Comparing Two Matrices by Means of Isometric Projections*

Thomas Cason, Pierre-Antoine Absil, and Paul Van Dooren

Université catholique de Louvain, Department of Mathematical Engineering
Bâtiment Euler, avenue Georges Lemaître 4, 1348 Louvain-la-Neuve, Belgium
http://www.inma.ucl.ac.be/{∼cason, ∼absil, ∼vandooren}

Abstract

In this paper, we go over a number of optimization problems defined on a manifold in order to compare two matrices, possibly of different order. We consider several variants and show how these problems relate to various specific problems from the literature.

1 Introduction

When comparing two matrices A and B it is often natural to allow for a class of transformations acting on these matrices. For instance, when comparing adjacency matrices A and B of two graphs with an equal number of nodes, one can allow symmetric permutations $P^T AP$ on one matrix in order to compare it to B, since this is merely a relabelling of the nodes of A. The so-called comparison then consists in finding the best match between A and B under this class of transformations.

A more general class of transformations would be that of unitary similarity transformations Q^*AQ, where Q is a unitary matrix. This leaves the eigenvalues of A unchanged but rotates its eigenvectors, which will of course play a role in the comparison between A and B. If A and B are of different order, say m and n, one may want to consider their restriction on a lower dimensional subspace:

$$U^*AU \quad \text{and} \quad V^*BV, \tag{1}$$

with U and V belonging to $St(k, m)$ and $St(k, n)$ respectively, and where $St(k, m) = \left\{U \in \mathbb{C}^{m\times k} : U^*U = I_k\right\}$ denotes the *compact Stiefel manifold*. This yields two square matrices of equal dimension $k \leq \min(m, n)$, which can again be compared.

But one still needs to define a measure of comparison between these restrictions of A and B which clearly depends on U and V. Fraikin et *al.* [1] propose in this context to maximize the inner product between the *isometric projections*,

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

U^*AU and V^*BV, namely:

$$\arg \max_{\substack{U^*U=I_k \\ V^*V=I_k}} \langle U^*AU, V^*BV \rangle := \Re \operatorname{tr}\big((U^*AU)^* (V^*BV)\big),$$

where $\Re$ denotes the real part of a complex number. They show this is also equivalent to

$$\arg \max_{\substack{X=VU^* \\ U^*U=I_k \\ V^*V=I_k}} \langle XA, BX \rangle = \Re \operatorname{tr}(A^*X^*BX),$$

and eventually show how this problem is linked to the notion of graph similarity introduced by Blondel et *al.* in [2]. The graph similarity matrix S introduced in that paper also proposes a measure of comparing two matrices A and B via the fixed point of a particular iteration. But it is shown in [3] that this is equivalent to the optimization problem

$$\arg \max_{\|S\|_F=1} \langle SA, BS \rangle = \Re \operatorname{tr}((SA)^*BS)$$

or also

$$\arg \max_{\|S\|_F=1} \langle S, BSA^* \rangle = \Re \operatorname{tr}((SA)^*BS).$$

Notice that S also belongs to a Stiefel manifold, since $\operatorname{vec}(S) \in \operatorname{St}(1, mn)$.

In this paper, we use a distance measure rather than an inner product to compare two matrices. As distance measure between two matrices M and N, we will use

$$\operatorname{dist}(M, N) = \|M - N\|_F = \sqrt{\operatorname{tr}((M-N)^*(M-N))}.$$

We will analyze distance minimization problems that are essentially the counterparts of the similarity measures defined above. These are

$$\arg \min_{\substack{U^*U=I_k \\ V^*V=I_k}} \operatorname{dist}^2(U^*AU, V^*BV),$$

$$\arg \min_{\substack{X=VU^* \\ U^*U=I_k \\ V^*V=I_k}} \operatorname{dist}^2(XA, BX),$$

and

$$\arg \min_{\substack{X=VU^* \\ U^*U=I_k \\ V^*V=I_k}} \operatorname{dist}^2(X, BXA^*),$$

for the problems involving two isometries U and V. Notice that these three distance problems are not equivalent although the corresponding inner product problems are equivalent.

Similarly, we will analyze the two problems

$$\arg \min_{\|S\|_F=1} \operatorname{dist}^2(SA, BS) = \operatorname{tr}\big((SA-BS)^* (SA-BS)\big)$$

and

$$\arg \min_{\|S\|_F=1} \operatorname{dist}^2(S, BSA^*) = \operatorname{tr}\big((S-BSA^*)^* (S-BSA^*)\big),$$

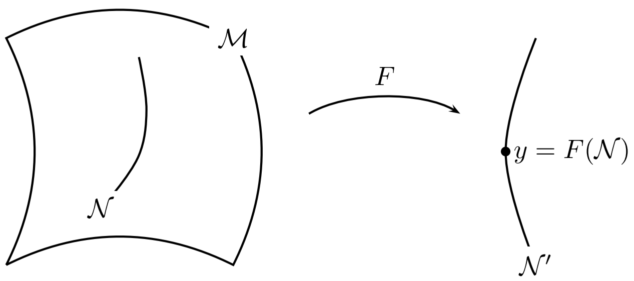
for the problems involving a single matrix S. Again, these are not equivalent in their distance formulation although the corresponding inner product problems are equivalent.

We will develop optimality conditions for those different problems, indicate their relations with existing problems from the literature and give an analytic solution for particular matrices A and B.

2 The Problems and their Geometry

All those problems are defined on feasible sets that have a *manifold* structure. Roughly speaking, this means that the feasible set is locally smoothly identified with $\mathbb{R}^d$, where d is the dimension of the manifold. Optimization on a manifold generalizes optimization in $\mathbb{R}^d$ while retaining the concept of smoothness. We refer the reader to [4, 5] for details.

A well known and largely used class of manifolds is the class of embedded submanifolds. The submersion theorem gives a useful sufficient condition to prove that a subset of a manifold $\mathcal{M}$ is an embedded submanifold of $\mathcal{M}$. If there exists a smooth mapping $F : \mathcal{M} \to \mathcal{N}'$ between two manifolds of dimension d_m and $d'_n (< d_m)$ and $y \in \mathcal{N}'$ such that the rank of F is equal to d'_n at each point of $\mathcal{N} := F^{-1}(y)$, then $\mathcal{N}$ is a embedded submanifold of $\mathcal{M}$ and the dimension of $\mathcal{N}$ is $d_m - d'_n$.



Example The unitary group $\mathrm{U}(n) = \{Q \in \mathbb{C}^{n\times n} : Q^*Q = I_n\}$ is an embedded submanifold of $\mathbb{C}^{n\times n}$. Indeed, consider the function

$$F : \mathbb{C}^{n\times n} \to \mathcal{S}_{Her}(n) : Q \mapsto Q^*Q - I_n$$

where $\mathcal{S}_{Her}(n)$ denotes the set of Hermitian matrices of order n. Clearly, $\mathrm{U}(n) = F^{-1}(0_n)$. It remains to show for all $\hat{H} \in \mathcal{S}_{Her}(k)$, there exists an $H \in \mathbb{C}^{n\times n}$ such that $\mathrm{D}F(Q)\cdot H = Q^*H + H^*Q = \hat{H}$. It is easy to see that $\mathrm{D}F(Q)\cdot(Q\hat{H}/2) = \hat{H}$, and according to the submersion theorem, it follows that $\mathrm{U}(n)$ is an embedded submanifold of $\mathbb{C}^{n\times n}$. The dimension of $\mathbb{C}^{n\times n}$ and $\mathcal{S}_{Her}(n)$ are $2n^2$ and n^2 respectively. Hence $\mathrm{U}(n)$ is of dimension n^2.

In our problems, embedding spaces are matrix-Euclidean spaces $\mathbb{C}^{m\times k} \times \mathbb{C}^{n\times k}$ and $\mathbb{C}^{n\times m}$ which have a trivial manifold structure since $\mathbb{C}^{m\times k} \times \mathbb{C}^{n\times k} \simeq \mathbb{R}^{2mnk^2}$ and $\mathbb{C}^{n\times m} \simeq \mathbb{R}^{2mn}$. For each problem, we further analyze whether or not the feasible set is an embedded submanifold of their embedding space.

When working with a function on a manifold $\mathcal{M}$, one may be interested in having a local linear approximation of that function. Let M be an element of

$\mathcal{M}$ and $\mathfrak{F}_M(\mathcal{M})$ denote the set of smooth real-valued functions defined on a neighborhood of M.

Definition 1 *A tangent vector ξ_M to a manifold $\mathcal{M}$ at a point M is a mapping from $\mathfrak{F}_M(\mathcal{M})$ to $\mathbb{R}$ such that there exists a curve γ on $\mathcal{M}$ with $\gamma(0) = M$, satisfying*

$$\xi_M f = \left.\frac{\mathrm{d}f(\gamma(t))}{\mathrm{d}t}\right|_{t=0}, \qquad \forall f \in \mathfrak{F}_M(\mathcal{M}).$$

Such a curve γ is said to realize the tangent vector ξ_x.

So, the only thing we need to know about a curve γ in order to compute the first-order variation of a real-value function f at $\gamma(0)$ along γ is the tangent vector ξ_x realized by γ. The tangent space to $\mathcal{M}$ at M, denoted by $T_M\mathcal{M}$, is the set of all tangent vectors to $\mathcal{M}$ at M and it admits a structure of vector space over $\mathbb{R}$. When considering an embedded submanifold in a Euclidean space $\mathcal{E}$, any tangent vector ξ_M of the manifold is equivalent to a vector E of the Euclidean space. Indeed, let $\hat{f}$ be any a differentiable continuous extension of f on $\mathcal{E}$, we have

$$\xi_M f := \left.\frac{\mathrm{d}f(\gamma(t))}{\mathrm{d}t}\right|_{t=0} = \mathrm{D}\hat{f}(M) \cdot E, \tag{2}$$

where E is $\dot{\gamma}(0)$ and D is the directional derivative operator

$$\mathrm{D}\hat{f}(M) \cdot E = \lim_{t \to 0} \frac{\hat{f}(M + tE) - \hat{f}(M)}{t}.$$

The tangent space reduces to a linear subspace of the original space $\mathcal{E}$.
Example Let $\gamma(t)$ be a curve on the unitary group $\mathrm{U}(n)$ passing through Q at $t = 0$, *i.e.* $\gamma(t)^*\gamma(t) = I_n$ and $\gamma(0) = Q$. Differentiating with respect to t yields

$$\dot{\gamma}(0)^* Q + Q^* \dot{\gamma}(0) = 0_n.$$

One can see from equation (2) that the tangent space to $\mathrm{U}(n)$ at Q is contained in

$$\left\{E \in \mathbb{C}^{n \times n} : E^* Q + Q^* E = 0_n\right\} = \left\{Q\Omega \in \mathbb{C}^{n \times n} : \Omega^* + \Omega = 0_n\right\}. \tag{3}$$

Moreover, this set is a vector space over $\mathbb{R}$ of dimension n^2, and hence is the tangent space itself.

Let g_M be an inner product defined on the tangent plane $T_M\mathcal{M}$. The gradient of f at M, denoted $\operatorname{grad} f(M)$, is defined as the unique element of the tangent plane $T_M\mathcal{M}$, that satisfies

$$\xi_M f = g_M(\operatorname{grad} f(M), \xi_M), \qquad \forall \xi_M \in T_M\mathcal{M}.$$

The gradient, together with the inner product, fully characterizes the local first order approximation of a smooth function defined on the manifold. In the case of an embedded manifold of a Euclidean space $\mathcal{E}$, since $T_M\mathcal{M}$ is a linear subspace of $T_M\mathcal{E}$, an inner product $\hat{g}_M$ on $T_M\mathcal{E}$ generates by restriction an inner product g_M on $T_M\mathcal{M}$. The orthogonal complement of $T_M\mathcal{M}$ with respect to $\hat{g}_M$ is called the normal space to $\mathcal{M}$ at M and denoted by $(T_M\mathcal{M})^\perp$. The gradient of a smooth function $\hat{f}$, defined on the embedding manifold may be decomposed into its orthogonal projection on the tangent and normal space, respectively

$$\mathrm{P}_M \operatorname{grad} \hat{f}(M) \qquad \text{and} \qquad \mathrm{P}_M^\perp \operatorname{grad} \hat{f}(M),$$

and it follows that the gradient of f (the restriction of $\hat{f}$ on $\mathcal{M}$) is the projection on the tangent space of the gradient of $\hat{f}$

$$\operatorname{grad} f(M) = \mathrm{P}_M \operatorname{grad} \hat{f}(M).$$

Example Let A and B, two Hermitian matrices. We define

$$\hat{f} : \mathbb{C}^{n\times n} \to \mathbb{R} : Q \mapsto \Re\operatorname{tr}(Q^* AQB),$$

and f its restriction on the unitary group $\mathrm{U}(n)$. We have

$$\mathrm{D}\hat{f}(Q) \cdot E = 2\Re\operatorname{tr}(E^* AQB).$$

We endow the tangent space $T_Q\mathbb{C}^{n\times n}$ with an inner product

$$\hat{g} : T_Q\mathbb{C}^{n\times n} \times T_Q\mathbb{C}^{n\times n} \to \mathbb{R} : E, F \mapsto \Re\operatorname{tr}(E^* F),$$

and the gradient of $\hat{f}$ at Q is then given by $\operatorname{grad} \hat{f}(Q) = 2AQB$. One can further define an orthogonal projection on $T_Q\mathrm{U}(n)$

$$\mathrm{P}_Q E := E - Q \operatorname{Her}(Q^* E),$$

and the gradient of f at Q is given by $\operatorname{grad} f(Q) = \mathrm{P}_Q \operatorname{grad} \hat{f}(Q)$.

Those relations are useful when one wishes to analyze optimization problems, and will hence be further developed for the problems we are interested in.

Below we look at the various problems introduced earlier and focus on the first problem to make these ideas more explicit.

Problem 1 *Given $A \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times n}$, let*

$$\hat{f} : \mathbb{C}^{m\times k} \times \mathbb{C}^{n\times k} \to \mathbb{C} : (U, V) \mapsto \hat{f}(U, V) = \operatorname{dist}^2(U^*AU, V^*BV),$$

find the minimizer of

$$f : \mathrm{St}\,(k, m) \times \mathrm{St}\,(k, n) \to \mathbb{C} : (U, V) \mapsto f(U, V) = \hat{f}(U, V),$$

where

$$\mathrm{St}\,(k, m) = \left\{U \in \mathbb{C}^{m\times k} : U^* U = I_k\right\}$$

denotes the compact Stiefel manifold.

Let $A = (A_1, A_2)$ and $B = (B_1, B_2)$ be pairs of matrices. We define the following useful operations:

- an entrywise product, $A \diamond B = (A_1 B_1, A_2 B_2)$,
- a contraction product, $A \star B = A_1 B_1 + A_2 B_2$, and
- a conjugate-transpose operation, $A^* = (A_1^*, A_2^*)$.

The definitions of the binary operations, $\diamond$ and $\star$, are (for readability) extended to single matrices when one has to deal with pairs of identical matrices. Let, for instance, $A = (A_1, A_2)$ be a pair of matrices and B be a single matrix, we define

$$A \diamond B = (A_1, A_2) \diamond B = (A_1, A_2) \diamond (B, B) = (A_1 B, A_2 B)$$
$$A \star B = (A_1, A_2) \star B = (A_1, A_2) \star (B, B) = A_1 B + A_2 B.$$

The feasible set of Problem 1 is given by the cartesian product of two compact Stiefel manifolds, namely $\mathcal{M} = \mathrm{St}\,(k,m) \times \mathrm{St}\,(k,n)$ and is hence a manifold itself (*cf.* [5]). Moreover, we can prove that $\mathcal{M}$ is an embedded submanifold of $\mathcal{E} = \mathbb{C}^{m\times k} \times \mathbb{C}^{n\times k}$. Indeed, consider the function

$$F : \mathcal{E} \to \mathcal{S}_{Her}(k) \times \mathcal{S}_{Her}(k) : M \mapsto M^* \diamond M - (I_k, I_k)$$

where $\mathcal{S}_{Her}(k)$ denotes the set of Hermitian matrices of order k. Clearly, $\mathcal{M} = F^{-1}(0_k, 0_k)$. It remains to show that each point $M \in \mathcal{M}$ is a regular value of F which means that F has full rank, *i.e.* for all $\hat{Z} \in \mathcal{S}_{Her}(k)\times\mathcal{S}_{Her}(k)$, there exists $Z \in \mathcal{E}$ such that $\mathrm{D}F(M)\cdot Z = \hat{Z}$. It is easy to see that $\mathrm{D}F(M)\cdot(M \diamond \hat{Z}/2) = \hat{Z}$, and according to the submersion theorem, it follows that $\mathcal{M}$ is an embedded submanifold of $\mathcal{E}$.

The tangent space to $\mathcal{E}$ at a point $M = (U,V) \in \mathcal{E}$ is the embedding space itself (*i.e.* $T_M\mathcal{E} \simeq \mathcal{E}$), whereas the tangent space to $\mathcal{M}$ at a point $M = (U,V) \in \mathcal{M}$ is given by

$$\begin{aligned} T_M\mathcal{M} &:= \{\dot\gamma(0) : \gamma, \text{ differentiable curve on } \mathcal{M} \text{ with } \gamma(0) = M\} \\ &= \{\xi = (\xi_U, \xi_V) : \mathrm{Her}(\xi^* \diamond M) = 0\} \\ &= \left\{ M \diamond \begin{pmatrix} \Omega_U \\ \Omega_V \end{pmatrix} + M_\perp \diamond \begin{pmatrix} K_U \\ K_V \end{pmatrix} : \Omega_U, \Omega_V \in \mathcal{S}_{s-Her}(k) \right\}, \end{aligned}$$

where $M_\perp = (U_\perp, V_\perp)$ with $U_\perp$ and $V_\perp$ any orthogonal complement of respectively U and V, where $\mathrm{Her}(\cdot)$ stands for

$$\mathrm{Her}(\cdot) : X \mapsto (X + X^*)/2,$$

and where $\mathcal{S}_{s-Her}(k)$ denotes the set of skew-Hermitian matrices of order k. We endow the tangent space $T_M\mathcal{E}$ with an inner product:

$$\hat{g}_M(\cdot,\cdot) : T_M\mathcal{E} \times T_M\mathcal{E} \to \mathbb{C} : \xi, \zeta \mapsto \hat{g}_M(\xi,\zeta) = \Re\,\mathrm{tr}(\xi^* \star \zeta),$$

and define its restriction on the tangent space $T_M\mathcal{M}(\subset T_M\mathcal{E})$:

$$g_M(\cdot,\cdot) : T_M\mathcal{M} \times T_M\mathcal{M} \to \mathbb{C} : \xi, \zeta \mapsto g_M(\xi,\zeta) = \hat{g}_M(\xi,\zeta).$$

One may now define the normal space to $\mathcal{M}$ at a point $M \in \mathcal{M}$:

$$\begin{aligned} T_M^\perp\mathcal{M} &:= \{\xi : \hat{g}_M(\xi,\zeta) = 0,\ \forall \zeta \in T_M\mathcal{M}\} \\ &= \{M \diamond (H_U, H_V) : H_U, H_V \in \mathcal{S}_{Her}(k)\}, \end{aligned}$$

where $\mathcal{S}_{Her}(k)$ denotes the set of Hermitian matrices of order k.

Problem 2 *Given $A \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times n}$, let*

$$\hat{f} : \mathbb{C}^{n\times m} \to \mathbb{C} : X \mapsto \hat{f}(X) = \mathrm{dist}^2\,(XA, BX)$$

find the minimizer of

$$f : \mathcal{M} \to \mathbb{C} : X \mapsto f(X) = \hat{f}(X),$$

where $\mathcal{M} = \{VU^* \in \mathbb{C}^{n\times m} : (U,V) \in \mathrm{St}\,(k,m) \times \mathrm{St}\,(k,n)\}$.

$\mathcal{M}$ is a smooth and connected manifold. Indeed, let $\Sigma := \left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right]$ be an element of $\mathcal{M}$. Since every $X \in \mathcal{M}$ is congruent to Σ by the congruence action $((\tilde{U}, \tilde{V}), X) \mapsto \tilde{V}^* X \tilde{U}$, $(\tilde{U}, \tilde{V}) \in \mathrm{U}(m) \times \mathrm{U}(n)$, where $\mathrm{U}(n) = \{U \in \mathbb{C}^{n \times n} : U^* U = I_n\}$ denotes the unitary group of degree n. The set $\mathcal{M}$ is an orbit of this smooth complex algebraic Lie group action of $\mathrm{U}(m) \times \mathrm{U}(n)$ on $\mathbb{C}^{n \times m}$ and therefore a smooth manifold [6, App. C]. $\mathcal{M}$ is the image of the connected subset $\mathrm{U}(m) \times \mathrm{U}(n)$ of the continuous (and in fact smooth) map $\pi : \mathrm{U}(m) \times \mathrm{U}(n) \to \mathbb{C}^{n \times m}, \pi(\tilde{U}, \tilde{V}) = \tilde{V}^* X \tilde{U}$, and hence is also connected.
The tangent space to $\mathcal{M}$ at a point $X = VU^* \in \mathcal{M}$ is

$$
\begin{aligned}
T_X \mathcal{M} &:= \{\dot{\gamma}(0) : \gamma \text{ curve on } \mathcal{M} \text{ with } \gamma(0) = X\} \\
&= \{\xi_V U^* + V \xi_U^* : \mathrm{Her}(V^* \xi_V) = \mathrm{Her}(U^* \xi_U) = 0_k\} \\
&= \{V \Omega U^* + V K_U^* U_\perp^* + V_\perp K_V U^* : \Omega \in \mathcal{S}_{s-Her}(k)\}.
\end{aligned}
$$

We endow the tangent space $T_X \mathbb{C}^{n \times m} \simeq \mathbb{C}^{n \times m}$ with an inner product:

$$
\hat{g}_X(\cdot, \cdot) : T_X \mathbb{C}^{n \times m} \times T_X \mathbb{C}^{n \times m} \mapsto \mathbb{C} : \xi, \zeta \to \hat{g}_X(\xi, \zeta) = \Re \operatorname{tr}(\xi^* \zeta),
$$

and define its restriction on the tangent space $T_X \mathcal{M} (\subset T_X \mathcal{E})$:

$$
g_X(\cdot, \cdot) : T_X \mathcal{M} \times T_X \mathcal{M} \mapsto \mathbb{C} : \xi, \zeta \to g_X(\xi, \zeta) = \hat{g}_X(\xi, \zeta).
$$

One may now define the normal space to $\mathcal{M}$ at a point $X \in \mathcal{M}$:

$$
\begin{aligned}
T_X^\perp \mathcal{M} &:= \{\xi : \hat{g}_X(\xi, \zeta) = 0, \ \forall \zeta \in T_X \mathcal{M}\} \\
&= \{V H U^* + V_\perp K U_\perp^* : H \in \mathcal{S}_{Her}(k)\}.
\end{aligned}
$$

Problem 3 *Given $A \in \mathbb{C}^{m \times m}$ and $B \in \mathbb{C}^{n \times n}$, let*

$$
\hat{f} : \mathbb{C}^{n \times m} \to \mathbb{C} : X \mapsto \hat{f}(X) = \mathrm{dist}^2 (X, BXA^*)
$$

find the minimizer of

$$
f : \mathcal{M} \to \mathbb{C} : X \mapsto f(X) = \hat{f}(X),
$$

where $\mathcal{M} = \{VU^* \in \mathbb{C}^{n \times m} : (U, V) \in \mathrm{St}(k, m) \times \mathrm{St}(k, n)\}$.

Since they have the same feasible set, topological developments obtained for Problem 2 hold also for Problem 3.

Problem 4 *Given $A \in \mathbb{C}^{m \times m}$ and $B \in \mathbb{C}^{n \times n}$, let*

$$
\hat{f} : \mathbb{C}^{n \times m} \to \mathbb{C} : S \mapsto \hat{f}(S) = \mathrm{dist}^2(SA, BS)
$$

find the minimizer of

$$
f : \mathcal{M} \to \mathbb{C} : X \mapsto f(X) = \hat{f}(X),
$$

where $\mathcal{M} = \{S \in \mathbb{C}^{n \times m} : \|S\|_F = 1\}$.

The tangent space to $\mathcal{M}$ at a point $S \in \mathcal{M}$ is

$$T_S\mathcal{M} = \{\xi : \Re\,\mathrm{tr}(\xi^* S) = 0\}.$$

We endow the tangent space $T_S\mathbb{C}^{n\times m} \simeq \mathbb{C}^{n\times m}$ with an inner product:

$$\hat{g}_S(\cdot,\cdot) : T_S\mathbb{C}^{n\times m} \times T_S\mathbb{C}^{n\times m} \mapsto \mathbb{C} : \xi, \zeta \to \hat{g}_S(\xi,\zeta) = \Re\,\mathrm{tr}(\xi^*\zeta),$$

and define its restriction on the tangent space $T_S\mathcal{M}(\subset T_S\mathcal{E})$:

$$g_S(\cdot,\cdot) : T_S\mathcal{M} \times T_S\mathcal{M} \mapsto \mathbb{C} : \xi, \zeta \to g_S(\xi,\zeta) = \hat{g}_S(\xi,\zeta).$$

One may now define the normal space to $\mathcal{M}$ at a point $S \in \mathcal{M}$:

$$T_S^{\perp}\mathcal{M} := \{\xi : \hat{g}_S(\xi,\zeta) = 0,\ \forall \zeta \in T_S\mathcal{M}\} = \{\alpha S : \alpha \in \mathbb{R}\}$$

Problem 5 *Given $A \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times n}$, let*

$$\hat{f} : \mathbb{C}^{n\times m} \to \mathbb{C} : S \mapsto \hat{f}(S) = \mathrm{dist}^2(S, BSA^*)$$

find the minimizer of

$$f : \mathcal{M} \to \mathbb{C} : X \mapsto f(X) = \hat{f}(X),$$

where $\mathcal{M} = \{S \in \mathbb{C}^{n\times m} : \|S\|_F = 1\}$.

Since they have the same feasible set, topological developments obtained for Problem 4 also hold for Problem 5.

3 Optimality conditions

Our problems are optimization problems of smooth functions defined on a compact domain $\mathcal{M}$, and therefore there always exists an optimal solution $M \in \mathcal{M}$ where the first order optimality condition is satisfied,

$$\mathrm{grad}\, f(M) = 0. \tag{4}$$

We study the stationary points of Problem 1 in detail, and we show how the other problem can be tackled.

Problem 1

We first analyze this optimality condition for Problem 1. For any $(W,Z) \in T_M\mathcal{E}$, we have

$$\begin{aligned}
&\mathrm{D}\hat{f}(U,V)\cdot(W,Z)\\
&\quad = 2\Re\,\mathrm{tr}\begin{pmatrix}(W^*AU + U^*AW - Z^*BV - V^*BZ)^*\\ (U^*AU - V^*BV)\end{pmatrix}\\
&\quad = \hat{g}_{(U,V)}\left(2\begin{pmatrix}AU\Delta_{AB}^* + A^*U\Delta_{AB}\\ BV\Delta_{BA}^* + B^*V\Delta_{BA}\end{pmatrix}, (W,Z)\right),
\end{aligned} \tag{5}$$

with $\Delta_{AB} := U^*AU - V^*BV =: -\Delta_{BA}$, and hence the gradient of $\hat{f}$ at a point $(U,V) \in \mathcal{E}$ is

$$\mathrm{grad}\,\hat{f}(U,V) = 2\begin{pmatrix}AU\Delta_{AB}^* + A^*U\Delta_{AB}\\ BV\Delta_{BA}^* + B^*V\Delta_{BA}\end{pmatrix}. \tag{6}$$

Since the normal space $T_M^\perp\mathcal{M}$ is the orthogonal complement of the tangent space $T_M\mathcal{M}$, one can, for any $M \in \mathcal{M}$, decompose any $E \in \mathcal{E}$ into its orthogonal projections on $T_M\mathcal{M}$ and $T_M^\perp\mathcal{M}$:

$$\mathrm{P}_M E := E - \mathrm{P}_M^\perp E \quad \text{and} \quad \mathrm{P}_M^\perp E := M \diamond \mathrm{Her}(M^* \diamond E). \tag{7}$$

For any $(W,Z) \in T_M\mathcal{M}$, (5) which yields

$$\mathrm{D}\hat{f}(U,V)\cdot(W,Z) = g_{(U,V)}\left(\mathrm{P}_M \mathrm{grad}\,\hat{f}(U,V),(W,Z)\right),$$

and the gradient of f at a point $(U,V) \in \mathcal{M}$ is

$$\mathrm{grad}\, f(U,V) = \mathrm{P}_M \mathrm{grad}\,\hat{f}\,(M). \tag{8}$$

For our problem, the first order optimality condition (4) yields, by means of (6), (7) and (8)

$$\begin{pmatrix} AU\Delta_{AB}^* + A^*U\Delta_{AB} \\ BV\Delta_{BA}^* + B^*V\Delta_{BA} \end{pmatrix} = \begin{pmatrix} U \\ V \end{pmatrix} \diamond \mathrm{Her}\begin{pmatrix} U^*AU\Delta_{AB}^* + U^*A^*U\Delta_{AB} \\ V^*BV\Delta_{BA}^* + V^*B^*V\Delta_{BA} \end{pmatrix}. \tag{9}$$

Observe that f is constant on the equivalence classes

$$[U,V] = \{(U,V)\diamond Q : Q \in \mathrm{U}(k)\},$$

and that any point of $[U,V]$ is a stationary point of f whenever (U,V) is.
We consider the special case where U^*AU and V^*BV are simultaneously diagonalizable by a unitary matrix at all stationary points (U,V) (it follows from (9) that this happens when A and B are both Hermitian), *i.e.* eigendecomposition of U^*AU and V^*BV are respectively WD_AW^* and WD_BW^*, with $W \in \mathrm{U}(k)$ and $D_A = \mathrm{diag}\left(\theta_1^A,\cdots,\theta_k^A\right)$, $D_B = \mathrm{diag}\left(\theta_1^B,\cdots,\theta_k^B\right)$.
The cost function at stationary points simply reduces to $\sum_{i=1}^{k}\left|\theta_i^A - \theta_i^B\right|^2$ and the minimization problem roughly consists in finding the isometric projections U^*AU, V^*BV such that their eigenvalues are as equal as possible.
More precisely, the first optimality condition becomes

$$\left[\begin{pmatrix} A \\ B \end{pmatrix} \diamond \begin{pmatrix} U \\ V \end{pmatrix} \diamond W - \begin{pmatrix} U \\ V \end{pmatrix} \diamond W \diamond \begin{pmatrix} D_A \\ D_B \end{pmatrix}\right] \diamond \begin{pmatrix} D_A - D_B \\ D_B - D_A \end{pmatrix} = 0, \tag{10}$$

that is,

$$\left[\begin{pmatrix} A \\ B \end{pmatrix} \diamond \begin{pmatrix} \bar{U}_i \\ \bar{V}_i \end{pmatrix} - \begin{pmatrix} \bar{U}_i \\ \bar{V}_i \end{pmatrix} \diamond \begin{pmatrix} \theta_i^A \\ \theta_i^B \end{pmatrix}\right] \diamond \begin{pmatrix} \theta_i^A - \theta_i^B \\ \theta_i^B - \theta_i^A \end{pmatrix} = 0, \quad i = 1,\ldots,k \tag{11}$$

where $\bar{U}_i$ and $\bar{V}_i$ denotes the i^{th} column of $\bar{U} = UW$ and $\bar{V} = VW$ respectively. This implies that for all $i = 1,\ldots,k$ either $\theta_i^A = \theta_i^B$ or $\left(\theta_i^A, \bar{U}_i\right)$ and $\left(\theta_i^B, \bar{V}_i\right)$ are eigenpairs of respectively A and B. If A and B are Hermitian matrices and $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_m$ and $\beta_1 \le \beta_2 \le \cdots \le \beta_m$ are their respective eigenvalues, the Cauchy interlacing theorem yields after reordering of the θ_i^A and θ_i^B in an increasing fashion

$$\theta_i^A \in [\alpha_i, \alpha_{i-k+m}] \quad \text{and} \quad \theta_i^B \in [\beta_i, \beta_{i-k+n}], \quad i = 1,\ldots,k.$$

Definition 2 *For S_1 and S_2, two non-empty subsets of a metric space, we define*

$$e_d\left(S_1, S_2\right) = \inf_{\substack{s_1 \in S_1 \\ s_2 \in S_2}} d(s_1, s_2)$$

When considering two non-empty subsets $[\alpha_1, \alpha_2]$ and $[\beta_1, \beta_2]$ of $\mathbb{R}$, one can easily see that

$$e_d\left([\alpha_1, \alpha_2], [\beta_1, \beta_2]\right) = \max\left(0, \alpha_1 - \beta_2, \beta_1 - \alpha_2\right).$$

Theorem 3.1 *Let $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ be the eigenvalues of Hermitian matrices respectively A and B. The solution of Problem 1 is bounded below by*

$$\sum_{i=1}^{k} \left(e_d\left([\alpha_i, \alpha_{i-k+m}], [\beta_i, \beta_{i-k+n}]\right)\right)^2$$

with d the Euclidean norm.

Proof Recall that when A and B are Hermitian matrices, the cost function at stationary points of Problem 1 reduces to

$$\sum_{i=1}^{k} \left|\theta_i^A - \theta_i^B\right|^2,$$

where $\theta_1^A, \ldots, \theta_k^A$ and $\theta_1^B, \ldots, \theta_k^B$ are the eigenvalues of U^*AU and V^*BV, respectively. It follows from the Cauchy interlacing theorem that the minimum of this function is bounded below by

$$\min_{\theta_i^A, \theta_i^B} \min_{\pi} \sum_{i=1}^{k} \left(\theta_{\pi(i)}^A - \theta_i^B\right)^2 \tag{12}$$

such that

$$\theta_1^A \leq \theta_2^A \leq \cdots \leq \theta_k^A, \quad \theta_1^B \leq \theta_2^B \leq \cdots \leq \theta_k^B, \tag{13}$$

$$\theta_i^A \in [\alpha_i, \alpha_{i-k+m}], \quad \theta_i^B \in [\beta_i, \beta_{i-k+n}], \tag{14}$$

and $\pi(\cdot)$ is a permutation of $1, \ldots, k$.

Let $\theta_1^A, \ldots, \theta_k^A$, and $\theta_1^B, \ldots, \theta_k^B$ satisfy (13). Then, the identity permutation $\pi(i) = i$ is optimal for problem (12). Indeed, if π is not the identity, then there exists i and j such that $i < j$ and $\pi(i) > \pi(j)$, and we have

$$\begin{aligned}
&\left(\theta_i^A - \theta_{\pi(i)}^B\right)^2 + \left(\theta_j^A - \theta_{\pi(j)}^B\right)^2 - \left[\left(\theta_j^A - \theta_{\pi(i)}^B\right)^2 + \left(\theta_i^A - \theta_{\pi(j)}^B\right)^2\right] \\
&\quad = 2\left(\theta_j^A - \theta_i^A\right)\left(\theta_{\pi(i)}^B - \theta_{\pi(j)}^B\right) \leq 0.
\end{aligned}$$

Since the identity permutation is optimal, our minimization problem simply reduces to

$$\sum_{i=1}^{k} \min_{(13)(14)} \left(\theta_i^A - \theta_i^B\right)^2.$$

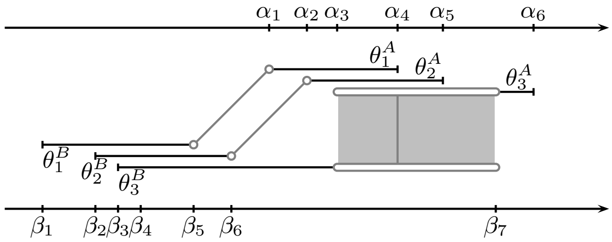


Figure 1: Let the α_i and β_i be the eigenvalues of the Hermitian matrices A and B, and $k = 3$. Problem 1 is then equivalent to $\sum_{i=1}^{3} \min_{\theta_i^A, \theta_i^B} \left(\theta_i^A - \theta_i^B\right)^2$ such that $\theta_i^A \in [\alpha_i, \alpha_{i+3}]$, $\theta_i^B \in [\beta_i, \beta_{i+4}]$. The two first terms of this sum have strictly positive contributions whereas the third one can be reduced to zero within a continuous set of values for θ_3^A and θ_3^B in $[\alpha_3, \beta_7]$.

We now show that (13) can be relaxed. Indeed, assume there is an optimal solution that does not satisfy the ordering condition, *i.e.* there exist i and j, $i < j$ such that $\theta_j^A \leq \theta_i^A$. One can see that the following inequalities hold

$$\alpha_i \leq \alpha_j \leq \theta_j^A \leq \theta_i^A \leq \alpha_{i-k+m} \leq \alpha_{j-k+m}.$$

Since θ_i^A belongs to $[\alpha_j, \alpha_{j-k+m}]$ and θ_j^A belongs to $[\alpha_i, \alpha_{i-k+m}]$, one can switch i and j and build an ordered solution that does not change the cost function and hence remains optimal.

It follows that $\sum_{i=1}^{k} \min_{(13)(14)} \left(\theta_i^A - \theta_i^B\right)^2$ is equal to $\sum_{i=1}^{k} \min_{(14)} \left(\theta_i^A - \theta_i^B\right)^2$. This result is precisely what we were looking for. $\square$

We conjecture that the Cauchy interlacing condition are "tight" in the following sense:

Conjecture 1 *Let $A \in \mathcal{S}_{Her}(m)$ with eigenvalues $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_m$. Let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$ satisfy the interlacing conditions*

$$\theta_i \in [\alpha_i, \alpha_{i-k+m}], \quad i = 1, \ldots, k.$$

Then there exists $U \in \mathrm{St}(k, m)$ such that $\{\theta_1, \theta_2, \ldots, \theta_k\}$ is the spectrum of $U^* A U$.

If this conjecture holds, then Theorem 3.1 holds without the words "bounded below by".

Figure 1 gives an example of optimal matching.

Problem 2

For all $Y \in T_X\mathbb{C}^{n\times m} \simeq \mathbb{C}^{n\times m}$, we have

$$\mathrm{D}\hat{f}(X) \cdot Y = 2\Re\,\mathrm{tr}(Y^* (XAA^* - B^*XA - BXA^* + B^*BX)), \tag{15}$$

and hence the gradient of $\hat{f}$ at a point $X \in \mathbb{C}^{n\times m}$ is

$$\operatorname{grad} \hat{f}(X) = 2(XAA^* - B^*XA - BXA^* + B^*BX). \tag{16}$$

Since the normal space $T_X^{\perp}\mathcal{M}$ is the orthogonal complement of the tangent space $T_X\mathcal{M}$, one can, for any $X = VU^* \in \mathcal{M}$, decompose any $E \in \mathbb{C}^{n\times m}$ into its orthogonal projections on $T_X\mathcal{M}$ and $T_X^{\perp}\mathcal{M}$:

$$\begin{aligned} \mathrm{P}_X E &= E - V\,\mathrm{Her}(V^*EU)\,U^* - (I_n - VV^*)\,E\,(I_m - UU^*), \text{ and} \\ \mathrm{P}_X^{\perp} E &= V\,\mathrm{Her}(V^*EU)\,U^* + (I_n - VV^*)\,E\,(I_m - UU^*). \end{aligned} \tag{17}$$

For any $Y \in T_X\mathcal{M}$, (15) hence yields

$$\mathrm{D}\hat{f}(X) \cdot Y = \mathrm{D}f(X) \cdot Y = g_X\left(\mathrm{P}_X \operatorname{grad} \hat{f}(X), Y\right),$$

and the gradient of f at a point $X = VU^* \in \mathcal{M}$ is

$$\operatorname{grad} f(X) = \mathrm{P}_X \operatorname{grad} \hat{f}\,(X). \tag{18}$$

Problem 3

This problem is very similar to Problem 2. We have

$$\mathrm{D}\hat{f}(X) \cdot Y = 2\Re\operatorname{tr}(Y^*\,(X - B^*XA - BXA^* + B^*BXA^*A)), \tag{19}$$

for all $Y \in T_X\mathbb{C}^{n\times m} \simeq \mathbb{C}^{n\times m}$, and hence the gradient of $\hat{f}$ at a point $X \in \mathbb{C}^{n\times m}$ is

$$\operatorname{grad} \hat{f}(X) = 2(X - B^*XA - BXA^* + B^*BXA^*A).$$

The feasible set is the same as in Problem 2. Hence the orthogonal decomposition (17) holds, and the gradient of f at a point $X = VU^* \in \mathcal{M}$ is

$$\operatorname{grad} f(X) = \mathrm{P}_X \operatorname{grad} \hat{f}\,(X).$$

Problem 4

For all $T \in T_S\mathbb{C}^{n\times m} \simeq \mathbb{C}^{n\times m}$, we have

$$\mathrm{D}\hat{f}(S) \cdot T = 2\Re\operatorname{tr}(T^*\,(SAA^* - B^*SA - BSA^* + B^*BS)), \tag{20}$$

and hence the gradient of $\hat{f}$ at a point $S \in \mathbb{C}^{n\times m}$ is

$$\operatorname{grad} \hat{f}(S) = 2(SAA^* - B^*SA - BSA^* + B^*BS). \tag{21}$$

Since the normal space, $T_S^{\perp}\mathcal{M}$, is the orthogonal complement of the tangent space, $T_S\mathcal{M}$, one can, for any $S \in \mathcal{M}$, decompose any $E \in \mathbb{C}^{n\times m}$ into its orthogonal projections on $T_S\mathcal{M}$ and $T_S^{\perp}\mathcal{M}$:

$$\mathrm{P}_S E = E - S\,\Re\operatorname{tr}(S^*E) \quad \text{and} \quad \mathrm{P}_S^{\perp} E = S\,\Re\operatorname{tr}(S^*E). \tag{22}$$

For any $T \in T_S\mathcal{M}$, (20) then yields

$$\mathrm{D}\hat{f}(S) \cdot T = \mathrm{D}f(S) \cdot T = g_S\left(\mathrm{P}_S \operatorname{grad} \hat{f}(S), T\right),$$

and the gradient of f at a point $S \in \mathcal{M}$ is $\operatorname{grad} f(S) = \mathrm{P}_S \operatorname{grad} \hat{f}(S)$.

For our problem, (4) yields, by means of (21) and (22)

$$\lambda S = (SA - BS)A^* - B^*(SA - BS)$$

where $\lambda = \operatorname{tr}((SA - BS)^*(SA - BS)) \equiv \hat{f}(S)$. Its equivalent vectorized form is

$$\lambda \operatorname{vec}(S) = \left(A^T \otimes I - I \otimes B\right)^* \left(A^T \otimes I - I \otimes B\right) \operatorname{vec}(S).$$

Hence, the stationary points of Problem 4 are given by the eigenvectors of $\left(A^T \otimes I - I \otimes B\right)^* \left(A^T \otimes I - I \otimes B\right)$. The cost function f simply reduces to the corresponding eigenvalue and the minimal cost is then the smallest eigenvalue.

Problem 5

This problem is very similar to Problem 4. A similar approach yields

$$\lambda S = (S - BSA^*) - B^*(S - BSA^*)A$$

where $\lambda = \operatorname{tr}((S - BSA^*)^*(S - BSA^*))$. Its equivalent vectorized form is

$$\lambda \operatorname{vec}(S) = \left(I \otimes I - \bar{A} \otimes B\right)^* \left(I \otimes I - \bar{A} \otimes B\right) \operatorname{vec}(S),$$

where $\bar{A}$ denotes the complex conjugate of A.

Hence, the stationary points of Problem 5 are given by the eigenvectors of $\left(I \otimes I - \bar{A} \otimes B\right)^* \left(I \otimes I - \bar{A} \otimes B\right)$, and the cost function f again simply reduces to the corresponding eigenvalue and the minimal cost is then the smallest eigenvalue.

4 Relation to the Crawford Number

The field of values of a square matrix A is defined as the set of complex numbers

$$\mathcal{F}(A) := \{x^*Ax : x^*x = 1\},$$

and is known to be a closed convex set [7]. The Crawford number is defined as the distance from that compact set to the origin

$$Cr(A) := \min \{|\lambda| : \lambda \in \mathcal{F}(A)\},$$

and can be computed *e.g.* with techniques described in [7]. One could define the *generalized Crawford number* of two matrices A and B as the distance between $\mathcal{F}(A)$ and $\mathcal{F}(B)$, *i.e.*

$$Cr(A, B) := \min \{|\lambda - \mu| : \lambda \in \mathcal{F}(A), \mu \in \mathcal{F}(B)\}.$$

Clearly, $Cr(A, 0) = Cr(A)$ which thus generalizes the concept. Moreover this is a special case of our problem since

$$Cr(A, B) = \min_{U^*U = V^*V = 1} \|U^*AU - V^*BV\|.$$

One can say that Problem 1 is a k-dimensional extension of this problem.

References

[1] C. Fraikin, Y. Nesterov, and P. Van Dooren. Optimizing the coupling between two isometric projections of matrices. *SIAM Journal on Matrix Analysis and Applications*, 2007.

[2] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren. A measure of similarity between graph vertices: applications to synonym extraction and Web searching. *SIAM Review*, 46(4):647–666, 2004.

[3] C. Fraikin and P. Van Dooren. Graph matching with type constraints on nodes and edges. In Andreas Frommer, Michael W. Mahoney, and Daniel B. Szyld, editors, *Web Information Retrieval and Linear Algebra Algorithms*, number 07071 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI).

[4] J. M. Lee. *Introduction to Smooth Manifolds.* Springer, New York, 2006.

[5] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, New Jersey, January 2008.

[6] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994. With a foreword by R. Brockett.

[7] R. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, New York, 1991.