A NOTE ON THE COMPLEX MATRIX PROCRUSTES PROBLEM

J. Kiskiras* **and G.D. Halikias***

* Control Engineering Research Centre, School of Engineering and Mathematical Sciences, City University, Northampton Square, London EC1V 0HB, U.K.
e-mail addresses: kiskiras@msn.com, g.halikias@city.ac.uk

Keywords: Matrix Procrustes problem, Cone of positive semi-definite matrices, Least-squares approximation.

Abstract

This note outlines an algorithm for solving the complex "matrix Procrustes problem". This is a least-squares approximation over the cone of positive semi-definite Hermitian matrices, which has a number of applications in the areas of Optimization, Signal Processing and Control. The work generalises the method of [1], who obtained a numerical solution to the real-valued version of the problem. It is shown that, subject to an appropriate rank assumption, the complex problem can be formulated in a real setting using a matrix dilation technique, for which the method of [1] is applicable. However, this transformation results in an over-parametrisation of the problem and, therefore, convergence to the optimal solution is slow. Here an alternative algorithm is developed for solving the complex problem, which exploits fully the special structure of the dilated matrix. The advantages of the modified algorithm are demonstrated via a numerical example.

1 Notation

Most of the notation used is standard and is summarised here for convenience. Additional notation is introduced at various sections of the paper via appropriate definitions.

$\mathcal{R}^n(\mathcal{C}^n)$	n-dimensional real (complex) vector space.
$\mathcal{R}^{m\times n}(\mathcal{C}^{m\times n})$	Space of real (complex) m by n matrices.
$\mathcal{H}^n(\mathcal{S}^n)$	Space of n by n Hermitian (Symmetric) matrices.
$\mathcal{S}^n_{\geq}(\mathcal{S}^n_{>})$	Cone of n by n Symmetric Positive Semi-definite (definite) matrices.
$\mathcal{H}^n_{\geq}(\mathcal{H}^n_{>})$	Cone of n by n Hermitian Positive Semi-definite (definite) matrices.
A'	Transpose of matrix A.
A^*	Complex conjugate transpose of matrix A.
A^{-1}	Inverse of nonsingular matrix A.
det(A)	Determinant of matrix A.
rank(A)	Rank of matrix A.
tr(A)	Trace of matrix A.
vec(A)	The rows of A stacked in a column vector.
$\sigma_i(A)$	i-th singular value of A (indexed in non-increasing order of magnitude).
$\lambda_i(A)$	i-th eigenvalue of A.
$\otimes$	Kronecker product of two matrices.
$\|\bullet\|$	Euclidian norm of vector, largest singular value of matrix $\sigma_1(\cdot)$.
$\|\bullet\|_F$	Frobenious norm of matrix.

In addition the following notation is used: The range and null-space of a matrix A are written as Range(A) and Null(A), respectively. The convex hull of a nonempty set X is denoted by conv$[X]$ and the conical hull of X, i.e. the set conv$[\{\alpha x : \alpha \geq 0, \; x \in X\}]$ by cone$[X]$. If $A \in \mathcal{C}^{m\times n}$ and $X \subseteq \mathcal{C}^n$ then AX denotes the set $\{Ax : x \in X\} \subseteq \mathcal{C}^m$. The line $\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$ between two points x and y is written as line$\{x, y\}$. The unique point x in a closed, convex set X which minimises $\|d - x\|$ with respect to $x \in X$ is denoted by minpoint$[d, X]$ and the corresponding minimal distance $\|d - \hat{x}\|$ by mindist$[d, X]$.

2 Introduction

The present work develops an algorithm for solving a least squares approximation problem over the cone of Hermitian positive semi-definite matrices. This is a generalisation of the work in [1] to the complex case. The motivation for solving this approximation problem (known in the literature as the "Procrustes" problem) initially arose in the area of Optimisation, in the context of developing methods for estimating the inverse Hessian matrix in quasi-Newton algorithms. Recently, some novel applications have appeared in the areas of Signal processing and System Identification of elastic structures [8]. Variations to the approximation problem, in which the optimisation is carried out over alternative matrix sets (e.g. symmetric or permutation matrices) have also appeared in the literature and are motivated by certain types of statistical estimation problems. [2, 3, 4, 9, 7].

The problem addressed in [1] is the following: For arbitrary real matrices A and C, solve:

$$\min_{X\in\mathcal{S}^n_{\geq}} \|A - XC\|_F \tag{1}$$

It is shown in [1] that the solution to this problem exists, i.e. the infimum is attained for some $X \in \mathcal{S}^n_{\geq}$; furthermore the solution is unique if C has full row rank.

The technique of [1] for solving (1) is based on a novel characteristion of the set of real symmetric positive semidefinite matrices. Using this characterisation, together with certain properties of the Kronecker product of two matrices and other linear algebraic techniques, it was shown in [1] that (1) reduces to a minimum-distance problem of a vector from the conical hull of a closed convex set; an algorithm was finally presented for solving this distance problem.

In this note we generalise the method of [1] to the complex case. Specifically, the following problem is addressed:

Given matrices $A, C \in \mathcal{C}^{n\times m}$ with $m \geq n$ and $\text{Rank}(C) = n$, solve:

$$\min_{X\in\mathcal{H}^n_{\geq}} \|A - XC\|_F \tag{2}$$

It is first shown that (2) can be re-formulated in a real setting by inflating the dimension of the problem. This leads to an equivalent real problem of the same form as (1) which can be solved via existing techniques [1]. In this form, however, the problem is over-parametrised and convergence to the (unique) optimal solution is slow. By exploiting the structure of the inflated problem, however, a significantly improved algorithm can be obtained.

Many derivations included in this note follow closely those in [1]; in fact it is shown that most of the intermediate results of [1] involving properties of the sets of real symmetric ($\mathcal{S}^n$) and positive semi-definite matrices ($\mathcal{S}^n_{\geq}$) are inherited by two appropriately defined subsets of $\mathcal{S}^n$ and $\mathcal{S}^n_{\geq}$, arising naturally from the structure of the inflated problem. Results which deviate from those in [1] and which specifically apply to the complex problem are highlighted in our presentation. Due to limited space most proofs have been omitted.

3 Main results

We first briefly review the characterisation of the set of positive semi-definite matrices given in [1]. Recall that $\mathcal{S}^n$ denotes the set of $n \times n$ real symmetric matrices and $\mathcal{S}^n_{\geq}$ the set of $n \times n$ positive semi-definite real symmetric matrices. Now,

$$\text{vec}(\mathcal{S}^n) = \{\text{vec}(A) : A \in \mathcal{S}^n\} \subseteq \mathcal{R}^{n^2}$$

is a linear subspace of $\mathcal{S}^n$ since it is closed under addition and scalar multiplication. Clearly, $\text{vec}(\mathcal{S}^n)$ has dimension $r = n(n+1)/2$, since a symmetric matrix is fully described via its diagonal and upper triangular (or lower triangular) elements. Let $\{w_1, w_2, \ldots, w_r\}$, be an orthonormal basis set for $\text{vec}(S^n)$ and define $W_{\mathcal{S}} = [w_1 \ w_2 \ \ldots w_r]$. For each $A \in \mathcal{S}^n$ the column vector of co-ordinates of $\text{vec}(A)$ with respect to $\{w_1, w_2, \ldots, w_r\}$ is denoted by $\overline{\text{vec}}_{\mathcal{S}}(A)$. Clearly we have that:

$$\text{vec}(A) = W_{\mathcal{S}}\overline{\text{vec}}_{\mathcal{S}}(A) \Rightarrow \overline{\text{vec}}_{\mathcal{S}}(A) = W'_{\mathcal{S}}\text{vec}(A)$$

Also, $W'_{\mathcal{S}}W_{\mathcal{S}} = I_r$, $\text{Range}[W'_{\mathcal{S}}] = \mathcal{R}^r$ and $\text{Range}[W_{\mathcal{S}}] = \text{vec}(\mathcal{S}^n)$.

The characterization of positive semi-definiteness in [1] is based on the fact that $A \in \mathcal{S}^n_{\geq}$ can be written (e.g. via its spectral decomposition) as $A = \alpha B^2$ for some $B = B'$ and $\alpha \geq 0$. Let:

$$\mathcal{U}_{\mathcal{S}} := \{B \in \mathcal{R}^{n\times n} : B = B' \ \text{ and } \ \|B\|_F = 1\} \subseteq \mathcal{S}^n$$

Also define:

$$\Psi_{\mathcal{S}} := \{\text{vec}(B^2) : \ \ B \in \mathcal{U}_{\mathcal{S}}\} \subseteq \mathcal{R}^{n^2} \ \text{ and } \ \Omega_{\mathcal{S}} = \text{conv}[\Psi_{\mathcal{S}}]$$

Then it is shown in [1] that: (i) $\text{vec}(\mathcal{S}^n_{\geq}) = \text{cone}[\Omega_{\mathcal{S}}]$. (ii) $\overline{\text{vec}}_{\mathcal{S}}(\mathcal{S}^n_{\geq}) = \text{cone}[W'_{\mathcal{S}}\Omega_{\mathcal{S}}]$, and (iii) $\Psi_{\mathcal{S}}$ is a compact set, $\Omega_{\mathcal{S}}$ is a non empty convex compact set with $\text{mindist}(0, \Omega_{\mathcal{S}}) = 1/\sqrt{n}$ and $\text{cone}[\Omega_{\mathcal{S}}]$ is a nonempty closed convex cone. This characterisation of $\mathcal{S}^n_{\geq}$ is subsequently used in [1] to reduce (1) to a distance problem.

Next, consider the complex problem (2). The following lemma allows us to formulate condition $X \in \mathcal{H}^n_{\geq}$ in a real setting:

Lemma 3.1: Let $X = X_r + jX_i \in \mathcal{C}^{n\times n}$ where $X_r, X_i \in \mathcal{R}^{n\times n}$. Then the following two statements are equivalent:

(i) $X \in \mathcal{H}^n_{\geq}$.

(ii) $P := \begin{pmatrix} X_r & X_i \\ -X_i & X_r \end{pmatrix} \in \mathcal{S}^{2n}_{\geq}$

Lemma 3.2: (i) Let $A = A_r + jA_i \in \mathcal{C}^{n\times m}$ with $A_r, A_i \in \mathcal{R}^{n\times m}$. Then,

$$\begin{aligned}\|A\|_F^2 &= \|A_r\|_F^2 + \|A_i\|_F^2 = \|[A_r \ A_i]\|_F^2 \\ &= \frac{1}{2}\left\|\begin{pmatrix} A_r & A_i \\ A_r & A_i \end{pmatrix}\right\|_F^2 = \frac{1}{2}\left\|\begin{pmatrix} A_r & A_i \\ -A_i & A_r \end{pmatrix}\right\|_F^2\end{aligned}$$

(ii) For any four complex matrices A, B, C, D of compatible dimensions,

$$\left\|\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right\|_F^2 = \|A\|_F^2 + \|B\|_F^2 + \|C\|_F^2 + \|D\|_F^2$$

The following result shows that optimization problem (2) can be transformed to the real setting of problem (1), which can be solved using the algorithm of [1]. Before stating this result we define the following two "structured" subsets of $\mathcal{S}^{2n}$ and $\mathcal{S}^{2n}_{\geq}$:

$$\mathcal{Q}^{2n} = \left\{\begin{pmatrix} X & Y \\ -Y & X \end{pmatrix} \in \mathcal{S}^{2n}\right\} \tag{3}$$

and

$$\mathcal{Q}^{2n}_{\geq} = \left\{\begin{pmatrix} X & Y \\ -Y & X \end{pmatrix} \in \mathcal{S}^{2n}_{\geq}\right\} \tag{4}$$

Clearly $\mathcal{Q}^{2n}_{\geq} \subseteq \mathcal{Q}^{2n} \subseteq \mathcal{S}^{2n}$ and $\mathcal{Q}^{2n}_{\geq} \subseteq \mathcal{S}^{2n}_{\geq} \subseteq \mathcal{S}^{2n}$.

Theorem 3.3 Let $A, C \in \mathcal{C}^{n\times m}$, where $m \geq n$ and $\text{Rank}(C) = n$. Write $A = A_r + jA_i$, $C = C_r + jC_i$, where $A_r, A_i, C_r, C_i \in \mathcal{R}^{n\times m}$, and consider the two optimization problems,

$$\gamma_o = \min_{X\in\mathcal{H}^n_{\geq}} \|A - XC\|_F \tag{5}$$

and

$$\gamma_1 = \min_{P\in\mathcal{S}^{2n}_{\geq}} \left\| \begin{pmatrix} A_r & A_i \\ -A_i & A_r \end{pmatrix} - P \begin{pmatrix} C_r & C_i \\ -C_i & C_r \end{pmatrix} \right\|_F \quad (6)$$

Then: (i) The minimum in both (5) and (6) exists, and the corresponding optimum solutions $X_o \in \mathcal{H}^n_{\geq}$ of (5) and $P_o \in \mathcal{S}^{2n}_{\geq}$ of (6) are unique. Further $\gamma_1 = \sqrt{2}\gamma_o$.

(ii) Write P_o as

$$P_o = \begin{pmatrix} P_1 & P_2' \\ P_2 & P_3 \end{pmatrix}$$

where $P_1, P_2, P_3 \in \mathcal{R}^{n\times n}$. Then $P_1 = P_3$, $P_2 + P_2' = 0$ and $X_o = P_1 + jP_2$.

Proof: Consider first (6). Theorem 3.2 in [1] shows that the minimum exists. To show that the corresponding optimum solution P_o is unique, note that:

$$\text{Rank}(C) = n \Leftrightarrow \text{Rank}\begin{pmatrix} C_r & C_i \\ -C_i & C_r \end{pmatrix} = 2n \quad (7)$$

Next, consider the optimization problem (5). Defining $G(X) = A - XC = G_r(X) + jG_i(X)$ with $G_r(X),\ G_i(X) \in \mathcal{R}^{n\times m}$, this may be written as:

$$\gamma_o^2 = \min_{X=X_r+jX_i\geq 0} \|G(X)\|_F^2 \quad (8)$$

$$= \frac{1}{2} \min_{X=X_r+jX_i\geq 0} \left\| \begin{pmatrix} G_r(X) & G_i(X) \\ -G_i(X) & G_r(X) \end{pmatrix} \right\|_F^2 \quad (9)$$

using Lemma 3.2 part (i). On noting that:

$$G(X) = A_r - X_rC_r + X_iC_i + j(A_i - X_rC_i - X_iC_r)$$

we obtain:

$$G_r(X) = A_r - X_rC_r + X_iC_i \quad (10)$$

and

$$G_i(X) = A_i - X_rC_i - X_iC_r \quad (11)$$

Substituting (10) and (11) into (9) and using Lemma 3.1 gives:

$$\gamma_o^2 = \frac{1}{2} \min_{P\in\mathcal{Q}^{2n}_{\geq}} \left\| \begin{pmatrix} A_r & A_i \\ -A_i & A_r \end{pmatrix} - P \begin{pmatrix} C_r & C_i \\ -C_i & C_r \end{pmatrix} \right\|_F^2$$

Since $\mathcal{Q}^{2n}_{\geq} \subseteq \mathcal{S}^{2n}_{\geq}$, it follows that $\gamma_1 \leq \sqrt{2}\gamma_o$. It is shown next that this inequality is actually an equality since the (unique) minimizer of (6) actually belongs to $\mathcal{Q}^{2n}_{\geq}$. Let,

$$P_o = \begin{pmatrix} P_1 & P_2 \\ P_2' & P_3 \end{pmatrix}$$

be the (unique) minimizer of (6), where $P_1, P_2, P_3 \in \mathcal{R}^{n\times n}$. Then,

$$\gamma_1^2 = \left\| \begin{pmatrix} A_r & A_i \\ -A_i & A_r \end{pmatrix} - \begin{pmatrix} P_1 & P_2 \\ P_2' & P_3 \end{pmatrix} \begin{pmatrix} C_r & C_i \\ -C_i & C_r \end{pmatrix} \right\|_F^2$$

Next, consider the matrix,

$$\widehat{P}_o = \begin{pmatrix} P_3 & -P_2' \\ -P_2 & P_1 \end{pmatrix}$$

Note that $P_o \in \mathcal{S}^{2n}_{\geq}$ if and only if $\widehat{P}_o \in \mathcal{S}^{2n}_{\geq}$. Moreover, it is straightforward to verify using Lemma 3.2 part (ii) that

$$\gamma_1^2 = \left\| \begin{pmatrix} A_r & A_i \\ -A_i & A_r \end{pmatrix} - \begin{pmatrix} P_3 & -P_2' \\ -P_2 & P_1 \end{pmatrix} \begin{pmatrix} C_r & C_i \\ -C_i & C_r \end{pmatrix} \right\|_F^2$$

Hence both P_o and $\widehat{P}_o$ are optimal solutions of (6). Since, however, the optimal solution is unique, we must have $P_o = \widehat{P}_o$, which implies that $P_1 = P_3$ and that $P_2 + P_2' = 0$. This shows that the (unique) optimal solution of (7) is actually a member of the set $\mathcal{Q}^{2n}_{\geq} \subseteq \mathcal{S}^{2n}_{\geq}$, so that $\gamma_1 = \sqrt{2}\gamma_o$ and that $X_o = P_1 + jP_2$ is the unique solution of (5). □

Remark: Theorem 3.3 shows that the complex optimization (5) can be recast in the real setting of (6) by inflating the problem. In this setting, any real algorithm (e.g. the algorithm in [1] can be used to obtain a numerical solution to (6). Note however that the transformed problem is over-parametrised, since the special structure of

$$P = \begin{pmatrix} X_r & X_i \\ -X_i & X_r \end{pmatrix} \quad (12)$$

is ignored. In particular, a real $2n \times 2n$ positive semi definite matrix involves $2n^2 + n$ independent variables, whereas the matrix in the right hand side of equation (12) involves only n^2 independent variables. In the sequel we exploit this redundancy to derive a more efficient algorithm for solving the optimization problem (6).

In [1] the characterization of positive-definiteness of real symmetric matrices relies on the fact that any such matrix may be written in the form aB^2, where $a \geq 0$, B is symmetric and $\|B\|_F = 1$. It is shown next that a similar result holds for the set $\mathcal{Q}^{2n}_{\geq}$.

Lemma 3.4: (i) If $B \in \mathcal{Q}^{2n}$ then $B^2 \in \mathcal{Q}^{2n}_{\geq}$. (ii) Every matrix $A \in \mathcal{Q}^{2n}_{\geq}$, can be written as $A = aB^2$ where $a \geq 0$, $B \in \mathcal{Q}^{2n}$ and $\|B\|_F = 1$.

From the definition (5) it is clear that $\mathcal{Q}^{2n}$ is a subspace of $\mathcal{R}^{2n}$. Consider the subspace of $\mathcal{R}^{4n^2}$, $\text{vec}(\mathcal{Q}^{2n}) = \{\text{vec}(A) : A \in \mathcal{Q}^{2n}\}$. This has dimension $r = n^2$. Suppose that $\{w_1, w_2, ..., w_r\}$ is an orthonormal basis set for $\text{vec}(\mathcal{Q}^{2n})$. Then,

$$W_{\mathcal{Q}} = \begin{pmatrix} w_1 & w_2 & \ldots & w_r \end{pmatrix} \in \mathcal{R}^{4n^2\times r}$$

is a basis matrix for $\text{vec}(\mathcal{Q}^{2n})$; further, $W_{\mathcal{Q}}'W_{\mathcal{Q}} = I_r$, $\text{Range}(W_{\mathcal{Q}}') = \mathcal{R}^r$ and $\text{Range}(W_{\mathcal{Q}}) = \text{vec}(\mathcal{Q}^{2n})$. For $A \in \mathcal{Q}^{2n}$, let $\overline{\text{vec}}_{\mathcal{Q}}(A)$ denote the vector of co-ordinates of $\text{vec}(A)$ with respect to the columns of $W_{\mathcal{Q}}$. Then, $\text{vec}(A) = W_{\mathcal{Q}}\overline{\text{vec}}_{\mathcal{Q}}(A) \in \mathcal{R}^{4n^2}$ and $\overline{\text{vec}}_{\mathcal{Q}}(A) = W_{\mathcal{Q}}'\text{vec}(A) \in \mathcal{R}^r$.

Note that with the exception of dimensions, the construction of $W_{\mathcal{Q}}$ and the subsequent definitions are essentially those in [1]. Next, we define the following sets:

$$U_{\mathcal{Q}} = \{B \in \mathcal{Q}^{2n} : \|B\|_F = 1\} \quad (13)$$

and

$$\Psi_{\mathcal{Q}} = \{\text{vec}(B^2) : B \in \mathcal{U}_{\mathcal{Q}}\}, \ \ \Omega_{\mathcal{Q}} := \text{conv}[\Psi_{\mathcal{Q}}] \quad (14)$$

Also define the inverse functions $\text{vec}^{-1} : \mathcal{R}^{4n^2} \to \mathcal{R}^{2n\times 2n}$ and $\overline{\text{vec}}_{\mathcal{Q}}^{-1} : \mathcal{R}^{n^2} \to \mathcal{Q}^{2n}$ which will be used later.

In the sequel we make use of the following two results:

Lemma 3.5: (i) If X =conv(X), then cone$[X] = \{\alpha x : \alpha \geq 0, x \in X\}$. (ii) If $\Omega \subseteq \mathcal{C}^m$ is convex and $A \in \mathcal{C}^{p\times m}$, then $A\Omega$ is convex. (iii) If $\Omega \subseteq \mathcal{C}^m$ is convex and $A \in \mathcal{C}^{p\times m}$, then Acone$[\Omega]$ =cone$[A\Omega]$.

Lemma 3.6: Suppose that $\Psi \subseteq \mathcal{R}^n$ is a compact set and $\Omega =$ conv(Ψ). Then $\min_{\omega\in\Omega} g'\omega = \min_{\omega\in\Psi} g'\omega$.

Theorem 3.7 below characterises the "structured" set $\mathcal{Q}^{2n}_{\geq}$ as the conical hull of a nonempty convex compact set. This is similar to the characterisation of $\mathcal{S}^n_{\geq}$ in [1]. In fact, the proof of parts (1)-(4) is a simple generalisation of the corresponding results in [1]. Theorem 3.7 part (5) solves an optimisation problem involving a linear function over a convex set which is one of the steps of the algorithm for solving the distance problem and is outlined later in the section.

Theorem 3.7:

1. vec($\mathcal{Q}^{2n}_{\geq}$) =cone$[\Omega_{\mathcal{Q}}]$.

2. $\overline{\text{vec}}$($\mathcal{Q}^{2n}_{\geq}$) =cone$[W'_{\mathcal{Q}}\Omega_{\mathcal{Q}}]$.

3. $\Psi_{\mathcal{Q}}$ is a compact set, $\Omega_{\mathcal{Q}}$ is a non empty convex compact set with mindist$(0, \Omega_{\mathcal{Q}}) = \frac{1}{\sqrt{2n}}$ and cone$[\Omega_{\mathcal{Q}}]$ is a nonempty closed convex cone.

4. mindist$(0, LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}) \geq \frac{\sigma(L)}{\sqrt{2n}}$.

5. For $g \in \mathcal{R}^r$ and $L \in \mathcal{R}^{r\times r}$, $\min_{\gamma\in LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}} g'\gamma = \lambda_{\min}(Z)$ and a minimizing γ is $\hat{\gamma} = LW'_{\mathcal{Q}}\text{vec}(\hat{B}^2)$. Here Z =vec$^{-1}(\overline{g}) \in \mathcal{Q}^{2n}$ where $\overline{g} = W_{\mathcal{Q}}L'g$,

$$\hat{B}^2 = \frac{1}{2}\begin{pmatrix}\alpha_n & -\beta_n \\ \beta_n & \alpha_n\end{pmatrix}\begin{pmatrix}\alpha'_n & \beta'_n \\ -\beta'_n & \alpha'_n\end{pmatrix} \in \mathcal{Q}^{2n}_{\geq}$$

   in which $(\alpha'_n \ \ \beta'_n)'$ and $(-\beta'_n \ \ \alpha'_n)'$ are two normalized eigenvectors of Z corresponding to $\lambda_{\min}(Z)$, the smallest eigenvalue of Z (which is repeated).

Proof: The proof of part 5 is only given here. Note first that the minimum exists since the constraint set is compact. Now,

$$\begin{aligned}\min_{\gamma\in LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}} g'\gamma &= \min_{\omega\in\Omega_{\mathcal{Q}}} g'LW'_{\mathcal{Q}}\omega \\ &= \min_{\omega\in\Omega_{\mathcal{Q}}} \overline{g}'\omega \\ &= \min_{B\in\mathcal{U}_{\mathcal{Q}}} \overline{g}'\text{vec}(B^2) \\ &= \min_{B\in\mathcal{U}_{\mathcal{Q}}} \text{trace}\{\text{vec}^{-1}(\overline{g})B^2\} \\ &= \min_{B\in\mathcal{U}_{\mathcal{Q}}} \text{trace}\{B\text{vec}^{-1}(\overline{g})B\}\end{aligned}$$

where the fact that the minimum of a linear function over a compact set is equal to the minimum of the function over the set's convex hull (see Lemma (3.6) and Theorem 3.7 part 3). Note that since $\overline{g} \in$ Range$(W_{\mathcal{Q}})$, vec$^{-1}(\overline{g}) \in \mathcal{Q}^{2n}$. Now consider the minimization,

$$\min_{B\in\mathcal{U}_{\mathcal{S}}} \text{trace}\{B\text{vec}^{-1}(\overline{g})B\}$$

for some $\overline{g} \in$ Range$(W_{\mathcal{S}})$. From [1] Theorem 2.1(v), the minimum to this problem is given by $\lambda_{\min}(\text{vec}^{-1}(\overline{g}))$ where $\lambda_{\min}$ denotes the minimum eigenvalue of a matrix. Since Range$(W_{\mathcal{Q}}) \subseteq$ Range$(W_{\mathcal{S}})$ and $\mathcal{U}_{\mathcal{Q}} \subseteq \mathcal{U}_{\mathcal{S}}$ we have,

$$\min_{\gamma\in LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}} g'\gamma \geq \lambda_{\min}(\text{vec}^{-1}(\overline{g})) \quad (15)$$

Next we construct an explicit $B \in \mathcal{U}_{\mathcal{Q}}$ for which this lower bound is attained. Let $(\alpha'_n \ \ \beta'_n)'$ be a normalized eigenvector of

$$\text{vec}^{-1}(\overline{g}) := \begin{pmatrix}X & Y \\ -Y & X\end{pmatrix} \in \mathcal{Q}^{2n}$$

corresponding to the smallest eigenvalue, so that

$$\begin{pmatrix}X & Y \\ -Y & X\end{pmatrix}\begin{pmatrix}\alpha_n \\ \beta_n\end{pmatrix} = \lambda_{\min}(\text{vec}^{-1}(\overline{g}))\begin{pmatrix}\alpha_n \\ \beta_n\end{pmatrix} \quad (16)$$

This can be rearranged as

$$\begin{pmatrix}X & Y \\ -Y & X\end{pmatrix}\begin{pmatrix}-\beta_n \\ \alpha_n\end{pmatrix} = \lambda_{\min}(\text{vec}^{-1}(\overline{g}))\begin{pmatrix}-\beta_n \\ \alpha_n\end{pmatrix} \quad (17)$$

which shows that the smallest eigenvalue (in fact every eigenvalue) has multiplicity at least equal to two, since the two eigenvectors in (16) and (17) are orthogonal. Next define

$$\begin{aligned}\hat{B} &= \frac{1}{\sqrt{2}}\begin{pmatrix}\alpha_n & -\beta_n \\ \beta_n & \alpha_n\end{pmatrix}\begin{pmatrix}\alpha'_n & \beta'_n \\ -\beta'_n & \alpha'_n\end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix}\alpha_n\alpha'_n + \beta_n\beta'_n & \alpha_n\beta'_n - \beta_n\alpha'_n \\ \beta_n\alpha'_n - \alpha_n\beta'_n & \alpha_n\alpha'_n + \beta_n\beta'_n\end{pmatrix}\end{aligned}$$

It is clear that $\hat{B} \in \mathcal{Q}^{2n}_{\geq}$ and $\|\hat{B}\|_F = 1$, so that $\hat{B} \in \mathcal{U}_{\mathcal{Q}}$. A straightforward calculation also shows that trace$[\hat{B}\text{vec}^{-1}(\overline{g})\hat{B}] = \lambda_{\min}(\text{vec}^{-1}(\overline{g}))$ which proves the result. □

The following Theorem shows that (1) can be reduced to a minimum-distance problem from a vector to a convex cone.

The Theorem is a straightforward generalisation of a parallel result in [1].

Theorem 3.8: Consider the minimization problem

$$\gamma_o = \min_{X \in \mathcal{Q}^{2n}_{\geq}} \|F - XG\|_F \tag{18}$$

in which $F \in \mathcal{R}^{2nm}$, $(m \geq n)$, and $\text{Rank}(G) = 2n$. Let $J = [I_{2n\times 2n} \otimes G']W \in \mathcal{R}^{4mn\times n^2}$. Then $\text{Rank}(J) = n^2$. Factor J as $J = P[L' \quad 0]'$ for some invertible matrix $L \in \mathcal{R}^{n^2\times n^2}$ with $P \in \mathcal{R}^{4mn\times 4mn}$ orthogonal. Using P factor $f :=\text{vec}(F) \in \mathcal{R}^{4mn}$ as $f = P[u' \quad l']'$ with $u \in \mathcal{R}^{n^2}$. Then $\hat{X}$ solves (18) if and only if

$$\hat{X} = \overline{\text{vec}}_{\mathcal{Q}}^{-1}(L^{-1}\hat{k})$$

where $\hat{k}$ is the unique solution of

$$\min_{k\in\mathcal{K}} \|u - k\|$$

and

$$\|F - \hat{X}G\|_F^2 = \|u - \hat{k}\|^2 + \|l\|^2$$

Here $\mathcal{K}$ is the convex cone

$$\mathcal{K} = \text{cone}[LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}]$$

Further,

$$\hat{k} = u \quad \text{iff} \quad \overline{\text{vec}}_{\mathcal{Q}}^{-1}(L^{-1}u) \geq 0$$

and

$$\hat{X} \notin \mathcal{Q}^{2n}_{>} \quad \text{if} \quad \hat{k} \neq u \tag{19}$$

where $\mathcal{Q}^{2n}_{>}$ denotes the subset of $\mathcal{Q}^{2n}_{\geq}$ with positive definite matrix elements.

The following theorem provides a lower bound to the optimal cost γ_0 at the end of each iteration. This is almost identical to a parallel result in [1].

Theorem 3.9: Suppose that F, G, L, u and l are as defined in Theorem 3.8, with $\text{Rank}(G) = 2n$. Consider any real $\epsilon \geq 0$. Then, if $\overline{k} \in \mathcal{K}$ satisfies,

$$\|u - \overline{k}\|^2 + \|l\|^2 \leq (1+\epsilon)^2(\|u - \hat{k}\|^2 + \|l\|^2) \tag{20}$$

and $\overline{X} = \overline{\text{vec}}_{\mathcal{Q}}^{-1}(L^{-1}\overline{k})$, then $\overline{X}$ approximates $\hat{X}$ of Theorem 3.8, in that :

1. $\overline{X} \in \mathcal{Q}^{2n}_{\geq}$
2. $\|\overline{X} - \hat{X}\|_F \leq \sqrt{2\epsilon + \epsilon^2}\ \|L^{-1}\|\ \|F - \hat{X}G\|_F$
3. $\|F - \overline{X}G\|_F \leq (1+\epsilon)\|F - \hat{X}G\|_F$.

As shown in Theorem 3.8, problem (1) can be reduced to the calculation of the shortest distance between a vector u and the conical hull of a closed convex set, $\mathcal{K} = \text{cone}(\Gamma)$, where $\Gamma = LW'_{\mathcal{Q}}\Omega_{\mathcal{Q}}$. This can be restricted to a bounded convex set S without affecting the optimal solution. As expected, the truncation is u-dependent and is summarised below:

Lemma 3.10 Let $\eta = \|u\|/\pi$ where $\pi = \frac{\sigma_{\min}[L]}{\sqrt{2n}}$ and define $S = \{\alpha\gamma : \alpha \in [0,\eta], \gamma \in \Gamma\}$. Then S is a convex set and $\text{minpoint}[u, S] = \text{minpoint}[u, \text{cone}[\Gamma]]$.

The modified optimisation problem $\text{mindist}[u, S]$ may now be solved using the algorithm of [1]. In fact, the relevant proof may be reproduced, almost word for word, to show that the algorithm converges to the optimal solution. However, a number of modifications in the implementation of certain steps of the algorithm are necessary, to account for the special structure of the problem in the complex case. These are described at the end of the section.

Algorithm 3.1 [1]: The following algorithm finds an ϵ-suboptimal approximation $\overline{k}$ to $\hat{k}$ by minimizing $v(k) = \|u - k\|^2$ over $\mathcal{K}$ as proposed in Theorem 3.8 and Lemma 3.10.

0. *Select parameters:* Choose $\epsilon \in (0,\infty)$ (ϵ defines the degree of sub-optimality acceptable in $\overline{k}$) and $\overline{k}_0$ $\in$cone$[\Gamma]$ (an initial estimate of $\hat{k}$).

I. *Initialize variables:*
k_0 := minpoint$[u,$ cone$[\overline{k}_0]]$ (the point nearest u in the ray through 0 and $\overline{k}_0$), $\hat{b}_{-1} := 0$ (the best lower bound for $v(\hat{k})$ available so far), $i := 0$.

II. *Decide when to stop iterating:*
Find a $y_i \in \arg\min_{y\in S} \nabla v(k_i')(y_i - k_i)$ by finding:

$$\gamma_i \in \arg\min_{\gamma\in\Gamma} \nabla v(k_i)'\gamma$$

and setting:

$$y_i = \eta\gamma_i \ \text{ if } \ \nabla v(k_i)'\gamma_i < 0 \ \text{ and } \ y_i = k_i \ \text{ otherwise.}$$

Compute a lower bound b_i for $v(\hat{k})$:

$$b_i := v(k_i) + \nabla v(k_i)'(y_i - k_i).$$

Compute $\hat{b}_i$, the best lower bound for $v(k)$ found so far:

$$\hat{b}_i := \max\{\hat{b}_{i-1}, b_i\}.$$

If

$$[v(k_i) + \|l\|^2] \leq (1+\epsilon)^2[\hat{b}_i + \|l\|^2]$$

then set $\overline{k} = k_i$ and stop; else continue. III. *Choose the next iterand:*

$$k_{i+1} := \text{minpoint}(u, \text{cone}[\text{line}\{k_i, y_i\}]),$$
$$i := i + 1.$$

Go to II. □

Algorithm 3.1 can be used to solve the complex approximation problem. However, the implementation of certain steps of the algorithm needs to be modified in this case. The required modifications are:

- The construction of the basis matrix $W_{\mathcal{Q}}$ is more complicated and needs to take into account the block structure of set $\mathcal{Q}^{2n}$.

- The minimisation $\gamma_i \in \arg\min_{\gamma\in\Gamma} \nabla v(k_i)'\gamma$ in step II is now solved using Theorem 3.7 part 5.

- Although any $\overline{k}_0 \in \mathrm{cone}(\Gamma)$ can be selected as an initial estimate of $\hat{k}$ in step 0 of the algorithm, a good approximation to $\hat{k}$ enhances considerable the convergence properties of the algorithm. In [1] it is argued that the most appropriate choice is to take $\overline{k}_0 = L\overline{\mathrm{vec}}_{\mathcal{S}}(\hat{X})$ where $\hat{X}$ minimises $\|\overline{\mathrm{vec}}_{\mathcal{S}}^{-1}(L^{-1}u) - X\|_F$ with respect to $X \in \mathcal{S}^n_{\geq}$. The optimal $\hat{X}$ in this case is given as $\hat{X} = [\overline{\mathrm{vec}}_{\mathcal{S}}^{-1}(L^{-1}u)]_0$ where $[M]_0$ denotes the result of changing to zero all negative eigenvalues in the spectral form of M. The main advantage of this choice is that the algorithm converges immediately (i.e. in one iteration) to the optimal solution, when this happens to be positive definite. In the complex case, a similar argument may be used to show that $\overline{k}_0$ should be selected as: $\overline{k}_0 = L\overline{\mathrm{vec}}_{\mathcal{Q}}([\overline{\mathrm{vec}}_{\mathcal{Q}}^{-1}(L^{-1}u)]_0)$. This follows from the fact that $[\overline{\mathrm{vec}}_{\mathcal{Q}}^{-1}(L^{-1}u)]_0 \in \mathcal{Q}^{2n}_{\geq}$ which is established in the Lemma below.

Lemma 3.11 If $A \in \mathcal{Q}^{2n}$ then $[A]_0 \in \mathcal{Q}^{2n}_{\geq}$.

Proof: It follows from the proof of Lemma 3.7 part 5 that every $A \in \mathcal{Q}^{2n}$ has eigenvalues which appear in pairs; further each eigenvalue λ_i corresponds to an eigenvector pair of the form:

$$\left\{ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}, \begin{pmatrix} -\beta_i \\ \alpha_i \end{pmatrix} \right\}$$

for $i = 1, 2, \ldots, n$. Let A have a spectral decomposition $A = U\Lambda U'$, where $UU' = I$ and where the eigenvalues have been ordered in non-increasing order, i.e. $\lambda_i \geq \lambda_{i+1}$ for $i = 1, 2, \ldots, n-1$. Then,

$$\begin{aligned} A &= \sum_{i=1}^{n} \begin{pmatrix} \alpha_i & -\beta_i \\ \beta_i & \alpha_i \end{pmatrix} \begin{pmatrix} \lambda_i & 0 \\ 0 & \lambda_i \end{pmatrix} \begin{pmatrix} \alpha_i' & \beta_i' \\ -\beta_i' & \alpha_i' \end{pmatrix} \\ &= \sum_{i=1}^{n} \lambda_i \begin{pmatrix} \alpha_i\alpha_i' + \beta_i\beta_i' & \alpha_i\beta_i' - \beta_i\alpha_i' \\ \beta_i\alpha_i' - \alpha_i\beta_i' & \alpha_i\alpha_i' + \beta_i\beta_i' \end{pmatrix} = \sum_{i=1}^{n} \lambda_i Q_i \end{aligned}$$

where, $Q_i \in \mathcal{Q}^{2n}_{\geq}$ and have rank equal to 2. Thus,

$$[A]_0 = \sum_{\{i:\lambda_i\geq 0\}}^{k} \lambda_i Q_i \in \mathcal{Q}^{2n}_{\geq}$$

as required. □

4 Numerical example

The paper has presented two methods for solving problem (2). The first is based on Theorem 3.3 and optimises the cost over a (real) positive semi-definite dilated matrix, without imposing any special structure. The second uses the special structure of the dilated matrix and the techniques developed in the later parts of the paper. A numerical example involving the two methods will be presented at the Conference. Computational experience suggests that the second algorithm always converges much faster to the optimum solution compared to the first, both in terms of iterations and computation time. This is not surprising, since the second algorithm removes the over-parametrisation of the problem, by restricting the search for the optimal solution from an $(2n^2 + n)$-dimensional space to an n^2-dimensional space.

5 Conclusions

This paper has presented an algorithm for solving the complex version of the Procrustes problem. This is a least-squares approximation problem over the cone of Hermitian, positive semidefinite matrices. It was first shown that it is possible to transform the problem to a real setting by inflating its dimension. Although in this case the real algorithm of [1] is directly applicable, the problem is over-parametrised and hence the convergence properties of the algorithm are poor. The necessary modifications to this algorithm have been indicated, resulting in significantly accelerated convergence.

References

[1] Allwright J.C., Positive semidfinite matrices: Characterization via conical hulls and least squares solution of a matriux equation, *SIAM J Control and Optimization*, 26(3):537-556, 1988.

[2] Golub G.H. and Van Loan C.F., Matrix Computations, *The John Hopkins University Press*, 1996.

[3] Green B.F., The orthogonal approximation of an oblique structure in factor analysis, *Psychometrica*, 17(4):429-440, 1952.

[4] Higham N.J., The symmetric Procrustes problem, *Psychometrica*, 28:133-143, 1988.

[5] Horn R. and Johnson C., Topics in matrix analysis, *Cambridge University Press*, 1991.

[6] Rockafellar R.T., Convex Analysis, *Princeton Landmarks in Mathematics, Princeton University Press*, 1997.

[7] Schonemann P.H., A generalized solution of the orthogonal Procrustes problem, *Psychometrica*, 31:1-10, 1966.

[8] Woodgate K.G., A new algorithm for the positive semidefinite Procrustes problem, *Proc. 32nd Conf. on Decision & Control*, San Antonio, Texas, 3596-3601, 1993.

[9] Wahba G., Problem 65-1: A least squares estimate of sattelite altitude, *SIAM Review*, 7:409, 1965. Solutions in 8:384-386, 1966.