Linear Algebra and its Applications 348 (2002) 153–174

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Least-squares inner product shaping

Yonina C. Eldar

Research Laboratory of Electronics, Room 36-615, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 13 January 2001; accepted 6 November 2001

Submitted by H. Schneider

Abstract

We develop methods that construct an optimal set of vectors with a specified inner product structure, from a given set of vectors in a complex Hilbert space. The optimal vectors are chosen to minimize the sum of the squared norms of the errors between the constructed vectors and the given vectors. Four special cases are considered. In the first, the constructed vectors are orthonormal. In the second, they are orthogonal. In the third, the Gram matrix of inner products of the constructed vectors is a circulant matrix. As we show, the vectors form a cyclic set. In the fourth, the Gram matrix has the property that the rows are all permutations of each other. The constructed vectors are shown to be geometrically uniform. © 2002 Elsevier Science Inc. All rights reserved.

AMS classification: 65F25; 15A04; 15A23; 15A99

Keywords: Orthogonalization; Polar decomposition; Least-squares; Circulant matrices; Geometric uniformity; Generalized Fourier transform

1. Introduction

Constructing a set of orthonormal (or orthogonal) vectors $\{h_k, 1 \leqslant k \leqslant N\}$ from a given set of vectors $\{s_k, 1 \leqslant k \leqslant N\}$ is a well-known problem. We may view this problem as a special case of an inner product shaping problem, in which we construct a set of vectors from a given set of vectors such that the inner products between vectors in the set have some specified structure.

E-mail address: yonina@mit.edu (Y.C. Eldar).

PII: S0024-3795(01)00575-4

There are many ways to construct vectors with specified inner products. In this paper we consider new methods that construct vectors that are closest in a least-squares sense to a given set of vectors. Specifically, the constructed vectors are chosen to minimize the sum of the squared norms of the error vectors between the constructed vectors and the given vectors. The development considers both the case in which the given vectors are linearly independent and the case in which they are linearly dependent.

Orthonormalization procedures that have some form of optimality have been suggested in [5,15,31]. Constructing least-squares orthonormal vectors can be viewed as a special case of an orthogonal Procrustes problem [17,18,20,27,32]. In Section 3 we review these known results, and provide new insight into the optimal solution in the case in which the given vectors are linearly dependent. We refer to the closest orthonormal vectors as the orthonormal least-squares vectors (OLSV). Section 4 generalizes these results to allow for unequal weighting of the squared norms of the error vectors. The resulting vectors are referred to as the weighted orthonormal least-squares vectors (WOLSV).

Section 5 considers the problem of constructing orthogonal vectors that are not constrained to have equal norm, that are closest to the given vectors in a least-squares sense. The optimal vectors are referred to as the orthogonal least-squares vectors (OGLSV). Obtaining a closed form analytical expression for the optimal vectors in this case is in general a difficult problem. We first consider a special case for which an analytical solution is derived. We then propose an iterative algorithm to construct the OGLSV in the general case.

In Section 6 we consider the problem of constructing an optimal vector set $\{g_k, 1 \leqslant k \leqslant N\}$ with circulant inner product structure $\langle g_m, g_k \rangle = a_{k-m \bmod N}$. As we show, these vectors form a cyclic set; the optimal vectors are therefore referred to as the cyclic least-squares vectors (CLSV). Section 7 generalizes these results to allow for arbitrary permuted inner product shaping. In this case the vectors $\{g_k\}$ are constructed so that $\{\langle g_m, g_k \rangle, 1 \leqslant k \leqslant N\}$ is an arbitrary permutation of $\{\langle g_1, g_k \rangle, 1 \leqslant k \leqslant N\}$ for all m. We show that the constructed vectors are geometrically uniform (GU) [16]. The optimal vectors of this form are therefore referred to as the GU least-squares vectors (GULSV).

Extensions of this work to more general forms of least-squares inner product shaping are considered in [9, Chapter 8].

Before proceeding to the detailed development, in Section 2 we first provide an overview of the notation and a formulation of our problem.

2. Problem formulation

We denote vectors in a complex Hilbert space $\mathscr{H}$ by lowercase letters. Vectors in $\mathbb{C}^m$ are denoted by boldface lowercase letters, and matrices in $\mathbb{C}^{m \times m}$ by boldface uppercase letters. General linear transformations are denoted by uppercase letters.

The Frobenius norm of a linear transformation is denoted by $\|A\|_{\mathrm{F}}^2 = \mathrm{Tr}(A^*A)$. $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, $P_{\mathscr{U}}$ denotes the orthogonal projection onto the subspace $\mathscr{U}$, $(\cdot)^*$, $(\cdot)^\dagger$ and $\mathrm{Tr}(\cdot)$ denote the adjoint, the Moore–Penrose pseudoinverse, and the trace, respectively. δ_{km} denotes the Kronecker delta function where $\delta_{mk} = 1$ if $m = k$, and 0 otherwise.

To facilitate our derivations throughout the paper we introduce the following definition. Let $\{x_k, 1 \leqslant k \leqslant N\}$ denote a set of N vectors in $\mathscr{H}$. The *set transformation* (*ST*) $X: \mathbb{C}^N \to \mathscr{H}$ corresponding to these vectors is given by

$$X\mathbf{a} = \sum_{k=1}^{N} x_k a_k$$

for any vector $\mathbf{a} \in \mathbb{C}^N$, where a_k denotes the kth component of $\mathbf{a}$. From the definition of the adjoint $X^*: \mathscr{H} \to \mathbb{C}^N$, if $\mathbf{a} = X^*y$ then $a_k = \langle x_k, y\rangle$.

Suppose we are given a set of N vectors $\{s_k, 1 \leqslant k \leqslant N\}$ and a corresponding ST S in a complex Hilbert space $\mathscr{H}$, with inner product $\langle x, y\rangle$ for any $x, y \in \mathscr{H}$. The vectors $\{s_k\}$ span an M-dimensional subspace $\mathscr{U} \subseteq \mathscr{H}$. If the vectors are linearly independent, then $M = N$; otherwise $M < N$. Our objective is to construct a set of vectors with a specified inner product structure, from the given vectors $\{s_k, 1 \leqslant k \leqslant N\}$. We consider the following problems:

(1) Construct a set of orthonormal vectors $\{h_k, 1 \leqslant k \leqslant N\}$ so that $\langle h_m, h_k\rangle = \delta_{mk}$ (Section 3).
(2) Construct a set of orthogonal vectors $\{h_k, 1 \leqslant k \leqslant N\}$ so that $\langle h_m, h_k\rangle = c_k^2\delta_{mk}$ for some $c_k \geqslant 0$ (Section 5). We consider both the case of constrained norms (Section 5.1) and the case of unconstrained norms (Sections 5.2–5.4).
(3) Construct a set of cyclic vectors $\{g_k, 1 \leqslant k \leqslant N\}$ so that $\langle g_m, g_k\rangle = a_{k-m \bmod N}$ for some $\{a_k, 0 \leqslant k \leqslant N-1\}$. We consider both the case of constrained inner products (Section 6.1) and the case of unconstrained inner products (Section 6.2).
(4) Construct a set of GU vectors $\{g_k, 1 \leqslant k \leqslant N\}$ so that $\{\langle g_m, g_k\rangle, 1 \leqslant k \leqslant N\} = P\{a_k, 0 \leqslant k \leqslant N-1\}$ for all m, where $P\{a_k, 0 \leqslant k \leqslant N-1\}$ is a permutation of $\{a_k, 0 \leqslant k \leqslant N-1\}$ (Section 7).

In all the problems above we seek the vectors that are "closest" to the vectors s_k in the least-squares sense, so that they minimize the least-squares error

$$E = \sum_{k=1}^{N} \langle e_k, e_k\rangle = \|\Upsilon\|_{\mathrm{F}}^2 \tag{1}$$

subject to the appropriate constraint on the inner products. Here $e_k = s_k - h_k$ or $e_k = s_k - g_k$ and $\Upsilon = S - H$ or $\Upsilon = S - G$ is the ST corresponding to the vectors e_k, where H and G are the STs corresponding to the vectors h_k and g_k, respectively. The optimizing vectors that minimize E subject to the constraints (1)–(4) above are referred to, respectively, as the OLSV, OGLSV, CLSV, and GULSV.

Least-squares inner product shaping has potential applications to a variety of problems. One application, explored in [6], is to a detection problem in quantum mechanics. In this context, a set of orthonormal vectors defines a measurement that can be performed on a quantum system. The problem then is to construct a measurement, or equivalently a set of orthonormal vectors, that is optimized to distinguish between nonorthogonal vectors. It turns out that the OLSV lead to measurements that have many desirable properties.

Another application is to a generic classical detection problem in which one of a set of known signals is transmitted, and the objective is to detect the transmitted signal from the signal which has been received over an additive noise channel. A generic receiver typically used in such problems is the well-known matched filter (MF) receiver which consists of correlating the received signal with the possible transmitted signals. We can improve the performance over the MF receiver in many cases by correlating the received signal with a set of signals with a specified inner product structure, tailored to the specific problem, that are closest in a least-squares sense to the transmitted signals [7–9]. Similar applications to the problem of suppressing interference in multiuser wireless communication systems have also been explored [9–12].

Finally, we note that most signals used in digital communications are GU [16,33]. Such signal sets have strong symmetry properties that are desirable in various applications such as channel coding [16,29,33], and multiple description source coding [13,19]. It may therefore be useful to have a method for constructing optimal signal sets of this form.

3. Least-squares orthonormalization

In this section we consider the problem of constructing a set of orthonormal vectors $\{h_k, 1 \leqslant k \leqslant N\}$ with corresponding ST H, that minimize $E = \|S - H\|_F^2$ subject to $\langle h_m, h_k \rangle = \delta_{mk}$, or $H^*H = \mathbf{I}_N$. Thus H is constrained to be a partial isometry.[1]

If $\mathscr{H} = \mathbb{C}^n$ for some $n \geqslant N$, then $s_k = \mathbf{s}_k, h_k = \mathbf{h}_k$ are vectors in $\mathbb{C}^n$, and $S = \mathbf{S}, H = \mathbf{H}$ are the $n \times N$ matrices of columns $\mathbf{s}_k$ and $\mathbf{h}_k$, respectively. The least-squares orthonormalization problem reduces to finding the partial isometry $\mathbf{H}$ that minimizes $\|\mathbf{S} - \mathbf{H}\|_F^2$. The optimal $\mathbf{H}$, denoted $\widehat{\mathbf{H}}$, is well known (see, e.g., [15,24,34]) and is the partial isometry in the polar decomposition [24,26] of $\mathbf{S}$. The OLSV are then the columns of $\widehat{\mathbf{H}}$.

Let $\mathbf{S}$ have a singular value decomposition (SVD) $\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$, where $\mathbf{U}$ is an $n \times N$ partial isometry, $\boldsymbol{\Sigma}$ is an $N \times N$ diagonal matrix and $\mathbf{V}$ is an $N \times N$ unitary matrix. Then

$$\widehat{\mathbf{H}} = \mathbf{U}\mathbf{V}^*.$$

[1] A partial isometry is a transformation T that satisfies $T^*T = \mathbf{I}$.

If the vectors $\mathbf{s}_k$ are linearly independent, then

$$\widehat{\mathbf{H}} = \mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1/2}.$$

The OLSV can also be obtained as the solution to an *orthogonal Procrustes problem* [18,20,32]. In this problem, we are given two $n \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$, and we want to rotate $\mathbf{B}$ into $\mathbf{A}$ by seeking a unitary matrix $\widehat{\mathbf{Z}}$ to minimize $\|\mathbf{A} - \mathbf{B}\mathbf{Z}\|_{\mathrm{F}}^2$. By choosing $\mathbf{A} = \mathbf{S}$ and $\mathbf{B}$ as a matrix whose columns form an orthonormal basis for $\mathscr{U}$, the minimization problem of (1) reduces to an orthogonal Procrustes problem. The OLSV are then the columns of $\mathbf{B}\widehat{\mathbf{Z}}$.

If $\mathscr{H}$ is an arbitrary Hilbert space, then the optimal ST $\widehat{H}$ can be found by identifying $\mathscr{H}$ with $\mathbb{C}^N$ using an appropriate choice of coordinates, resulting in

$$\widehat{H} = U\mathbf{V}^*,$$

where now $S = U\mathbf{\Sigma}\mathbf{V}^*$ is the SVD of S, where $U: \mathbb{C}^N \to \mathscr{H}$ is a partial isometry, and $\mathbf{\Sigma}$ and $\mathbf{V}$ are defined as before. The OLSV are then $\hat{h}_k = \widehat{H}\mathbf{i}_k$, where $\mathbf{i}_k \in \mathbb{C}^N$ is the vector with mth component δ_{mk}. If the vectors s_k are linearly independent, then

$$\widehat{H} = S(S^*S)^{-1/2}.$$

Evidently, the OLSV $\{\hat{h}_k\}$ have the property that they do not depend on the order of the vectors $\{s_k\}$.

From the properties of the polar decomposition it follows that $\widehat{H}$ is unique if and only if the vectors s_k are linearly independent [15,34]. We now try to gain some insight into the OLSV in the linearly dependent case. Our problem is to find a set of N orthonormal vectors that are as close as possible to the vectors $\{s_k, 1 \leqslant k \leqslant N\}$ that span an M-dimensional subspace $\mathscr{U}$. When $M < N$, there are at most M orthonormal vectors in $\mathscr{U}$. Therefore, the optimal orthonormal vectors h_k must lie partly in the orthogonal complement $\mathscr{U}^\perp$, and consequently span an N-dimensional subspace $\mathscr{V}$, where $\mathscr{U} \subset \mathscr{V}$. Each vector has a component in $\mathscr{U}$, $h_k^{\mathscr{U}} = P_{\mathscr{U}} h_k$, and a component in $\mathscr{U}^\perp$, $h_k^{\mathscr{U}^\perp} = P_{\mathscr{U}^\perp} h_k$. We may now rewrite the error E of (1) as

$$\begin{aligned} E &= \sum_{k=1}^{N} \left\langle s_k - h_k^{\mathscr{U}} - h_k^{\mathscr{U}^\perp}, s_k - h_k^{\mathscr{U}} - h_k^{\mathscr{U}^\perp} \right\rangle \\ &= \sum_{k=1}^{N} \left(\left\langle s_k - h_k^{\mathscr{U}}, s_k - h_k^{\mathscr{U}} \right\rangle + \left\langle h_k^{\mathscr{U}^\perp}, h_k^{\mathscr{U}^\perp} \right\rangle \right), \end{aligned}$$

since $\langle s_k - h_k^{\mathscr{U}}, h_k^{\mathscr{U}^\perp} \rangle = 0$. For any choice of orthonormal vectors h_k,

$$\sum_{k=1}^{N} \left\langle h_k^{\mathscr{U}}, h_k^{\mathscr{U}} \right\rangle = \mathrm{Tr}\left(\left(H^{\mathscr{U}}\right)^* H^{\mathscr{U}} \right) = \mathrm{Tr}\left(P_{\mathscr{U}} H H^*\right) = \mathrm{Tr}\left(P_{\mathscr{U}} P_{\mathscr{V}}\right) = \mathrm{Tr}\left(P_{\mathscr{U}}\right) = M,$$

where $H^{\mathscr{U}} = P_{\mathscr{U}} H$ is the ST corresponding to the vectors $h_k^{\mathscr{U}} = P_{\mathscr{U}} h_k$. So

$$\sum_{k=1}^{N}\left\langle h_k^{\mathscr{U}^\perp}, h_k^{\mathscr{U}^\perp}\right\rangle = \sum_{k=1}^{N}\left(\langle h_k, h_k\rangle - \left\langle h_k^{\mathscr{U}}, h_k^{\mathscr{U}}\right\rangle\right) = N - M.$$

Thus, minimization of E is equivalent to minimization of

$$E' = \sum_{k=1}^{N}\left\langle s_k - h_k^{\mathscr{U}}, s_k - h_k^{\mathscr{U}}\right\rangle.$$

We conclude that when the vectors s_k are linearly dependent, choosing a set of orthonormal vectors which are as close as possible to the vectors s_k is equivalent to choosing a set of orthonormal vectors whose projections onto $\mathscr{U}$ are closest to these vectors.[2] The closest projections are unique, and are the vectors corresponding to

$$\widehat{H}^{\mathscr{U}} = P_{\mathscr{U}}\widehat{H} = U\tilde{\mathbf{I}}_M\mathbf{V}^* = S\left((S^*S)^{1/2}\right)^{\dagger}, \tag{2}$$

where $\tilde{\mathbf{I}}_M$ is a diagonal matrix whose first M diagonal elements are equal to 1 and whose remaining diagonal elements are equal to 0.

The optimal vectors $\hat{h}_k$ satisfy

$$\langle \hat{h}_k, s_k\rangle = \langle P_{\mathscr{U}}\hat{h}_k, s_k\rangle = [\widehat{H}^* P_{\mathscr{U}} S]_{kk} = \alpha[S^*S]_{kk}^{1/2}, \tag{3}$$

where $[\cdot]_{mk}$ denotes the mkth element of the corresponding matrix. This relation may be used to derive bounds on the inner products $\langle \hat{h}_k, s_k\rangle$ in terms of the inner products $\langle s_k, s_m\rangle$; see [21].

We summarize our results regarding the OLSV in the following theorem.

Theorem 1 (Orthonormal least-squares vectors (OLSV)). *Let $\{s_k, 1 \leqslant k \leqslant N\}$ be a set of N vectors in a Hilbert space $\mathscr{H}$, that span an M-dimensional subspace $\mathscr{U} \subseteq \mathscr{H}$. Let $\{\hat{h}_k, 1 \leqslant k \leqslant N\}$ denote N optimal vectors that are orthonormal and minimize the least-squares error defined by* (1). *Let $S = U\mathbf{\Sigma}\mathbf{V}^*$ and $\widehat{H}$ denote the set transformations corresponding to the vectors s_k and $\hat{h}_k$, respectively. Then the optimal $\widehat{H}$ can be chosen as*

$$\widehat{H} = U\mathbf{V}^*.$$

The corresponding vectors $\{\hat{h}_k = \widehat{H}\mathbf{i}_k, 1 \leqslant k \leqslant N\}$ are defined as the OLSV. In addition,

(1) *If $M = N$, then*
  (a) $\widehat{H} = S(S^*S)^{-1/2}$;
  (b) *the OLSV are the unique orthonormal vectors that minimize the least-squares error.*

(2) *If $M < N$, then*

[2] These projections constitute a normalized tight frame [1] for $\mathscr{U}$. The problem of least-squares tight frame construction is also considered in [14].

(a) *the projection of* $\widehat{H}$ *onto* $\mathscr{U}$ *is unique and is given by* $P_{\mathscr{U}}\widehat{H} = S((S^*S)^{1/2})^{\dagger}$;
(b) *the vectors* $\{P_{\mathscr{U}}\hat{h}_k, 1 \leqslant k \leqslant N\}$ *are the closest projections onto* $\mathscr{U}$ *of a set of orthonormal vectors to the vectors* $\{s_k, 1 \leqslant k \leqslant N\}$, *in the least-squares sense.*

4. Weighted least-squares orthonormalization

In Section 3 we sought a set of orthonormal vectors $\{h_k, 1 \leqslant k \leqslant N\}$ to minimize the sum of the squared errors (1), where we are assigning equal weights to the different errors. We now seek the orthonormal vectors h_k that minimize a weighted squared error,

$$E_w = \sum_{k=1}^{N} w_k \langle s_k - h_k, s_k - h_k \rangle, \tag{4}$$

where $w_k \geqslant 0$ is the weight given to the kth squared norm error.

To derive the solution to this minimization problem let $S_w = S\mathbf{W}$, where $\mathbf{W}$ is an $N \times N$ diagonal matrix with diagonal elements w_k. Then

$$\begin{aligned} E_w &= \mathrm{Tr}\left((S-H)^*(S-H)\mathbf{W}\right) \\ &= \mathrm{Tr}\left((S_w-H)^*(S_w-H)\right) + \mathrm{Tr}\left((\mathbf{W}-I_N)H^*H\right) \\ &\quad + \mathrm{Tr}\left(\mathbf{W}(I_N-\mathbf{W})S^*S\right) \\ &= \mathrm{Tr}\left((S_w-H)^*(S_w-H)\right) + K, \end{aligned}$$

where K is independent of the choice of H, since $H^*H = \mathbf{I}_N$ for any H. Thus minimizing E_w is equivalent to minimizing $E'_w = \|S_w - H\|_{\mathrm{F}}^2$, which is equivalent to least-squares minimization (1), with S_w substituted for S.

Employing the SVD $S_w = U_w \mathbf{\Sigma}_w \mathbf{V}_w^*$, the optimal ST $\widehat{H}_w$ follows from Theorem 1,

$$\widehat{H}_w = U_w \mathbf{V}_w^*.$$

If the vectors s_k are linearly independent and $w_k > 0$ for all k, then

$$\widehat{H}_w = S_w\left(S_w^* S_w\right)^{-1/2} = S\mathbf{W}(\mathbf{W}^*S^*S\mathbf{W})^{-1/2}.$$

The WOLSV are then given by $\hat{h}_k^w = \widehat{H}_w \mathbf{i}_k$.

5. Least-squares orthogonalization

We now consider the problem of constructing optimal *orthogonal* vectors $\{h_k, 1 \leqslant k \leqslant N\}$ from the given vectors $\{s_k, 1 \leqslant k \leqslant N\}$. We may wish to constraint the vectors $\{h_k\}$ to have some specified norm, e.g., $\langle h_k, h_k \rangle = \langle s_k, s_k \rangle$, or we may choose the vectors $\{h_k\}$ to be orthogonal and to minimize the least-squares error $E = \|S - H\|_{\mathrm{F}}^2$. In this case $\langle h_k, h_k\rangle$ will be such that E is minimized.

5.1. Orthogonalization with constrained norms

We first consider the problem of constructing a set of vectors $\{h_k, 1 \leqslant k \leqslant N\}$ that minimize $E = \|S - H\|_{\mathrm{F}}^2$ subject to the constraint

$$\langle h_m, h_k\rangle = c_k^2 \delta_{mk},$$

where the scalars $c_k \geqslant 0$ are specified. We assume without loss of generality that $c_k > 0$ for $1 \leqslant k \leqslant P$. We may now write the error E as

$$E = \sum_{k=1}^{N} \langle s_k - h_k, s_k - h_k\rangle = \sum_{k=1}^{N} c_k^2 \langle \tilde{s}_k - \tilde{h}_k, \tilde{s}_k - \tilde{h}_k\rangle, \tag{5}$$

where the vectors $\tilde{h}_k$ are orthonormal vectors such that $\tilde{h}_k = (1/c_k)h_k$, $1 \leqslant k \leqslant P$, and the vectors $\tilde{s}_k$ are defined by $\tilde{s}_k = (1/c_k)s_k$, $1 \leqslant k \leqslant P$ and $\tilde{s}_k = 0$, $P + 1 \leqslant k \leqslant N$. Comparing (5) with (4), the orthonormal vectors $\hat{\tilde{h}}_k$ that minimize (5) are the WOLSV corresponding to the vectors $\{\tilde{s}_k\}$ with weights $w_k = c_k^2$. Thus, $\hat{\tilde{h}}_k = \widehat{\widetilde{H}}\mathbf{i}_k$, where

$$\widehat{\widetilde{H}} = \widetilde{U}\widetilde{\mathbf{V}}^*.$$

Here $\widetilde{S}\mathbf{C}^2 = \widetilde{U}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^*$ is the SVD of $\widetilde{S}\mathbf{C}^2$, $\widetilde{S} = S\mathbf{C}^{\dagger}$ is the ST corresponding to the vectors $\tilde{s}_k$, and $\mathbf{C}$ is the diagonal matrix with diagonal elements c_k. The OGLSV are then $\hat{h}_k = c_k\hat{\tilde{h}}_k = \widehat{H}\mathbf{i}_k$, where

$$\widehat{H} = \widetilde{U}\widetilde{\mathbf{V}}^*\mathbf{C}. \tag{6}$$

Note that the vectors $\hat{h}_k$ lie in the space $\mathscr{V}$ spanned by the first P columns of $\widetilde{U}$. This follows from the fact that with $T = S\mathbf{C}$, T^*T is a block diagonal matrix whose lower $(N - P) \times (N - P)$ block is a 0 matrix with all 0 entries, so that $\widetilde{\mathbf{V}}$ is also a block diagonal matrix. Then the the last $N - P$ elements of each of the first P columns of the matrix $\widetilde{\mathbf{V}}\mathbf{C}$ are all equal 0, and the remaining columns of $\widetilde{\mathbf{V}}\mathbf{C}$ are all 0.

Therefore, as we now show, if $P \leqslant M$, then we can always choose $\widehat{H}$ so that the vectors $\hat{h}_k$ lie in the M-dimensional space $\mathscr{U}$ spanned by the vectors s_k. Specifically, if $L = \mathrm{rank}(S\mathbf{C})$ is equal to P, then $\mathscr{V}$ is equal to the space spanned by the first P columns of $S\mathbf{C}$ so that $\mathscr{V} \subseteq \mathscr{U}$. Since $\hat{h}_k \in \mathscr{V}$, it follows immediately that $\hat{h}_k \in \mathscr{U}$. If $L < P$ then only the first L columns of $\widetilde{\mathscr{U}}$ are specified, and the remaining columns can be chosen arbitrarily. Since these L columns span a subset of $\mathscr{U}$, we can always choose the next $P - L$ columns so that the space $\mathscr{V}$ spanned by the first P columns of $\widetilde{\mathscr{U}}$ is also a subset of $\mathscr{U}$. Then since $\hat{h}_k \in \mathscr{V}$ and $\mathscr{V} \subseteq \mathscr{U}$, $\hat{h}_k \in \mathscr{U}$.

If the vectors s_k are linearly independent and $c_k > 0$ for all k, then

$$\widehat{H} = S\mathbf{C}(\mathbf{C}S^*S\mathbf{C})^{-1/2}\mathbf{C}. \tag{7}$$

5.2. Unconstrained orthogonalization

We now consider the orthogonalization problem with unconstrained norms. Thus, we seek a set of vectors h_k that minimize $E = \|S - H\|_{\mathrm{F}}^2$ subject to

$$\langle h_m, h_k\rangle = 0 \quad \text{for } m \neq k.$$

Expressing the error E as

$$E = \sum_{k=1}^{N} \left(\langle s_k, s_k\rangle + \langle h_k, h_k\rangle - 2\Re\left\{\langle h_k, s_k\rangle\right\}\right),$$

and denoting $\tilde{h}_k = (1/b_k)h_k$, where $b_k^2 = \langle h_k, h_k\rangle$, it follows that minimization of E is equivalent to minimization of

$$E' = \sum_{k=1}^{N} \left(\langle h_k, h_k\rangle - 2\Re\left\{\langle h_k, s_k\rangle\right\}\right) = \sum_{k=1}^{N} \left(b_k^2 - 2b_k\Re\left\{\langle \tilde{h}_k, s_k\rangle\right\}\right). \tag{8}$$

To determine the optimal vectors h_k we have to minimize E' with respect to b_k and $\tilde{h}_k$. Fixing $\tilde{h}_k$ and minimizing with respect to b_k, the optimal value of b_k, denoted $\hat{b}_k$, is given by

$$\hat{b}_k = \Re\left\{\langle \tilde{h}_k, s_k\rangle\right\}. \tag{9}$$

Substituting $\hat{b}_k$ back into (8), the vectors $\tilde{h}_k$ are chosen to maximize

$$R_{hs} = \sum_{k=1}^{N} \Re^2\left\{\langle \tilde{h}_k, s_k\rangle\right\} \tag{10}$$

subject to the constraint $\langle \tilde{h}_m, \tilde{h}_k\rangle = \delta_{mk}$.

Obtaining a closed form analytical expression for the orthonormal vectors $\tilde{h}_k$ that maximize (10) is a difficult problem (in the general case). In fact, we can show that this problem is similar to a detection problem in quantum mechanics for which there is no known analytical solution in general [6,22]. However, as we show in Section 5.3, in the special case where the vectors $\{s_k\}$ have a symmetry property referred to as geometric uniformity [16], an analytical solution can be obtained. An iterative algorithm for constructing a set of orthonormal vectors that maximize (10) for arbitrary vectors $\{s_k\}$ is considered in Section 5.4.

5.3. Maximizing R_{hs} for geometrically uniform vector sets

To obtain a more convenient expression for R_{hs}, let $S = U\mathbf{\Sigma}\mathbf{V}^*$ and $\widetilde{H}$ denote the STs corresponding to s_k and $\tilde{h}_k$, respectively. Since $\{\tilde{h}_k\}$ are proportional to the vectors closest to the vectors $\{s_k\}$, the space spanned by the vectors s_k is a subspace of the space spanned by the vectors $\tilde{h}_k$. In addition, the vectors $\tilde{h}_k$ are orthonormal. Consequently, $\widetilde{H}$ has an SVD of the form $\widetilde{H} = U\mathbf{Q}$ for some $N \times N$ unitary matrix $\mathbf{Q}$. Let $\mathbf{v}_k$ and $\mathbf{q}_k$ denote the columns of $\mathbf{V}^*$ and $\mathbf{Q}^*$, respectively. Then we can express R_{hs} as

$$R_{hs} = \sum_{k=1}^{N} \Re^2\{\langle \tilde{h}_k, s_k\rangle\} = \sum_{k=1}^{N} \Re^2\{\langle U^*\tilde{h}_k, U^*s_k\rangle\} = \sum_{k=1}^{N} \Re^2\{\langle \mathbf{q}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle\}.$$

Our problem then reduces to finding a set of orthonormal vectors $\mathbf{q}_k$ that maximize $\sum_{k=1}^{N} \Re^2\{\langle \mathbf{q}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle\}$, where the vectors $\mathbf{v}_k$ are also orthonormal. Using the Cauchy–Schwarz inequality,

$$\begin{aligned} R_{hs} &= \sum_{k=1}^{N} \Re^2\{\langle \mathbf{\Sigma}^{1/2}\mathbf{q}_k, \mathbf{\Sigma}^{1/2}\mathbf{v}_k\rangle\} \leqslant \sum_{k=1}^{N} \left|\langle \mathbf{\Sigma}^{1/2}\mathbf{q}_k, \mathbf{\Sigma}^{1/2}\mathbf{v}_k\rangle\right|^2 \\ &\leqslant \sum_{k=1}^{N} \langle \mathbf{q}_k, \mathbf{\Sigma}\mathbf{q}_k\rangle\langle \mathbf{v}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle, \end{aligned} \tag{11}$$

with equality if and only if $\mathbf{\Sigma}^{1/2}\mathbf{q}_k = x_k\mathbf{\Sigma}^{1/2}\mathbf{v}_k$ for some $x_k > 0$ and all k. In particular, we have equality for $\mathbf{q}_k = \mathbf{v}_k$. In general the bound of (11) depends on the unknown vectors $\mathbf{q}_k$. However, in the special case where $\langle \mathbf{v}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle = C$ independent of k, (11) reduces to

$$R_{hs} \leqslant C \sum_{k=1}^{N} \langle \mathbf{q}_k, \mathbf{\Sigma}\mathbf{q}_k\rangle = C\,\mathrm{Tr}(\mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^*) = C\,\mathrm{Tr}(\mathbf{\Sigma})$$

with equality if $\mathbf{v}_k = \mathbf{q}_k$. Since $\langle \mathbf{v}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle$ is the kth diagonal element of $(S^*S)^{1/2}$ we conclude that if the diagonal elements of $(S^*S)^{1/2}$ are equal, then the optimal orthonormal vectors maximizing R_{hs} correspond to $U\mathbf{V}^*$, and are therefore just the OLSV with respect to $\{s_k\}$. We note that like the OLSV, if the vectors $\{s_k\}$ are linearly dependent, then the vectors maximizing R_{hs} are not unique. However, their projections onto $\mathscr{U}$ are always unique.

In [6] it was shown that when the vectors $\{s_k\}$ are geometrically uniform (GU),[3] the components of the vectors $\mathbf{v}_k$ have equal magnitude $1/N$ for all k, so that $\langle \mathbf{v}_k, \mathbf{\Sigma}\mathbf{v}_k\rangle = 1/N \sum_j \sigma_j$ independent of k. Therefore, in this case the OLSV max-

---

[3] A set of vectors is GU if given any two vectors s_k and s_l in the set, there is an isometry T_{kl} that transforms s_k into s_l while leaving the set invariant [16]. A more detailed discussion on GU vector sets is given in Section 7.

imize R_{hs}. From (9), $\hat{b}_k = [(S^*S)^{1/2}]_{kk} = (1/N)\sum_j \sigma_j$ for all k so that the optimal orthogonal vectors have equal norm.

5.4. Iterative algorithm maximizing R_{hs} for arbitrary vector sets

For simplicity of exposition, we assume throughout this section that $\mathscr{H} = \mathscr{R}^n$ for some $n \geqslant N$ so that $R_{hs} = \sum_{k=1}^{N} \langle \tilde{\mathbf{h}}_k, \mathbf{s}_k \rangle^2$.

The proposed algorithm proceeds as follows. Starting with an arbitrary matrix with orthonormal columns, at each iteration we construct a new matrix with orthonormal columns by multiplying the current matrix by an orthogonal matrix. The orthogonal matrix is chosen so that R_{hs} does not decrease form iteration to iteration, where at each iteration R_{hs} is computed using the orthonormal columns of the new matrix. Since R_{hs} is bounded above for any choice of orthonormal vectors, the iterations are guaranteed to converge.

The algorithm is initialized by choosing an arbitrary matrix $\mathbf{H}^{(0)}$ with orthonormal columns $\mathbf{h}^{(0)}$. A good choice is the matrix $\mathbf{H}^{(0)} = \mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1/2}$, where $\mathbf{S}$ is the matrix of columns $\mathbf{s}_k$, So that the columns $\mathbf{h}_k^{(0)}$ are the closest orthonormal vectors in a least-squares sense to the vectors $\mathbf{s}_k$. Then, for $j = 0, 1, 2, \ldots,$ we choose an orthogonal matrix $\mathbf{Q}^{(j)}$, and set $\mathbf{H}^{(j+1)} = \mathbf{H}^{(j)}\mathbf{Q}^{(j)}$. If the columns of $\mathbf{H}^{(j)}$ are orthonormal, and $\mathbf{Q}^{(j)}$ is an orthogonal matrix, then the columns of $\mathbf{H}^{(j+1)}$ are also orthonormal. Since the columns of $\mathbf{H}^{(0)}$ are orthonormal, $\mathbf{H}^{(j)}$ will have orthonormal columns for all j.

Suppose we can choose $\mathbf{Q}^{(j)}$ so that for all j,

$$R_{hs}^{(j+1)} \geqslant R_{hs}^{(j)}. \tag{12}$$

Here $R_{hs}^{(j)} = \sum_{k=1}^{N} (d_{kk}^{(j)})^2$ where $d_{kk}^{(j)} = \langle \mathbf{h}_k^{(j)}, \mathbf{s}_k \rangle$, and $\mathbf{h}_k^{(j)}$ denotes the kth column of $\mathbf{H}^{(j)}$. Then, since $R_{hs}^{(j+1)} \leqslant \sum_{k=1}^{N} \langle s_k, s_k \rangle$ for all j, the iterations are guaranteed to converge to a local maximum.

Thus, the crux of the algorithm is choosing $\mathbf{Q}^{(j)}$ so that (12) is satisfied for all j. This can be accomplished by choosing $\mathbf{Q}^{(j)}$ to be an "optimal" *Givens rotation* [17]. A Givens rotation $\mathbf{J}(r, l, \theta)$ is an orthogonal matrix that is equal to the identity matrix except for the entries

$$\begin{bmatrix} j_{rr} & j_{rl} \\ j_{lr} & j_{ll} \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \triangleq \mathbf{J}(\theta), \tag{13}$$

where j_{mk} is the mkth element of $\mathbf{J}(r, l, \theta)$, $c = \cos(\theta)$, $s = \sin(\theta)$, $1 \leqslant r, l \leqslant N$ and $r \neq l$.

Now, let $\mathbf{Q}^{(j)} = \mathbf{J}(r, l, \hat{\theta})$, where $\hat{\theta}$ is chosen to maximize $R_{hs}^{(j+1)}$. Since for $\theta = 0$, $\mathbf{J}(r, l, 0) = \mathbf{I}_N$ and $R_{hs}^{(j+1)} = R_{hs}^{(j)}$, we are guaranteed that for an optimal choice of θ, (12) is satisfied. Note that if $\mathbf{H}^{(j+1)} = \mathbf{H}^{(j)}\mathbf{J}(r, l, \theta)$, then $d_{kk}^{(j+1)} = d_{kk}^{(j)}$ for $k \neq l, r$. Therefore, choosing θ to maximize $R_{hs}^{(j+1)}$ is equivalent to choosing θ to maximize

$$R_{hs}'^{(j+1)} = \left(d_{ll}^{(j+1)}\right)^2 + \left(d_{rr}^{(j+1)}\right)^2.$$

Let j, r, l be fixed, and let $\mathbf{D}^{(j)} = (\mathbf{H}^{(j)})^*\mathbf{S}$. Denote by $\mathbf{D}'^{(j)}$ the 2×2 matrix

$$\mathbf{D}'^{(j)} = \begin{bmatrix} \mathbf{D}^{(j)}(r,r) & \mathbf{D}^{(j)}(r,l) \\ \mathbf{D}^{(j)}(l,r) & \mathbf{D}^{(j)}(l,l) \end{bmatrix} \triangleq \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Then $R_{hs}'^{(j+1)}$ is equal to the sum of the squares of the diagonal elements of $\mathbf{D}'^{(j+1)} = \mathbf{J}^*(\theta)\mathbf{D}'^{(j)}$, where $\mathbf{J}(\theta)$ is defined by (13). Thus,

$$R_{hs}'^{(j+1)} = (cb_{11} + sb_{21})^2 + (cb_{22} - sb_{12})^2.$$

In Appendix A we show that $R_{hs}'^{(j+1)}$ is maximized when

$$\hat{\theta} = \begin{cases} \frac{1}{2}\tan^{-1}\left(\frac{y}{x}\right) \text{ and } x\cos(2\hat{\theta}) > 0, & x \neq 0, \\ \pi/4, & x = 0, y > 0, \\ -\pi/4, & x = 0, y < 0, \\ 0, & x = y = 0, \end{cases} \tag{14}$$

where $x = b_{11}^2 + b_{22}^2 - b_{21}^2 - b_{12}^2$ and $y = b_{11}b_{21} - b_{22}b_{12}$.

The iterations are continued until convergence, where in each iteration we choose different values of r and l.

We now summarize the iterative algorithm:

(1) Choose the vectors $\mathbf{h}_k^{(0)}$ as the columns of $\mathbf{H}^{(0)} = \mathbf{S}(\mathbf{S}^*\mathbf{S})^{-1/2}$, and set $j = 0$;
(2) choose r and l arbitrarily so that $1 \leqslant r, l \leqslant N$ and $r \neq l$;
(3) compute $b_{11} = \langle \mathbf{h}_r^{(j)}, \mathbf{s}_r \rangle$, $b_{12} = \langle \mathbf{h}_r^{(j)}, \mathbf{s}_l \rangle$, $b_{21} = \langle \mathbf{h}_l^{(j)}, \mathbf{s}_r \rangle$ and $b_{22} = \langle \mathbf{h}_l^{(j)}, \mathbf{s}_l \rangle$;
(4) compute $\mathbf{H}^{(j+1)} = \mathbf{H}^{(j)}\mathbf{J}(r, l, \hat{\theta})$, where $\hat{\theta}$ is given by (14);
(5) set $j = j + 1$ and go to step (2).

Our iterative algorithm can be shown to be equivalent to the algorithm developed in [23] in the context of quantum detection, which is derived using quantum-mechanical ideas and concepts. However, since our algorithm does not invoke such considerations, its derivation is more straightforward.

Based on results derived in that context it can be shown that the vectors $\mathbf{h}_k$ maximizing R_{hs} are unique (up to multiplication by -1) [22,23]. Furthermore, the optimal vectors $\mathbf{h}_k$ must be such that the matrix Ψ, defined by $[\Psi]_{ik} = \langle \mathbf{h}_i, \mathbf{s}_k \rangle \langle \mathbf{h}_k, \mathbf{s}_k \rangle$, is nonnegative definite [23]. Therefore, upon convergence of the algorithm we can test whether or not the vectors $\mathbf{h}_k$ are the optimal vectors maximizing R_{hs} or whether the algorithm converged to a local maximum, by checking if Ψ is nonnegative definite. If the algorithm converged to a local minimum, then we may slightly rotate the matrix $\mathbf{H}$ of columns $\mathbf{h}_k$, i.e., multiply $\mathbf{H}$ by a unitary matrix $\mathbf{U}$, such that the rotated vectors $\mathbf{h}_k' = \mathbf{U}\mathbf{h}_k$ result in a higher R_{hs}; these vectors form the initial conditions $\mathbf{h}_k^{(0)}$ for resumption of the main algorithm. A formula for determining $\mathbf{U}$ can be found in [23, Appendix I]. We note that when the initial conditions for the algorithm are chosen as the OLSV, $R_{hs}^{(0)}$ will be pretty close to the maximal value, and the algorithm is unlikely to converge to a local maximum.

6. Least-squares circulant inner product shaping

In the previous sections we discussed various forms of least-squares orthogonalization, which can be viewed as a special case of the more general problem of constructing least-squares vectors with a specified inner product structure. We may view the least-squares inner product shaping problem as a Procrustes problem described in Section 3, with $\mathbf{B} = \mathbf{I}$ and a suitable constraint on the Gram matrix $\mathbf{Z}^*\mathbf{Z}$. Problems of this form with different constraints on $\mathbf{Z}$ have been studied previously in the literature [18]. For example, in [2] the authors consider the case where $\mathbf{Z}$ is constrained to have one of the following forms: Nonnegative elements, symmetric and nonnegative elements, symmetric nonnegative definite (also considered in [25]), or a specific set of nonzero elements. In [30] the matrix $\mathbf{Z}$ is constrained to be normal, and in [4,17] the columns of $\mathbf{Z}$ are constrained to be normalized. The problem of constrained inner products corresponding to a constraint on the off-diagonal elements of $\mathbf{Z}^*\mathbf{Z}$ has not been previously addressed in the literature.

In this section we consider circulant inner product shaping, in which we seek a set of vectors $\{g_k, 1 \leqslant k \leqslant N\}$ corresponding to an ST G that minimizes $\|S - G\|_{\mathrm{F}}^2$, and such that $\langle g_m, g_k\rangle$ depends only on $k - m \bmod N$. Then $\{\langle g_m, g_k\rangle, 1 \leqslant k \leqslant N\}$ is a cyclic permutation of $\{\langle g_1, g_k\rangle, 1 \leqslant k \leqslant N\}$ for all m, and the Gram matrix G^*G is circulant.[4] We may wish to constraint the values $\{\langle g_1, g_k\rangle, 1 \leqslant k \leqslant N\}$, or we may choose these values so that the least-squares error is minimized.

6.1. Constrained circulant inner product shaping

We first consider the problem of minimizing

$$E = \sum_{k=1}^{N} \langle s_k - g_k, s_k - g_k\rangle = \|S - G\|_{\mathrm{F}}^2 \tag{15}$$

subject to

$$\langle g_m, g_k\rangle = a_{k-m \bmod N}, \tag{16}$$

where the sequence $\{a_k, 0 \leqslant k \leqslant N - 1\}$ is specified. The minimizing vectors are referred to as the cyclic least-squares vectors (CLSV). This terminology will be justified in Section 6.1.1 where we show that a set of vectors with inner products satisfying (16) forms a cyclic set. The CLSV are then derived in Section 6.1.2.

[4] A circulant matrix is a matrix where every row (or column) is obtained by a right circular shift (by one position) of the previous row (or column).

6.1.1. Cyclic vector sets

Definition 1. Let $\{g_k, 1 \leqslant k \leqslant N\}$ be a set of vectors that span an M-dimensional subspace $\mathscr{V}$. Then the vectors g_k form a *cyclic set* if $\mathbf{g}_k = \mathbf{U}^{k-1}\mathbf{g}_1$ for $1 \leqslant k \leqslant N$, where $\mathbf{g}_k$ is the representation of g_k in an orthonormal basis for $\mathscr{V}$, and $\mathbf{U}$ is an $M \times M$ unitary matrix with $\mathbf{U}^N = \mathbf{I}$.

We note that the choice of basis in Definition 1 is immaterial; if the representation of g_k in some orthonormal basis for $\mathscr{V}$ satisfies the required conditions, then the representation of g_k in any orthonormal basis for $\mathscr{V}$ will satisfy these conditions since any two representations are related by a unitary transformation.

Vector sets with circulant inner product structure are conveniently characterized by the following theorem.

Theorem 2. *A set of vectors $\{g_k, 1 \leqslant k \leqslant N\}$ has a circulant inner product structure if and only if the vectors form a cyclic set.*

The proof of this theorem relies on the following well-known result.

Lemma 1. *Let G be a set transformation corresponding to vectors $\{g_k, 1 \leqslant k \leqslant N\}$. Then G^*G is circulant if and only if the vectors $\{\overline{g}_k, 1 \leqslant k \leqslant N\}$ corresponding to $\overline{G} = G\mathbf{F}$ are orthogonal, where $\mathbf{F}$ is the $N \times N$ Fourier transform (FT) matrix with mk*th *element $\mathbf{F}_{mk} = 1/\sqrt{N}\mathrm{e}^{-\mathrm{j}2\pi(m-1)(k-1)/N}$. In this case $\langle \overline{g}_k, \overline{g}_k\rangle = A_k$, where $A_k = \sum_{l=0}^{N-1} a_l \mathrm{e}^{-\mathrm{j}2\pi(k-1)l/N}$.*

Proof of Theorem 2. Suppose the vectors g_k form a cyclic set. Then

$$\langle g_m, g_k\rangle = \langle \mathbf{g}_m, \mathbf{g}_k\rangle = \left\langle \mathbf{U}^{m-1}\mathbf{g}_1, \mathbf{U}^{k-1}\mathbf{g}_1\right\rangle = \left\langle \mathbf{g}_1, (\mathbf{U}^{m-1})^*\mathbf{U}^{k-1}\mathbf{g}_1\right\rangle.$$

Since $\mathbf{U}$ is unitary, $(\mathbf{U}^{m-1})^* = \mathbf{U}^{1-m}$ and $\langle g_m, g_k\rangle = \langle \mathbf{g}_1, \mathbf{U}^{k-m}\mathbf{g}_1\rangle$, which depends only on $k-m$ mod N since $\mathbf{U}^{k-m} = \mathbf{U}^{k-m \bmod N}$.

Now, let the vectors g_k corresponding to G have a circulant inner product structure. Then using Lemma 1 we may decompose G as $G = Q\mathbf{A}\mathbf{F}^*$, where Q is an ST corresponding to an orthonormal set of vectors q_k, and $\mathbf{A}$ is a diagonal $N \times N$ matrix with kth diagonal element $\sqrt{A_k}$. Then $g_k = G\mathbf{i}_k = Q\mathbf{A}\mathbf{F}^*\mathbf{i}_k = Q\mathbf{A}\mathbf{f}_k$, where $\mathbf{f}_k$ denotes the kth column of $\mathbf{F}^*$. From the definition of $\mathbf{F}$, $\mathbf{f}_k = \mathbf{D}_k\mathbf{f}_1$ where $\mathbf{D}_k$ is a diagonal matrix with mth diagonal element $\mathrm{e}^{\mathrm{j}2\pi(m-1)(k-1)/N}$. We may express $\mathbf{D}_k$ as $\mathbf{D}_k = \mathbf{U}^{k-1}$, where $\mathbf{U}$ is a diagonal unitary matrix with mth diagonal element $u_m = \mathrm{e}^{\mathrm{j}2\pi(m-1)/N}$. Then

$$g_k = Q\mathbf{A}\mathbf{f}_k = Q\mathbf{A}\mathbf{U}^{k-1}\mathbf{f}_1 = Q\mathbf{U}^{k-1}\mathbf{A}\mathbf{f}_1, \tag{17}$$

where the last equality follows from the fact that diagonal matrices commute.

Let $\mathscr{V}$ denote the M-dimensional subspace spanned by the vectors g_k, and K be the set of indices for which $A_k \neq 0$. Then $\mathscr{V}$ is spanned by the M vectors $\{q_k, k \in K\}$. From (17) it then follows that the components of g_k in this bases for $\mathscr{V}$ are the nonzero elements of $\mathbf{U}^{k-1}\mathbf{A}\mathbf{f}_1$. Thus, $\mathbf{g}_k = \mathbf{U}'^{k-1}\mathbf{g}_1$, where $\mathbf{U}'$ is the $M \times M$ diagonal matrix with diagonal elements $u_m, m \in K$. Since $\mathbf{U}'$ is a unitary matrix with $\mathbf{U}'^N = \mathbf{I}$, the vectors g_k form a cyclic set. □

In summary, a set of vectors $\{g_k\}$ has a circulant inner product structure if and only if the vectors form a cyclic set. In this case the FT matrix diagonalizes the corresponding Gram matrix. This property is the key element in the derivation of the CLSV.

6.1.2. Least-squares cyclic set

Our approach to determining the CLSV is to perform a unitary change of coordinates $\mathbf{F}$ so that S is mapped to $\overline{S} = S\mathbf{F}$ and G is mapped to $\overline{G} = G\mathbf{F}$. Since $\mathbf{F}$ is unitary, $\|S - G\|_{\mathrm{F}}^2 = \|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$. Thus, we may first solve the circulant inner product shaping problem in the new coordinate system. The CLSV are then the vectors corresponding to $\widehat{\overline{G}}\mathbf{F}^*$, where $\widehat{\overline{G}}$ minimizes $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$ subject to the constraint

$$\langle \overline{g}_m, \overline{g}_k \rangle = A_k \delta_{km} \tag{18}$$

with

$$A_k = \sum_{l=0}^{N-1} a_l \mathrm{e}^{-\mathrm{j}2\pi(k-1)l/N}, \quad 1 \leqslant k \leqslant N, \tag{19}$$

the $(k-1)$th coefficient of the DFT of $\{a_k, 0 \leqslant k \leqslant N-1\}$, given by (16). Note that $A_k \geqslant 0$ since the sequence a_k is nonnegative definite.

Minimization of $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$ subject to (18) is equivalent to the least-squares orthogonalization problem with constrained norms, discussed in Section 5.1. Thus from (6), the optimal orthogonal vectors correspond to

$$\widehat{\overline{G}} = \overline{U}\,\overline{\mathbf{V}}^*\mathbf{A}, \tag{20}$$

where $\overline{U}$ and $\overline{\mathbf{V}}^*$ are the right-hand unitary set transformation and the left-hand unitary matrix in the SVD of $\overline{S}\mathbf{A} = S\mathbf{F}\mathbf{A}$, and $\mathbf{A}$ is the diagonal matrix with diagonal elements $\sqrt{A_k}$, where A_k is defined in (19).

The CLSV that minimize $\|S - G\|_{\mathrm{F}}^2$ subject to (16) are $\hat{g}_k = \widehat{G}\mathbf{i}_k$, where

$$\widehat{G} = \widehat{\overline{G}}\mathbf{F}^* = \overline{U}\,\overline{\mathbf{V}}^*\mathbf{A}\mathbf{F}^*. \tag{21}$$

If $\operatorname{rank}(\mathbf{A}) \leqslant M$, then the vectors $\hat{g}_k$ can always be chosen to lie in $\mathscr{U}$.

If the vectors s_k are linearly independent and $A_k > 0$ for all k, then the vectors $\hat{g}_k$ are unique and

$$\widehat{G} = S\mathbf{F}\mathbf{A}(\mathbf{A}\mathbf{F}^* S^* S\mathbf{F}\mathbf{A})^{-1/2}\mathbf{A}\mathbf{F}^*.$$

If the vectors s_k are linearly dependent, then the optimal cyclic vectors are not unique, however their projections onto $\mathscr{U}$ are unique and are given by $P_{\mathscr{U}}\hat{g}_k = P_{\mathscr{U}}\widehat{G}\mathbf{i}_k$, where

$$P_{\mathscr{U}}\widehat{G} = S\mathbf{FA}((\mathbf{AF}^* S^* S\mathbf{FA})^{1/2})^{\dagger}\mathbf{AF}^*.$$

6.2. Unconstrained circulant inner product shaping

We now seek the vectors $\{g_k, 1 \leqslant k \leqslant N\}$ that minimize $\|S - G\|_{\mathrm{F}}^2$ subject to

$$\langle g_m, g_k \rangle = a_{k-m \bmod N}$$

for some sequence $\{a_k, 0 \leqslant k \leqslant N-1\}$ where the a_k's are to be determined.

In analogy to the previous section, we may solve this minimization problem by performing a change of coordinates and solving an optimal orthogonalization problem in the new coordinate system. However, now the orthogonal vectors are not norm constrained. Thus our problem reduces to seeking the orthogonal vectors $\overline{g}_k$ corresponding to $\overline{G}$ that minimize $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$ subject to

$$\langle \overline{g}_m, \overline{g}_k \rangle = 0 \quad \text{for } m \neq k. \tag{22}$$

This minimization problem is equivalent to the unconstrained least-squares orthogonalization problem discussed in Section 5.2. We may solve this problem using the iterative algorithm described in Section 5.4. The optimal cyclic vectors are then given by $\hat{g}_k = \widehat{G}\mathbf{i}_k$, where $\widehat{G} = \widehat{\overline{G}}\mathbf{F}^*$, and $\widehat{\overline{G}}$ is the ST that minimizes $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$ subject to (22).

We summarize our results regarding least-squares circulant inner product shaping in the following theorem.

Theorem 3 (Cyclic least-squares vectors (CLSV)). *Let $\{s_k, 1 \leqslant k \leqslant N\}$ be a set of N vectors in a Hilbert space $\mathscr{H}$, that span an M-dimensional subspace $\mathscr{U} \subseteq \mathscr{H}$. Let $\{\hat{g}_k, 1 \leqslant k \leqslant N\}$ denote N optimal vectors that minimize the least-squares error defined by* (15) *subject to* (16). *Let S and $\widehat{G}$ denote the set transformations corresponding to the vectors s_k and $\hat{g}_k$, respectively, and let $\mathbf{F}$ be the $N \times N$ FT matrix with mk*th *element $\mathbf{F}_{mk} = 1/\sqrt{N}\mathrm{e}^{-\mathrm{j}2\pi(m-1)(k-1)/N}$. Then*:

(1) *If the inner products in* (16) *are given, then let $\mathbf{A}$ be the diagonal matrix with diagonal elements $\sqrt{A_k}$, where A_k is the* $(k-1)$th *coefficient of the DFT of the sequence a_k, and let $S\mathbf{FA} = \overline{U}\,\overline{\Sigma}\,\overline{\mathbf{V}}^*$. Then $\widehat{G}$ can be chosen as*

$$\widehat{G} = \overline{U}\,\overline{\mathbf{V}}^*\mathbf{AF}^*.$$

The corresponding vectors $\{\hat{g}_k = \widehat{G}\mathbf{i}_k, 1 \leqslant k \leqslant N\}$ form a cyclic set and are defined as the cyclic least-squares vectors. In addition,

(a) *If* $\mathrm{rank}(\mathbf{A}) \leqslant M$, *then the CLSV can always be chosen to lie in $\mathscr{U}$.*

(b) *If $M = N$, then*

(i) $\widehat{G} = S\mathbf{FA}(\mathbf{AF}^* S^* S\mathbf{FA})^{-1/2}\mathbf{AF}^*$;

(ii) *the CLSV are the unique cyclic vectors that minimize the least-squares error.*

(c) *If* $M < N$, *then the projection of* $\widehat{G}$ *onto* $\mathscr{U}$ *is unique and is given by* $P_{\mathscr{U}}\widehat{G} = S\mathbf{FA}((\mathbf{AF}^*S^*S\mathbf{FA})^{1/2})^{\dagger}\mathbf{AF}^*$.

(2) *If the inner products are not specified, then* $\widehat{G}$ *can be chosen as* $\widehat{G} = \widehat{\overline{G}}\mathbf{F}^*$, *where* $\widehat{\overline{G}}$ *is the set transformation corresponding to an orthogonal set of vectors that minimize* $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$ *with* $\overline{S} = S\mathbf{F}$, *and can be computed using the iterative algorithm of Section* 5.4.

7. Least-squares permuted inner-product shaping

In this section we seek a set of vectors $\{g_k, 1 \leqslant k \leqslant N\}$ corresponding to G that minimize $\|S - G\|_{\mathrm{F}}^2$ subject to the constraint

$$\left\{\langle g_m, g_k\rangle, 1 \leqslant k \leqslant N\right\} = P\left\{a_k, 0 \leqslant k \leqslant N-1\right\} \quad \text{for all } m, \tag{23}$$

where $P\{a_k, 0 \leqslant k \leqslant N-1\}$ is an arbitrary permutation of $\{a_k, 0 \leqslant k \leqslant N-1\}$. In this case, every row (or column) of the Gram matrix of inner products $\langle g_m, g_k\rangle$ is a permutation of the first row (or column). We refer to such a matrix as a permuted matrix.

Throughout this section we restrict our attention to real Hilbert spaces.

As in Section 6, we first characterize a set of vectors with a real permuted Gram matrix. In particular, in Section 7.2 we show that the vectors are *geometrically uniform* (*GU*) [16], and their Gram matrix is diagonalized by a generalized FT matrix defined on a direct product of cyclic groups. The derivation of the GULSV is then analogous to the derivation of the CLSV where the generalized FT matrix is substituted for the FT matrix. Before deriving the GULSV, in the next section we define the generalized FT.

7.1. Generalized Fourier transform

Let $T = G^*G$ be a real permuted Gram matrix,[5] so that each row of T is a permutation of the first row whose elements are labeled by $\{a_k, 1 \leqslant k \leqslant N\}$. As we now show, we may regard the matrix T as the multiplication table of an abelian group with elements $\mathscr{A} = \{a_k, 1 \leqslant k \leqslant N\}$ so that $a_k a_m = T_{km}$. Indeed, $T_{1m} = a_1 a_m = a_m$ for $1 \leqslant m \leqslant N$ and consequently a_1 is the identity on $\mathscr{A}$. Furthermore, a_1 appears exactly once in each row of T so that there is exactly one element a_m that satisfies $a_k a_m = a_1$ for each a_k; this element is the inverse of a_k. Finally, $a_k a_m = T_{km} = T_{mk} = a_m a_k$ so that multiplication in the group is commutative. We thus conclude

[5] The results hold in the more general case in which the Gram matrix G is symmetric so that $G = G^T$.

that the group $\mathscr{A}$ with multiplication defined above as the group operation is an abelian group: $\mathscr{A}$ contains the identity, every element has an inverse, and $a_k a_m = a_m a_k$ for all m, k.

To define the generalized FT matrix we rely on the well-known theorem (see e.g., [3]) that states that every finite abelian group $\mathscr{A}$ with N elements is isomorphic [6] to a direct product A of a finite number of cyclic groups: $\mathscr{A} \cong A = \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_p}$, where $\mathbb{Z}_{N_m}$ is the cyclic additive group of integers modulo N_m, and $N = \prod_{m=1}^{p} N_m$. Thus every element $a_k \in \mathscr{A}$ can be associated with an element $a^{(k)} \in A$ of the form $a^{(k)} = (a_1^{(k)}, a_2^{(k)}, \ldots, a_p^{(k)})$, where $a_m^{(k)} \in \mathbb{Z}_{N_m}$. We denote this one-to-one correspondence by $a_k \leftrightarrow a^{(k)}$. Because the correspondence is an isomorphism, it follows that if $a_k \leftrightarrow a^{(k)}$, $a_m \leftrightarrow a^{(m)}$, $a_l \leftrightarrow a^{(l)}$ and $a_k = a_m a_l$, then $a^{(k)} = a^{(m)} + a^{(l)}$, where the addition of $a^{(m)} = (a_1^{(m)}, a_2^{(m)}, \ldots, a_p^{(m)})$ and $a^{(l)} = (a_1^{(l)}, a_2^{(l)}, \ldots, a_p^{(l)})$ is performed by component-wise addition modulo the corresponding N_m.

The generalized FT matrix $\mathbf{F}_g$ over the group $\mathscr{A}$ is defined in terms of the additive group A [6,28]: The *mk*th element of $\mathbf{F}_g$ is $1/\sqrt{N}\langle a^{(m)}, a^{(k)}\rangle$, where

$$\left\langle a^{(m)}, a^{(k)} \right\rangle = \prod_{s=1}^{p} \mathrm{e}^{-\mathrm{j}2\pi a_s^{(m)} a_s^{(k)}/N_s}.$$

Here $a_s^{(m)}$ and $a_s^{(k)}$ are the sth components of $a^{(m)}$ and $a^{(k)}$, respectively, and the product $a_s^{(m)} a_s^{(k)}$ is taken as an ordinary integer modulo N_s. We may readily verify that $\mathbf{F}_g$ is unitary. In the special case where $T = G^*G$ is a circulant matrix, $\mathscr{A} \cong A = Z_N$ and $\langle a^{(m)}, a^{(k)}\rangle = \mathrm{e}^{-\mathrm{j}2\pi mk/N}$, where $m, k \in \mathbb{Z}_N$ and the product mk is taken as an ordinary integer modulo N; thus the generalized FT matrix $\mathbf{F}_g$ reduces to the FT matrix $\mathbf{F}$.

To better acquaint ourselves with the concepts introduced, we consider a simple example. (The same example was given in [6].) Suppose that

$$T = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_1 & a_4 & a_3 \\ a_3 & a_4 & a_1 & a_2 \\ a_4 & a_3 & a_2 & a_1 \end{bmatrix}. \tag{24}$$

We define the group $\mathscr{A} = \{a_1, a_2, a_3, a_4\}$ with multiplication table given by the Gram matrix T:

	a_1	a_2	a_3	a_4
a_1	a_1	a_2	a_3	a_4
a_2	a_2	a_1	a_4	a_3
a_3	a_3	a_4	a_1	a_2
a_4	a_4	a_3	a_2	a_1

(25)

[6] Two groups $\mathscr{A}$ and $\mathscr{A}'$ are *isomorphic*, denoted $\mathscr{A} \cong \mathscr{A}'$, if there is a bijection (one-to-one and onto map) $\varphi : \mathscr{A} \to \mathscr{A}'$ satisfying $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in \mathscr{A}$.

If we define the correspondence

$$a_1 \leftrightarrow (0,0),\quad a_2 \leftrightarrow (0,1),\quad a_3 \leftrightarrow (1,0),\quad a_4 \leftrightarrow (1,1),$$

then this table becomes the addition table of $A = \mathbb{Z}_2 \times \mathbb{Z}_2$:

	(0, 0)	(0, 1)	(1, 0)	(1, 1)
(0, 0)	(0, 0)	(0, 1)	(1, 0)	(1, 1)
(0, 1)	(0, 1)	(0, 0)	(1, 1)	(1, 0)
(1, 0)	(1, 0)	(1, 1)	(0, 0)	(0, 1)
(1, 1)	(1, 1)	(1, 0)	(0, 1)	(0, 0)

(26)

Only the way in which the elements are labeled distinguishes the table of (26) from the table of (25); thus $\mathscr{A} \cong A$. The FT matrix $\mathbf{F}_g$ over $\mathscr{A}$ is the Hadamard matrix

$$\mathbf{F}_g = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

7.2. *Least-squares geometrically uniform set*

Definition 2. Let $\{g_k, 1 \leqslant k \leqslant N\}$ be a set of vectors that span an M-dimensional subspace $\mathscr{V}$. Then the vectors g_k form a *geometrically uniform* (*GU*) set if $\mathbf{g}_k = \mathbf{U}_k\mathbf{g}_1$ for $1 \leqslant k \leqslant N$ where $\mathbf{g}_k$ is the representation of g_k in any orthonormal basis for $\mathscr{V}$, and the matrices $\{\mathbf{U}_k, 1 \leqslant k \leqslant N\}$ are unitary, and form an abelian group[7] $\mathscr{G}$.

Note that a cyclic vector set defined in Definition 1 is a special case of a GU vector set, where $\mathbf{U}_k = \mathbf{U}^{k-1}$ for some unitary $\mathbf{U}$ that satisfies $\mathbf{U}^N = \mathbf{I}$.

Real vector sets with permuted inner product structure are conveniently characterized by the following theorem.

Theorem 4. *A set of real vectors $\{g_k, 1 \leqslant k \leqslant N\}$ has a permuted Gram matrix if and only if the vectors form a geometrically uniform set.*

The proof of the theorem relies on Lemma 2 below. The direct part of the lemma is proved in [6], while the proof of the converse is straightforward and is therefore omitted.

Lemma 2. *Let G be a set transformation corresponding to real vectors $\{g_k, 1 \leqslant k \leqslant N\}$. Then G^*G is a permuted matrix if and only if the vectors $\{\overline{g}_k, 1 \leqslant k \leqslant N\}$*

[7] That is, $\mathscr{G}$ contains the identity matrix I; if $\mathscr{G}$ contains U_k, then it also contains its inverse $U_k^{-1} = U_k^*$; the product U_kU_m of any two elements of $\mathscr{G}$ is in $\mathscr{G}$; and $U_kU_m = U_mU_k$ for any two elements in $\mathscr{G}$.

corresponding to $\overline{G} = G\mathbf{F}_g$ *are orthogonal, where* $\mathbf{F}_g$ *is the generalized FT matrix corresponding to the group* $\{a_k = \langle g_1, g_k\rangle, 1 \leqslant k \leqslant N\}$. *In this case,* $\langle \overline{g}_k, \overline{g}_k\rangle = A_{gk}$, *where* A_{gk} *is the* kth *component of* $\sqrt{N}\mathbf{F}_g\mathbf{a}$, *and* $\mathbf{a}$ *is the vector of components* a_k.

Proof of Theorem 4. Suppose the vectors g_k form a GU set generated by the abelian group $\mathscr{G}$ of unitary matrices $\{\mathbf{U}_k, 1 \leqslant k \leqslant N\}$. Then since $\mathbf{U}_m^* = \mathbf{U}_m^{-1}$,

$$\langle g_m, g_k\rangle = \langle \mathbf{g}_m, \mathbf{g}_k\rangle = \langle \mathbf{U}_m\mathbf{g}_1, \mathbf{U}_k\mathbf{g}_1\rangle = \langle \mathbf{g}_1, \mathbf{U}_m^{-1}\mathbf{U}_k\mathbf{g}_1\rangle = a\left(\mathbf{U}_m^{-1}\mathbf{U}_k\right),$$

where $a(\mathbf{U}_k) = \langle \mathbf{g}_1, \mathbf{U}_k\mathbf{g}_1\rangle$. For fixed k, the set $\mathbf{U}_k^{-1}\mathscr{G} = \{\mathbf{U}_k^{-1}\mathbf{U}_m, \mathbf{U}_m \in \mathscr{G}\}$ is a permutation of $\mathscr{G}$ since $\mathbf{U}_k^{-1}\mathbf{U}_m \in \mathscr{G}$ for all k, m [3]. Therefore $\{a(\mathbf{U}_k^{-1}\mathbf{U}_m), 1 \leqslant m \leqslant N\}$ is a permutation of $\{a(\mathbf{U}_k), 1 \leqslant k \leqslant N\}$. Consequently, every row of the Gram matrix G^*G of elements $\langle g_m, g_k\rangle$ is a permutation of $\{a(\mathbf{U}_k), 1 \leqslant k \leqslant N\}$, which is the first row of G^*G.

The proof of the converse is analogous to the proof of Theorem 2 with $\mathbf{F}_g$ substituted for $\mathbf{F}$. □

From Lemma 2 we conclude that instead of seeking the GU vectors $\{g_k\}$ corresponding to G that minimize $\|S - G\|_{\mathrm{F}}^2$, we may seek the orthogonal vectors $\{\overline{g}_k\}$ corresponding to $\overline{G}$ that minimize $\|\overline{S} - \overline{G}\|_{\mathrm{F}}^2$, where $\overline{S} = S\mathbf{F}_g$. This problem was discussed in Section 6.1.2. The optimal GULSV then follow from Theorem 3 with $\mathbf{F}_g$ and $\mathbf{A}_g$ substituted for $\mathbf{F}$ and $\mathbf{A}$, respectively, where $\mathbf{A}_g$ is a diagonal matrix with diagonal elements $\sqrt{A_{gk}}$, and A_{gk} is defined in Lemma 2.

Acknowledgements

The author wishes to thank Prof. D. Forney and Prof. A. Oppenheim for many fruitful discussions, Prof. G. Strang and Prof. D. O'Leary for their helpful comments, Dr. A. Yeredor for suggesting the use of Givens rotations, and the anonymous reviewer for many valuable suggestions that greatly improved the presentation.

Appendix A. Maximizing $R_{hs}'^{(j+1)}$

Recall that,

$$\begin{aligned} R_{hs}'^{(j+1)} &= (cb_{11} + sb_{21})^2 + (cb_{22} - sb_{12})^2 \\ &= c^2(b_{11}^2 + b_{22}^2) + s^2(b_{21}^2 + b_{12}^2) + 2cs(b_{11}b_{21} - b_{22}b_{12}), \end{aligned} \tag{A.1}$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$. Differentiating (A.1) with respect to θ,

$$\frac{\mathrm{d}R_{hs}'^{(j+1)}}{\mathrm{d}\theta} = -2cs(b_{11}^2 + b_{22}^2) + 2cs(b_{21}^2 + b_{12}^2) + 2(c^2 - s^2)(b_{11}b_{21} - b_{22}b_{12})$$

$$= -x\sin(2\theta) + 2y\cos(2\theta), \tag{A.2}$$

where $x = b_{11}^2 + b_{22}^2 - b_{21}^2 - b_{12}^2$, $y = b_{11}b_{21} - b_{22}b_{12}$, and we used the relations $c^2 - s^2 = \cos(2\theta)$, $2cs = \sin(2\theta)$. Equating (A.2) to 0 yields

$$x\sin(2\theta) = 2y\cos(2\theta). \tag{A.3}$$

Note that $\mathbf{J}(r, l, \theta) = \mathbf{J}(r, l, \theta + 2\pi k)$ for any integer k. Thus it is sufficient to consider solutions $\theta \in (-\pi, \pi]$. If $x \neq 0$, then the solutions to (A.3) are

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2y}{x}\right). \tag{A.4}$$

If $x = 0$ and $y \neq 0$, then the solutions to (A.3) are $\theta = \pm\pi/4$. If $x = y = 0$, then from (A.1) we see that $R_{hs}^{\prime(j+1)}$ does not depend on θ, and we choose $\theta = 0$.

Taking the second derivative of $R_{hs}^{\prime(j+1)}$ with respect to θ yields

$$\frac{\mathrm{d}^2 R_{hs}^{\prime(j+1)}}{\mathrm{d}\theta^2} = -2x\cos(2\theta) - 4y\sin(2\theta).$$

Thus, a solution θ of (A.3) maximizes $R_{hs}^{\prime(j+1)}$ if

$$x\cos(2\theta) + 2y\sin(2\theta) > 0,$$

which for $x \neq 0$ reduces to

$$x\cos(2\theta)(1 + 4y^2/x^2) > 0.$$

Thus if $x \neq 0$, then $\hat{\theta}$ is a maximum of $R_{hs}^{\prime(j+1)}$ if $\hat{\theta}$ has the form (A.4) and $\cos(2\hat{\theta})$ and x have the same sign. Similarly, for $x = 0$, $\sin(2\hat{\theta})$ and y must have the same sign. So, for $y > 0$, $\hat{\theta} = \pi/4$, and for $y < 0$, $\hat{\theta} = -\pi/4$.

References

[1] A. Aldroubi, Portraits of frames, Proc. Amer. Math. Soc. 123 (1995) 1661–1668.
[2] L.-E. Andersson, T. Elfving, A constrained Procrustes problem, SIAM J. Matrix Anal. Appl. 18 (1997) 124–139.
[3] M.A. Armstrong, Groups and Symmetry, Springer, New York, 1988.
[4] J.M.F. ten Berge, K. Nevels, A general solution to Mosier's oblique Procrustes problem, Psychometrika 42 (1977) 593–600.
[5] S. Chaturvedi, A.K. Kapoor, V. Srinivasan, A new orthogonalization procedure with an extremal property, J. Phys. A 31 (1998) 367–370.
[6] Y.C. Eldar, G.D. Forney Jr., On quantum detection and the square-root measurement, IEEE Trans. Inform. Theory 47 (2001) 858–872.
[7] Y.C. Eldar, A.V. Oppenheim, D. Egnor, Orthogonal and projected orthogonal matched filter detection, IEEE Trans. Signal Process. (2001) (submitted).
[8] Y.C. Eldar, A.V. Oppenheim, Orthogonal matched filter detection, in: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2001), 2001.
[9] Y.C. Eldar, Quantum signal processing, Ph.D. thesis, MIT, Cambridge, MA, 2001.

[10] Y.C. Eldar, A.V. Oppenheim, Orthogonal multiuser detection, Signal Process. (2001) (to appear).
[11] Y.C. Eldar, A.M. Chan, Orthogonal and projected orthogonal multiuser detection, IEEE Trans. Inform. Theory (2001) (submitted).
[12] Y.C. Eldar, A.V. Oppenheim, Covariance shaping least-squares estimation, IEEE Trans. Signal Process. (2001) (submitted).
[13] Y.C. Eldar, H. Bölcskei, Geometrically uniform frames, IEEE Trans. Inform. Theory (2001) (submitted). Available from http://arXiv.org/abs/math.FA/0108096.
[14] Y.C. Eldar, G.D. Forney Jr., Optimal tight frames and quantum measurement, IEEE Trans. Inform. Theory (2001) (to appear). Available from http://xxx.lanl.gov/abs/quant-ph/0106070.
[15] K.Y. Fan, A.J. Hoffman, Some metric inequalities in the space of matrices, Proc. Amer. Math. Soc. 6 (1955) 111–116.
[16] G.D. Forney Jr., Geometrically uniform codes, IEEE Trans. Inform. Theory 37 (1991) 1241–1260.
[17] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, Baltimore MD, 1996.
[18] J.C. Gower, Multivariate analysis: ordination multidimensional scaling and allied topics, Handbook of Applicable Mathematics, vol. 6, Wiley, Chichester, UK, 1984, pp. 727–781.
[19] V.K. Goyal, J. Kovačević, J.A. Kelner, Quantized frame expansions with erasures, Appl. Comp. Harmonic Anal. 10 (2001) 203–233.
[20] B.F. Green, The orthogonal approximation of an oblique structure in factor analysis, Psychometrika 17 (1952) 429–440 (1952).
[21] P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, W.K. Wootters, Classical information capacity of a quantum channel, Phys. Rev. A 54 (1996) 1869–1876.
[22] C.W. Helstrom, Quantum Detection and Estimation Theory, Academic Press, New York, 1976.
[23] C.W. Helstrom, Bayes-cost reduction algorithm in quantum hypothesis testing, IEEE Trans. Inform. Theory 28 (1982) 359–366.
[24] N.J. Higham, Computing the polar decomposition—with applications, SIAM J. Sci. Statist. Comput. 7 (1986) 1160–1174.
[25] N.J. Higham, Computing a nearest symmetric positive semidefinite matrix, Linear Algebra Appl. 103 (1988) 103–118.
[26] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[27] J.R. Hurley, R.B. Cattell, Producing direct rotation to test a hypothesized factor structure, Behavioral Sci. 7 (1962) 258–262.
[28] D. Maslen, D. Rockmore, Generalized FFTs—a survey of some recent results, Proceedings of the DIMACS Workshop in Groups and Computation 28 (1995) 183–238.
[29] D. Raphaeli, On multidimensional coded modulations having uniform error property for generalized decoding and flat-fading channels, IEEE Trans. Comm. 46 (1998) 34–40.
[30] A. Ruhe, Closest normal matrix finally found!, BIT 27 (1987) 585–598.
[31] H.C. Schweinler, E.P. Wigner, Orthogonalization methods, J. Math. Phys. 11 (1970) 1693–1694.
[32] P.H. Schonemann, A generalized solution of the orthogonal Procrustes problem, Psychometrika 31 (1966) 1–10.
[33] G. Ungerboeck, Channel coding with multilevel/phase signals, IEEE Trans. Inform. Theory 28 (1982) 55–67.
[34] P. Zielinski, K. Zietak, The polar decomposition—properties applications and algorithms, Mat. Stosowana 38 (1995) 23–40.