Technical report: Exact Recovery with Symmetries for Procrustes Matching

Nadav Dym Yaron Lipman

In this technical report we prove the relaxation properties and exactness results mentioned in this paper. We also show equivalence to the problem of exact graph matching in the appendix of this manuscript. We do this using the terminology of [Lasserre, 2000], which gives a useful interpetation for SDP relaxations. We begin by presenting this terminology, and then proceed to the proofs.

Notation We denote the feasible set of PM by $G = \mathcal{O}(d) \times \Pi_n$.

We denote the j-th column of a matrix $X \in \mathbb{R}^{n\times n}$ by $X_j \in \mathbb{R}^{n\times 1}$, and the i-th row by $X_{i\star} \in \mathbb{R}^{1\times n}$.

Expressions such as X_j^T should be interpreted as $(X_j)^T$ (as opposed to $(X^T)_j$).

We denote by $\mathbf{1}$ the vector $\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^{n\times 1}$.

We denote polynomials of degree $\leq r$ in $x = (x_1, \dots, x_\ell)$ by $\mathcal{P}_r(x)$.

1 Moment interpetation

We recall that our relaxation is of the form

$$\min_{x,Y} \mathcal{L}\left[f_0\right](x, Y) \tag{1a}$$

$$\text{s.t.} \quad \mathcal{L}\left[f_s\right](x, Y) = 0, \quad \forall s = 1 \dots S \tag{1b}$$

$$\begin{bmatrix} 1 & \left(x^{(j)}\right)^T \\ x^{(j)} & Y^{(j)} \end{bmatrix} \succeq 0 \tag{1c}$$

where f_k are the polynomials participating in the relaxation, and $x^{(j)} = (R_{11}, \dots, R_{dd}, X_{1j}, \dots, X_{nj})$.

We present an alternative way to write (1) in the spirit of [Lasserre, 2000]. A linear functional μ on $\mathcal{P}_2(x)$ is uniquely defined by the values it assigns to members of the standard basis of monomials with degree ≤ 2. We assume $\mu(1) = 1$ and denote

$$\mu(x_i) = \chi_i \quad ; \quad \mu(x_i x_j) = \mathcal{Y}_{ij}$$

Let μ_j denote the restriction of μ to $\mathcal{P}_2(x^{(j)})$. We say that $\mu_j \succeq 0$ if the subvectors $\chi^{(j)}$ of χ and $\mathcal{Y}^j$ of $\mathcal{Y}$ satisfy (1c). Note that applying μ to a polynomial f is the same as evaluating $\mathcal{L}[f]$ at $(\chi, \mathcal{Y})$, *i.e.*,

$$\mu(f) = \mathcal{L}[f](\chi, \mathcal{Y})$$

and therefore an equivalent formulation of (1) is given by

$$\min_{\mu} \mu(f_0) \tag{2a}$$

$$\text{s.t.} \quad \mu(f_s) = 0, \quad \forall s = 1 \dots S \tag{2b}$$

$$\mu_j \succeq 0 \tag{2c}$$

If $F : \mathbb{R}^N \to \mathbb{R}^{a\times b}$ is a function with $F_{ij} \in \mathcal{P}_2(x), \forall i, j$, then we define $\mu(F) \in \mathbb{R}^{a\times b}$ by $\mu(F)_{ij} = \mu\left(F_{ij}\right)$. PM-SDP in this notation takes the form:

$$\min_{\mu} \quad \mu\left(\|RP - QX\|_F^2\right) \tag{3a}$$

$$\text{s.t.} \quad \mu\left(X\mathbf{1}\right) = \mathbf{1} \quad , \quad \mu\left(\mathbf{1}^T X\right) = \mathbf{1}^T \tag{3b}$$

$$\mu\left(X_j X_j^T\right) = \mu\left(\text{diagX}_\text{j}\right) \quad j = 1 \dots n \tag{3c}$$

$$\mu\left(RR^T\right) = I_d, \quad \mu\left(R^T R\right) = I_d \tag{3d}$$

$$\mu_j \succeq 0 \tag{3e}$$

We denote the set of feasible μ by $\mathcal{M}_F$.

1.1 Properties of moment relaxation

The purpose of this subsection is to explain the implication of the constraint $\mu_j \succeq 0$ on the functional μ_j and state some additional results we will need. These are summarised in the following proposition.

Proposition 1. *Let μ be a linear functional on $\mathcal{P}_2(x)$ and let $h, g \in \mathcal{P}_1(x_j)$, then*

1. *$\mu_j \succeq 0$ if and only if all $p \in \mathcal{P}_1(x^{(j)})$ satisfy $\mu_j(p^2) \geq 0$.*
2. *If $\mu_j \succeq 0$ and $\mu_j(g^2) = 0$ then $\mu_j(hg) = 0$.*
3. *If $\mu_j \succeq 0$ then $g(\chi^{(j)})^2 \leq \mu(g^2)$.*

Proof. We recall that $\mu_j \succeq 0$ means that

$$M_j \equiv \begin{bmatrix} 1 & \left(\chi^{(j)}\right)^T \\ \chi^{(j)} & \mathcal{Y}^{(j)} \end{bmatrix} \succeq 0$$

where as before $\chi^{(j)}, \mathcal{Y}^{(j)}$ describe the moments of μ_j.

The linearity of μ implies that for matrices A, B,

$$\mu(AFB) = A\mu(F)B \tag{4}$$

Define $v_j(x) = (1, x_j)^T \in \mathbb{R}^{(|I_j|+1)\times 1}$ and note that $\mu\left(v_j v_j^T\right) = M_j$. polynomials $h, g \in \mathcal{P}_1\left(x^{(j)}\right)$ can be written as

$$h(x) = \mathbf{h}^T v_j(x), \quad g(x) = \mathbf{g}^T v_j(x)$$

where $\mathbf{h}, \mathbf{g}$ are the coeffecients of h, g. It follows that

$$\mu(hg) = \mu\left(\mathbf{h}^T v_j v_j^T \mathbf{g}\right) = \mathbf{h}^T \mu\left(v_j v_j^T\right)\mathbf{g} = \mathbf{h}^T M_j \mathbf{g} \tag{5}$$

The first part of the theorem follows immediately by choosing $g = h$.

If $0 = \mu(g^2) = \mathbf{g}^T M_j \mathbf{g}$ then since M_j is positive semi-definite $M_j\mathbf{g} = 0$ and therefore

$$\mu(hg) = \mathbf{h}^T M_j \mathbf{g} = 0$$

which proves the second claim.

To prove the third claim, note that $g(\chi)^2 = \delta_\chi(g^2)$. It is sufficient to prove that the matrix $\tilde{M}_j$ generated by the moments of the functional $\mu - \delta_\chi$ is positive semi-definite. This matrix is given by

$$\tilde{M}_j = \begin{bmatrix} 0 & 0 \\ 0 & \mathcal{Y}^{(j)} - \chi^{(j)} \left(\chi^{(j)}\right)^T \end{bmatrix}$$

$\tilde{M}_j \succeq 0$ iff $\mathcal{Y}^{(j)} \succeq \chi^{(j)} \left(\chi^{(j)}\right)^T$, and this follows from applying Schur's complement to $M_j \succeq 0$. □

1.2 Properties of relaxation.

We begin with presenting two consequences of our discussion in the previous section.

Proposition 2. *The objective of* (3) *is bounded from below by zero.*

Proof of proposition 2 . We note that $\|RP - QX\|_F^2$ can be rewritten as $\sum_{i,j} f_{i,j}^2$ for $f_{i,j} = R_{i\star}P_j - Q_{i\star}X_j$, and each $f_{i,j}$ is a linear polynomials in the variables $x^{(j)}$. Thus since $\mu_j \succeq 0$,

$$\sum_{i,j} \mu\left(f_{i,j}^2\right) = \sum_{i,j} \mu_j\left(f_{i,j}^2\right) \geq 0$$

□

For a solution μ of (3) we denote $\mu(X) = \mathcal{X}$ and $\mu(R) = \mathcal{R}$.

Proposition 3. *$\mathcal{R}, \mathcal{X}$ are in the convex hull of the orthogonal transformations and permutations, respectively. Moreover*

$$\|\mathcal{R}P - Q\mathcal{X}\|_F^2 \leq \mu\left(\|RP - QX\|_F^2\right) \tag{6}$$

In particular, if $\mu\left(\|RP - QX\|_F^2\right) = 0$ then $\mathcal{R}P = Q\mathcal{X}$.

Proof of proposition 3 . We begins by showing that $\mathcal{X}$ is in the convex hull of Π_n, *i.e.*, that $\mathcal{X}$ is doubly stochastic. The rows and columns of $\mathcal{X} = \mu(X)$ are constrained to sum to one, and each coordinate of $\mathcal{X}$ must be non-negative since

$$\mathcal{X}_{i,j} = \mu\left(X_{i,j}\right) = \mu\left(X_{i,j}\right) = \mu\left(X_{i,j}^2\right) \geq 0$$

where the inequality on the r.h.s. follows from $\mu_j \succeq 0$. The convex hull of orthogonal matrices are matrices whose 2-norm is not larger than one. For and $v \in \mathbb{R}^d$ define $g_v(R) = \|Rv\|_2^2$. Then $\|\mathcal{R}\|_2 \leq 1$ since

$$g_v(\mathcal{R}) \leq^{\text{prop. 1}} \mu(g_v) = \mu\left(\left\langle v, R^T R v\right\rangle\right) = \mu\left(\langle v, v\rangle\right) = \|v\|_2^2$$

(6) is an immediate consequence of proposition 1. □

Invariance to choice of representative. The procrustres distance $d(P,Q)$ is defined on equivalence classes: We say that $P \sim \hat{P}$ if $P = R_0\hat{P}X_0^T$, $(R_0, X_0) \in G$. If $P \sim \hat{P}$ and $Q \sim \hat{Q}$ then $d(\hat{P},\hat{Q}) = d(P,Q)$. We now show that our convex approximation $\underline{d}$ of d also satisfies $\underline{d}(\hat{P},\hat{Q}) = \underline{d}(P,Q)$. Moreover, there is a natural one-to-one correspondence between the optimal set of $\mu \in \mathcal{M}_F$ which attain the minimal value $\underline{d}(P,Q)$ and the optimal set of $\mu \in \mathcal{M}_F$ which attain the minimal value $\underline{d}(\hat{P},\hat{Q})$. This property will allow us to rotate and relabel P and Q without losing generality, which will simplify notation later on.

To show this, we define for fixed $(g_0, g_1) \in G \times G$ (where $g_i = (R_i, X_i)$),

$$(g_0, g_1)_\star \mu\left(p(R,X)\right) = \mu\left(p(R_0^T R R_1, X_0^T X X_1)\right)$$

This defines an action of $G \times G$ on the space $\mathcal{M}$ of linear functionals on $\mathcal{P}_2$. We claim

Lemma 1 (proof in appendix)**.** $(G \times G)\mathcal{M}_F = \mathcal{M}_F$

Accordingly, for $P = R_1\hat{P}X_1^T$, $Q = R_0\hat{Q}X_0^T$ and $\mu \in \mathcal{M}_F$,

$$\mu\left(\|RP - QX\|_F^2\right) = \mu\left(\left\|RR_1\hat{P}X_1^T - R_0\hat{Q}X_0^T X\right\|_F^2\right)$$
$$= \mu\left(R_0^T R R_1 \hat{P} - \hat{Q}X_0^T X X_1\right) = (g_0, g_1)_\star \mu\left(\left\|R\hat{P} - \hat{Q}X\right\|_F^2\right)$$

which easily implies that $\underline{d}(P,Q) = \underline{d}(\hat{P},\hat{Q})$, with the one-to-one correspondence between optimal sets given by $(g_0, g_1)_\star$.

2 Simple spectrum

In this section we prove theorem 1. We begin by introducing some more notation. We define

$$\text{ISO}(P,Q) = \{(R,X) \in G |\ RP = QX\}$$
$$\text{ISO}_R(P,Q) = \{R |\ (R,X) \in \text{ISO}(P,Q) \text{ for some } X \in \Pi_n\}$$

We also define

$$\mathcal{N}_F(P,Q) = \{\mu \in \mathcal{M}_F |\ \mu\left(\|RP - QX\|_F^2\right) = 0\}$$
$$\mathcal{N}_F^R(P,Q) = \{\mathcal{R} \in \text{conv}\,\mathcal{O}(d) |\ \mathcal{R} = \mu(R) \text{ for some } \mu \in \mathcal{M}_F\}$$

We note that ISO_R is a subset of $\mathcal{N}_F^R$, and the map $g \mapsto \delta_g$ is a one-to-one mapping of $\text{ISO}(P,Q)$ into $\mathcal{N}_F(P,Q)$.

It is known that when PP^T has simple spectrum, matrices $R \in \text{ISO}_R(P,P)$ represented according to a basis of eigenvectors of PP^T are diagonal with diagonal entries in $\{-1,1\}$. We denote the group of all such diagonal matrices by $\{-1,1\}^d$.

We denote

$$\text{Orb}_R(P,j) = \{R \in \mathcal{O}(d) |\ RP_j \text{ is a column of } P\}$$

and note that for any j, $\text{ISO}_R(P,P) \subseteq \text{Orb}_R(P,j)$. In general this inclusion can be strict. If j is an index for which the inclusion is in fact an equality, we say that P_j is *faithful*. The weak assumption mentioned in the article is that P has at least one column which is faithful.

In addition to assuming simple spectrum, we also assume that P has at least one column that is faithful. It seems that this assumption is fulfilled in many problems of interests, such as point cloud created by sampling living creatures. We now prove the theorem.

Proof. Our goal is to show that $\text{ISO}_R(P,Q)$ are the extreme points of $\mathcal{N}_F^R(P,Q)$.

The general strategy is as follows: We use our knowledge of the behavior of exact solutions $g = (R,X) \in \text{ISO}(P,Q)$ in the simple spectrum case (equivalently, exact solutions

$\delta_g \in \mathcal{N}_F$), to derive generalizations which hold for all $\mu \in \mathcal{N}_F$. These generalizations enable the proof of the theorem.

We divide the proof into two parts. In the first part we present an overview of the proof highlighting this strategy, and in the second part we prove the lemmas stated without proof in the first part.

Overview. Due to lemma (3) we can w.l.o.g. rotate P so that its principle axes are the standard basis of $\mathbb{R}^d$, *i.e.*, so that PP^T is diagonal. Since $d(P,Q) = 0$ we can also rotate and relabel Q so that $P = Q$ (nonetheless, we only replace Q with P in the proof when necessary, so as to make the notation easier to follow). By assumption the eigenvalues of PP^T, which are also its diagonal entries, are of the form $\lambda_1 > \lambda_2 > \ldots > \lambda_d$.

For the unrelaxed problem it can be easily shown that for all exact solutions $(R, X) \in G$, we have $RPP^T = QQ^TR$. This implies that the eigenspaces of $PP^T = QQ^T$ are invariant under R, which in the simple spectrum case implies that R is diagonal and $R_{jj} \in \{-1, 1\}$. This can be generalized to the relaxed problem as follows:

Lemma 2. *Assume $\mu \in \mathcal{N}_F$. Then*

1. $\mu\left(\left\|RPP^T - QQ^TR\right\|_F^2\right) = 0.$
2. $\mu\left(R_{ij}^2\right) = \delta_{ij}.$
3. *$\mathcal{R}$ is diagonal, and $\mathcal{R}_{jj} \in [-1, 1]$.*

For exact solutions of the unrelaxed problem $|(RP)_{ij}| = |P_{ij}|$. Therefore if the permutation component of the solution takes the column P_j to the column Q_ℓ, necessarily the columns are identical up to sign changes. A similar statement can be made for the relaxed problem as well:

Lemma 3. *Assume $\mu \in \mathcal{N}_F$. Then $\mathcal{X}_{\ell,j} > 0$ implies that for all i, $P_{i,j}^2 = Q_{i,\ell}^2$.*

Equivalently, the lemma says that $\mathcal{X}_{\ell,j} = 0$ unless Q_ℓ is in the orbit of P_j under $\{-1, 1\}^d$.

For ℓ, i, j satisfying $|Q_{\ell i}| = |P_{\ell j}| > 0$ define

$$r_{\ell\ell}(ij) = \frac{Q_{\ell i}}{P_{\ell j}} \tag{7}$$

Then the previous lemma implies that

Lemma 4. *If $P_{\ell j} \neq 0$ and $\mu \in \mathcal{N}_F$, then*

$$\mathcal{R}_{\ell\ell} = \sum_{i:|Q_{\ell i}|=|P_{\ell j}|} \mathcal{X}_{ij} r_{\ell\ell}(i,j)$$

Proof of lemma 4. We consider the (ℓ, j) coordinate of the equation $\mathcal{R}P - Q\mathcal{X} = 0$ to obtain

$$0 = (\mathcal{R}P - Q\mathcal{X})_{\ell,j} = \mathcal{R}_{\ell\ell}P_{\ell j} - \sum_{i:|Q_{\ell i}|=|P_{\ell j}|} Q_{\ell i}\mathcal{X}_{i1}$$

$$= P_{\ell j}\left(\mathcal{R}_{\ell\ell} - \sum_{i:|Q_{\ell i}|=|P_{\ell j}|} r_{\ell\ell}(i,j)\mathcal{X}_{ij}\right)$$

□

These three lemmas can now be combined to prove the theorem. According to our assumption P has a faithful column. w.l.o.g. we take that column to be P_1, and assume that the first ρ columns of P are the columns in P obtained by applying $\mathrm{ISO}_R(P,Q)$ to P_1.

Note that either the ℓ-th row of the first ρ columns of $P = Q$ are identically zero, or all elements in this row are non-zero. Without loss of generality assume that the rows with zero are the last s rows. This implies that $\mathrm{ISO}_R(P,Q)$ is of the form $H_{d-s} \times \{-1,1\}^s$ for some subgroup $H_{d-s} \subseteq \{-1,1\}^{d-s}$. Therefore, $\mathrm{conv}(\mathrm{ISO}_R(P,Q)) = \mathrm{conv}\,(H_{d-s}) \times [-1,1]^s$. Our goal is to prove that $\mathcal{R} \in \mathrm{conv}(\mathrm{ISO}_R(P,Q))$. By lemma 2 we know that $\mathcal{R}$ is diagonal and that for $j > d-s$, we have $\mathcal{R}_{jj} \in [-1,1]$. Thus, if we denote by $\bar{\mathcal{R}}$ the $(d-s)\times(d-s)$ matrix obtained by restricting $\mathcal{R}$ to the first $d-s$ coordinated, it remains to show that $\bar{\mathcal{R}} \in \mathrm{conv}\,(H_{d-s})$.

For every $1 \leq i \leq \rho$ define $r(i,1)$ to be the diagonal matrix in $\mathbb{R}^{(d-s)\times(d-s)}$ whose diagonal elements $r_{\ell\ell}(i,1)$ are defined as in (7). By assumption, a matrix $R \in \mathcal{O}(d)$ taking P_1 to Q_i is in $\mathrm{ISO}_R(P)$, and its restriction to the first $d-s$ coordinates can be seen to be $r(i,1)$. Thus, $r(i,1) \in H_{d-s}$. According to lemma 4

$$\bar{\mathcal{R}} = \sum_{i=1}^{\rho} \mathcal{X}_{i1} r(i,1)$$

Since $\mathcal{X}$ is doubly stochastic, and $\mathcal{X}_{i1} = 0$ for all $i > \rho$ according to lemma 3, we showed that $\bar{\mathcal{R}} \in \mathrm{conv}\,(H_{d-s})$ as required.

Proof of lemmas.

proof of lemma 2.

1. We leave the proof to the appendix.
2. According to the first part, and since $P = Q$ so that in particular P, Q have the same eigenvalues,

$$0 = \mu\left(\left\|RPP^T - QQ^TR\right\|_F^2\right) = \sum_{i,j} \mu\left(R_{ij}^2(\lambda_j - \lambda_i)^2\right) \tag{8}$$

Since $\mu\left(R_{ij}^2\right) \geq 0$, necessarily $\mu(R_{ij}^2) = 0$ unless $i = j$, in which case

$$\mu\left(R_{jj}^2\right) = \sum_i \mu(R_{i,j}^2) = \mu\left((R^TR)_{jj}\right) = (I_d)_{jj} = 1 \tag{9}$$

3. We know that for off diagonal elements $\mu(R_{ij}^2) = 0$ which implies that $\mathcal{R}_{ij} = 0$. Proposition 3 says that $\|\mathcal{R}\|_2 \leq 1$ and so all diagonal elements of $\mathcal{R}$ satisfy $|\mathcal{R}_{jj}| \leq 1$.

□

For the proof of lemma 3 we will need the following simple lemma:

Lemma 5. *If $Xv = w$, where X is doubly stochastic and $\|v\|_2 = \|w\|_2$ then there is a permutation taking v to w. Moreover, if $X_{ij} > 0$ then $w_i = v_j$.*

Proof. Write X as a convex combination of permutations X_k, $X = \sum \theta_k X_k$. Then

$$\|w\|_2 = \|Xv\|_2 \leq^{(*)} \sum_k \theta_k \|X_k v\|_2 \leq \|v\|_2 = \|w\|_2$$

Thus inequality $(*)$ is an equality, which is possible only if $X_1 v = X_2 v = \ldots$, which in turn implies that $X_k v = Xv = w$. Since X_k are permutations, if $w_i \neq v_j$ then the (i,j) coordinate of X_k must be zero, and so the same holds for $X = \sum \theta_k X_k$. □

proof of lemma 3. For a row vector $v \in \mathbb{R}^{1\times n}$ we define by $v^{(2)}$ the row vector $(v_1^2, \ldots, v_n^2)$. We will show that

$$P^{(2)}_{i\star} = Q^{(2)}_{i\star}\mathcal{X} \tag{10}$$

The proof is then completed by applying lemma 5 to the transpose of this equation, since by assumption $P_{i\star} = Q_{i\star}$ and in particular $P^{(2)}_{i\star}$ and $Q^{(2)}_{i\star}$ have the same norm.

Note that

$$\mu\left((R_{i\star}P_j)^2\right) = \mu\left(\left(\sum_\ell R_{i\ell}P_{\ell j}\right)^2\right) =^{(*)} P^2_{ij} \tag{11}$$

Where $(*)$ follows from the fact that $\mu(R^2_{ij}) = \delta_{ij}$ and therefore all expressions which include an off-diagonal element of R cancel out. On the other hand, since $\mu \in \mathcal{N}_F$, we can use the decomposition of the objective from proposition 2 to show that $\mu\left((R_{i\star}P_j - Q_{i\star}X_j)^2\right) = 0$ which implies

$$\mu\left((R_{i\star}P_j)^2\right) = \mu\left(R_{i\star}P_jQ_{i\star}X_j\right) = \mu\left((Q_{i\star}X_j)^2\right)$$

Using this we obtain

$$\mu\left((R_{i\star}P_j)^2\right) = \mu\left(\left(\sum_\ell Q_{i\ell}X_{\ell,j}\right)^2\right) =^{(3c)} \sum_\ell Q^2_{i\ell}\mathcal{X}_{\ell j}$$

The last equation and (11) hold for all j, which proves the correctness of (10).

□

3 Projection

To obtain the extreme points of $\mathcal{N}_F^R$ we randomly choose a matrix $W \in \mathbb{R}^{d\times d}$ with normal distribution, and then solve the problem of maximizing $\mathrm{tr}(W^T R)$ subject to the constraint $\mu \in \mathcal{N}_F$. We prove this algorithm satisfies the properties stated in theorem 2.

Let us first consider the case $P = Q$. In this case $\mathrm{ISO}_R(P,P)$ is a group and we index the group so that $R_0 = I_d$.

R_i is the unique maximizer of $\mathrm{tr}(W^T R)$ iff W is a member of the set

$$\mathcal{A}_i = \{W| \quad \mathrm{tr}(W^T R_i) > \mathrm{tr}(W^T R_j), \forall j \neq i, 0 \leq j \leq L\}$$

clearly the union of $\mathcal{A}_i$ is a disjoint union, and has probability one. We note that $R_\ell \mathcal{A}_0 = \mathcal{A}_\ell$. Additionally, the probability of the sets $\mathcal{A}_\ell$ is preserved under linear isometries of the space $\mathbb{R}^{d\times d}$ endowed with the frobenious inner product, and the map $W \mapsto R_\ell W$ is such an isometry. Therefore the sets $\mathcal{A}_\ell = R_\ell \mathcal{A}_0$ all have the same probability.

In the general case where $P \neq Q$, we note that $R_0^T \mathrm{ISO}_R(P,Q) = \mathrm{ISO}_R(P,P)$. Denote $\mathcal{A}_i$ as before and $\tilde{\mathcal{A}}_i$ to be the sets defined as above but replacing R_i with $R_0^T R_i$. Then we know that $\tilde{\mathcal{A}}_i$ all have the same probability, and since $\mathcal{A}_i = R_0\tilde{\mathcal{A}}_\rangle$ we see that $\mathcal{A}_i$ have the same probability as well.

We note that once $R \in \mathrm{ISO}_R(P,Q)$ is found, there is a unique $X \in \Pi_n$ satisfying $RP = QX$ (assuming all points of P are distinct), and X can be found by solving a linear program over the space of doubly stochastic matrices.

4 Correctness in the almost-exact case

We prove corollary 1. Assume $P_n \to P_0, Q_n \to Q_0$ and P_0, Q_0 are asymmetric fulfilling the conditions of Theorem 1. Let μ_n, μ_0 be the solutions of $PM - SDP(P_n, Q_n)$ and $PM - SDP(P_0, Q_0)$ whose R, X coordinates we denote by $\mathcal{R}_n, \mathcal{X}_n$ and $\mathcal{R}_0, \mathcal{X}_0$. Let R_n, X_n and $R_0, X_0 (= \mathcal{R}_0, \mathcal{X}_0)$ be the elements of G obtained by μ_n and μ_0 by our projection process. Our aim is to show that for large enough n, R_n, X_n are the minimizers of $PM(P_n, Q_n)$.

A standard argument shows that μ_n converge to μ_0. In particular $\mathrm{tr}\mathcal{R}_n P_n \mathcal{X}_n^T Q_n^T \to \mathrm{tr}R_0 P_0 X_0^T Q_0^T$. We note that from the construction of our projection, $\mathrm{tr}\mathcal{R}_n P_n \mathcal{X}_n^T Q_n^T \leq \mathrm{tr}R_n P_n X_n^T Q_n^T$. Using compactness, this can be shown to imply that $\mathrm{tr}R_n P_n X_n^T Q_n^T \to \mathrm{tr}R_0 P_0 X_0^T Q_0^T$ as well.

We define the continuous function

$$F(X,P,Q) = \max_{R\in\mathcal{O}(d)} \mathrm{tr}RPX^TQ^T = \max_{R\in \mathrm{conv}\,\mathcal{O}(d)} \mathrm{tr}RPX^TQ^T$$

as X_0 is a unique minimizer of PM it follows that $F(X_0, P_0, Q_0) > F(X, P_0, Q_0)$ for all permutations X. Thus for large enough n we have

$$\mathrm{tr}R_n P_n X_n^T Q_n^T > F(X, P_n, Q_n), \quad \forall X_0 \neq X \in \Pi_n$$
$$F(X_0, P_n, Q_n) > F(X, P_n, Q_n), \quad \forall X_0 \neq X \in \Pi_n$$

and so $X_n = X_0$ which is the X coordinate of the optimal solution of $PM(P_n, Q_n)$. It follows that R_n is a maximizer of $\mathrm{tr}RP_n X_0^T Q_n^T$ and hence R_n, X_n are the optimal solution of $PM(P_n, Q_n)$.

References

[Lasserre, 2000] Lasserre, J. B. (2000). Global optimization with polynomials and the problem of moments. *SIAM J. on Optimization*, 11(3):796–817.

Appendix A Equivalence with exact graph matching

The exact graph matching problem is the problem of checking whether an isomorphism exists between two graphs, and if so, computing it. Mathematically, the graphs can be represented by symmetric matrices, called their adjacency matrices $G, H \in \mathbb{R}^{n\times n}$, and a graph isomorphism exists if

$$\min_{X\in\Pi_n} \|XG - HX\|_F^2 = 0$$

so that the objective of exact graph matching is to check whether the minimum of the optimization problem above is zero, and if so return a minimizer. The exact graph matching problem is a well known problem in the theory of computer science. It is know to be NP, but it is not known to be either polynomial or NP-hard, and is therefore considered a class of its own right.

We show the equivalence of this problem to our problem in the exact case, when we no longer assume d to be a fixed small number, but allow any $d \leq n$: Let us define as the *exact shape matching* problem, the problem of checking whether

$$\min_{X,R\in\Pi_n\times\mathcal{O}(d)} \|RP - QX\|_F^2 = 0$$

and if so returning a minimizer.

Theorem. *There is a polynomial time reduction between the exact shape matching problem, and the exact graph matching problem, and vice versa.*

Proof. We begin by showing a reduction from the graph matching problem to the shape matching problem.

Let $P, Q \in \mathbb{R}^{d\times n}$ be the input point clouds. Consider the Gram matrices of P and Q,

$$G = P^TP\ ;\ H = Q^TQ$$

and assume we are given $X \in \Pi_n$ satisfying

$$XG = HX$$

Denote the column space of P by V and the column space of Q by W. Now choose $\dim V = \ell$ columns of P which form a basis for V. For simplicity of notation we assume the first ℓ columns $P_1, \ldots, P_\ell$ are this basis, and that $X = I$, so that $P^TP = Q^TQ$.

We define a linear map $R : V \to W$ by specifying its value on basis elements:

$$RP_i = Q_i,\ \forall 1 \leq i \leq \ell$$

Note that

$$\left(P^TP\right)_{ij} = \langle RP_i, RP_j\rangle = \langle Q_i, Q_j\rangle = \left(Q^TQ\right)_{ij} = \left(P^TP\right)_{ij},\ \forall 1 \leq i, j \leq \ell$$

so that $R : V \to R(V)$ is an isometry. It remains to show that $Rv_m = u_m$ also for $m > \ell$.

Denote

$$\bar{P} = (P_1, \ldots, P_\ell) \quad ; \quad \bar{Q} = (Q_1, \ldots, Q_\ell)$$
$$\bar{G} = (G_{ij})_{1\leq i,j\leq\ell} = \bar{P}^T\bar{P} \quad ; \quad \bar{H} = (H_{ij})_{1\leq i,j\leq\ell} = \bar{Q}^T\bar{Q}$$

Since $P_1, \ldots, P_\ell$ is a basis, $\bar{G} = \bar{H}$ is invertible, and thus $u_1, \ldots, u_\ell$ is a basis as well. Therefore, for any $m > \ell$ we have

$$v_m = \sum_{i=1}^{\ell} a_i v_i\ ;\ u_m = \sum_{i=1}^{\ell} b_i u_i$$

and our goal is to show that $Rv_m = u_m$, or equivalently

$$\mathbf{a} = (a_1, \ldots, a_\ell)^T = (b_1, \ldots, b_\ell)^T = \mathbf{b}$$

Since

$$\bar{G}\mathbf{a} = \bar{P}^Tv_m = G_m \quad ; \quad \bar{H}\mathbf{b} = \bar{Q}^Tu_m = H_m$$

we obtain

$$\mathbf{a} = \bar{G}^{-1}G_m = \bar{H}^{-1}H_m = \mathbf{b}$$

Thus, we have shown that $RP = QX$ as required.

We now show a reduction from the shape matching problem to the graph matching problem.

Let G, H be the input adjacency matrices. We note that for all $\lambda \in \mathbb{R}$ and all $X \in \Pi_n$,

$$XG - HX = X(G + \lambda I) - (H + \lambda I)X$$

Thus it is sufficient to solve the exact graph matching problem for $G + \lambda I, H + \lambda I$, where we choose $\lambda = \max\{\|G\|_2, \|H\|_2\}$ so that

$$G + \lambda I, H + \lambda I \succeq 0$$

Choose some factorization

$$P^TP = G + \lambda I \; ; \; Q^TQ = H + \lambda I$$

and assume we are given $X, R \in \Pi_n \times \mathcal{O}(d)$ with

$$RP = QX$$

multiplying this equation with its transpose $P^TR^T = X^TQ^T$ we obtain

$$P^TP = X^TQ^TQX$$

and so multiplying by X from the left we obtain

$$X(G + \lambda I) = XP^TP = Q^TQX = (H + \lambda I)X$$

□

Appendix B Proof of lemma 3

Proof. Since $G \times G$ acts on $\mathcal{M}$, it is sufficient to show that for $\mu \in \mathcal{M}_F$ and $(g_0, g_1) \in G \times G$, the induced linear functional $(g_0, g_1)_*\mu = \nu$ is in $\mathcal{M}_F$. The first step is to verify that each of the equality constraints $\nu(f_i) = 0$ are satisfied. We show two examples:

1. $\nu(R^TR) = I_d$ since
$$\nu(R^TR) = \mu\left(\left(R_0^TRR_1\right)^T R_0^TRR_1\right) = \mu\left(R_1^TR^TRR_1\right) = I_d$$
2. $\nu(X_{ij}X_{\ell j}) = \delta_{i\ell}\nu(X_{ij})$. Let τ_0, τ_1 be the permutations corresponding to X_0, X_1. Then for all $X \in \mathbb{R}^{n\times n}$ and $1 \le i, j \le n$,
$$\left(X_0^TXX_1\right)_{ij} = X_{\tau_0(i)\tau_1(j)}$$
Using this, we have
$$\nu(X_{ij}X_{\ell j}) = \mu\left(\left(X_0^TXX_1\right)_{ij}\left(X_0^TXX_1\right)_{\ell j}\right) = \mu\left(X_{\tau_0(i)\tau_1(j)}X_{\tau_0(\ell)\tau_1(j)}\right) = \delta_{\tau_0(i)\tau_0(\ell)}\mu\left(X_{\tau_0(i)\tau_1(j)}\right) = \delta_{i\ell}\nu(X_{ij})$$

The other equality constraints can be shown using similar arguments. The second step of the proof is verifying that the ν fulfills the semi-definite constraints. We note that if $p(R, X_j)$ is a linear polynomial in $x^{(j)} = (R, X_j)$, then $\tilde{p} = p(R_0^TRR_1, X_0^TX_jX_1)$ is a linear polynomial in $(R, X_{\tau_1(j)})$. Since $\mu_{\tau_1(j)} \succeq 0$, it follows that $\nu(p^2) = \mu(\tilde{p}^2) \ge 0$ and thus $\nu_j \succeq 0$ as required. □

Appendix C Proof of first part of lemma 2

We need to show that for $\mu \in \mathcal{M}_F$ with $\mu\left(\|RP - QX\|_F^2\right) = 0$, also $\mu\left(\left\|RPP^T - QQ^TR\right\|_F^2\right) = 0$. In our case it will suffice to prove for $P = Q$ which will be helpful at the end of this calculation, although this assumption is not necessary.

$$\begin{aligned}\mu\left(\left\|RPP^T - QQ^TR\right\|_F^2\right) &= \mu\left(\mathrm{tr}\left(PP^TR^TRPP^T\right)\right) + \mu\left(\mathrm{tr}\left(QQ^TRR^TQQ^T\right)\right) - 2\mu\left(\mathrm{tr}\left(RPP^TR^TQQ^T\right)\right) \\ &= \left\|PP^T\right\|_F^2 + \left\|QQ^T\right\|_F^2 - 2\mu\left(\mathrm{tr}\left(RPP^TR^TQQ^T\right)\right)\end{aligned} \tag{12}$$

We now deal only with the non-constant term. Note that

$$\mathrm{tr}\left(RPP^TR^TQQ^T\right) = \mathrm{tr}\left(P^TR^TQQ^TRP\right) = \sum_j e_j^TP^TR^TQQ^TRPe_j = \sum_j \left(RP_j\right)^TQQ^T(RP_j)$$

Using the fact that $\mu\|RP_j - QX_j\|_2^2 = 0$ we can exchange RP_j with QX_j in the expression above, to obtain

$$\begin{aligned}\mu\left(\mathrm{tr}\left(RPP^TR^TQQ^T\right)\right) &= \sum_j \mu\left(\left(RP_j\right)^TQQ^T(RP_j)\right) = \sum_j \mu\left(\left(QX_j\right)^TQQ^T(QX_j)\right) \\ &= \sum_j \mu\left(\mathrm{tr}\left(Q^TQQ^TQX_jX_j^T\right)\right) = \sum_j \mathrm{tr}\left(Q^TQQ^TQ\,\mu\left(X_jX_j^T\right)\right)\end{aligned}$$

According to (3c), $\mu\left(X_jX_j^T\right)$ is a diagonal matrix whose diagonal elements are the members of X_j, which implies

$$\sum_j \mathrm{tr}\left(Q^TQQ^TQ\,\mu\left(X_jX_j^T\right)\right) = \sum_i \sum_j \mathcal{X}_{ij}\left(Q^TQQ^TQ\right)_{ii} = \sum_i \left(Q^TQQ^TQ\right)_{ii} = \left\|QQ^T\right\|_F^2$$

Returning to (12) and using the fact that in our case $P = Q$, we obtain

$$\mu\left(\left\|RPP^T - QQ^TR\right\|_F^2\right) = \left\|PP^T\right\|_F^2 + \left\|PP^T\right\|_F^2 - 2\left\|PP^T\right\|_F^2 = 0$$