Chapter 1

Reduced-rank intelligent signal processing with application to radar

J.S. Goldstein, J.R. Guerci and I.S. Reed

1.1 Introduction

The technologies associated with radar signal processing have developed and advanced at a tremendous rate over the past sixty years. This evolution is driven by the desire to detect more stealthy targets in increasingly challenging noise environments. Two fundamental requirements on signal processing have developed as advanced radar systems strive to achieve these detection goals: 1) The dimensionality of the signal space is increased in order to find subspaces in which the targets can be discriminated from the noise; and 2) The bandwidth of each of these dimensions is increased to provide the degrees of freedom and resolution that are needed to accomplish this discrimination when the competing noise and the target are in close proximity. To be more precise, radar has developed from having only a spatial dimension to the utilization of a Doppler frequency (or slow-time) dimension to combat monostatic clutter, to a signal frequency (or fast-time) dimension to defeat terrain scattered interference, to multiple polarization dimensions for target discrimination, etc. The number of degrees of freedom required to separate the target from nearby competing noise within any one of these signal dimensions (i.e., the resolution or bandwidth in that dimension) grows as the target becomes smaller and the noise becomes more challenging. The total number of degrees of freedom for the radar is then given by the Cartesian product of the degrees of freedom for each individual dimension.

As radar signal processing evolved, it quickly became clear that one needed an estimate of the noise environment in order to realize detectors that worked well in the real-world. This concept led simultaneously to the development of adaptive radar signal processing and adaptive constant false-alarm rate (CFAR) detectors. The theory of adaptive arrays [1, 2] was developed at a time when the spatial dimension was predominantly used alone. The theory of adaptive radar was next advanced to simultaneously apply adaptivity to the spatial and Doppler dimensions [3], introducing the popular field called space-time adaptive processing (STAP). These adaptive techniques for radar signal processing were based upon second-order statistics for wide-sense stationary (WSS) random processes. The idea that the noise was Gaussian, as well as independent and identically distributed (IID) over range, served as fundamental assumptions which were embedded in these theoretical developments. These same assumptions were used in the tools which evaluated the performance of radar systems.

The estimation of the noise environment, with the above assumptions, requires a training region composed of a number of samples which are at least on the order of twice the number of the radar's degrees of freedom[1] [4]. This famous "2N" rule means that, for N degrees of freedom, the Gaussian noise field estimation requires a minimum of $2N$ IID samples. Where does this training data, or sample support for noise field estimation, come from? In radar, one tries to detect a target within some specified range cell. If there are N total degrees of

[1] More formally, a training region consisting of at least twice the number of the radar's degrees of freedom is required to obtain a statistical estimate of the noise which results in an output signal-to-interference plus noise ratio that is within 3 dB of that obtained with the true statistics.

freedom, then a data-cube consisting of at least $2N+1$ range cells is processed coherently. The group of $2N$ or more range cells which exclude the test cell are termed auxiliary range cells. The auxiliary range cells are used to estimate the statistics of the noise within the test range cell, thereby providing a target-free training set. The statistical estimate of interest here is the covariance matrix, which contains all of the second-order information that is needed when the underlying assumptions are satisfied. This estimate of the noise statistics is then used to compute an adaptive weight vector, or Wiener filter, which maximizes the output signal-to-interference plus noise ratio (SINR). This Wiener filter is equivalent to a normalized colored-noise matched filter.

The advancement of sensor technology easily allows radars to be constructed with large numbers of degrees of freedom in the spatial, Doppler, fast-time and polarization dimensions. The new fundamental problem for radar signal processing is that this large number of degrees of freedom makes it impossible for the IID, WSS, second-order assumptions embedded in adaptive signal processing to be valid. A moderate STAP radar design includes a minimum of a few hundred degrees of freedom. For example, the DARPA Mountain Top radar uses 16 Doppler pulses and 14 receive channels for a total of $N=224$ degrees of freedom. This implies a requirement for at least 448 range cells that contain noise which is IID with respect to that present in the range cell of interest. The IID assumption accompanies a spatial ergodicity argument with respect to the stationarity of the noise. However, with the parameters given above, the range extent of the sample support would cover approximately 338 kilometers. Given the topography of the earth, it is unreasonable to assume homogeneity over a region of this size or larger, and therefore the stationarity assumption on the data received at the radar cannot, in general, be valid for advanced radars.

The proposed solution to this problem requires intelligent signal processing for both subspace compression and training region selection. It is necessary to greatly enlarge the signal space in order to find that subspace which permits target detection. This drives the requirement for large numbers of antenna elements, large sets of tapped-delay-lines (at the relevant frequency spacing), many polarization channels, etc. However, this signal space enlargement also drives the need for large sample support and, as a consequence, stresses or breaks the underlying statistical assumptions. Since the space is enlarged, the true noise subspace is generally overmodeled in order to guarantee that a subspace for detection can be found. It is therefore necessary to perform intelligent signal representation so that the smallest possible subspace that contains the majority (if not all) of the noise power can be determined. This representation must take into account some information about the manifold of the noise subspace that it is desirable to estimate rather than to blindly estimate the entire noise subspace. This representation would of necessity allow optimal compression of the noise subspace, permit both rapid convergence and tracking of the noise statistics, and reduce the size of the required data region for statistical sample support.

It is also necessary to introduce intelligent training methods in order to determine which data in the auxiliary training set is most similar, in some appropriate statistical measure, to that present in the range gate of interest. Finally, intelligent signal processing approach should also be capable of utilizing prior or additional knowledge to incorporate information about the structure of roads or other man-made objects and geospatial information such as the U.S. National Imagery and Mapping Agency's digital terrain elevation database (DTED).

This chapter introduces a new method of intelligent signal representation and compression. This theory is presented in an application-independent manner since the concepts are valid in nearly every statistical signal processing problem. The radar application is then revisited to demonstrate the principles developed herein.

1.2 Background

Consider the representation of discrete-time, wide-sense stationary signals, which is fundamental in the many applications of statistical detection and estimation theory. For the purposes of this work, the efficiency of a signal representation is evaluated by its ability to compact useful signal energy as a function of rank. This criterion is equivalent to optimal signal compression.

The multiple-signal problem is considered herein, where a non-white signal of interest is only observed in the presence of at least one other generally non-white process[2]. Signal processing for multiple signals, under the above conditions, is described within the general framework of the discrete-time, finite-impulse response Wiener filter. The Wiener filter is a fundamental component in the solution to virtually every problem that is concerned with the optimality of linear filtering, detection, estimation, classification, smoothing and prediction in the framework of statistical signal processing in the presence of stationary random processes. This same approach provides the least-squares solution for the processing of collected data either in a deterministic framework or by invoking some form of ergodicity.

The fundamental issue in signal representation and compression is the determination of an optimal coordinate system. It is well-known that the eigenvectors associated with the covariance matrix of an N-dimensional WSS signal provide the basis-set for the Karhunen-Lóeve expansion of that signal. The minimax theorem establishes that this set of eigenvectors represent the particular basis for an N-dimensional space which is most efficient in the energy sense. This autocorrelation-based energy maximization for a single process satisfies the stated representation criterion, and the eigenvectors form the best basis representation for this single process. If the eigenvectors are ordered to correspond with the magnitude of the eigenvalues in a descending manner, then this enumeration is termed ordering by principal-components. Optimal signal representation and compression as a function of rank (or dimension) is then obtained by a truncation of the principal-components. In other words, the rule for optimal rank M basis selection ($M < N$) and compression is to choose those M eigenvectors which correspond with the largest M eigenvalues of the observed

[2]This scenario is very common; even if a true white noise field is observed by a sensor, for example a radar or a sonar receiver, the output from the sensor is a filtered process which, in general, will not be white.

signal covariance matrix.

However, there are many statistical decision problems where the criterion of interest is more general. This fact is readily verified by considering the popular problems of detection, estimation, or any of the many other statistical signal processing applications. Here, for the problem to be non-trivial, there are a minimum of two additive non-white signal processes: the signal process of interest and a process of colored-noise. If one now speaks of signal representation or compression of one process, the solution must take both processes into account in order to determine an optimal basis.

The goal of signal representation and compression for detection and estimation is to find an optimal basis *without prior knowledge* of the inverse of the covariance matrix. Optimal basis selection allows for signal compression, rank reduction, and a lower computational complexity in order to obtain the Wiener solution. Reduced sample support for estimating statistics and faster convergence of filter parameters are also obtained with rank reduction if the Wiener filter is implemented in an adaptive manner. Note that there is a subtle information-theoretic idea embedded in this goal; if the covariance matrix inverse is known *a priori*, then so is the Wiener filter, and signal representation is irrelevant.

Previous work in the area of signal representation has centered around principal component analysis for the single process case [5, 6] or canonical correlation analysis for the multiple process case [7, 8, 9, 10]. The vector Wiener filter is unique because, while it is amongst the most common filtering structures, neither of these analysis tools applies in this case to the optimization of performance as a function of rank. The solution to the canonical correlation analysis degenerates into the vector Wiener filter itself, and therefore provides no insight into the selection of an optimal basis. It is demonstrated explicitly in this paper that the principal-components is not the correct enumeration of the eigenvectors to achieve optimal representation for the Wiener filtering problem. Once this fact is established, it is shown herein that the standard Karhunen-Lóeve decomposition no longer provides a solution to optimal basis selection, and that a new basis set must be derived that takes into account the presence of other signals; it is called here a generalized joint-process Karhunen-Lóeve transformation.

Previous attempts to solve the rank reduction problem for the vector Wiener filter only result in solutions which, at best, dictate the computation of the Wiener filter itself [11, 12, 13]; they do not provide a basis set for the vector Wiener filter problem. A new approach is now presented that provides an optimal basis set through the natural extension of the Karhunen-Lóeve transform (KLT) for the Wiener filter.

Classical Karhunen-Lóeve analysis is briefly reviewed next in Sect. 1.3. The vector Wiener filtering model is then introduced in Sect. 1.4. The necessary modifications to the KLT are addressed in Sect. 1.5 to obtain an optimal Wiener filter with an eigenvector basis. In Sect. 1.6 a new method is developed to obtain an optimal basis without having the need for knowledge of the covariance matrix, its inverse, or the Wiener filter. Concluding remarks and a summary are presented in Sect. 1.8.

1.3 Karhunen-Lóeve Analysis

This section presents a review of the Karhunen-Lóeve transformation and signal expansion. These preliminaries set the stage for an analysis of signal representation and compression when multiple signals are present and the goal is to perform detection, estimation, classification or prediction of a signal of interest.

1.3.1 The Karhunen-Lóeve Transformation

Consider an N-dimensional, complex, WSS signal $\mathbf{x}_0$ with an $N \times 1$ mean-vector $\boldsymbol{\mu}_x$ (assumed without loss in generality to be the zero-vector) and a nonnegative definite, Hermitian, $N \times N$ covariance matrix $\mathbf{R}_{x_0}$. Let the covariance matrix $\mathbf{R}_{x_0}$ be represented by its Karhunen-Lóeve transformation,

$$\mathbf{R}_{x_0} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H = \sum_{i=1}^{N} \lambda_i \mathbf{v}_i \mathbf{v}_i^H, \tag{1.1}$$

where $(\cdot)^H$ denotes the complex Hermitian transpose operator, the $N \times N$ matrix $\mathbf{V}$ is composed of the N-unitary eigenvectors $\{\mathbf{v}_i\}_{i=1}^N$ and the diagonal matrix $\mathbf{\Lambda}$ is composed of the corresponding eigenvalues $\{\lambda_i\}_{i=1}^N$. It is assumed that the eigenvectors are ordered in a descending manner in accordance with the magnitude of the corresponding eigenvalues and, for convenience, that all of the eigenvalues are distinct.

The KLT of the covariance matrix $\mathbf{R}_{x_0}$, with the assumptions presented above, yields N orthonormal eigenvectors $\mathbf{v}_i$. Now denote the complex N-dimensional space spanned by the columns of $\mathbf{R}_{x_0}$ as $\mathcal{C}^N$. Then these eigenvectors form a basis for the space $\mathcal{C}^N$, and any vector $\mathbf{x}_0 \in \mathcal{C}^N$ can be represented by a linear combination of any basis vectors for $\mathcal{C}^N$.

1.3.2 The Karhunen-Lóeve Expansion

The Karhunen-Lóeve expansion of the N-vector $\mathbf{x}_0$ is obtained by its representation in terms of the basis generated by the eigenvectors of $\mathbf{R}_{x_0}$; that is,

$$\mathbf{x}_0 = \sum_{i=1}^{N} \alpha_i \mathbf{v}_i, \quad \text{where} \quad \alpha_i = \mathbf{v}_i^H \mathbf{x}_0. \tag{1.2}$$

It is easily verified that

$$\mathbf{E}\left[\alpha_i\right] = 0 \;\forall i, \tag{1.3}$$

and

$$\mathbf{E}\left[\alpha_i \alpha_j^*\right] = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}, \tag{1.4}$$

where $(\cdot)^*$ is the complex conjugation operator.

A k-dimensional subspace $\mathcal{C}^k \subset \mathcal{C}^N$ is formed by any arbitrary $\binom{N}{k}$ collection of basis vectors for the space spanned by the columns of $\mathbf{R}_{x_0}$. The k-dimensional principal-components subspace is defined to be that subspace spanned by the k principal eigenvectors and denoted $\mathcal{C}_{pc}^k$. A new reduced-rank N-dimensional vector, denoted $\mathbf{z}_0$, is given by the truncated series representation of $\mathbf{x}_0$ in (1.2) using only the k principal-components,

$$\mathbf{z}_0 = \sum_{j=1}^{k} \alpha_j \mathbf{v}_j. \tag{1.5}$$

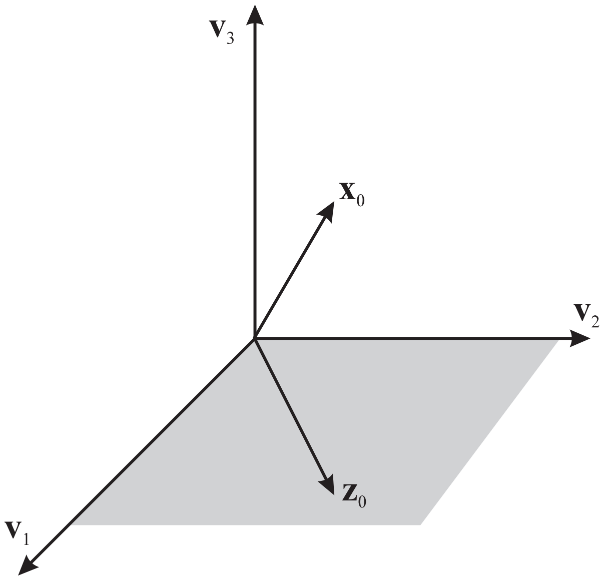

Figure 1.1: The representation of the vector $\mathbf{x}_0$ in $\Re^3$ and its projection $\mathbf{z}_0$ in $\Re^2$.

Note that in the k-dimensional subspace $\mathcal{C}^k_{pc}$, this same vector has the k-dimensional representation,

$$\mathbf{z}_0 = \mathbf{V}^H_{pc}\mathbf{x}_0, \tag{1.6}$$

where $\mathbf{V}_{pc}$ is the $N \times k$ matrix composed of the k principal eigenvectors[3]. Finally, from (1.5) and (1.6), it is seen that $\mathbf{z}_0$ is the projection of $\mathbf{x}_0$ onto the k-dimensional principal-components subspace $\mathcal{C}^k_{pc} \subset \mathcal{C}^N$, and therefore represents the compression of the N-dimensional vector to k coefficients. This geometrical relationship is depicted in Fig. 1.1 for $\Re^3$, the 3-dimensional vector space of real numbers. The principal-components method of compression and representation is optimal for a single signal in the sense that it provides the best representation of the full-rank space in terms of autocorrelation energy retained as a function of rank [5, 6].

[3] For this reason, the terms dimension and rank are used interchangeably in the rest of this paper.

1.3.3 Implementing the KLT

The KLT of the covariance matrix associated with discrete-time random processes is most often calculated by a two-step process which yields the eigendecomposition of the covariance matrix $\mathbf{R}_{x_0}$. The first step is a tridiagonalization, achieved through the use of Householder reduction. The second step is a QR, zero-chasing, iterative method which completes the diagonalization upon convergence [14, 15].

To visualize this process, consider the following $N \times N$ Hermitian covariance matrix,

$$\mathbf{R}_{x_0} = \begin{bmatrix} r_{1,1} & r_{2,1}^* & r_{3,1}^* & \cdots & r_{N,1}^* \\ r_{2,1} & r_{2,2} & r_{3,2}^* & \cdots & r_{N,2}^* \\ \vdots & & & \ddots & \vdots \\ r_{N,1} & r_{N,2} & r_{N,3} & \cdots & r_{N,N} \end{bmatrix}. \tag{1.7}$$

Define the (N-1)-dimensional vector $\mathbf{s}_1$ to be the first column of $\mathbf{R}_{x_0}$ excluding the top element,

$$\mathbf{s}_1 = \begin{bmatrix} r_{2,1} & r_{3,1} & \cdots & r_{N,1} \end{bmatrix}^T. \tag{1.8}$$

Next the vector $\mathbf{h}_1$ is given by

$$\mathbf{h}_1 = \frac{\mathbf{s}_1}{||\mathbf{s}_1||}, \tag{1.9}$$

where $||(\cdot)||$ represents the Euclidean norm and an orthonormal (N-2)$\times$(N-1) matrix $\mathbf{B}_1$ is computed such that its nullspace is $\mathbf{h}_1$. Then the first Householder reflection is given by the $N \times N$ unitary matrix operator $\mathbf{T}_1$,

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0}_{N-1,1}^H \\ \mathbf{0}_{N-1,1} & \begin{bmatrix} \mathbf{h}_1^H \\ \mathbf{B}_1 \end{bmatrix} \end{bmatrix}, \tag{1.10}$$

which initializes the tridiagonalization. The notation $\mathbf{I}_k$ is used to represent the rank-k identity matrix and $\mathbf{0}_{p,k}$ is the $p \times k$ zero matrix. The first-stage quadratic transformation of the covariance is then given by

$$\mathbf{T}_1\mathbf{R}_{x_0}\mathbf{T}_1^H = \begin{bmatrix} r_{1,1} & \tilde{r}_{2,1}^* & 0 & \cdots & 0 \\ \tilde{r}_{2,1} & \tilde{r}_{2,2} & \tilde{r}_{3,2}^* & \cdots & \tilde{r}_{N,2}^* \\ 0 & \tilde{r}_{3,2} & \tilde{r}_{3,3} & \cdots & \tilde{r}_{N,3}^* \\ \vdots & & & \ddots & \vdots \\ 0 & \tilde{r}_{N,2} & \tilde{r}_{N,3} & \cdots & \tilde{r}_{N,N} \end{bmatrix}, \tag{1.11}$$

Table 1.1: Unitary KLT Decomposition Recursion

$\mathbf{R}_t = \mathbf{R}_{x_0}$
$\mathbf{h} = \frac{\mathbf{R}_t(2:N,1)}{\lvert\lvert\mathbf{R}_t(2:N,1)\rvert\rvert}$
for k=1:N-2
$\quad\mathbf{B} = null(\mathbf{h})$
$\quad\mathbf{T}_k = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k,N-k} \\ \mathbf{0}_{N-k,k} & \begin{bmatrix} \mathbf{h}^H \\ \mathbf{B} \end{bmatrix} \end{bmatrix}$
$\quad\mathbf{R}_t = \mathbf{T}_k \mathbf{R}_t \mathbf{T}_k^H$
$\quad\mathbf{h} = \frac{\mathbf{R}_t(k+2:N,k+1)}{\lvert\lvert\mathbf{R}_t(k+2:N,k+1)\rvert\rvert}$
end

and it is easily verified from (1.7)-(1.11) that $\tilde{r}_{2,1} = ||r_{2,1}||$.

The above process is continued by defining the (N-2)-dimensional vector $\mathbf{s}_2$ to be the second column of the covariance matrix $\mathbf{T}_1\mathbf{R}_{x_0}\mathbf{T}_1^H$ excluding the top two elements. The vector $\mathbf{h}_2$, the orthonormal matrix $\mathbf{B}_2$ and the unitary matrix $\mathbf{T}_2$ are then calculated in a manner analogous to (1.9) and (1.10). This second stage replaces all but the first element of $\mathbf{s}_2$ with zeros. This iteration is repeated and at stage N-1 the covariance matrix is tridiagonalized by the product of the N-1 unitary operators $\mathbf{T}_k$. The algorithm which generates the tridiagonal covariance matrix $\mathbf{R}_t$ is described in Table 1.1.

Note that this operation produces a representation of the signal powers at different lags on the main-diagonal and the cross-correlations between the lags, compressed into positive scalars, on the upper and lower-diagonals. The tridiagonal matrix $\mathbf{R}_t$ is diagonalized by an iterative procedure using the QR algorithm which guarantees a unitary transfer function and results in the KLT due to the uniqueness theorem for unitary matrices. One popular version of the QR al-

Table 1.2: Unitary KLT QR Synthesis Recursion

$\mathbf{\Lambda}_0 = \mathbf{R}_t$
for k=1,2,...
$\quad\quad \mathbf{\Lambda}_{k-1} = \mathbf{Q}_k \mathbf{R}_k$
$\quad\quad \mathbf{\Lambda}_k = \mathbf{R}_k \mathbf{Q}_k$
end

gorithm is shown in Table 1.2. The QR factorization is first applied to the tridiagonal covariance matrix to yield a unitary matrix $\mathbf{Q}_1$ and an upper triangular matrix $\mathbf{R}_1$. These resulting factors are then multiplied in reverse order to update the covariance matrix. It is easily verified that, under suitable conditions, the QR algorithm converges and the tridiagonal covariance matrix is diagonalized via a sequence of unitary similarity transformations,

$$\mathbf{\Lambda}_k = \mathbf{Q}_k^H \mathbf{\Lambda}_{k-1} \mathbf{Q}_k. \tag{1.12}$$

Convergence is declared when the magnitude of the off-diagonal elements are within some acceptable tolerance of zero. Finally, let $\mathbf{T}$ represent the product of the $\mathbf{T}_k$ generated in Table 1.1 and $\mathbf{Q}$ represent the product of the $\mathbf{Q}_k$ generated in Table 1.2. Then the KLT is computed as follows,

$$\begin{aligned} \mathbf{\Lambda} &= \mathbf{Q}_N^H \cdots \mathbf{Q}_1^H \mathbf{T}_N \cdots \mathbf{T}_1 \mathbf{R}_{x_0} \mathbf{T}_1^H \cdots \mathbf{T}_N^H \mathbf{Q}_1 \cdots \mathbf{Q}_N \\ &= \mathbf{Q}^H \mathbf{T} \mathbf{R}_{x_0} \mathbf{T}^H \mathbf{Q} \\ &= \mathbf{V}^H \mathbf{R}_{x_0} \mathbf{V}, \end{aligned} \tag{1.13}$$

where $\mathbf{V} = \mathbf{T}^H \mathbf{Q}$ is the unitary eigenvector matrix and the sequence $\mathbf{\Lambda}_k$ converges to the diagonal matrix of eigenvalues $\mathbf{\Lambda}$.

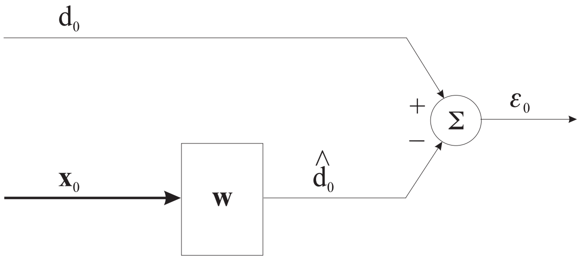

Figure 1.2: The Wiener filter.

1.4 The Multiple Signal Model and Wiener Filtering

In a more general setting, the classical problems of statistical signal processing are concerned with joint signal representation and compression. These problems are often characterized by the Wiener filter, depicted in Fig. 1.2, where there are two processes present. The N-dimensional process $\mathbf{x}_0$ is now considered to be the composite of potentially many processes, while d_0 is a scalar process which is correlated with $\mathbf{x}_0$.

The process d_0, normally termed a desired process, is representative of a signal of interest in some way, and the goal is to estimate d_0 from $\mathbf{x}_0$. For example, in the radar and sonar detection problem the "desired" signal is usually the output of a beamformer or matched-field processor, and the observed data vector $\mathbf{x}_0$ consists of data different from this signal that is received at a sensor array. In the communications application of multiuser detection and demodulation, the process d_0 may be generated by a known correlation with the signal of interest such as the code of a user in a CDMA wireless network. As a final example, in classification for automatic target recognition the desired signal may be a template image from training data, while the differing observed data is an image received by the fielded sensor. In general, the mechanism which generates the reference signal is application specific; however nearly every problem in linear statistical signal processing may be represented by the use of this model.

The problem at hand is the determination of optimal signal representation and compression for the observed data vector $\mathbf{x}_0$. The resulting basis will still span $\mathcal{C}^N$, the space spanned by the columns of $\mathbf{R}_{x_0}$, but the energy, which must be compactly represented, is now the estimation energy $\mathbf{E}[|\hat{d}_0|^2]$ and not the autocorrelation energy described by $\mathbf{R}_{x_0}$.

The signal model, introduced in Sect. 1.3 is extended here to two jointly-stationary zero-mean processes. The process d_0 is a zero-mean scalar process with variance $\sigma^2_{d_0}$ and $\mathbf{x}_0$ is an observed N-dimensional signal, which itself may be a composite random process, with covariance $\mathbf{R}_{x_0}$. The filter to be defined, $\mathbf{w}$, processes the observed-data to form an estimate of the desired signal $\hat{d}_0 = \mathbf{w}^H \mathbf{x}_0$. The error process ε_0,

$$\varepsilon_0 = d_0 - \mathbf{w}^H \mathbf{x}_0, \tag{1.14}$$

is the signal which characterizes the performance of the filter, and the optimal Wiener filter minimizes the mean-square value of this error signal.

The minimum mean-square error (MMSE) optimization criterion is formally stated as follows:

$$\min_{w} \ \mathbf{E}[\,|\varepsilon_0|^2\,] = \min_{w} \ \left\{\sigma^2_{d_0} - \mathbf{w}^H \mathbf{r}_{x_0 d_0} - \mathbf{r}^H_{x_0 d_0}\mathbf{w} + \mathbf{w}^H \mathbf{R}_{x_0}\mathbf{w}\right\}, \tag{1.15}$$

where the N-vector $\mathbf{r}_{x_0 d_0} = \mathbf{E}[\mathbf{x}_0 d_0^*]$ is the cross-correlation between the processes d_0 and $\mathbf{x}_0$. The well-known solution to (1.15) is the Wiener filter, which is computed as follows:

$$\mathbf{w} = \mathbf{R}^{-1}_{x_0}\mathbf{r}_{x_0 d_0}. \tag{1.16}$$

The MMSE is calculated by substituting (1.16) into the expression for the mean-square value of the error,

$$\xi_0 = \sigma^2_{d_0} - \mathbf{w}^H \mathbf{R}_{x_0}\mathbf{w} \tag{1.17}$$

where the far right-hand term in (1.17) is the optimal value of the estimation energy, given by

$$\mathbf{E}[|\hat{d}_0|^2] = \mathbf{w}^H \mathbf{R}_{x_0}\mathbf{w} = \mathbf{r}^H_{x_0 d_0}\mathbf{R}^{-1}_{x_0}\mathbf{r}_{x_0 d_0}. \tag{1.18}$$

It is now apparent that the optimal basis vector to select is the Wiener filter. However, this result requires the solution to the full-rank problem and provides no insight with respect to the selection of an optimal basis set when the Wiener filter (or the inverse of the covariance matrix) is unknown *a priori*. The desired information-theoretic goal is to achieve optimal basis selection, and therefore optimal compression, without complete prior knowledge. Previous attempts to solve this problem for the vector Wiener filter only resulted in a solution which, at best, dictated the computation of the Wiener filter itself [11, 12, 13]. This previous work is also related to canonical correlation analysis, the solution to which degenerates into the vector Wiener filter in this case as well (see, for example, [7, 8, 9, 10]). Therefore these previous attempts to extend the KLT and the related canonical correlation analysis are not further discussed. Instead, in Sect. 1.5, the necessary modifications to the KLT are addressed to obtain optimality with an eigenvector basis, which is the natural extension to the KLT previously sought by other researchers. A method to obtain an optimal basis without complete knowledge of the covariance matrix, its inverse, or the Wiener filter is then developed in Sect. 1.6.

1.5 The Signal-Dependent KLT for Statistical Signal Processing

The role of the KLT and the principal-components in signal representation and compression is now examined within the framework of Wiener filtering for statistical signal processing, detection and estimation. The result developed in this section demonstrates that the principal-components procedure is a suboptimal basis selection rule for detection and estimation problems. This fact motivates the question of whether the KLT and its eigenvector basis are optimal for these statistical signal processing applications, and the answer to this question serves as the topic of Sect. 1.6.

1.5.1 The KLT and Principal-Components

A low-dimensional numerical example serves as a valuable tool for understanding the behavior of the KLT and the principal-components in the statistical signal processing framework. Consider a simple example in $\Re^2$, where there is a 2-dimensional observed-data vector $\mathbf{x}_0$ with covariance matrix $\mathbf{R}_{x_0}$,

$$\mathbf{R}_{x_0} = \begin{bmatrix} 10 & 4 \\ 4 & 10 \end{bmatrix}, \tag{1.19}$$

and a desired signal d_0 with a variance $\sigma^2_{d_0} = 10$. The two processes, d_0 and $\mathbf{x}_0$, are assumed to be zero-mean, jointly stationary, and correlated. The cross-correlation between the two processes is given by the vector $\mathbf{r}_{x_0 d_0}$,

$$\mathbf{r}_{x_0 d_0} = \begin{bmatrix} 9 \\ 1 \end{bmatrix}. \tag{1.20}$$

The Wiener-Hopf equation in (1.16) may also be expressed in the form,

$$\mathbf{R}_{x_0}\mathbf{w} = \mathbf{r}_{x_0 d_0}, \tag{1.21}$$

which explicitly demonstrates that the optimal filter $\mathbf{w}$ is that particular linear combination of the columns of $\mathbf{R}_{x_0}$ which yields $\mathbf{r}_{x_0 d_0}$. It also means that $\mathbf{w}$ is in the space spanned by the columns of $\mathbf{R}_{x_0}$, and therefore that efficient signal representation is still equivalent to optimal basis selection in $\mathcal{C}^N$. These facts indicate that the KLT be considered.

The KLT provides the eigendecomposition of the matrix $\mathbf{R}_{x_0}$:

$$\begin{aligned} \mathbf{R}_{x_0} &= \begin{bmatrix} 10 & 4 \\ 4 & 10 \end{bmatrix} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 14 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \frac{1}{2}\left(14\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} + 6\begin{bmatrix} 1 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & -1 \end{bmatrix}\right), \end{aligned} \tag{1.22}$$

which demonstrates that one eigenvalue is significantly greater than the other. The KLT takes into account this self-directional preference of the signal. The Rayleigh quotient [16],

$$\Psi = \mathbf{e}^H\mathbf{R}_{x_0}\mathbf{e} = \sum_{i=1}^{N} \lambda_i \mid \mathbf{v}_i^H\mathbf{e} \mid^2, \tag{1.23}$$

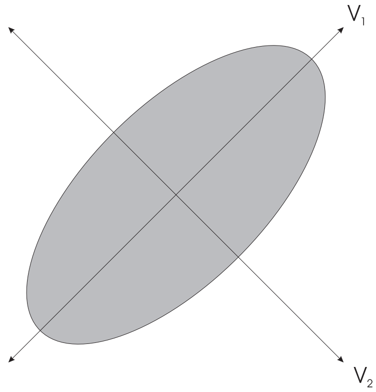

Figure 1.3: The Rayleigh quotient ellipsoid.

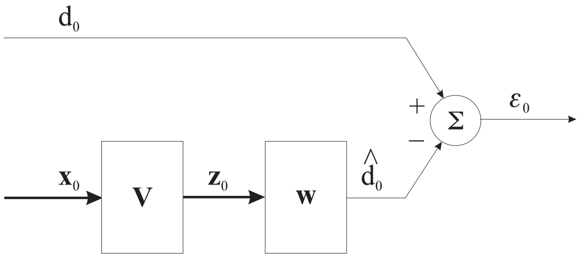

Figure 1.4: The transform domain Wiener filter.

mathematically describes this self-directional preference, where $\mathbf{e}$ is a unit-norm direction-varying vector. The Rayleigh quotient is maximized when $\mathbf{e} = \mathbf{v}_{max}$, the eigenvector corresponding with the largest eigenvalue. The Rayleigh quotient values form an ellipsoid, where the eigenvectors serve as the principal axes and the length of the ellipsoid along each eigenvector is proportional to the magnitude of the corresponding eigenvalue. For the example under consideration, this ellipsoid is depicted in Fig. 1.3. It is evident that the KLT provides the most efficient representation of the autocorrelation energy in the signal.

The question at hand, however, is the determination of the *best basis representation for* $\mathbf{x}_0$ *in terms of estimating* d_0. To explore whether the KLT and the principal-components are the best basis choice, consider preprocessing the observed process $\mathbf{x}_0$ with a filter composed of the eigenvectors of $\mathbf{R}_{x_0}$. This situation is depicted in Fig. 1.4, where the new process $\mathbf{z}_0$,

$$\mathbf{z}_0 = \mathbf{V}^H \mathbf{x}_0, \tag{1.24}$$

has a diagonal covariance matrix given by $\mathbf{\Lambda}$. Also, the cross-correlation between $\mathbf{z}_0$ and d_0 is now given by,

$$\mathbf{r}_{z_0 d_0} = \mathbf{V}^H \mathbf{r}_{x_0 d_0}. \tag{1.25}$$

The MMSE performance of the eigenvector-basis as a function of rank is now evaluated. The full-rank solution is identical regardless of basis representation since the performance measure is invariant to invertible transformations of the observed signal $\mathbf{x}_0$,

$$\xi_0 = \sigma_{d_0}^2 - \mathbf{r}_{z_0 d_0}^H \mathbf{R}_{z_0}^{-1} \mathbf{r}_{z_0 d_0} = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{R}_{x_0}^{-1} \mathbf{r}_{x_0 d_0}, \tag{1.26}$$

where $\mathbf{R}_{z_0} = \mathbf{V}^H \mathbf{R}_{x_0} \mathbf{V}$. However the results are different for this example in the rank-1 case, where one of the two eigenvectors which compose $\mathbf{V}$ is discarded. The principal-components algorithm states that the eigenvector corresponding with the largest eigenvalue should be retained, and that the eigenvector corresponding with the smaller eigenvalue should be discarded. Here, the largest eigenvalue magnitude corresponds with the first eigenvector, and the MMSE for this case is given by,

$$\xi_{PC} = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{v}_1 \lambda_1^{-1} \mathbf{v}_1^H \mathbf{r}_{x_0 d_0}. \tag{1.27}$$

The MMSE performance, converted to decibels, for the full-rank Wiener filter is 0.3951 dB and that for the rank-1 principal-components Wiener filter is 8.0811 dB. Thus, there is a loss of 7.6860 dB in reducing the rank from 2 to 1.

To answer the question of whether the principal-components is optimal for signal representation in this problem, first evaluate the MMSE which results if the smaller eigenvector is retained in this example:

$$\xi_{test} = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{v}_2 \lambda_2^{-1} \mathbf{v}_2^H \mathbf{r}_{x_0 d_0}, \tag{1.28}$$

yielding an MMSE of 6.6901 dB. Here the MMSE loss is approximately 1.4 dB less than that experienced by the principal-components selection; that is, a performance enhancement is obtained by selecting a different eigenvector than that indicated by the principal-components.

1.5.2 The Cross-Spectral Metric: An Intelligent and Signal-Dependent KLT

It is necessary to explain what, at first exposure, appears to be a breakdown in theory. Thus far it may seem strange that the selection of the principal-components is not optimal for signal representation and compression in detection and estimation problems. An insight is now developed to demonstrate why the principal-components is the wrong performance measure for these problems. It is also shown that the result of this example is not an anomaly. Finally, the necessary modification to the principal-components algorithm is provided so that optimal signal representation and compression is again gained if one restricts signal representation to an eigenvector basis: an intelligent KLT which takes all available statistical information into account to select the best eigenvectors.

The optimal solution for the performance measure of interest[4] in (1.17) is given by

$$\xi_0 = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{R}_{x_0}^{-1} \mathbf{r}_{x_0 d_0} = \sigma_{d_0}^2 - \sum_{i=1}^{N} \frac{\left| \mathbf{v}_i^H \mathbf{r}_{x_0 d_0} \right|^2}{\lambda_i}. \qquad (1.29)$$

The expression in (1.29) demonstrates that the presence of d_0 induces another directional preference through $\mathbf{r}_{x_0 d_0}$ in the spectrum of $\mathbf{x}_0$. This induced directional preference must be taken into account to achieve the optimal subspace selection for detection and estimation. Here it is evident that minimizing the performance measure requires maximizing the summation on the right-hand side of (1.29). This is accomplished by evaluating the square of a normalized projection of the cross-correlation vector upon the KLT basis vectors and, by (1.18), retaining those eigenvectors which maximize the estimation energy. Note that the result in the middle term of (1.29) may be interpreted as the difference between the desired signal variance and the mean-squared value of the whitened

[4]The MMSE performance measure is actually very general, for example under many conditions it is also equivalent to the maximum likelihood, maximum output signal-to-interference plus noise ratio and maximum mutual information performance measures, amongst others.

cross-correlation (or whitened matched-filter replica vector), $|\mathbf{R}_{x_0}^{-1/2}\mathbf{r}_{x_0 d_0}|^2$.

This new metric chooses those eigenvectors which correspond with the largest values of a ratio which takes into account both the directional preference of the process $\mathbf{x}_0$ and the impact of the correlated portion of the spectrum due to the process d_0, namely the projection of the cross-correlation along the i-th basis vector,

$$\frac{\left|\mathbf{v}_i^H \mathbf{r}_{x_0 d_0}\right|^2}{\lambda_i}. \tag{1.30}$$

Accordingly, consider a different enumeration of the eigenvectors, in a descending order based upon the largest magnitude of those N terms in (1.30). This ranking is called the cross-spectral metric [17], and keeping those k eigenvectors which maximize this cross-spectral metric provide the lowest mean-square error as a function of k, the dimension of the eigenvector basis. The subspace spanned by the eigenvectors selected by the cross-spectral metric is denoted $\mathcal{C}^k_{csm}$, and is in general different from the subspace $\mathcal{C}^k_{pc}$; especially for small k. The cross-spectral metric is the magnitude-squared value of a direction-cosine in $\mathcal{C}^N$ which measures the 'closeness' of the basis vector $\mathbf{v}_i$ to the cross-correlation vector [17]. Recall from (1.21) that the Wiener-Hopf equation yields the optimal coefficient vector for linearly combining the basis vectors to obtain this cross-correlation vector.

An interesting fact is realized by constraining the vector $\mathbf{w}$ in (1.18) to be unit-norm and allowing it to vary over all possible values. The quadratic form of the Rayleigh's quotient in (1.23) is again obtained, with the optimal solution not being an eigenvector but the Wiener filter in (1.16). This is another interpretation of the change in directional preference induced by the observation of multiple signals.

The example concerned with the selection of rank-1 Wiener filters is now revisited in order to complete the KLT analysis. The magnitude of the cross-spectral metric for the first eigenvector, corresponding with the larger eigenvalue of magnitude 14, is 3.5714. The magnitude of the cross-spectral metric for the second eigenvector, corresponding with the smaller eigenvalue of magnitude 6,

Table 1.3: The efficiency of KLT-based signal representation.

	$\lvert\lambda_i\rvert$	$\frac{\lvert \mathbf{v}_i^H \mathbf{r}_{x_0 d_0} \rvert^2}{\lambda_i}$	ξ_{PC}	ξ_{CSM}
i=1	14	3.5714	8.0811 dB	
i=2	6	5.3333		6.6901 dB

is 5.3333. Thus, $\xi_{test} = \xi_{CSM}$ in (1.28), and it is verified that the cross-spectral metric provides the better solution. A summary of the KLT analysis for this example is provided in Table 1.3.

1.6 Intelligent Signal Representation for Statistical Signal Processing

The above results demonstrate that the principal-components method is not the optimal metric for signal representation and compression in Wiener filtering problems. From a pedagogical perspective, it has now been demonstrated that the principal-components must be modified to take into account the directional preference induced by other signals to retain optimality (with respect to the KLT basis) for signal representation in detection and estimation problems. However, it is reasonable to argue that knowledge of the eigen-structure of $\mathbf{R}_{x_0}$ is equivalent to knowledge of $\mathbf{R}_{x_0}^{-1}$, and therefore a different solution is desired.

This section derives a generalization of the KLT which naturally includes this cross-spectral information to generate a new basis for signal representation with optimal properties when the covariance matrix inverse is unknown. In particular, the generalized joint-process KLT is demonstrated to be optimal with respect to a maximization of the 'cross-correlation' energy, subject to a unity response filter gain. This may be directly interpreted as a matched filter criterion which maximizes the signal response energy.

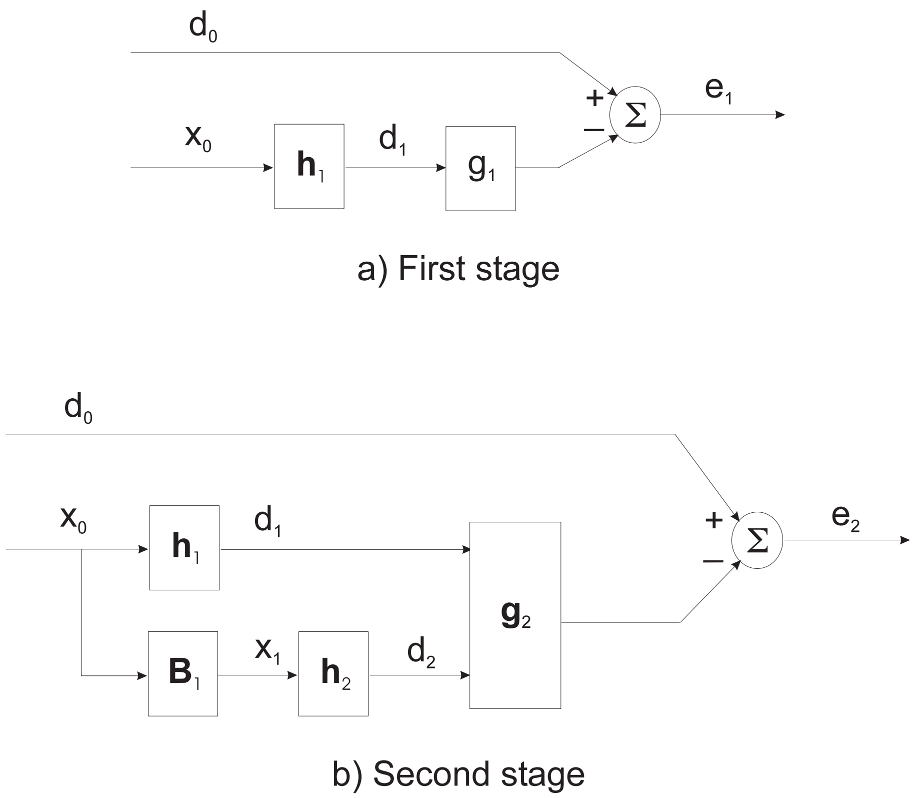

Figure 1.5: The first two stages of a matched filter decomposition.

1.6.1 A New Criterion for Signal Representation and its Implementation

A multistage Wiener filter, which derives a signal-dependent basis for compression [18, 19, 20, 21, 22, 23, 24, 25], is seen in this section to achieve the joint-process KLT from a "whitening" or innovations perspective [26]. The objective here is to sequentially optimize each rank-one basis selection so as to both further whiten the error residue and compress (compactly represent) the colored portion of the observed-data subspace, thereby achieving the desired signal-dependent rank-reduction.

Consider the selection of the first stage of the multistage Wiener filter depicted in Fig. 1.5a. At each stage, a rank-one basis is chosen to both reduce the mean-square error and compactly represent the correlated portion of the observed random process. In so doing, however, it is desired to avoid the obvious Wiener solution, which requires knowledge of the full-rank covariance matrix associated with the observed process $\mathbf{x}_0$. Note that, in general, this requirement also eliminates the usual KLT-based methods. However, in order to contribute to a reduction of the mean-square error, the basis selection process should produce signals which are, in some appropriate measure, maximally correlated with the desired process d_0. Therefore consider the following basis selection procedure: At stage 1 (see Fig. 1.5a), select that rank-one subspace $\mathbf{h}_1 \in \mathcal{C}^N$, which is maximally correlated with the desired signal d_0. More specifically, since $d_1 = \mathbf{h}_1^H \mathbf{x}_0$, the following optimization problem is obtained,

$$\max_{\mathbf{h}_1} \ \mathbf{E}\left[\left|\mathbf{h}_1^H \mathbf{x}_0 d_0^*\right|^2\right] \text{ subject to } \mathbf{h}_1^H \mathbf{h}_1 = 1. \tag{1.31}$$

A direct application of Schwarz's inequality with $\mathbf{R}_{x_0}$ unknown readily yields the selection, viz.,

$$\mathbf{h}_1 = \frac{\mathbf{r}_{\mathbf{x}_0 d_0}}{||\mathbf{r}_{\mathbf{x}_0 d_0}||}. \tag{1.32}$$

Note that, as desired, the solution does not depend on full knowledge of $\mathbf{R}_{x_0}$. Since $\mathbf{h}_1$ is not, in general, colinear with the Wiener solution (unless $\mathbf{R}_{x_0}$ is white), further mean-square error reduction (whitening) is possible by adding additional stages. Thus, consider the error residue, e_1, that results from the first stage basis selection:

$$e_1 = d_0 - g_1^* d_1 = d_0 - g_1^* \mathbf{h}_1^H \mathbf{x}_0 \tag{1.33}$$

where $g_1 \in \mathcal{C}$, is the optimal scalar Wiener weight (filter) for linearly estimating d_0 from d_1.

All of the information required for estimating d_0 from $\mathbf{x}_0$ which is not in the direction of $\mathbf{r}_{x_0 d_0}$ is contained in its (N-1)-dimensional nullspace. The second stage, therefore, is introduced by first filtering $\mathbf{x}_0$ with the operator $\mathbf{B}_1$, an (N-1)$\times N$ matrix whose rows form an orthonormal basis for the nullspace of $\mathbf{r}_{x_0 d_0}$. This operator is again easily realized by one stage of a Householder reflection. Now, $\mathbf{h}_2$ is chosen by an optimization which is identical in form to (1.31), thereby

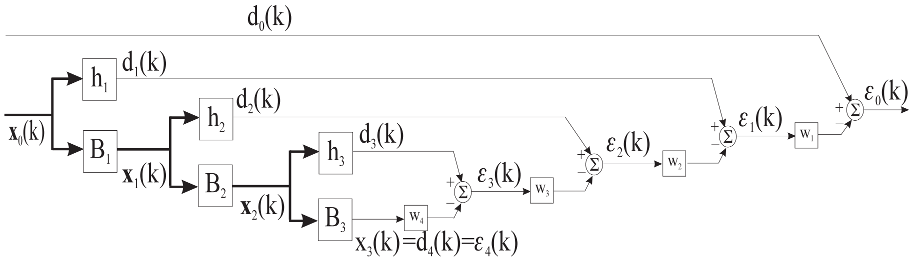

Figure 1.6: The multistage Wiener filters for N=4.

resulting in the selection,

$$\mathbf{h}_2 = \frac{\mathbf{r}_{x_1 e_1}}{||\mathbf{r}_{x_1 e_1}||}, \tag{1.34}$$

which is the matched filter for estimating the error residual from the remaining information available in $\mathbf{x}_0$ (see Fig. 1.5b). This decomposition is continued and a new basis is constructed that is based on maximal correlation.

In general, the objective at the i-th stage, $2 \leq i \leq N$-1, is to select that $\mathbf{h}_i \in \mathcal{C}^{N-i+1}$ which is maximally correlated with the residue e_{i-1} from all previous stages, i.e.,

$$\max_{\mathbf{h}_i} \mathbf{E}\left[\left|\mathbf{h}_i^H \mathbf{x}_{i-1} e_{i-1}^*\right|^2\right] \text{ subject to } \mathbf{h}_i^H \mathbf{h}_i = 1, \tag{1.35}$$

without knowledge of $\mathbf{R}_{x_0}$, where

$$e_i = d_0 - \mathbf{g}_i^H \mathbf{d}_i, \tag{1.36}$$

$\mathbf{d}_i = \begin{bmatrix} d_1 & d_2 & \ldots & d_i \end{bmatrix}^T \in \mathcal{C}^i$, and $\mathbf{g}_i \in \boldsymbol{\mathcal{C}}^i$ is the optimal weight vector in the transformed coordinates. For example, Fig. 1.5b shows filtering structure for the case where i=2.

A significant simplification in solving for $\mathbf{h}_i$, which eliminates the need for explicitly computing the weight vector $\mathbf{g}_i$ in the basis selection process, results by recognizing that $\mathbf{r}_{x_i e_i}$ is colinear with $\mathbf{r}_{x_i d_i}$. Thus,

$$\mathbf{h}_i = \frac{\mathbf{r}_{\mathbf{x}_i e_i}}{||\mathbf{r}_{\mathbf{x}_i e_i}||} = \frac{\mathbf{r}_{\mathbf{x}_i d_i}}{||\mathbf{r}_{\mathbf{x}_i d_i}||}, \tag{1.37}$$

and the equivalent filter structure can be represented in the form of a multistage Wiener filter [18, 19, 21]. This filter, depicted in Fig. 1.6 for the N=4 case, is readily interpreted as a nested chain of scalar Wiener filters following a bank of white-noise matched filters whose composite form solves the colored-noise matched filtering problem. Define $\xi_i = \mathbf{E}[|\varepsilon_i|^2]$ to be the error variance for the i-th stage and $\delta_N = r_{x_N d_N}$. Then the scalar weights in Fig. 1.6 are computed as follows:

$$w_i = \xi_i^{-1}\delta_i, \quad 1 \leq i \leq N, \tag{1.38}$$

where

$$\delta_i = \mathbf{h}_i^H \mathbf{r}_{x_{i-1}d_{i-1}}, \quad 1 \leq i < N. \tag{1.39}$$

Note that the error signal for the last stage ε_N is defined by

$$\varepsilon_N = d_N = x_{N-1}, \tag{1.40}$$

as depicted in Fig. 1.6.

It is now emphasized that the filterbank-based multistage Wiener filter in Fig. 1.6 does not require that the covariance matrix be estimated, inverted or decomposed. This structure and the vector Wiener filter yield identical solutions, implying that the multistage Wiener filter solves the colored-noise match filtering problem via a sequence of white noise match filters which do not require a matrix whitening operator or a matrix inversion. These white noise matched filters are determined by a sequence of correlation vectors, which may be estimated directly. A significant advantage of this structure, emphasized in Sect. 1.6.2, is that in general all stages of this filter need not be computed to obtain excellent performance.

1.6.2 A Generalized Joint-Process KLT: Nonunitary Diagonalization of the Covariance

The implementation of the KLT is now modified to motivate its generalization for the induced directional preference which occurs in detection and estimation. The Householder reduction of a covariance matrix to tridiagonal form [14, 15] is the first step for most algorithms which solve the eigenvalue problem, as noted in Sect. 1.3.3. The second step is the diagonalization of the tridiagonal covariance matrix in a unitary manner, yielding the eigenvalues and eigenvectors. The use of Householder reflectors to tridiagonalize the $N \times N$ covariance matrix $\mathbf{R}_x$ is again considered here, where a modification to the presentation in Sect. 1.3.3 is introduced to provide the desired directional preference. It is seen that this modification to the KLT results directly in the multistage Wiener filter.

The $N \times N$ reflector $\mathbf{T}_1$ is now defined to satisfy the following relation [21],

$$\mathbf{T}_1 \mathbf{r}_{x_0 d_0} = \left[\begin{array}{cccc} \delta_1 & 0 & \cdots & 0 \end{array} \right]^T, \tag{1.41}$$

where $\delta_1 = ||\mathbf{r}_{x_0 d_0}||$. This reflector is found by normalizing the cross-correlation vector,

$$\mathbf{h}_1 = \frac{\mathbf{r}_{x_0 d_0}}{||\mathbf{r}_{x_0 d_0}||}. \tag{1.42}$$

and solving the relation,

$$\mathbf{B}_1 = null(\mathbf{h}_1). \tag{1.43}$$

The matrix $\mathbf{T}_1$ in (1.41) is given by,

$$\mathbf{T}_1 = \left[\begin{array}{c} \mathbf{h}_1^H \\ \mathbf{B}_1 \end{array} \right], \tag{1.44}$$

and the operation of $\mathbf{T}_1$ on $\mathbf{x}_0$ yields a covariance matrix $\mathbf{R}_{x_1}$,

$$\mathbf{R}_{x_1} = \mathbf{T}_1 \mathbf{R}_{x_0} \mathbf{T}_1^H = \left[\begin{array}{cc} \sigma_{d_1}^2 & \mathbf{r}_{x_1 d_1}^H \\ \mathbf{r}_{x_1 d_1} & \mathbf{R}_{x_2} \end{array} \right], \tag{1.45}$$

where the scalar $\sigma_{d_1}^2 = \mathbf{h}_1^H \mathbf{R}_{x_0} \mathbf{h}_1$, the $(N-1) \times (N-1)$-dimensional matrix $\mathbf{R}_{x_1} = \mathbf{B}_1 \mathbf{R}_{x_0} \mathbf{B}_1^H$, and the $(N-1) \times 1$ vector $\mathbf{r}_{x_1 d_1} = \mathbf{B}_1 \mathbf{R}_{x_0} \mathbf{h}_1$. This operation is then repeated as described in Table 1.1 of Sect. 1.3.3 and shown here in Table 1.4. Note that the operations in Table 1.4 are identical to those in Sect. 1.3.3, with the exception that the first reflection (or pivot) uses the directional information from the desired process. Also note that available algorithms for Householder reflections can solve the tridiagonalization, initialized by the matrix multiplications in (1.11) and (1.45), without explicitly forming the product of the matrices $\mathbf{T}_i$.

This unitary reduction of the covariance matrix $\mathbf{R}_{x_0}$ results in the matrix $\mathbf{R}_t$ having a tridiagonal form. For example, the following decomposition is realized

Table 1.4: Modified Unitary Decomposition Recursion

$\mathbf{R}_t = \mathbf{R}_{x_0}$

$\mathbf{h} = \frac{\mathbf{r}_{x_0 d_0}}{||\mathbf{r}_{x_0 d_0}||}$

for k=1:N-1

$$\mathbf{B} = null(\mathbf{h})$$

$$\mathbf{T}_k = \left[\begin{array}{cc} \mathbf{I}_{k-1} & \mathbf{0}_{k-1,N-k+1} \\ \mathbf{0}_{N-k+1,k-1} & \left[\begin{array}{c} \mathbf{h}^H \\ \mathbf{B} \end{array} \right] \end{array} \right]$$

$$\mathbf{R}_t = \mathbf{T}_k \mathbf{R}_t \mathbf{T}_k^H$$

$$\mathbf{h} = \frac{\mathbf{R}_t(k+1:N,k)}{||\mathbf{R}_t(k+1:N,k)||}$$

end

for $N = 3$:

$$\mathbf{R}_{x_0} \stackrel{\mathbf{T}_1}{\Longrightarrow} \left[\begin{array}{cc} \sigma_{d_1}^2 & \mathbf{r}_{x_1 d_1}^H \\ \mathbf{r}_{x_1 d_1} & \mathbf{R}_{x_1} \end{array} \right] \stackrel{\mathbf{T}_2}{\Longrightarrow} \left[\begin{array}{ccc} \sigma_{d_1}^2 & \delta_2^* & 0 \\ \delta_2 & \sigma_{d_2}^2 & \delta_3^* \\ 0 & \delta_3 & \sigma_{d_3}^2 \end{array} \right] = \mathbf{R}_t, \qquad (1.46)$$

where $\mathbf{h}_1 = \frac{\mathbf{r}_{x_0 d_0}}{||\mathbf{r}_{x_0 d_0}||}$, $\mathbf{h}_2 = \frac{\mathbf{r}_{x_1 d_1}}{||\mathbf{r}_{x_1 d_1}||}$, $\mathbf{R}_{x_0} \in \mathcal{C}^{3\times 3}$, $\mathbf{R}_{x_1} \in \mathcal{C}^{2\times 2}$, $\mathbf{r}_{x_1 d_1} \in \mathcal{C}^{2\times 1}$, and the remaining variables are all scalars with $\delta_2 = ||\mathbf{r}_{x_1 d_1}||$, $\delta_3 = r_{x_2 d_2}$, and $\sigma_{d_3}^2 = R_{x_2}$.

The decomposition is equivalently represented as an analysis filterbank, shown in Fig. 1.6, which transforms the observed-data N-vector $\mathbf{x}_0$ to a new N-vector $\mathbf{d}_N$,

$$\mathbf{d}_N = \mathbf{T}\mathbf{x}_0 = \left[\begin{array}{cccc} d_1 & d_2 & \cdots & d_N \end{array} \right]^T, \qquad (1.47)$$

where $\mathbf{T}$ is the product of the matrices $\mathbf{T}_i$ from each stage [21]. The covariance matrix associated with vector $\mathbf{d}_N$ is tridiagonal, as shown in (1.46). Because each of the $\mathbf{T}_i$ are unitary, the matrix $\mathbf{T}$ is unitary, and the MMSE is not modified. However, the basis vectors are now ordered in a manner based upon the maximum correlation between the desired and observed processes. If this

Table 1.5: Nonunitary Synthesis Recursion

$\xi_N = \sigma^2_{d_N}$
for p=N-1:-1:1
$\xi_p = \sigma^2_{d_p} - \xi^{-1}_{p+1}\lvert\delta_{p+1}\rvert^2$
end

matrix is diagonalized using a unitary operator, then the KLT is obtained, the ordering is lost, and the basis is altered. It is therefore necessary to consider the diagonalization of the covariance matrix *in a manner which is not restricted to be unitary.*

In particular, to minimize the residual and generate an innovations process in cross-correlation, let the error-synthesis filterbank minimize the mean-square error recursively at each stage of the analysis filterbank. This nonunitary error-synthesis filterbank generates the recursion described in Table 1.5 and depicted in Fig. 1.6. The nonunitary diagonalization of $\mathbf{R}_{x_0}$ may be expressed in the following matrix equation:

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 & & & \\ & \xi_2 & & \\ & & \ddots & \\ & & & \xi_N \end{bmatrix} = \mathbf{QTR}_{x_0}\mathbf{T}^H\mathbf{Q}^H = \mathbf{E}^H\mathbf{R}_{x_0}\mathbf{E}, \tag{1.48}$$

where $\boldsymbol{\xi}$ is diagonal and the coefficients of the upper-triangular Gram matrix $\mathbf{Q}$ are found directly from the recursion in Table 1.5. Matrix $\mathbf{T}$ is always unitary when formed in this manner while matrix $\mathbf{Q}$ is not necessarily unitary. Therefore matrix $\mathbf{E} = \mathbf{T}^H\mathbf{Q}^H$ in (1.48) is generally nonunitary and invertible. The diagonalization of the covariance matrix in (1.48), using the algorithm described in Tables 1.4 and 1.5, is obtained in a numerically stable manner. Furthermore, the diagonalization of the tridiagonal covariance matrix in Table 1.5 is guaranteed to be complete in N-1 recursions, as opposed to the unknown and varying convergence property of the QR algorithm (Table 1.2) used in the KLT.

The rationale for developing this nonunitary diagonalization is that a correlation-based ranking of the subspaces can be imposed, where the unitary diagonalization would destroy the induced directional preference. Thus, reduced-rank signal representation or compression can be obtained by pruning the decomposition. This implies that the matrix $\mathbf{R}_t$ is reduced from $N \times N$ to $k \times k$ by discarding the last $N-k$ rows and columns (or never computing them). The recursion in Table 1.5 is then implemented with the value for N replaced by the value for k. Finally, the MMSE ξ_0 is found by applying the recursion in Table 1.5 one last time:

$$\xi_0 = \sigma_{d_0}^2 - \xi_1^{-1}|\delta_1|^2. \tag{1.49}$$

This solution has an interesting interpretation. The KLT results in a unitary diagonalizing transformation for one process which most compactly represents the modal or component signal energy for that process using the fewest spectral coefficients. The generalized joint-process KLT produces a nonunitary diagonalization, where the spectral coefficients are the *modal mean-square error values* between the observed composite process and the selected reference process. Recall that the motivation for the joint-process KLT is to redirect the original emphasis on the compaction of signal energy in $\mathbf{x}_0$ to the compaction of the estimation energy which minimizes the mean-square error. Therefore, this result seems intuitively satisfying, and it is especially worth noting the fact that an *a priori* knowledge of $\mathbf{R}_{x_0}^{-1}$ is never used.

Now return to the previous numerical example in $\Re^2$, with the 2×2 matrix $\mathbf{R}_{x_0}$ in (1.19), the 2×1 vector $\mathbf{r}_{x_0 d_0}$ in (1.20), and $\sigma_{d_0}^2 = 10$. The vectors $\mathbf{h}_1$ and $\mathbf{B}_1$ are then defined by (1.20), (1.42) and (1.43) to yield,

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{h}_1^H \\ \mathbf{B}_1 \end{bmatrix} = \frac{1}{\sqrt{82}} \begin{bmatrix} 9 & 1 \\ -1 & 9 \end{bmatrix}. \tag{1.50}$$

The correct tridiagonal form is then found after one application of the recursion:

$$\mathbf{R}_t = \mathbf{T}_1 \mathbf{R}_{x_0} \mathbf{T}_1^H = \begin{bmatrix} \sigma_{d_1}^2 & \delta_2^* \\ \delta_2 & \sigma_{d_2}^2 \end{bmatrix} = \frac{1}{82} \begin{bmatrix} 892 & 320 \\ 320 & 748 \end{bmatrix}. \tag{1.51}$$

Table 1.6: The MMSE (dB) for each Wiener filter.

$\xi_0 \in \Re^2$	$\xi_{JKLT} \in \Re^1$	$\xi_{CSM} \in \Re^1$	$\xi_{PC} \in \Re^1$
0.3951	3.9127	6.6901	8.0811

The diagonal matrix $\boldsymbol{\xi}$ is then given by,

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{bmatrix} = \begin{bmatrix} \sigma_{d_1}^2 - \frac{|\delta_2|^2}{\sigma_{d_2}^2} & 0 \\ 0 & \sigma_{d_2}^2 \end{bmatrix} = \frac{1}{82}\begin{bmatrix} 755.1 & 0 \\ 0 & 748 \end{bmatrix}. \tag{1.52}$$

It is simple in this case to verify (1.48),

$$\boldsymbol{\xi} = \mathbf{Q}\mathbf{T}_1\mathbf{R}_{x_0}\mathbf{T}_1^H\mathbf{Q}^H, \tag{1.53}$$

where the upper-triangular matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{bmatrix} 1 & -\frac{\delta_2^*}{\sigma_{d_2}^2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -0.4278 \\ 0 & 1 \end{bmatrix}, \tag{1.54}$$

is easily found from the recursion in (1.14) ,(1.16), (1.17) and Table 1.5. The MMSE is now computed by,

$$\xi_0 = \sigma_{d_0}^2 - \frac{|\delta_1|^2}{\xi_1}, \tag{1.55}$$

which yields 0.3951 dB in agreement with the previous full-rank MMSE result. Finally, the rank-k joint-process KLT (JKLT) chooses a rank-k subspace $\mathcal{C}_{jklt}^k$ with an associated MMSE ξ_{JKLT}. Here, k=1, and

$$\xi_{JKLT} = \sigma_{d_0}^2 - \frac{|\delta_1|^2}{\sigma_{d_1}^2}, \tag{1.56}$$

which yields an MMSE of 3.9127 dB. As shown in Table 1.6, this represents an improvement of nearly 3 dB over the cross-spectral metric and slightly over 4 dB compared to the principal-components.

1.6.3 Analysis of the JKLT

The basis vectors of the JKLT are determined as a function of both the self-directional preference due to the eigenvectors that are associated with the observed-data covariance matrix $\mathbf{R}_{x_0}$ and the induced directional preference due to the

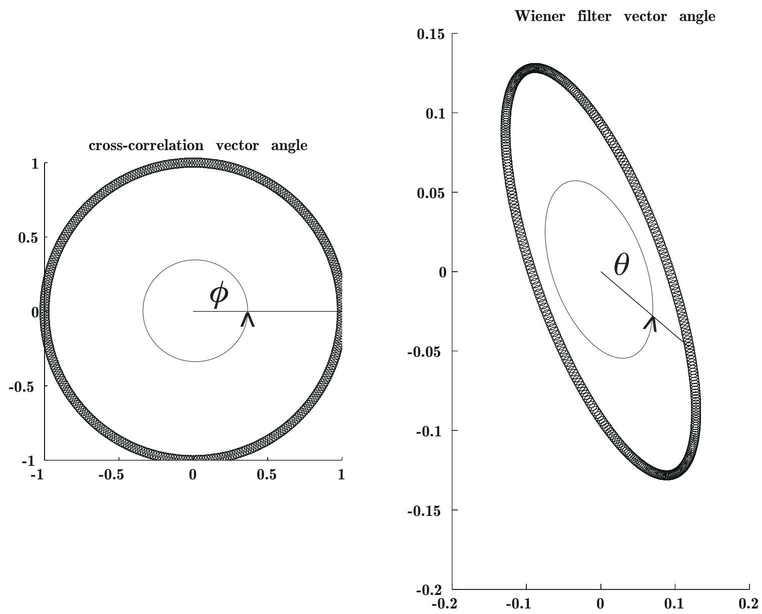

Figure 1.7: The angle representation of the cross-correlation and Wiener filter vectors in $\Re^2$.

effect of the cross-correlation $\mathbf{r}_{x_0 d_0}$. This fact implies the ability to perform low-rank subspace tracking in non-stationary signal environments and optimal reduced-rank Wiener filtering in stationary signal environments. Here optimality is with respect to the minimization of the average MMSE, as a function of rank, over all possible signal environments without knowledge of the full-rank Wiener solution (or of $\mathbf{R}_{x_0}^{-1}$).

To demonstrate these facts, consider again revisiting the numerical 2×2 example. Here, the eigenvectors in (1.22) are vectors at $45°$ and $-45°$ in the x-y plane. Now let $\mathbf{r}_{x_0 d_0}$ be a unit vector which varies from $0°$ to $360°$ in the plane. The resulting Wiener filter $\mathbf{w}$ maps out an ellipse as the cross-correlation vector varies counter-clockwise in the plane. These mappings are depicted in Fig. 1.7.

The optimal rank-2 Wiener filter and the different rank-1 solutions are compared as the statistics vary in Fig. 1.8, where the angle associated with $\mathbf{r}_{x_0 d_0}$ varies from $\phi = 0°$ to $\phi = 360°$. Note that this experiment tells the entire

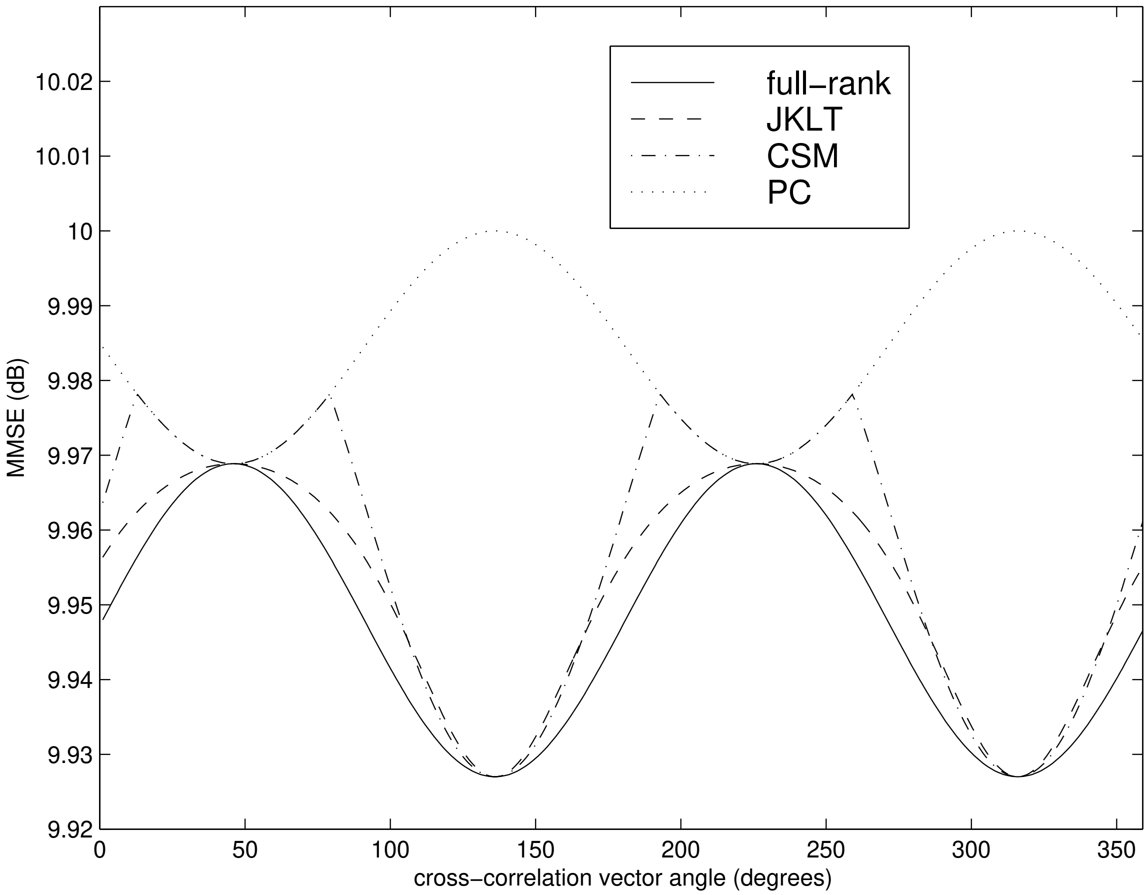

Figure 1.8: The performance as a function of statistical variation.

story for $\Re^2$; that is, the performance comparison is now made over all possible desired signals via the inclusion of all possible induced cross-correlation unit vectors.

The principal-components technique fixes the rank-1 basis vector as that eigenvector which corresponds with the largest eigenvalue. This choice results in poor performance except for in the small region where the Wiener filter is approximately colinear with that eigenvector. As seen in Fig. 1.8, the principal-components MMSE obtains the full-rank optimum only twice over 360°, exactly when the Wiener filter is colinear.

The cross-spectral metric introduces the ability to change eigenvectors depending upon the optimal Wiener solution. It is seen in Fig. 1.8 that the cross-spectral metric obtains the optimal solution twice as often, whenever the Wiener filter is colinear with either eigenvector. Thus the cross-spectral metric "switches" between the eigenvectors to optimize the MMSE.

Interestingly, both the principal-components and the cross-spectral metric require *a priori* knowledge of the eigenvectors. This is computationally on the same order as calculating $\mathbf{R}_{x_0}^{-1}$, and therefore the Wiener solution. However, neither of these techniques make very good use of this information. The JKLT has no such requirement for prior knowledge, instead only using the cross-correlation vector or its estimate. With this available information, a Wiener filter and the resulting MMSE for any rank can be computed efficiently. However, the merit of this method is realized in Fig. 1.8. It is seen that the JKLT represents the signal environment using a basis set which tracks the Wiener solution (or converges to it) very efficiently as a function of rank. Note that this is done without ever computing the full-rank Wiener solution. When the slope of the optimal weight vector dynamics goes to zero, corresponding to the dynamics slowing down, the JKLT becomes a unitary decomposition, the basis selection converges to the eigenvectors, and all of the methods obtain the optimal MMSE.

A few notes are now in order with respect to performance trends. First, the cross-spectral metric is demonstrated to always perform as well or better than the principal-components algorithm in terms of the resulting MMSE as a function of rank. The JKLT always performs as well or better than the principal-components and nearly always provides a better MMSE performance than the cross-spectral metric. The case where the cross-spectral metric may outperform the JKLT is over small regions where a relatively low-magnitude cross-correlation vector is contained in a subspace spanned by the eigenvectors that correspond with small eigenvalues. Here, the cross-spectral metric can take advantage of the additional information it has due to knowledge of the dominating self-directional preference. When this occurs, the cross-spectral metric can "switch" to a subspace spanned by the low-magnitude eigenvectors faster

Table 1.7: The approximate average MMSE loss (dB) for each Wiener filter.

$\int_{0^\circ}^{360^\circ} \lvert\xi_0(\omega) - \xi_{JKLT}(\omega)\rvert d\omega$	$\int_{0^\circ}^{360^\circ} \lvert\xi_0(\omega) - \xi_{CSM}(\omega)\rvert d\omega$	$\int_{0^\circ}^{360^\circ} \lvert\xi_0(\omega) - \xi_{PC}(\omega)\rvert d\omega$
3.58	6.68	30.00

than the JKLT can "learn" the correct subspace. Notice that this situation does not affect the performance of the algorithms along a principal-components subspace.

Finally, the performance from the numerical example used throughout the manuscript may be exploited to demonstrate some general properties relative to the optimality of the JKLT. The JKLT optimizes the MMSE performance on average over all possible signal environments. Here, the area between the full-rank MMSE curve and the other reduced-rank MMSE curves in Fig. 1.8 represents the average MMSE performance loss over all possible signal environments in $\Re^2$, and these measurements are presented in Table 1.7. These results depict that, with a cross-correlation unit vector in $\Re^2$, on average the JKLT outperforms the principal-components by approximately 26.5 dB and the cross-spectral metric by approximately 3 dB.

1.7 Radar Example

A radar detection example is now presented to demonstrate the application of intelligent reduced-rank signal processing to a practical problem of interest. Sensor signal processing introduces a unique sample support interpretation due to the dependency of the training data on range. Ergodicity, in the context of sensor signal processing, implies replacing the ensemble statistical average with an average over range rather than time. Therefore the IID and stationarity assumptions on the joint statistics of the training and test data are replaced with clutter and noise homogeneity assumptions throughout the data collected by the radar. Intelligent rank reduction offers the opportunity to reduce the sample support requirements without greatly degrading performance. Intelli-

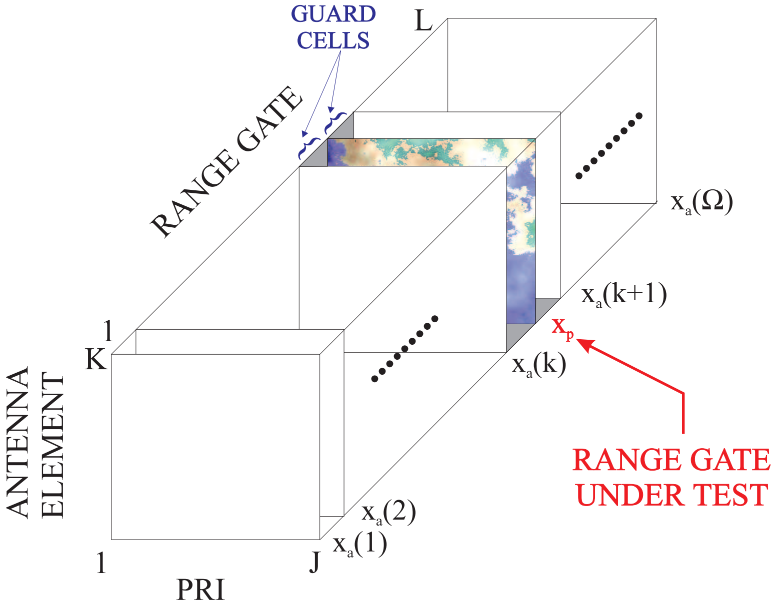

Figure 1.9: The 3-D STAP data cube.

gent training methods allow the statistical estimation to concentrate on the clutter and noise which are most correlated with that present in the test data set, thereby optimizing the detection problem. The intelligent subspace selection also reduces the sample support requirements when the stationary and IID assumptions are valid, allowing rapid adaptation.

1.7.1 Radar Signal Processing

Consider now the use of an airborne space-time adaptive processing (STAP) radar for the detection of a target at a particular range. The STAP radar collects angle-Doppler data over multiple range cells, as shown in Fig. 1.9. The use of STAP is required because the two-dimensional extent of ground clutter exhibits both spatial and Doppler dependency, as further explained in Sect. 1.7.3. A STAP processor is composed of K antenna elements which provide spatial degrees of freedom and a J-tap Doppler filterbank with time lags that correspond to the radar pulse repetition interval (PRI) in order to provide the spectral degrees of freedom. The total number of adaptive space-time degrees of freedom is then $N{=}KJ$.

The $N \times 1$ space-time steering vector $\mathbf{s}$ forms a beam at an angle-Doppler location where target presence or absence is going to be tested. The $N \times 1$ space-time snapshot formed from the radar data collected at the range gate of interest is denoted $\mathbf{x}_p$. Assume for the moment that the (unavailable) clairvoyant clutter and noise covariance matrix $\mathbf{R}$ is known *a priori*. Then the optimal STAP weight vector for Gaussian colored noise detection (simultaneous detection and clutter/noise mitigation) is given by [3, 27],

$$\mathbf{w}_a = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}} = \mathbf{s} - \mathbf{B}^H\mathbf{w}, \tag{1.57}$$

where the full row-rank matrix $\mathbf{B}$ is the nullspace of vector $\mathbf{s}$ such that $\mathbf{Bs} = 0$. The weight vector $\mathbf{w}$ is the optimal weight vector for estimating the clutter and noise present in the beamformer output, $d_0 = \mathbf{s}^H\mathbf{x}_p$, from that present in the data outside the radar look-direction, $\mathbf{x}_0 = \mathbf{B}\mathbf{x}_p$. This weight vector is computed by the Wiener-Hopf equation,

$$\mathbf{w} = \mathbf{R}_{x_0}^{-1}\mathbf{r}_{x_0 d_0}, \tag{1.58}$$

where $\mathbf{R}_{x_0} = \mathbf{B}\mathbf{R}\mathbf{B}^H$ and $\mathbf{r}_{x_0 d_0} = \mathbf{B}\mathbf{R}\mathbf{s}$. The weight vector in the center of Eq. (1.57) is termed the direct form, while the weight vector on the far right is called the filtered-data form.

The STAP filter output is given by,

$$\varepsilon_0 = \mathbf{w}_a^H\mathbf{x}_p = \frac{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{x}_p}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}} = \left(\mathbf{s}^H - \mathbf{w}^H\mathbf{B}\right)\mathbf{x}_p. \tag{1.59}$$

The direct form output noise power (under the null hypothesis) is computed as follows,

$$P = \mathbf{w}_a^H\mathbf{R}\mathbf{w}_a = \frac{1}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}. \tag{1.60}$$

This may also be expressed in the filtered-data form,

$$P = \sigma_{d_0}^2 - \mathbf{w}^H\mathbf{R}_{x_0}\mathbf{w} = \sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H\mathbf{R}_{x_0}^{-1}\mathbf{r}_{x_0 d_0}, \tag{1.61}$$

where the STAP beamformer output noise power is $\sigma_{d_0}^2 = \mathbf{s}^H\mathbf{R}\mathbf{s}$. It is now seen via Fig. 1.10 that these STAP variables are directly related to the Wiener

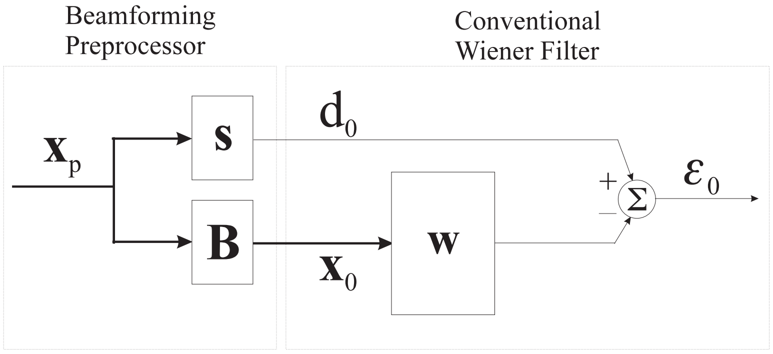

Figure 1.10: The filtered-data form of the STAP processor.

filter studied extensively earlier in this text. It is also of interest to note that the beamforming prepocessor in Fig. 1.10 represents the first stage of the JKLT which accounts for directional preference, as presented earlier. This structure therefore fits naturally within the theoretical framework of Wiener filtering and the JKLT algorithm.

The detection problem can be evaluated using these Wiener filter variables. One popular adaptive constant false alarm rate (CFAR) detection test, called the adaptive matched filter (AMF) [28, 29], is given by,

$$\Lambda = \frac{|\varepsilon_0|^2}{P} = \frac{|\mathbf{s}^H \mathbf{R}^{-1} \mathbf{x}_p|^2}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{1.62}$$

where H_1 and H_0 are the target present and target absent hypotheses, respectively. The performance of the AMF CFAR test is a function of the SINR and the false-alarm probability. The SINR is therefore the most frequently evaluated performance measure in assessing STAP performance. The optimization of the weight vector to maximize the SINR results in a locus of solutions with the form $\mathbf{w}_a = \alpha \mathbf{R}^{-1} \mathbf{s}$. The Wiener filters in Eqs. (1.57) and (1.58) represent the operating point which provides a distortionless response of the target test cell out of this locus of solutions. The optimal SINR is calculated as follows:

$$\xi_o = \frac{|\mathbf{w}_a^H \mathbf{s}|^2}{\mathbf{w}_a^H \mathbf{R} \mathbf{w}_a} = \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} = \frac{1}{\sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{R}_{x_0}^{-1} \mathbf{r}_{x_0 d_0}} = \frac{1}{P}. \tag{1.63}$$

The traditional method of applying rank reduction for STAP processing is to perform a KLT of the covariance matrix $\mathbf{R}$ and retain the principal components [30, 31, 32, 33]. However, an analysis of the SINR in either the direct or the filtered-data form yields the relations,

$$\xi_o = \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} = \sum_{i=1}^{N} \frac{|\mathbf{E}_i^H \mathbf{s}|^2}{\lambda_i}, \tag{1.64}$$

and

$$\xi_o = \frac{1}{\sigma_{d_0}^2 - \mathbf{r}_{x_0 d_0}^H \mathbf{R}_{x_0}^{-1} \mathbf{r}_{x_0 d_0}} = \frac{1}{\sigma_{d_0}^2 - \sum_{i=1}^{N-1} \frac{|\mathbf{F}_i^H \mathbf{r}_{x_0 d_0}|^2}{d_i}}, \tag{1.65}$$

respectively. The KLT of the two pertinent covariance matrices in Eqs. (1.64) and (1.65) are $\mathbf{R} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$, with diagonal matrix $\mathbf{\Lambda}$ composed of elements λ_i, and $\mathbf{R}_{x_0} = \mathbf{F}\mathbf{D}\mathbf{F}^H$, with diagonal matrix $\mathbf{D}$ composed of elements d_i. It is evident that the cross-spectral metric directly maximizes the reduced-rank SINR expression in Eq. (1.64) relative to the basis vectors $\mathbf{E}_i$[5]. The reduced-rank maximization of the SINR in Eq. (1.65) requires the minimization of a mean-square error expression in the denominator. The optimal KLT-based reduced-rank solution here is also given by a cross-spectral metric with basis vectors $\mathbf{F}_i$.

Finally, the JKLT algorithm can be applied to this problem to introduce a better type of intelligent rank reduction. As previously noted, that the first stage of the JKLT is exactly the transformation from the direct form to the filtered-data form of the Wiener filter using matrix filters $\mathbf{s}$ and $\mathbf{B}$. The implementation of the JKLT algorithm is then identical to that presented earlier for the standard Wiener filter, as evident through the comparison of Figs. 1.2 and 1.10.

1.7.2 Estimation of the Statistics and Sample Support

The unknown $N \times N$ noise covariance matrix is estimated in practice by the analysis of $\Omega = L - 2g - 1$ auxiliary range gates, where g is the number of *guard* range gates on each side of the test range gate as shown in Fig. 1.9. These guard

[5]This optimization for the direct form may, however, lead to a poor choice of subspaces for estimating statistics with small sample support. This is due to the selection of a non-dominant subspace associated with the noise. A modified direct form cross-spectral metric is presented in [34] which properly selects a stable subspace.

cells are excluded to ensure that the target of interest does not extend into the training region. The maximum likelihood estimate of the covariance matrix is calculated from the Ω auxiliary range gates by evoking a Gaussian assumption and the aforementioned range ergodicity condition:

$$\hat{\mathbf{R}} = \frac{1}{\Omega} \sum_{k=1}^{\Omega} \mathbf{x}_a(k) \mathbf{x}_a^H(k), \tag{1.66}$$

where $\mathbf{x}_a(k)$ is the $N \times 1$ space-time auxiliary data snapshot from range cell $k \in \{1, 2, ..., \Omega\}$. The STAP filter can then be calculated with Eqs. (1.57) and (1.58), where the clairvoyant covariance $\mathbf{R}$ is replaced with the maximum likelihood sample covariance $\hat{\mathbf{R}}$. Similarly, the AMF CFAR test can then be calculated using either the filtered-data variables or by the direct substitution of the sample covariance matrix in Eq. (1.62).

It is assumed that $\Omega >> 2N$ to satisfy the sample support requirements for the estimation of the covariance matrix in Eq. (1.66). This further assumes that the range gate snapshots $\{\mathbf{x}_a(k)\}_{k=1}^{\Omega}$ contain stationary and IID samples of the clutter and noise present in the test snapshot $\mathbf{x}_p$. The earth is not, in general, homogeneous over the large range extent required by STAP radars. This may preclude the existence of $2N$ IID stationary samples in the auxiliary data set. Therefore rank reduction and an intelligent signal processing measure, such as the JKLT, are needed to provide both a lower sample support and a reasonable amount of training data from the STAP data cube. It is also noted that analogous estimation issues are present even in the stationary Gaussian case with finite training data. Intelligent signal processing techniques are required under these conditions to determine the smallest stable subspace which maximizes the SINR. The reduced sample support requirements then map directly to rapid convergence and the ability to track nonstationarities.

An SINR loss is defined by

$$\xi = \frac{|\hat{\mathbf{w}}_a^H \mathbf{s}|^2}{(\hat{\mathbf{w}}_a^H \mathbf{R} \hat{\mathbf{w}}_a)(\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s})}, \tag{1.67}$$

where $\hat{\mathbf{w}}_a$ represents the weight vector formed by substituting Eq. (1.66) into Eqs. (1.57) and (1.58). The SINR loss is the normalized SINR first presented in

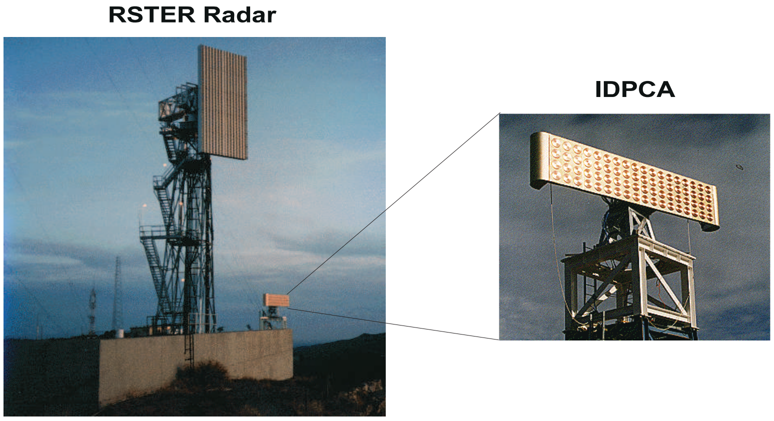

Figure 1.11: The Mountain Top radar system.

[4], and the interest here is in evaluating the loss due to reduced-rank versions of the estimated Wiener filter $\hat{\mathbf{w}}_a$. The SINR loss is called the *region of convergence for adaptivity* (ROC) when it is evaluated as a function of both the effective rank (as determined by the signal representation) and the amount of training data supplied for sample support [35]. This framework provides an informative way to analyze the potential benefit of intelligent signal processing methods for radar detection.

1.7.3 Simulation

The DARPA Mountain Top radar is simulated to demonstrate the ROC performance of the intelligent signal processing algorithms as a function of rank and sample support. The Mountain Top radar employs the Radar Surveillance Technology Experimental Radar (RSTER) and the Inverse Displaced-Phase Center Array (IDPCA), both colocated at the same site [36], as shown in Fig. 1.11. The radar consists of $K = 14$ half-wavelength spaced elements and $J = 16$ pulses in the PRI. The elevation angle is fixed (pre-beamformed) and the azimuth angle represents the only free parameter. The dimension of the adaptive processor is $KJ = 224$.

The scenario of interest consists of one target and returns from ground clut-

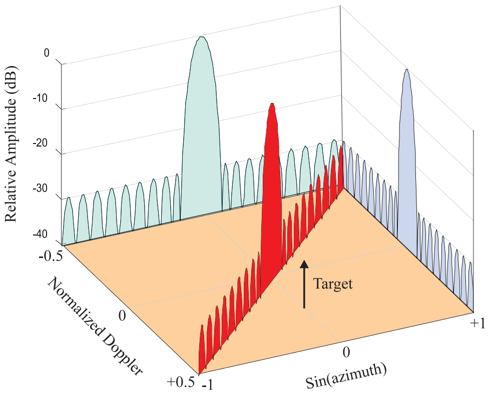

Figure 1.12: The STAP power spectrum.

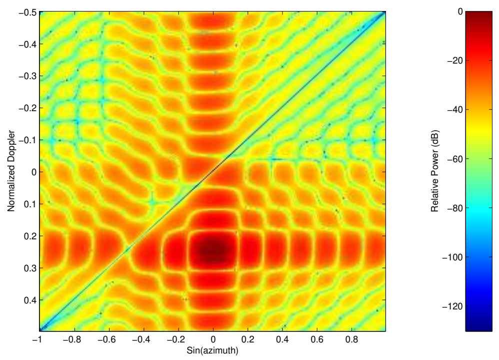

Figure 1.13: The STAP Wiener filter.

ter, as depicted in Fig. 1.12. The ground clutter has a normalized Doppler frequency which is a linear function of the spatial frequency. The locus in angle-Doppler space where the clutter is present is termed the clutter ridge. The slope of the clutter ridge, whose value is denoted β [37], dictates the number of times that the clutter Doppler spectrum aliases into the unambiguous Doppler space. The parameters selected for this simulation correspond to the clutter exactly filling the Doppler space once, or $\beta = 1$. The portion of Doppler space that the clutter ridge spans depends upon the platform velocity, the radar pulse repetition frequency and the radar operating wavelength. The two-dimensional extent of the clutter means that the mainbeam target competes with mainbeam clutter in the angle domain and the sidelobe clutter in the Doppler domain.

The optimal two-dimensional spectrum of the STAP weight vector, calculated using Eq. (1.57) for this example, is depicted in Fig. 1.13. It is readily seen that the optimal filter applies a unity gain at the angle-Doppler look-direction while simultaneously placing a null on the clutter ridge.

A Monte Carlo analysis is now considered to analyze the region of convergence for adaptivity which is obtainable using the principal-components, the cross-spectral metric and the JKLT algorithms. The effective rank r_c of the

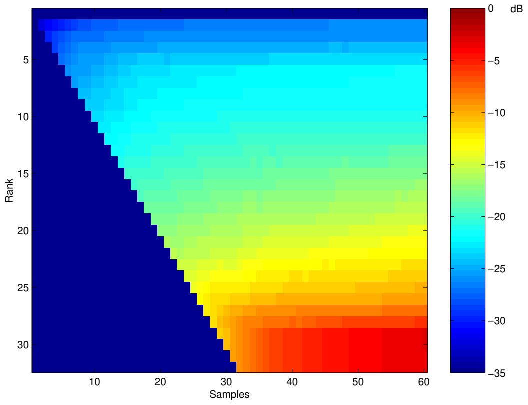

Figure 1.14: The ROC for the principal-components.

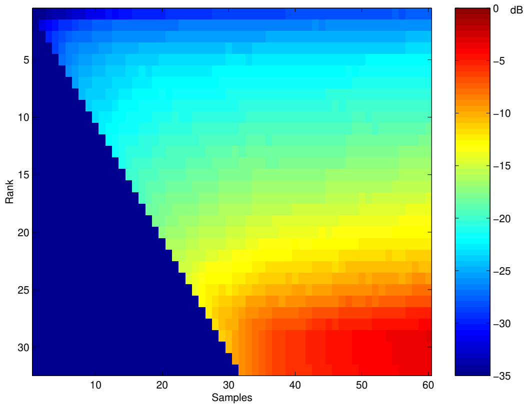

Figure 1.15: The ROC for the cross-spectral metric.

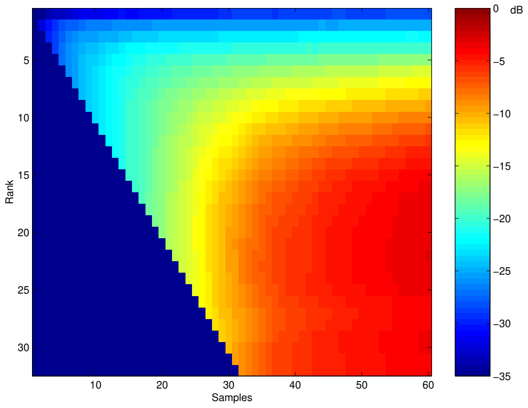

Figure 1.16: The ROC for the JKLT.

clutter can be estimated by Brennan's rule [37],

$$r_c = K + (J - 1)\beta. \tag{1.68}$$

The Mountain Top radar system, with $\beta = 1$, results in $r_c = 29$. It is therefore expected that the principal-components algorithm would require a rank on the order of 29, and that approximately $2r_c = 58$ samples would be required for convergence. The Monte Carlo simulations consist of 50 independent realizations.

The principal-components ROC is shown in Fig. 1.14. The area of this plot where the sample-support is less than the rank (the lower-left triangular region) represents the region where the reduced-rank covariance matrix is numerically unstable. The area where the SINR loss is negligible (the lower-right region defined by high sample support and high rank) represents the region of convergence for adaptivity with respect to the principal-components method. The slope or roll-off of the performance surface depicts both the robustness and the sensitivity of the processor as a function of rank and sample support. It is evident in Fig. 1.14 that the principal-components ROC does indeed require a rank of approximately $r_c = 29$, although the necessary sample support is only a little greater than r_c, or about half of the upper-bound calculated in [4]. The reduction of rank is seen to result in a quick loss of performance. This is an indication of a strong sensitivity to variations in rank and sample support.

The ROC for the cross-spectral metric is presented in Fig. 1.15. The cross-spectral metric ROC covers a slightly greater area where the sample support is high. This demonstrates that a greater rank reduction may be possible as long as sufficient sample support is available for adaptively estimating the statistics. The robustness and sensitivity to reductions in rank and sample support is nearly equivalent to that of the principal-components.

Finally, the ROC for the JKLT is depicted in Fig. 1.16. Here it is seen that the area covered by the ROC is significantly greater than that possible with the principal-components or the cross-spectral metric . The JKLT ROC includes the region of both lower sample support and significantly lower rank. The importance of this result for radar detection in the real-world is readily apparent when one considers the greatly reduced range homogeneity requirements for the low-rank JKLT. The dominating area of convergence for adaptivity in Fig. 1.16 represents much better robustness and sensitivity properties in comparison with the principal-components and cross-spectral metric algorithms.

1.8 Conclusions

This paper addresses intelligent signal representation and basis selection for detection and estimation problems. The Karhunen-Lóeve transformation and expansion are examined with emphasis on the relevant conditions of optimality. It is demonstrated that the principal-components do not provide the best enumeration of the eigenvectors corresponding with the pertinent covariance matrix for Wiener filtering. The use of intelligent subspace selection with the KLT results in the cross-spectral metric. The cross-spectral metric corrects the signal representation optimization relative to the eigenvector basis while simultaneously raising information-theoretic questions about the use of this basis itself. It is then demonstrated that a consequence of the principal component's lack of optimality results in the KLT not being optimal for those problems concerned with signal representation for detection and estimation. Finally, a new intelligent signal processing approach to signal representation, termed the joint Karhunen-Lóeve transform, is derived that uses less prior knowledge than other known techniques. The reduced-rank JKLT is demonstrated to be capable of outperforming the eigen-based algorithms even though the algorithm complexity and the amount of required information are both significantly reduced.

Acknowledgements

The authors would like thank Dr. Dan Dudgeon of MIT Lincoln Laboratory for his significant influence on the ideas presented in this work, particularly the matched filter perspective. The authors would also like to thank Dr. Peter Zulch and Lt. Col. Jan North of the Department of the Air Force, Dr. John Tague of the Office of Naval Research, Mr. Steven Huang of SAIC, and Prof. Michael Zoltowski of Purdue University for many interesting discussions and comments relating to this work.

Bibliography

[1] S.P. Applebaum. Adaptive arrays. Technical Report SPL TR 66-1, Syracuse University Research Corporation, August 1966.

[2] B. Widrow, P.E. Mantey, L.J. Griffiths, and B.B. Goode. Adaptive antenna systems. *Proc. IEEE*, 55:2143–2159, December 1967.

[3] L.E. Brennan and I.S. Reed. Theory of adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.*, 9:237–251, March 1973.

[4] I.S. Reed, J.D. Mallett, and L.E. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.*, AES-10(6):853–863, November 1974.

[5] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24:417–441 and 498–520, 1933.

[6] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrica*, 1:211–218, 1936.

[7] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[8] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, NY, 1958.

[9] M.L. Eaton. *Multivariate Statistics: A Vector Space Approach.* John Wiley & Sons, New York, NY, 1983.

[10] R.J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, NY, 1982.

[11] L.L. Scharf. *Statistical Signal Processing.* Addison-Wesley, Reading, MA, 1991.

[12] Y. Hua and W. Liu. Generalized Karhunen-Loeve transform. *IEEE Signal Processing Lett.*, 5(6):141–142, June 1998.

[13] Y. Yamashita and H. Ogawa. Relative Karhunen-Loeve transform. *IEEE Trans. Signal Processing*, 44(2):371–378, February 1996.

[14] G.H. Golub and C.F. Van Loan. *Matrix Computations.* John Hopkins University Press, Baltimore, MD, 1990.

[15] D.S. Watkins. *Fundamentals of Matrix Computations.* John Wiley & Sons, New York, NY, 1991.

[16] D.H. Johnson and D.E. Dudgeon. *Array Signal Processing.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[17] J.S. Goldstein and I.S. Reed. Reduced rank adaptive filtering. *IEEE Trans. Signal Processing*, 45(2):492–496, February 1997.

[18] J.S. Goldstein and I.S. Reed. Multidimensional Wiener filtering using a nested chain of orthogonal scalar Wiener filters. Technical Report CSI-96-12-04, University of Southern California, December 1996.

[19] J.S. Goldstein, I.S. Reed, and L.L. Scharf. A new method of Wiener filtering. In *Proc. AFOSR/DSTO Workshop on Defence Applications of Signal Processing*, Victor Harbor, Australia, June 1997.

[20] J.S. Goldstein. *Optimal Reduced-Rank Statistical Signal Processing, Detection and Estimation Theory.* PhD thesis, University of Southern California, November 1997.

[21] J.S. Goldstein, I.S. Reed, and L.L. Scharf. A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Information Theory*, 44(7):2943–2959, November 1998.

[22] J.S. Goldstein and I.S. Reed. A new method of Wiener filtering and its application to interference mitigation for communications. In *Proc. IEEE MILCOM*, Monterey, CA, November 1997.

[23] J.S. Goldstein, I.S. Reed, L.L. Scharf, and J.A. Tague. A low-complexity implementation of adaptive Wiener filters. In *Proc. 31th Asilomar Conf. Signals, Syst. Comput.*, volume 1, pages 770–774, Pacific Grove, CA, November 1997.

[24] J.S. Goldstein, I.S. Reed, and P.A. Zulch. Multistage partially adaptive STAP detection algorithm. *IEEE Trans. Aerosp. Electron. Syst.*, 35(2):645–661, April 1999.

[25] J.S. Goldstein, J.R. Guerci, and I.S. Reed. An optimal generalized theory of signal representation. In *Proc. IEEE ICASSP*, Pheonix, AZ, March 1999.

[26] T. Kailath. The innovations approach to detection and estimation theory. *Proc. IEEE*, 58:680–695, May 1970.

[27] J.S. Goldstein and I.S. Reed. Theory of partially adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.*, 33(4):1309–1325, October 1997.

[28] W.S. Chen and I.S. Reed. A new CFAR detection test for radar. *Digital Signal Proc.*, 4:198–214, October 1991.

[29] F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.*, 28(1):208–216, January 1992.

[30] D.W. Tufts, R. Kumaresan, and I. Kirsteins. Data adaptive signal estimation by singular value decomposition of a data matrix. *Proc. IEEE*, 70(6):684–685, June 1982.

[31] A.A. Shah and D.W. Tufts. Determination of the dimension of a signal subspace from short data records. *IEEE Trans. Signal Processing*, 42(9):2531–2535, September 1994.

[32] I.P. Kirsteins and D.W. Tufts. Adaptive detection using a low rank approximation to a data matrix. *IEEE Trans. Aerosp. Electron. Syst.*, 30(1):55–67, January 1994.

[33] A.M. Haimovich and Y. Bar Ness. An eigenanalysis interference canceler. *IEEE Trans. Signal Processing*, 39(1):76–84, January 1991.

[34] J.R. Guerci, J.S. Goldstein, and I.S. Reed. Optimal and adaptive reduced-rank STAP. *IEEE Trans. Aerosp. Electron. Syst.*, April 2000.

[35] P.A. Zulch, J.R. Guerci, J.S. Goldstein, and I.S. Reed. Comparison of reduced-rank signal processing techniques. In *Proc. 31th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, November 1998.

[36] G.W. Titi. An overview of the ARPA/Navy Mountain Top Program. In *Proc. IEEE Adapt. Ant. Systems Symp.*, Long Island, NY, November 1994.

[37] L.E. Brennan and F.M. Staudaher. Subclutter visibility demonstration. Technical Report RL-TR-92-21, USAF Rome Laboratory, March 1992.