MMSE Whitening and Subspace Whitening

Yonina C. Eldar* and Alan V. Oppenheim†

June 18, 2002

Abstract

This paper develops a linear whitening transformation that minimizes the mean squared error (MSE) between the original and whitened data, *i.e.,* that results in a white output that is as close as possible to the input, in an MSE sense. When the covariance matrix of the data is not invertible, the whitening transformation is designed to optimally whiten the data on a subspace in which it is contained. The optimal whitening transformation is developed both for the case of finite length data vectors and infinite length signals.

Index Terms— whitening, subspace whitening, mean squared error whitening.

*Corresponding author. Research Laboratory of Electronics, Massachusetts Institute of Technology. Mailing address for correspondence: 34A Borochov, Givatayim 53223, Israel. Email: yonina@alum.mit.edu.

†Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-615, Cambridge, MA 02139. E-mail: avo@mit.edu.

This research was supported in part through collaborative participation in the Advanced Sensors Collaborative Technology Alliance (CTA) sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0008.

1 Introduction

Data whitening arises in a variety of contexts in which it is useful to decorrelate a data sequence either prior to subsequent processing, or to control the spectral shape after processing. Examples in which data whitening has been used to advantage include enhancing direction of arrival algorithms by pre-whitening [1, 2], and improving probability of correct detection in multi-signature systems [3, 4] and multiuser wireless communication systems [5] by pre-whitening.

Whitening of a random sequence parallels closely the concept of orthogonalization of a set of vectors. Specifically, orthogonalizing a set of vectors involves mapping the set of vectors to a new set of vectors through a linear transformation so that the inner products between any two vectors in the set is zero. Similarly, whitening a zero-mean random sequence involves mapping the sequence to a new sequence through a linear transformation so that the expectation of the product of any two elements in the sequence is zero. Since the expectation has similar mathematical properties as an inner product, the mathematics associated with whitening of a random sequence parallels the mathematics associated with orthogonalizing a set of vectors.

Just as there are many ways to construct an orthogonal set of vectors from some given set of vectors, any whitening transformation cascaded with a linear unitary transformation will result in a different whitening transformation, so that the linear transformation that whitens a data vector or infinite length signal is not unique. While in some applications of whitening certain conditions might be imposed on the whitening transformation such as causality or symmetry, there have been no general assertions of optimality for various choices of a linear whitening transformation.

Recently, the concept of least-squares orthogonalization has been introduced [6, 7] in which an orthogonal set of vectors is constructed from a given set of vectors in such a way that the orthogonal vectors are as close as possible in a least-squares sense to the given set of vectors. Least-squares orthogonalization was originally motivated by a detection problem in quantum mechanics [6], and later applied to the design of optimal frames [8, 9].

Paralleling the concept of least-squares orthogonalization, in this paper we develop an optimal linear whitening transformation. Our criterion for optimality is motivated by the fact that in general whitening a data vector or signal introduces distortion to the values of the data relative to the unwhitened data. In certain applications of whitening, it may be desirable to whiten the data while minimizing this distortion. Therefore, in this paper we propose choosing a linear whitening transformation that minimizes the mean squared error (MSE) between the original and whitened data, *i.e.,* that results in a white output that is as close as possible to the input, in an MSE sense. We refer to such a whitening transformation as a minimum MSE (MMSE) whitening transformation. Extensions of this concept to other forms of covariance shaping are considered in [4]. Applications of MMSE whitening and subspace whitening to matched filter detection, multiuser detection, and least-squares estimation are considered in [3, 5, 10].

In Section 2 we derive the linear MMSE whitening transformation for a finite-length data vector with positive definite covariance matrix. In Section 3 we consider optimal whitening for the case in which the covariance matrix is not positive definite, *i.e.,* is not invertible. In this case, whitening and optimal whitening are restricted to the subspace in which the random vector is contained with probability 1. In Section 4 we consider optimal whitening of infinite length stationary data, *i.e.,* stationary random processes, both in the case of positive definite and positive semi-definite covariance functions.

2 Optimal Whitening Transformation

We denote vectors in $\mathcal{R}^m$ (m arbitrary) by boldface lowercase letters, and matrices in $\mathcal{R}^{m\times m}$ by boldface uppercase letters. $P_{\mathcal{V}}$ denotes the orthogonal projection operator onto the subspace $\mathcal{V}$ and $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. The adjoint of a transformation is denoted by $(\cdot)^*$, and $\hat{(\cdot)}$ denotes an optimal vector or transformation. The cross-covariance of random variables a and b is denoted by $\mathrm{cov}(a,b)$, and $E(\cdot)$ denotes the expectation.

Let $\mathbf{a} \in \mathcal{R}^m$ denote a zero-mean[1] random vector with positive-definite covariance matrix $\mathbf{C}_a$. We wish to whiten[2] the vector $\mathbf{a}$ using a whitening transformation $\mathbf{W}$ to obtain the random vector $\mathbf{b} = \mathbf{W}\mathbf{a}$, where the covariance matrix of $\mathbf{b}$ is given by $\mathbf{C}_b = c^2\mathbf{I}_m$ for some $c > 0$. Thus we seek a transformation $\mathbf{W}$ such that

$$\mathbf{C}_b = \mathbf{W}\mathbf{C}_a\mathbf{W}^* = c^2\mathbf{I}_m, \tag{1}$$

for some $c > 0$. We refer to any $\mathbf{W}$ satisfying (1) as a whitening transformation.

Given a covariance matrix $\mathbf{C}_a$, there are many ways to choose a whitening transformation $\mathbf{W}$ satisfying (1), for example using the eigendecomposition or Cholesky factorization of $\mathbf{C}_a$ [11]. Although there are an unlimited number of whitening transformations satisfying (1), no general assertion of optimality is known for the output $\mathbf{b} = \mathbf{W}\mathbf{a}$ of these different transformations. In particular, the white random vector $\mathbf{b} = \mathbf{W}\mathbf{a}$ may not be "close" to the input vector $\mathbf{a}$. If the vector $\mathbf{b}$ undergoes some noninvertible processing, or is used as an estimator of some unknown parameters represented by the data $\mathbf{a}$, then we may wish to choose the whitening transformation in a way that $\mathbf{b}$ is close to $\mathbf{a}$ in some sense. This can be particularly important in applications in which $\mathbf{b}$ is the input to a detector, so that we may wish to whiten $\mathbf{a}$ prior to detection, but at the same time minimize the distortion to $\mathbf{a}$ by choosing $\mathbf{W}$ so that $\mathbf{b}$ is close to $\mathbf{a}$. Applications of this

[1] If the mean $E(\mathbf{a})$ is not zero, then we can always define $\mathbf{a}' = \mathbf{a} - E(\mathbf{a})$ so that the results hold for $\mathbf{a}'$.

[2] In this paper we define a random vector $\mathbf{a}$ to be white if the covariance of $\mathbf{a}$, denoted $\mathbf{C}_a$, is given by $\mathbf{C}_a = c^2\mathbf{I}$ for some $c > 0$.

type have been recently investigated in various contexts including matched-filter detection [3, 4] and multiuser detection [5]. We therefore propose a whitening transformation that is optimal in the sense that it results in a random vector $\mathbf{b}$ that is as close as possible to $\mathbf{a}$ in MSE. Specifically, among all possible whitening transformations we seek the one that minimizes the total MSE given by

$$\varepsilon_{\mathrm{MSE}} = \sum_{k=1}^{m} E\left((a_k - b_k)^2\right) = E\left((\mathbf{a}-\mathbf{b})^*(\mathbf{a}-\mathbf{b})\right), \tag{2}$$

subject to (1), where a_k and b_k are the kth components of $\mathbf{a}$ and $\mathbf{b}$ respectively. We may wish to constraint the constant c in (1), or may choose c such that the total MSE is minimized.

Our approach to determining the whitening transformation that minimizes (2) is to perform a unitary change of coordinates $\mathbf{U}$ so that in the new coordinate system, $\mathbf{a}$ is mapped to $\overline{\mathbf{a}} = \mathbf{U}\mathbf{a}$ with the elements of $\overline{\mathbf{a}}$ uncorrelated, and $\mathbf{b}$ is mapped to $\overline{\mathbf{b}} = \mathbf{U}\mathbf{b}$. Since $\mathbf{U}$ is unitary and $\mathbf{C}_b = c^2\mathbf{I}_m$, the covariance matrix of $\overline{\mathbf{b}}$ is $\mathbf{C}_{\overline{\mathbf{b}}} = c^2\mathbf{I}_m$, and the MSE defined by (2) between $\mathbf{a}$ and $\mathbf{b}$ is equal to the MSE between $\overline{\mathbf{a}}$ and $\overline{\mathbf{b}}$.

Such a unitary transformation is provided by the eigendecomposition of $\mathbf{C}_a$. Specifically, since $\mathbf{C}_a$ is assumed positive-definite, it has an eigendecomposition $\mathbf{C}_a = \mathbf{V}\mathbf{D}\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and $\mathbf{D}$ is a diagonal matrix with diagonal elements $d_k > 0$. If we choose $\overline{\mathbf{a}} = \mathbf{V}^*\mathbf{a}$, then the covariance matrix of $\overline{\mathbf{a}}$ is $\mathbf{V}^*\mathbf{C}_a\mathbf{V} = \mathbf{D}$ and $\mathrm{cov}(\overline{a}_k, \overline{a}_l) = d_k\delta_{kl}$, where $\overline{a}_k$ denotes the kth component of $\overline{\mathbf{a}}$.

Thus, we may first solve the optimal whitening problem in the new coordinate system. Then, with $\widehat{\overline{\mathbf{W}}}$ and $\widehat{\mathbf{W}}$ denoting the optimal whitening transformations in the new and original coordinate systems respectively, it is straightforward to show that

$$\widehat{\mathbf{W}} = \mathbf{U}^*\widehat{\overline{\mathbf{W}}}\mathbf{U}. \tag{3}$$

To determine $\widehat{\overline{\mathbf{W}}}$, we express ε_{MSE} of (2) as

$$\varepsilon_{\text{MSE}} = \sum_{k=1}^{m} E\left((\overline{a}_k - \overline{b}_k)^2\right) = \sum_{k=1}^{m} d_k + mc^2 - 2\sum_{k=1}^{m} E(\overline{a}_k \overline{b}_k), \tag{4}$$

where $d_k = E(\overline{a}_k^2)$ and $\overline{b}_k$ denotes the kth component of $\overline{\mathbf{b}}$. From the Cauchy-Schwarz inequality,

$$E(\overline{a}_k \overline{b}_k) \leq |E(\overline{a}_k \overline{b}_k)| \leq \left(E\left(\overline{a}_k^2\right) E\left(\overline{b}_k^2\right)\right)^{1/2}, \tag{5}$$

with equality if and only if $\overline{b}_k = \gamma_k \overline{a}_k$ for some non-negative deterministic constant γ_k, in which case we also have $E(\overline{b}_k^2) = \gamma_k^2 E(\overline{a}_k^2) = \gamma_k^2 d_k = c^2$, so $\gamma_k = c/\sqrt{d_k}$. Note, that $\overline{b}_k$ can always be chosen proportional to $\overline{a}_k$ since the variables $\overline{a}_k$ are uncorrelated. Thus, the optimal value of $\overline{b}_k$, denoted by $\widehat{\overline{b}}_k$, is $\widehat{\overline{b}}_k = c\overline{a}_k/\sqrt{d_k}$.

If the constant c in (4) is specified, then $\widehat{\overline{\mathbf{W}}} = c\mathbf{D}^{-1/2}$. The optimal whitening transformation then follows from (3),

$$\widehat{\mathbf{W}} = c\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^* = c\mathbf{C}_a{}^{-1/2}. \tag{6}$$

Alternatively, we may choose to further minimize (4) with respect to c. Substituting $\widehat{\overline{b}}_k$ back into (4), we choose c to minimize

$$mc^2 - 2c\sum_{k=1}^{m} \sqrt{d_k}. \tag{7}$$

The optimal value of c, denoted by $\hat{c}$, is therefore given by

$$\hat{c} = \frac{1}{m}\sum_{k=1}^{m} \sqrt{d_k}, \tag{8}$$

and the optimal whitening transformation is

$$\widehat{\mathbf{W}} = \hat{c}\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^* = \hat{c}\mathbf{C}_a{}^{-1/2}. \tag{9}$$

It is interesting to note that the MMSE whitening transformation has the additional property that it is the unique *symmetric* whitening transformation [12] (up to a possible minus sign). It is also proportional to the Mahalanobis transformation, that is frequently used in signal processing applications incorporating whitening (see *e.g.,* [13, 1, 2]).

The results above are summarized in the following theorem:

Theorem 1 (MMSE whitening transformation) *Let* $\mathbf{a} \in \mathcal{R}^m$ *be a random vector with positive-definite covariance matrix* $\mathbf{C}_a = \mathbf{V}\mathbf{D}\mathbf{V}^*$*, where* $\mathbf{D}$ *is a diagonal matrix and* $\mathbf{V}$ *is a unitary matrix. Let* $\widehat{\mathbf{W}}$ *be the optimal whitening transformation that minimizes the MSE defined by (2), between the input* $\mathbf{a}$ *and the output* $\mathbf{b} = \mathbf{W}\mathbf{a}$ *with covariance* $\mathbf{C}_b = c^2\mathbf{I}_m$ *where* $c > 0$*. Then*

$$\widehat{\mathbf{W}} = \alpha \mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^* = \alpha {\mathbf{C}_a}^{-1/2},$$

where

1. *if* c *is specified then* $\alpha = c$*;*

2. *if* c *is chosen to minimize the MSE then* $\alpha = (1/m)\sum_{k=1}^{m}\sqrt{d_k}$.

3 Optimal Subspace Whitening

Suppose now that $\mathbf{C}_a$ is not positive-definite, *i.e.,* $\mathbf{C}_a$ is not invertible. In this case there is no whitening transformation $\mathbf{W}$ such that $\mathbf{W}\mathbf{C}_a\mathbf{W}^* = c^2\mathbf{I}_m$. Instead, we propose whitening $\mathbf{a}$ on the space in which it is contained, which we refer to as *subspace whitening.*

3.1 Subspace Whitening

Let $\mathbf{a}$ be a zero-mean random vector in $\mathcal{R}^m$ with covariance matrix $\mathbf{C}_a$, where $\text{rank}(\mathbf{C}_a) = n < m$, and let $\mathcal{V} \subset \mathcal{R}^m$ denote the range space of $\mathbf{C}_a$. If $\mathbf{C}_a$ is not invertible, then the elements of $\mathbf{a}$ are linearly dependent with probability one (w.p. 1)[3], and consequently any realization of the random vector $\mathbf{a}$ lies in $\mathcal{V}$. This follows from the fact that for any $v \in \mathcal{V}^\perp$, $\mathbf{C}_a v = 0$, so that $v^*\mathbf{a} = 0$ w.p. 1 for any realization of $\mathbf{a}$. We may therefore consider whitening $\mathbf{a}$ on $\mathcal{V}$, which we refer to as subspace whitening.

First consider a zero mean random vector $\mathbf{q} \in \mathcal{R}^m$ with full-rank covariance matrix, and let $\mathbf{r} = \mathbf{W}\mathbf{q}$ where $\mathbf{W}$ is a whitening transformation, so that $\mathbf{r}$ is white. Then $\mathbf{r}$ and $\mathbf{q}$ lie in the same space $\mathcal{R}^m$. Furthermore, if $\mathbf{r}$ is white then the representation of $\mathbf{r}$ in terms of any orthonormal basis for $\mathcal{R}^m$ is also white. This follows from the fact that any two orthonormal bases for $\mathcal{R}^m$ are related through a unitary transformation. We define subspace whitening to preserve these two properties.

Let $\mathbf{a}$ be a random vector with covariance $\mathbf{C}_a$ with range space $\mathcal{V}$, and let $\mathbf{b}$ denote the output of a subspace whitening transformation of $\mathbf{a}$. Since $\mathbf{a} \in \mathcal{V}$ we require that $\mathbf{b} \in \mathcal{V}$. In addition, we require that the representation of $\mathbf{b}$ in terms of some orthonormal basis for $\mathcal{V}$ is white, which implies that the representation in terms of any orthonormal basis for $\mathcal{V}$ is white.

In Appendix A we translate the conditions such that a random vector $\mathbf{b}$ is white on $\mathcal{V}$, to

[3]Throughout this section when we say that the elements of a random vector are linearly dependent we mean w.p. 1; similarly, when we say that a random vector lies in a subspace we mean w.p. 1.

conditions on the covariance matrix $\mathbf{C}_b$. Specifically, we show that $\mathbf{C}_b$ must satisfy,

$$\mathbf{C}_b = c^2 P_{\mathcal{V}} = c^2 \mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*, \tag{10}$$

where $P_{\mathcal{V}}$ is the orthogonal projection operator onto $\mathcal{V}$, the first n columns of $\mathbf{V}$ form an orthonormal basis for $\mathcal{V}$, and

$$\tilde{\mathbf{I}} = \begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & 0 \end{bmatrix}. \tag{11}$$

3.2 MMSE Subspace Whitening Transformation

To restate the MMSE subspace whitening problem, let $\mathbf{a} \in \mathcal{R}^m$ be a random vector with kth component a_k and covariance matrix $\mathbf{C}_a = \mathbf{V}\mathbf{D}\mathbf{V}^*$ with $\text{rank}(\mathbf{C}_a) = n < m$. Here $\mathbf{V}$ is a unitary matrix with orthonormal columns $\mathbf{v}_k$, and $\mathbf{D}$ is a diagonal matrix with diagonal elements d_k where $d_k > 0, 1 \leq k \leq n$ and $d_k = 0, n+1 \leq k \leq m$. Let $\mathcal{V}$ denote the range of $\mathbf{C}_a$, spanned by the columns $\{\mathbf{v}_k, 1 \leq k \leq n\}$. We seek a subspace whitening transformation $\mathbf{W}_s$ such that the vector $\mathbf{b} = \mathbf{W}_s\mathbf{a}$ is white on $\mathcal{V}$, namely such that $\mathbf{b}$ has a covariance matrix $\mathbf{C}_b = c^2 P_{\mathcal{V}} = c^2\mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*$, where $\tilde{\mathbf{I}}$ is given by (11), $c > 0$, and is as close as possible to $\mathbf{a}$ in the MSE sense. Thus, we seek the transformation that minimizes (2) subject to the constraint

$$\mathbf{C}_b = \mathbf{W}_s\mathbf{C}_a\mathbf{W}_s^* = c^2\mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*. \tag{12}$$

The MMSE subspace whitening transformation, denoted by $\widehat{\mathbf{W}}_s$, is derived in Appendix B in an analogous manner to the derivation of the MMSE whitening transformation of Section 2, and is given by

$$\widehat{\mathbf{W}}_s = \alpha_s\mathbf{V}(\mathbf{D}^{1/2})^{\dagger}\mathbf{V}^* = \alpha_s({\mathbf{C}_a}^{1/2})^{\dagger}, \tag{13}$$

where $(\cdot)^\dagger$ denotes the *Moore-Penrose pseudo inverse* [14]. Here $\alpha_s = c$ in the case in which c in (12) is specified, and $\alpha_s = (1/n)\sum_{k=1}^{n}\sqrt{d_k}$, in the case in which c is chosen to minimize the MSE (2).

It is intuitively reasonable and follows from the proof in Appendix B that $\widehat{\mathbf{W}}_s$ is uniquely specified on $\mathcal{V}$, but can be arbitrary on $\mathcal{V}^\perp$. However, since the input $\mathbf{a}$ to the whitening transformation lies in $\mathcal{V}$ w.p. 1, the choice of $\widehat{\mathbf{W}}_s$ on $\mathcal{V}^\perp$ does not affect the output $\mathbf{b}$ (w.p. 1).

The results above are summarized in the following theorem:

Theorem 2 (MMSE subspace whitening) *Let $\mathbf{a} \in \mathcal{R}^m$ be a random vector with covariance matrix $\mathbf{C}_a = \mathbf{V}\mathbf{D}\mathbf{V}^*$ with rank$(\mathbf{C}_a) = n < m$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{V}$ is a unitary matrix. Let $\mathcal{V}$ denote the range space of $\mathbf{C}_a$. Let $\widehat{\mathbf{W}}_s$ be any subspace whitening transformation that minimizes the MSE defined by (2), between the input $\mathbf{a}$ and the output $\mathbf{b}$ with covariance $\mathbf{C}_b = c^2 P_{\mathcal{V}} = c^2\mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*$, where $\tilde{\mathbf{I}}$ is given by (11) and $c > 0$. Then*

1. *$\widehat{\mathbf{W}}_s$ is not unique;*
2. *$\widehat{\mathbf{W}}_s = \alpha_s \mathbf{V}(\mathbf{D}^{1/2})^\dagger\mathbf{V}^* = \alpha_s({\mathbf{C}_a}^{1/2})^\dagger$ is an optimal subspace whitening transformation where*
   - *(a) if c is specified then $\alpha_s = c$;*
   - *(b) if c is chosen to minimize the MSE then $\alpha_s = (1/n)\sum_{k=1}^{n}\sqrt{d_k}$;*
3. *define $\mathbf{W}_s^{\mathcal{V}} = \widehat{\mathbf{W}}_s P_{\mathcal{V}}$ where $P_{\mathcal{V}}$ is a projection onto $\mathcal{V}$ and $\widehat{\mathbf{W}}_s$ is any optimal subspace whitening transformation; then*
   - *(a) $\mathbf{W}_s^{\mathcal{V}}$ is unique, and is given by $\mathbf{W}_s^{\mathcal{V}} = \alpha_s \mathbf{V}(\mathbf{D}^{1/2})^\dagger\mathbf{V}^* = \alpha_s({\mathbf{C}_a}^{1/2})^\dagger$;*
   - *(b) $\widehat{\mathbf{W}}_s\mathbf{a} = \mathbf{W}_s^{\mathcal{V}}\mathbf{a}$ w.p. 1;*
   - *(c) $\mathbf{b} = \widehat{\mathbf{W}}_s\mathbf{a}$ is unique w.p. 1.*

4 Optimal Whitening of Stationary Random Processes

We now consider optimal whitening and subspace whitening of a stationary random process.

4.1 MMSE Whitening

Let $a[n]$ be a zero-mean stationary random process with positive-definite correlation function $R_a[n]$. Suppose we wish to whiten $a[n]$, *i.e.,* find an LTI filter with impulse response $w[n]$ such that the correlation function $R_b[n]$ of the filter output $b[n] = a[n] * w[n]$ is $R_b[n] = c^2\delta[n]$ for some $c > 0$.

We can express $R_b[n]$ in terms of the filter impulse response $w[n]$ and $R_a[n]$ as [13]

$$R_b[n] = R_a[n] * w[n] * w[-n]. \tag{14}$$

Denoting the Fourier transforms of $R_a[n], R_b[n]$ and $w[n]$ by $S_a(\omega), S_b(\omega)$ and $W(\omega)$ respectively, and taking the Fourier transform of (14),

$$S_b(\omega) = S_a(\omega)|W(\omega)|^2. \tag{15}$$

Since $R_b[n] = c^2\delta[n]$, $S_b(\omega) = c^2$, and $W(\omega)$ must satisfy

$$|W(\omega)|^2 = c^2 S_a^{-1}(\omega). \tag{16}$$

Note that $S(\omega) > 0$ for all ω since we assume that $R_a[n]$ is positive-definite. A filter with frequency response $W(\omega)$ satisfying (16) is a whitening filter.

Given a random process $a[n]$ with correlation function $R_a[n]$, there are many possible whitening filters with frequency response satisfying (16). From all possible whitening filters we seek the filter that results in $b[n]$ as close as possible to $a[n]$ in an MSE sense. Thus we seek the filter that minimizes

$$\varepsilon_{\text{MSE}} = E((a[n] - b[n])^2), \tag{17}$$

subject to the constraint (16).

Expanding (17) we have

$$\varepsilon_{\text{MSE}} = E(a^2[n]) + E(b^2[n]) - 2E(a[n]b[n]) = R_a[0] + c^2 - 2E(a[n]b[n]). \tag{18}$$

Minimizing (18) with respect to $b[n]$ is equivalent to maximizing

$$E(a[n]b[n]) = R_{ab}[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ab}(\omega)d\omega. \tag{19}$$

Here $R_{ab}[n]$ is the cross-correlation function between $a[n]$ and $b[n]$, and $S_{ab}(\omega)$ is the Fourier transform of $R_{ab}[n]$, which is related to $S_a(\omega)$ and $W(\omega)$ by [13]

$$S_{ab}(\omega) = W^*(\omega)S_a(\omega). \tag{20}$$

Substituting (20) into (19) and using (16) we have that,

$$\begin{aligned} E(a[n]b[n]) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} W^*(\omega)S_a(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} c^2 W^{-1}(\omega)d\omega \\ &\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} c^2 |W^{-1}(\omega)|d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} c S_a^{1/2}(\omega)d\omega. \end{aligned} \tag{21}$$

We have equality in (21) if and only if

$$W^{-1}(\omega) = |W^{-1}(\omega)| = \frac{1}{c}S_a^{1/2}(\omega), \tag{22}$$

or

$$W(\omega) = cS_a^{-1/2}(\omega). \tag{23}$$

The optimal value of $b[n]$ is then given by $\hat{b}[n] = \tilde{w}[n] * a[n]$ where $\tilde{w}[n]$ is the inverse Fourier transform of $W(\omega)$ given by (23).

If c is specified, then the MMSE whitening filter is given by

$$\widehat{W}(\omega) = cS_a^{-1/2}(\omega). \tag{24}$$

We may further wish to minimize the MSE with respect to c. Substituting $\hat{b}[n]$ into (18) and minimizing with respect to c we have that the optimal value of c, denoted $\hat{c}$, is given by

$$\hat{c} = E\left(a[n](a[n] * \tilde{w}[n])\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_a^{1/2}(\omega)d\omega, \tag{25}$$

and the MMSE whitening filter is given by

$$\widehat{W}(\omega) = \hat{c}S_a^{-1/2}(\omega). \tag{26}$$

Note that the MMSE whitening filter is the unique zero-phase filter that satisfies (16).

The results above are summarized in the following theorem:

Theorem 3 (MMSE whitening filter) *Let $a[n]$ denote a random process with positive-definite correlation function $R_a[n]$, whose Fourier transform is denoted by $S_a(\omega)$. Let $\widehat{W}(\omega)$ be the frequency response of the optimal whitening filter with impulse response $\hat{w}[n]$ that minimizes the MSE defined by (17), between the input $a[n]$ and the output $b[n] = a[n] * w[n]$ with correlation function $R_b[n] = c^2\delta[n]$ with $c > 0$. Then*

$$\widehat{W}(\omega) = \sigma S_a(\omega)^{-1/2},$$

where

1. *if c is specified then $\sigma = c$;*

2. *if c is chosen to minimize the MSE then $\sigma = (1/2\pi)\int_{-\pi}^{\pi} S_a^{1/2}(\omega)d\omega$.*

The MMSE whitening filter given by Theorem 3 is reminiscent of the MMSE whitening transformations given by Theorem 1. The optimal whitening transformation is proportional to the inverse-square root of the input covariance matrix, and is symmetric. Similarly, the Fourier transform of the optimal whitening filter is proportional to the inverse-square root of the input spectral density function, and has zero phase.

4.2 MMSE Subspace Whitening

In the previous section we assumed that $R_a[n]$ is positive definite. If it is not positive-definite then, as is well known, $S_a(\omega)$ will be zero for some ω.

Suppose now that $S_a(\omega) = 0$ on a set of frequencies $\{\omega_i\}$. If $a[n]$ is the input to the filter with impulse response $h[n] = e^{j\omega_i n}$, then the spectral density function $S_d(\omega)$ of the output $d[n] = a[n] * h[n]$ is given by

$$S_d(\omega) = (2\pi\delta(\omega - \omega_i))^2 S_a(\omega_i) = 0, \tag{27}$$

and consequently $d[n] = 0$ w.p. 1. But

$$d[n] = \sum_m a[m]h[n-m] = e^{j\omega_i n}\sum_m a[m]e^{-j\omega_i m}. \tag{28}$$

Thus, $A(\omega_i) = \sum_m a[m]e^{-j\omega_i m} = 0$ w.p. 1, and therefore the elements of the sequence $a[n]$ are linearly (deterministically) dependent. The same analysis holds true when $S_a(\omega) = 0$ over a set of frequency intervals.

As in the finite-dimensional case, we then propose whitening $a[n]$ on the subspace to which it is confined. This is equivalent to whitening $a[n]$ over the frequency intervals for which $S_a(\omega) \neq 0$.

Thus, the subspace whitening filter satisfies

$$|W(\omega)|^2 = \begin{cases} c^2 S_a^{-1}(\omega), & \omega \text{ such that } S_a(\omega) \neq 0; \\ \text{arbitrary}, & \omega \text{ such that } S_a(\omega) = 0. \end{cases} \tag{29}$$

The frequency response of the MMSE subspace whitening filter is given by Theorem 3 at frequencies for which $S_a(\omega) \neq 0$, and is arbitrary otherwise.

Appendix

A Subspace Whitening

We translate the conditions on a random vector $\mathbf{b}$ to be white on $\mathcal{V}$, to conditions on the covariance $\mathbf{C}_b$ of $\mathbf{b}$. The first condition on the vector $\mathbf{b}$ is that $\mathbf{b} \in \mathcal{V}$. Suppose that $\mathbf{b} \in \mathcal{V}$. Then $\mathbf{v}_k^* \mathbf{b} = 0, m+1 \leq k \leq n$ (w.p. 1), since the vectors $\{\mathbf{v}_k, m+1 \leq k \leq n\}$ span $\mathcal{V}^\perp$. This in turn implies that

$$\mathbf{C}_b \mathbf{v}_k = \mathbf{0},\ n+1 \leq k \leq m, \tag{30}$$

so that the null space of $\mathbf{C}_b$ contains $\mathcal{V}^\perp$. Conversely, suppose that the null space of $\mathbf{C}_b$ contains $\mathcal{V}^\perp$. Then (30) holds, and we have already shown that this implies that $\mathbf{b} \in \mathcal{V}$. We conclude that $\mathbf{b} \in \mathcal{V}$ if and only if the null space of $\mathbf{C}_b$ contains $\mathcal{V}^\perp$, so that $\mathbf{C}_b$ satisfies (30).

We now discuss the requirement that the representation of $\mathbf{b}$ in terms of any orthonormal basis for $\mathcal{V}$ is white. Let $\mathbf{V}_1$ denote the matrix of columns $\{\mathbf{v}_k, 1 \leq k \leq n\}$, that form a basis for $\mathcal{V}$. The representation of $\mathbf{b}$ in this basis for $\mathcal{V}$ is $\mathbf{b}_v = \mathbf{V}_1^* \mathbf{b}$, $\mathbf{b}_v \in \mathcal{R}^n$. We require that $\mathbf{b}_v$ is white, namely that the covariance matrix of $\mathbf{b}_v$ is equal to $c^2 \mathbf{I}_n$. Since the covariance of $\mathbf{b}_v$ is given by $\mathbf{V}_1^* \mathbf{C}_b \mathbf{V}_1$, our requirement on $\mathbf{C}_b$ is

$$\mathbf{V}_1^* \mathbf{C}_b \mathbf{V}_1 = c^2 \mathbf{I}_n, \tag{31}$$

for some $c > 0$. Thus the matrix $\mathbf{C}_b$ has to satisfy (30) and (31), which can be combined into the single condition

$$\mathbf{C}_b = c^2 P_{\mathcal{V}} = c^2 \mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*, \tag{32}$$

where $\mathbf{V}$ is the matrix of columns $\{\mathbf{v}_k, 1 \leq k \leq m\}$, and $\tilde{\mathbf{I}}$ is given by (11).

B Subspace MMSE Whitening

Let $\overline{\mathbf{a}} = \mathbf{V}^*\mathbf{a}$, and $\overline{\mathbf{b}} = \mathbf{V}^*\mathbf{b}$ where $\mathbf{b}$ is white on $\mathcal{V}$ so that $\mathbf{b}$ has covariance $\mathbf{C}_b = c^2\mathbf{V}\tilde{\mathbf{I}}\mathbf{V}^*$. The covariance of $\overline{\mathbf{a}}$ is then $\mathbf{V}^*\mathbf{C}_a\mathbf{V} = \mathbf{D}$, and the covariance of $\overline{\mathbf{b}}$ is $\mathbf{V}^*\mathbf{C}_b\mathbf{V} = c^2\tilde{\mathbf{I}}$. As in MMSE whitening, instead of seeking a subspace whitening transformation that minimizes the MSE between $\mathbf{a}$ and $\mathbf{b}$, we may seek a transformation $\widehat{\widehat{\mathbf{W}}}_s$ such that the vector $\overline{\mathbf{b}} = \widehat{\widehat{\mathbf{W}}}_s\overline{\mathbf{a}}$ is as close as possible to $\overline{\mathbf{a}}$, and such that $\overline{\mathbf{b}}$ has covariance $c^2\tilde{\mathbf{I}}$. From (3) it then follows that $\widehat{\mathbf{W}}_s = \mathbf{V}\widehat{\widehat{\mathbf{W}}}_s\mathbf{V}^*$, where $\widehat{\mathbf{W}}_s$ is the optimal subspace whitening transformation that minimizes the MSE between $\mathbf{a}$ and $\mathbf{b}$.

Using the Cauchy-Schwarz inequality it follows that $\widehat{\widehat{\mathbf{W}}}_s$ is such that $\overline{b}_k = c\overline{a}_k/\sqrt{d_k}$ for $1 \leq k \leq n$. Since the covariance of $\overline{b}$ must be equal to $c^2\tilde{\mathbf{I}}$, $\widehat{\widehat{\mathbf{W}}}_s$ must also be chosen so that $\operatorname{var}(\overline{b}_k) = 0$ for $n+1 \leq k \leq m$. Now, the covariance of $\overline{\mathbf{a}}$ is $\mathbf{D}$, where the kth diagonal element of $\mathbf{D}$ is equal to 0, for $n+1 \leq k \leq m$. Consequently, $\overline{a}_k = 0$ w.p. 1 for $n+1 \leq k \leq m$. Therefore, we conclude that $\widehat{\widehat{\mathbf{W}}}_s$ is block diagonal. The upper left $n \times n$ block is a diagonal matrix, with diagonal elements $c/\sqrt{d_k}$; the lower right block is arbitrary, since $\overline{b}_k = \overline{a}_k = 0$ regardless of the choice of this block. We therefore choose $\widehat{\widehat{\mathbf{W}}}_s$ to be a diagonal matrix with the first n diagonal elements equal to $c/\sqrt{d_k}$ and the remaining diagonal elements equal to 0. Thus $\widehat{\widehat{\mathbf{W}}}_s = c(\mathbf{D}^{1/2})^\dagger$, and

$$\widehat{\mathbf{W}}_s = c\mathbf{V}(\mathbf{D}^{1/2})^\dagger\mathbf{V}^* = c({\mathbf{C}_a}^{1/2})^\dagger. \tag{33}$$

If we choose to minimize the MSE with respect to c as well, then it is straightforward to show that the optimal value of c is given by $\alpha_s = 1/n\sum_{k=1}^n \sqrt{d_k}$.

References

[1] E. M. Friel and K. M. Pasala, “Direction finding with compensation for a near field scatterer,” in *International Symposium Antennas and Propagation Society*, 1995, pp. 106–109.

[2] R. J. Piechocki, N. Canagarajah, and J. P. McGeehan, “Improving the direction-of-arrival resolution via double code filtering in WCDMA,” in *First International Conference on 3G Mobile Communication Technologies*, Mar. 2000, pp. 204–207.

[3] Y. C. Eldar, A. V. Oppenheim, and D. Egnor, “Orthogonal and projected orthogonal matched filter detection,” submitted to *Signal Processing*, June 2002.

[4] Y. C. Eldar, *Quantum Signal Processing*, Ph.D. thesis, Massachusetts Institute of Technology, Dec. 2001, also available at `http://allegro.mit.edu/dspg/publications/TechRep/index.html`.

[5] Y. C. Eldar and A. M. Chan, “An optimal whitening approach to linear multiuser detection,” submitted to *IEEE Trans. Inform. Theory*, Jan. 2002.

[6] Y. C. Eldar and G. D. Forney, Jr., “On quantum detection and the square-root measurement,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 858–872, Mar. 2001.

[7] Y. C. Eldar, “Least-squares inner product shaping,” *Linear Algebra Appl.*, vol. 348, pp. 153–174, May 2002.

[8] Y. C. Eldar and G. D. Forney, Jr., “Optimal tight frames and quantum measurement,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 599–610, Mar. 2002.

[9] Y. C. Eldar and H. Bölcskei, “Geometrically uniform frames,” submitted to *IEEE Trans. Inform. Theory*, August 2001; also available at `http://arXiv.org/abs/math.FA/0108096`.

[10] Y. C. Eldar and A. V. Oppenheim, “Covariance shaping least-squares estimation,” submitted to *IEEE Trans. Signal Processing*, May 2002.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1985.

[12] J. J. Atick and A. N. Redlich, “Convergent algorithm for sensory receptive field development,” *Neurak Comp.*, vol. 5, pp. 45–60, 1993.

[13] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1992.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore MD: Johns Hopkins Univ. Press, third edition, 1996.