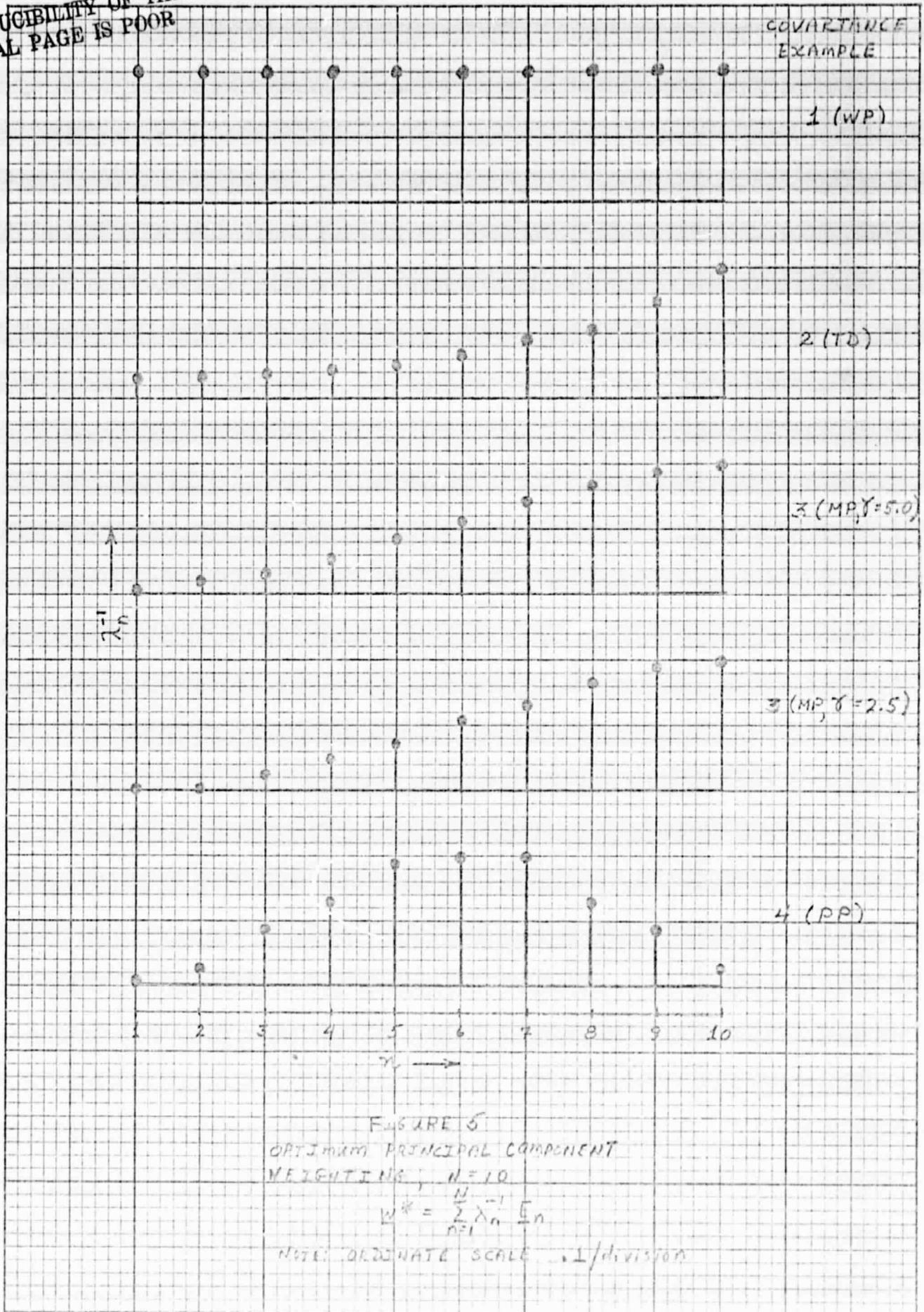


FIGURE 4
EXPECTED SIR VS.
SAMPLE SIZE FOR
GENERALIZED AND
TOEPLITZ COVARIANCE
ESTIMATORS, $N=10$
KEY
SAME AS FIGURE 2

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

COVARIANCE EXAMPLE



6. CONCLUDING REMARKS

We have examined two conjectures put forth in Section 1 pertaining to the suprisingly small sample size required to achieve good SIR performance with an adaptive signal detection filter. The first conjecture involves the possibility of convergence of the SIR well before convergence of the filter impulse response; the second, with the possibility that the mechanics of a certain adaptive filter algorithm are intelligent enough to capitalize on the structure of the input data stochastic process. Specifically, the structure referred to is the (finite) Toeplitz form of the input data covariance matrix.

The principal results of this work are as follows:

- (1) Equation (9), the expected SIR, conditioned on the estimator \hat{R} of the input data covariance matrix R ;
- (2) Equations (13) and (16), the expected SIR (under the stated assumptions) with the condition relieved for generalized \hat{R}_G and constrained Toeplitz \hat{R}_T estimators, respectively; and
- (3) Equations (33), (35), (36) and (37), synergistically providing a model of the behavior of the constrained Toeplitz estimator with dimension, sample size, true covariance matrix, and signal vector.

Use of the generalized estimator results in the expected SIR approaching the optimum SIR as $\approx (1 - \frac{N}{N_S} - \frac{4}{N_S})$, where N is the filter dimension and N_S is the sample size. In this case, $N_S \approx 5N + 20$ samples are needed to come within 1 dB of the optimum SIR; this is a relatively small number of samples, being considerably less than that required to estimate the $(N^2 + N)/2$ unique elements of a generalized covariance matrix. Use of the constrained Toeplitz estimator results in the expected SIR approaching the optimum SIR as $\approx [1 - \frac{\pi'(N;B)}{N_S} - \frac{7}{N N_S}]$ where $\pi'(N;B)$ behaves as $B \ln N$, $1 < B < N$, for large N . This result is derived from a combination of the data analyses presented in Section 5 for small-to-moderate N using the covariance matrix examples of Section 4, and heuristic asymptotic arguments. The importance of the results is evident. The constrained Toeplitz estimator operates with an "effective sample size" $N_S' = [\frac{N}{\pi'(N)}] N_S$ and therefore has the potential to provide high expected SIR at a sample size N_S for which the generalized estimator may provide exceedingly poor results. The reduction in required sample size reflects as a reduction in the amount of (real) time required for the adaptive algorithm performance to converge and, therefore, corresponds to an increase in the ability of the filter to accommodate realistic situations wherein the input data stochastic process is quasi-stationary.

The improved performance is naturally achieved with an adaptive algorithm, e.g., a first order, linearly constrained gradient search algorithm, operating at a so-called slow update rate, because the covariance matrix estimate formed is of constrained Toeplitz form. This observation suggests that the proposed scheme of Farden [9] for first forming a constrained Toeplitz form estimate of the input data covariance matrix and then inserting the estimate in the adaptive algorithm is unnecessary, if a slow update rate is employed. We do note that

there is a slight difference in the estimate utilized by a gradient search algorithm and that of Equation (14) employed for the calculations in this work in that the (first order) gradient search algorithm introduces exponential temporal weighting, whereas Equation (14) employs uniform temporal weighting. This difference is second order to the calculations presented here.

In many problems of interest, we deal with an adaptive filter which operates concurrently on a number, say K , of N -dimensional input data vectors $\underline{X}_k(i)$ $k=1, 2, \dots, K$, at the i^{th} sampling instant. The index on k may be a spatial reference, i.e., time records of length $N\Delta$ from K sensor locations are processed by the algorithm. If the (spatial) sequence of input data vectors are organized as in reference [4] and the resulting KN -dimensional input data vector stochastic process is assumed to be weakly stationary, then the input data covariance matrix is of (finite) block Toeplitz form. Extension of the results presented in this work to this case are immediate.

To broaden the areas of study that may benefit from this work, we mention the general area of pattern recognition. The problem discussed here corresponds to the situation where the form for the discriminant function is known and the samples are used to estimate the values of parameters of the classifier, i.e., the nonparametric problem. The results provided are also useful for relating sample size and dimensionality for the parametric problem, i.e., the form of the underlying probability distribution is known and the samples are used to estimate the values of parameters of the classifier [10].

REFERENCES

- [1] Widrow, B., P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive Antenna Systems", Proc. IEEE, Vol. 55, No. 12, Dec. 1967, pp. 2143 - 2159.
- [2] Griffiths, L., "A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays", Proc. IEEE, Vol. 57, No. 10, Oct. 1969, pp. 1696 - 1704.
- [3] Frost, O. L. III, "An Algorithm for Linearly Constrained Adaptive Array Processing", Proc. IEEE, Vol. 60, Aug. 1972, pp. 926 - 935.
- [4] Morgera, S. D., "Selective Spatial and Spectral Adaptive Processing of the Acoustic Field", Doctoral Dissertation, Division of Engineering, Brown University, June 1975.
- [5] Goodman, N. R., "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (an Introduction)", Annals of Math, Stat., Vol. 34, No. 1, 1963, pp. 152 - 177.
- [6] Grenander, U. and G. Szego, Toeplitz Forms and Their Applications, University of California, Berkeley, 1958.
- [7] Grenander, U. and M. Rosenblatt, Statistical Analysis of Stationary Time Series, John Wiley, New York, 1957.
- [8] Koopmans, L. H., The Spectral Analysis of Time Series, Academic Press, New York, 1974.

- [9] Farden, D., "Stochastic Approximation with Correlated Data",
Doctoral Dissertation, Engineering Department, Colorado State University,
Spring 1975.
- [10] Duda, R. O. and P. E. Hart, Pattern Classification and Scene Analysis,
John Wiley, New York, 1973.
- [11] Kanal, L. and B. Chandrasekaran, "On Dimensionality and Sample Size in
Statistical Pattern Classification", Proc. 1968 Nat. Elect. Conf., pp. 2 - 7.