

Exact recovery with symmetries for the Doubly-Stochastic relaxation

Nadav Dym

Weizmann Institute of Science

May 22, 2017

Abstract

Graph matching or quadratic assignment, is the problem of labeling the vertices of two graphs so that they are as similar as possible. A common method for approximately solving the NP-hard graph matching problem is relaxing it to a convex optimization problem over the set of doubly stochastic matrices. Recent analysis has shown that for almost all pairs of isomorphic and asymmetric graphs, the doubly stochastic relaxation succeeds in correctly retrieving the isomorphism between the graphs. Our goal in this paper is to analyze the case of symmetric isomorphic graphs. This goal is motivated by shape matching applications where the graphs of interest usually have reflective symmetry.

For symmetric problems the graph matching problem has multiple isomorphisms and so convex relaxations admit all convex combinations of these isomorphisms as viable solutions. If the convex relaxation does not admit any additional superfluous solution we say that it is convex exact.

We show that convex exactness depends strongly on the symmetry group of the graphs; For a fixed symmetry group G , either the DS relaxation will be convex exact for almost all pairs of isomorphic graphs with symmetry group G , or the DS relaxation will fail for *all* such pairs. We show that for reflective groups with at least one full orbit convex exactness holds almost everywhere, and provide some simple examples of non-reflective symmetry groups for which convex exactness always fails.

When convex exactness holds, the isomorphisms of the graphs are the extreme points of the convex solution set. We suggest an efficient algorithm for retrieving an isomorphism in this case. We also show that the "convex to concave" projection method will also retrieve an isomorphism in this case, and show experimentally that this projection method as well as the standard Euclidean projection will succeed in retrieving an isomorphism for near isomorphic graphs as well.

In certain cases it is sufficient to find the centroid of the set of isomorphisms, which gives a "fuzzy encoding" of the symmetries of the shape. We show that for *any* symmetry group G , the centroid solution can be recovered efficiently for almost all pairs of isomorphic graphs with symmetry group G . Additionally we show that for such isomorphic graphs interior-point solvers will generally return the centroid solution.

1 Introduction

Graph matching and *graph isomorphism* are classical problems in computer science. In this paper we will use the term *graph* for a pair (\mathbf{a}, A) , where $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ are the vertices of the graph, and A is a symmetric matrix encoding the relationship between the vertices. We will also sometimes refer to A alone as a graph. An *isomorphism* between graphs (\mathbf{a}, A) and (\mathbf{b}, B) is a relabeling of the vertices of B so that A and the relabeled B are identical. The graphs A and B are *isomorphic* if there is an isomorphism between them. In matrix notation, an isomorphism is a permutation matrix P such that $A = PB P^T$ or equivalently $AP = PB$. The problem of

deciding whether two graphs are isomorphic is known as the *Graph isomorphism problem* (GI). It is not known to be in P, but is also not known to be NP-hard. Recently [Babai, 2016] provided a quasi-polynomial time algorithm for GI. While no polynomial algorithm for the general GI problem is known, there are many families of graphs for which GI can be solved in polynomial time. One example which is relevant for this work is graphs with simple spectrum, or more generally bounded eigenvalue multiplicity [Babai et al., 1982].

The *graph matching* problem is the problem of determining how close two graphs are to being isomorphic by minimizing the graph matching energy over the set of permutation matrices which we denote by Π_n :

$$\min_{P \in \Pi_n} E(P) = \|AP - PB\|_F. \quad (1)$$

This optimization problem is also often referred to as the Koopmans-Beckmann *quadratic assignment* problem, and is usually phrased as the equivalent problem of maximizing $\text{tr}APB^TP^T$. In contrast to GI whose computational status is not fully known, global minimization of quadratic assignment, and even approximation to within a constant factor, is known to be NP-hard [Sahni and Gonzalez, 1976].

Graph matching problems have found many applications. See for example [Conte et al., 2004] for a survey on applications of graph matching for pattern recognition. Our work is motivated by shape matching applications: Shape matching is the problem of measuring how similar two given surfaces $\mathcal{S}_A, \mathcal{S}_B$ are. The notion of similarity between shapes is required to be invariant to shape preserving deformations such as rigid transformations for rigid objects (*e.g.*, chairs), and deformations which preserve geodesic distances for non-rigid objects (*e.g.*, humans). Accordingly shape matching problems are often modeled (*e.g.*, [Mémoli, 2007, Mémoli, 2011, Solomon et al., 2016]) as the problem of finding a mapping between two surfaces $\mathcal{S}_A, \mathcal{S}_B$ so that they are as isometric as possible. The metric on the shapes is typically either the extrinsic Euclidean metric for rigid shapes, or the intrinsic geodesic metric for non-rigid shapes.

Finding near-isometries between shapes can be phrased as a graph matching problem by selecting a finite sampling of the shapes to obtain vertices \mathbf{a}, \mathbf{b} on the two shapes, and taking A, B to be the distance matrices defined by the distances on the shapes, that is

$$A_{ij} = d_A(\mathbf{a}_i, \mathbf{a}_j), \quad B_{ij} = d_B(\mathbf{b}_i, \mathbf{b}_j)$$

In this setting an isomorphism between A and B corresponds to an isometry between the sampled metric spaces.

In this work we will focus on symmetric graphs, which are very relevant for shape matching applications since most natural shapes have intrinsic symmetries- that is, intrinsic isometries from the shape to itself other than the trivial identity mapping. Figure 1 shows some representative shapes from the [Giorgi et al., 2007] shape matching dataset. Typically natural shapes have a symmetry group with only two elements (bilateral symmetry) as in the left hand side of Figure 1, but there are interesting examples with larger symmetry groups as in the right hand side of Figure 1.

The doubly-stochastic relaxation In this paper we focus on analyzing the doubly-stochastic (DS) relaxation for graph matching. For a survey on other convex relaxation and combinatorial methods which have been proposed to achieve good solutions for quadratic assignment see [Loiola et al., 2007].

The doubly stochastic (DS) relaxation replaces the NP hard graph matching problem with a tractable optimization problem by relaxing the combinatorial set of permutations to its convex hull of doubly stochastic matrices:

$$\text{DS} = \{S \mid S\mathbf{1} = \mathbf{1}, \quad \mathbf{1}^T S = \mathbf{1}^T, \quad S \geq 0\},$$

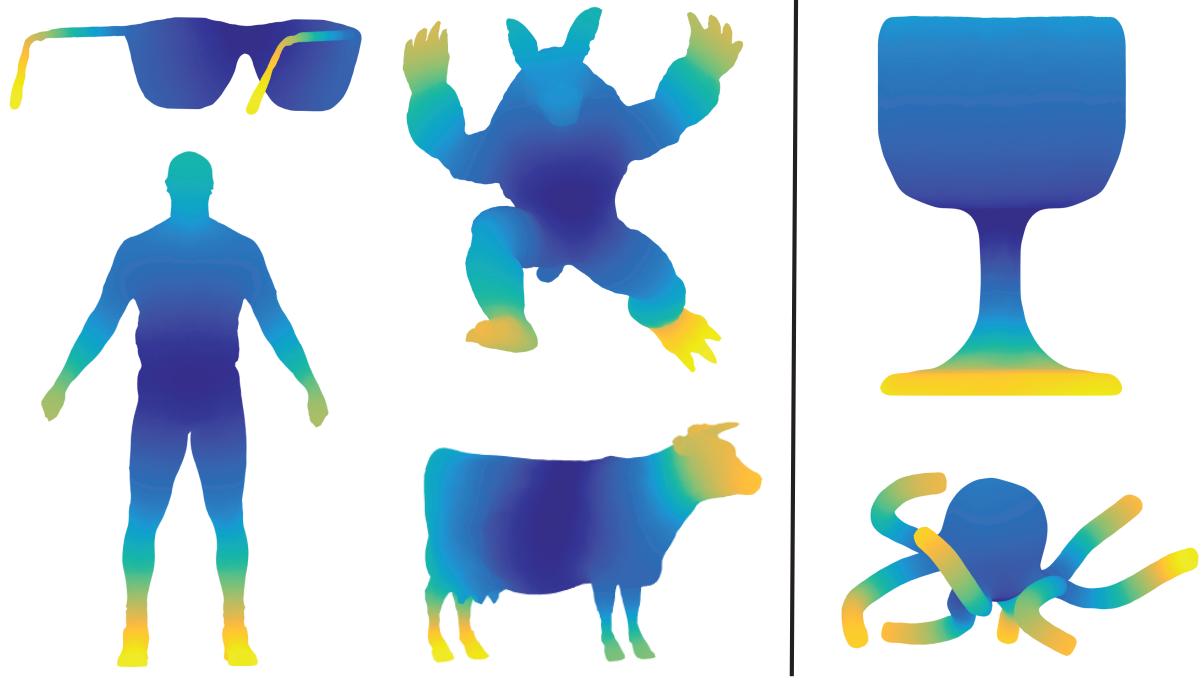


Figure 1: Shapes with bilateral symmetry (left) and with non reflective symmetries (right) from the SHREC dataset [Giorgi et al., 2007]. Our results suggest the DS relaxation is appropriate for solving problems with bilateral symmetry. For problems with non-reflective symmetries our results are non-conclusive (see Figure 2).

which leads to a convex quadratic program known as the *DS relaxation*:

$$\min_{S \in \text{DS}} E(S) = \|AS - SB\|_F. \quad (2)$$

We will refer to this optimization problem as $\text{DS}(A, B)$. Since the DS relaxation minimizes $E(\cdot)$ over a larger domain, its minimum value is a lower bound for the minimal value of the graph matching problem. As can be expected due to the hardness of the problem, the DS relaxation does not generally return the global minimum or minimizer of (1) [Lyzinski et al., 2016]. In particular, [Scheinerman and Ullman, 2011] characterizes all cases in which the minimum of (2) is zero even when the graphs are not isomorphic.

We will be interested in the case where A and B are isomorphic. Note that in this case the global minimum of (2) is zero and thus coincides with the global minimum of (1). The interesting question is whether the DS relaxation succeeds in returning a minimizer which is a permutation. Clearly we do not expect this will be the case for all graphs since this would provide us with a polynomial time algorithm to solve GI. On the other hand, since there are many families of graphs for which GI is tractable, we can hope that for many instances the DS relaxation will be successful in returning a permutation solution. The recent works of [Afifalo et al., 2015, Fiori and Sapiro, 2015] show that indeed this is the case. To state their results we introduce some notation:

Let us denote the set of isomorphisms of A, B by $\text{ISO}(A, B)$. We will say that $S \in \text{DS}$ is a *convex isomorphism* if it is a member of the set

$$\text{ISO}_{\text{conv}}(A, B) = \{S \in \text{DS} \mid AS = SB\}.$$

The inclusion

$$\text{ISO}(A, B) \subseteq \text{ISO}_{\text{conv}}(A, B) \quad (3)$$

is obvious. However it is possible that the DS relaxation will contain additional minimizers. We will say that $\text{DS}(A, B)$ is *exact* when this possibility does not occur and

$$\text{ISO}_{\text{conv}}(A, B) = \text{ISO}(A, B).$$

We note that the exactness property depends only on A : An isomorphism $P \in \text{ISO}(A, B)$ defines a linear bijection

$$S \mapsto SP^T$$

from $\text{ISO}_{\text{conv}}(A, B)$ to $\text{ISO}_{\text{conv}}(A, A)$ and from $\text{ISO}(A, B)$ to $\text{ISO}(A, A)$. Accordingly if A, B are isomorphic, $\text{DS}(A, B)$ is exact if and only if $\text{DS}(A, A)$ is exact. We will refer to (convex) isomorphisms in the case $A = B$ as (convex) *automorphisms*. We also denote:

$$\text{Aut}(A) = \text{ISO}(A, A), \quad \text{Aut}_{\text{conv}}(A) = \text{ISO}_{\text{conv}}(A, A), \quad \text{DS}(A) = \text{DS}(A, A).$$

We say that A is an *asymmetric* graph if the identity matrix is its only automorphism. Otherwise we say that A is a *symmetric* graph. A necessary condition for exactness of $\text{DS}(A)$ is that A is asymmetric. This is because if A has several automorphisms then due to the inclusion (3) and the convexity of $\text{Aut}_{\text{conv}}(A)$

$$\text{conv Aut}(A) \subseteq \text{Aut}_{\text{conv}}(A). \quad (4)$$

Thus, while A has a finite number of automorphisms, it has an infinite number of convex automorphisms. Even when A is asymmetric, exactness does not always occur. A simple counter example will be discussed in Section 2. However, [Aflalo et al., 2015] showed that for asymmetric A satisfying certain weak conditions exactness will hold. Their result was later shown to hold with even weaker conditions in [Fiori and Sapiro, 2015].

Convex exactness Our goal in this paper is to show that for certain kinds of symmetry groups the DS relaxation can still be successfully applied, by defining a suitable notion of *convex exactness*. A similar goal has recently been achieved by [Dym and Lipman, 2016] for a semi-definite programming relaxation of the Procrustes matching problem.

We say that $\text{DS}(A)$ is *convex exact* if equality holds in (4), or equivalently if for any B isomorphic to A ,

$$\text{conv ISO}(A, B) = \text{ISO}_{\text{conv}}(A, B). \quad (5)$$

Note that for asymmetric graphs, convex exactness and exactness coincide. When convex exactness holds an isomorphism can be extracted in a tractable manner as we will discuss in Section 6.

For every permutation subgroup $G \leq \Pi_n$ we define

$$\mathcal{A}(G) = \{A \in \mathcal{S}^n \mid \text{Aut}(A) = G\}.$$

In the asymmetric case $G = \{I_n\}$ we know that there are $A \in \mathcal{A}(G)$ such that $\text{DS}(A)$ is not (convex) exact, but also that (convex) exactness often does hold for asymmetric graphs. Our goal is to give a more precise notion of this claim by showing that for almost all asymmetric graphs (convex) exactness holds. More importantly, we would like to find non-trivial groups G for which $\text{DS}(A)$ will be convex exact for almost every $A \in \mathcal{A}(G)$. To do so we must first define a natural measure μ_G on $\mathcal{A}(G)$.

We will assume that $\mathcal{A}(G)$ is non-empty. Permutation groups $G \leq \Pi_n$, for which $\mathcal{A}(G)$ is empty do exist. A simple example is the cyclic group $G \leq \Pi_3$ generated by the permutation

$$\mathbf{a}_1 \mapsto \mathbf{a}_2, \quad \mathbf{a}_2 \mapsto \mathbf{a}_3, \quad \mathbf{a}_3 \mapsto \mathbf{a}_1.$$

Any $A \in \mathcal{A}(G)$ satisfies

$$A_{11} = A_{22} = A_{33} \text{ and } A_{12} = A_{23} = A_{31}.$$

Thus, all diagonal elements of A are identical and all off-diagonal elements of A are identical as well since $A = A^T$. It follows that $\text{ISO}(A) = \Pi_3$. If $\mathcal{A}(G)$ is non-empty we say that G is a *symmetry group*.

For a symmetry group G we consider the vector space

$$\mathcal{V}(G) = \{A \in \mathcal{S}^n \mid A = P^T AP \text{ for all } P \in G\}.$$

Since $\mathcal{V}(G)$ is a vector space of some dimension d it has a natural notion of measure- the d -dimensional Hausdorff measure on $\mathbb{R}^{n \times n}$ restricted to $\mathcal{V}(G)$, or equivalently the push forward of the Lebesgue measure on \mathbb{R}^d to $\mathcal{V}(G)$ via a linear isometry between the two spaces. We denote this measure by μ_G . Note that

$$\mathcal{A}(G) = \mathcal{V}(G) \setminus \bigcup_{G \subsetneq H \leq \Pi_n} \mathcal{V}(H).$$

Since by assumption $\mathcal{A}(G)$ is non-empty it follows that all the $\mathcal{V}(H)$ are strict subspaces of $\mathcal{V}(G)$ and therefore the complement of $\mathcal{A}(G)$ in $\mathcal{V}(G)$ has measure zero. Thus μ_G is a natural choice for a measure on $\mathcal{A}(G)$. We will say that a property is *generic*, or that it holds for almost every $A \in \mathcal{A}(G)$, if it holds for μ_G almost every $A \in \mathcal{A}(G)$.

We can now state our main results:

1.1 Main results

Reflective groups We show that convex exactness is a generic property for groups G fulfilling the following two conditions:

Definition 1. *We say that $G \leq \Pi_n$ is a reflection group if $P^2 = I_n$ for all $P \in G$.*

Any group $G \leq \Pi_n$ defines an action $(\sigma, \mathbf{a}_j) \mapsto \mathbf{a}_{\sigma(j)}$ on the set of vertices \mathbf{a} . We denote the orbit of \mathbf{a}_j by $[\mathbf{a}_j]$. In general we have that $|[\mathbf{a}_j]| \leq |G|$.

Definition 2. *We say that G has a full orbit if it has an orbit of length $|G|$.*

In shape matching applications the full orbit assumption is typically fulfilled; an orbit $[\mathbf{a}_i]$ will be full unless \mathbf{a}_i is on a symmetry axis of the shape. Under the full orbit and reflection group assumption, we prove:

Theorem 1. *Assume $G \leq \Pi_n$ is a reflection group with a full orbit. Then the DS relaxation is convex exact with respect to almost all $A \in \mathcal{A}(G)$.*

As a result we obtain that convex exactness is a generic property for the simplest but, in the context of shape matching applications, most important, symmetry groups:

Corollary 1. *If $G \cong \mathbb{Z}_2$ then the DS relaxation is convex exact with respect to almost all $A \in \mathcal{A}(G)$.*

General groups For general groups we provide a "zero-one probability" result:

Theorem 2. *For any symmetry group $G \leq \Pi_n$ one of the following holds:*

1. *The DS relaxation is convex exact with respect to almost every $A \in \mathcal{A}(G)$.*
2. *The DS relaxation is **not** convex exact for any $A \in \mathcal{A}(G)$.*

The proof of Theorem 2 is constructive in the sense that it enables checking which of the two mutually exclusive alternatives described in the theorem hold for a given symmetry group G . By using this strategy we can establish that there are quite simple non-reflective symmetry groups for which convex exactness fails. Figure 2 shows nine groups G_i represented by nine shapes whose symmetry group is G_i . For the first three groups (a)-(c) we found that convex exactness does not hold for any $A \in \mathcal{A}(G)$, while for the remaining groups convex exactness does hold for almost all $A \in \mathcal{A}(G)$. Note that all groups in the first column are isomorphic to \mathbb{Z}_3 , all groups in the second column are isomorphic to \mathbb{Z}_4 , and all groups in the last column are isomorphic to the dihedral group D_4 . Thus we see that while convex exactness is a generic property for any G isomorphic to \mathbb{Z}_2 , in general different permutation groups can behave very differently with respect to the DS relaxation even if they are isomorphic in the sense of group theory.

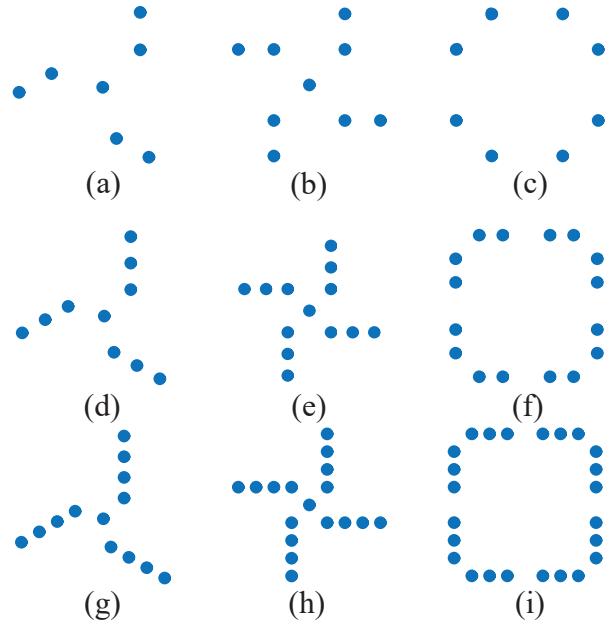


Figure 2: Non-reflective symmetry groups.
The figure contains nine sub-diagrams labeled (a) through (i). Each diagram consists of several blue dots representing vertices of a graph. (a) shows a triangular arrangement of dots. (b) shows a square-like arrangement. (c) shows a more scattered arrangement. (d) shows a zig-zag pattern. (e) shows a cross-like pattern. (f) shows a rectangular pattern. (g) shows a curved path of dots. (h) shows a more complex cross-like pattern. (i) shows a large rectangular loop of dots.

Additional results: Permutation solutions and centroid solution Since for symmetric problems the DS relaxation has an infinite number of convex isomorphisms, the question of achieving an "interesting" convex isomorphism arises. Naturally we would like to achieve a convex isomorphism which is a permutation. In the case of convex exactness this reduces to the problem of finding an extreme point (a "corner") of the set of convex isomorphisms, which is known to be a tractable problem. In Section 6 we describe two known methods to obtain extreme points. Additionally we provide a much faster algorithm for achieving all isomorphisms between A and B . This algorithm is valid for almost all graphs whose symmetry group G satisfy the conditions of Theorem 1.

A disadvantage of the methods mentioned above for finding permutation solutions is that they are constructed for perfectly isomorphic problems and are not suited for near isomorphic problems and are not used in practice. Instead, permutations are typically obtained using the L_2 projection or the more accurate, but more expensive, "convex to concave" projection. We prove that when convex exactness holds, the convex to concave projection is able to return an isomorphism. For symmetric problems with a small amount of noise, we show experimentally that both projection methods are generally able to retrieve an isomorphism, and the convex to concave method is often able to retrieve an isomorphism for higher noise levels as well.

An alternative "interesting" convex isomorphism which is easier to find than "corners" is the "centroid" of the set of isomorphisms:

$$S_c = \frac{1}{|G|} \sum_{P \in G} P, \quad (6)$$

where G is the set of isomorphisms between A and B . As advocated in [Solomon et al., 2012], finding S_c in the case of symmetric problems can potentially be useful as it gives an "encoding" of all isomorphisms of A, B . In Section 5 we show that the centroid solution is easier to find than corner solutions: In fact, for any symmetry group G , and almost every pair of isomorphic graphs $A, B \in \text{ISO}(G)$, the centroid solution can be achieved (almost always) for *any* symmetry group.

Additionally we show that for such A, B penalty based optimization methods will converge to S_c when solving $\text{DS}(A, B)$.

An illustration of the centroid solution is shown in Figure 3, for the problem of mapping a cylinder to itself, using as $A = B$ the Euclidean distance matrix of the cylinder. The right part of the figure shows the cylinder, colored so that points in the same orbit of the symmetry group of the cylinder share the same color. The left part of the figure shows the matrix S_c . Each yellow square in the left figure is a submatrix whose indices correspond to a circular section of the cylinder. It can be seen that the centroid solution assigns each point of the cylinder with equal probability to any other point in its orbit. We note that the centroid solution always has this property. Therefore different symmetry groups which have identical orbits will have the same centroid solution, and so the symmetry group cannot generally be reconstructed from the centroid solution.

The remainder of the paper is organized as follows: In Section 2 we define the notion of weak exactness which will be useful for the proofs presented later on. In Section 3 we prove convex exactness for reflective groups (Theorem 1). In Section 4 we prove our "zero-one probability result" (Theorem 2) and explain how to check which of the two alternatives described in the theorem apply for a given group. In Section 5 we discuss the issue of retrieving the centroid solutions and finally in Section 6 we discuss the issue of retrieving isomorphisms in the case that convex exactness holds.

2 Weak exactness

An important tool for the proofs we present later on is the concept of *weak exactness* which we will now define: For any set G of permutation matrices we define

$$\mathcal{N}(G) = \{T \in \mathbb{R}^{n \times n} \mid T_{ij} = 0 \text{ if } P_{ij} = 0 \text{ for all } P \in G\}.$$

We say that $\text{DS}(A)$ is weakly exact if all convex automorphisms of A are in $\mathcal{N}(\text{Aut}(A))$. Less formally, this means that the i, j coordinate of a convex automorphism S can be positive only if there is an automorphism taking \mathbf{a}_i to \mathbf{a}_j . If $\text{DS}(A)$ is weakly exact and B is isomorphic to A then all convex isomorphisms of A, B are in $\mathcal{N}(\text{ISO}(A, B))$. Weak exactness is guaranteed with full probability for any symmetry group G . We show this using the vector

$$s(A) = A\mathbf{1}.$$

The vector $s(A)$ is invariant under automorphisms, meaning that if $\mathbf{a}_j \in [\mathbf{a}_i]$ then $s_i(A) = s_j(A)$. We say that $s(A)$ is *discriminative* if for any i, j such that $\mathbf{a}_j \notin [\mathbf{a}_i]$, we have $s_i(A) \neq s_j(A)$. We prove

Theorem 3. *Let G be any symmetry group. Then*

1. *If $s(A)$ is discriminative then $\text{DS}(A)$ is weakly exact.*
2. *For almost every $A \in \mathcal{A}(G)$, the vector $s(A)$ is discriminative.*

We prove Theorem 3. We first prove that if $s(A)$ is discriminative then $\text{DS}(A)$ is weakly exact. If S is a convex isomorphism, then $s = s(A)$ is fixed by S because

$$Ss = SA\mathbf{1} = AS\mathbf{1} = A\mathbf{1} = s.$$

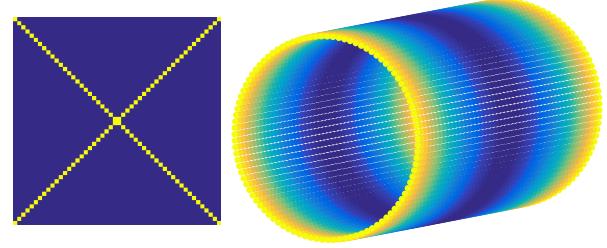


Figure 3: The centroid solution.

Write S as a convex combination of permutations $S = \sum_k \theta_k P(k)$. Using the fact that the operator norm of a permutation is one and the Cauchy-Schwartz inequality we obtain

$$\|s\|^2 = \langle Ss, s \rangle = \sum_k \theta_k \langle P(k)s, s \rangle \stackrel{(*)}{\leq} \sum_k \theta_k \|P(k)\| \|s\|^2 = \|s\|^2.$$

so $(*)$ is an equality, implying that $P(k)s = s$ for all k . Now if $\mathbf{a}_j \notin [\mathbf{a}_i]$ then $P_{ij}(k) = 0$ for all k and therefore $S_{ij} = 0$. Thus we have proven that $\text{DS}(A)$ is weakly exact when $s(A)$ is discriminative.

We now show that discriminativeness is a generic property. It is sufficient to show that for almost every $A \in \mathcal{V}(G)$ the claim holds since $\mathcal{A}(G)$ is a subset of $\mathcal{V}(G)$. Note that $s(A)$ is discriminative unless there are some (i, j) such that \mathbf{a}_j is not in $[\mathbf{a}_i]$ but A is in the vector space

$$\mathcal{V}_{i,j}(G) = \{A \in \mathcal{V}(G) | s_i(A) = s_j(A)\}.$$

Thus it is sufficient to show that all these spaces are strict subspaces of $\mathcal{V}(G)$, which we accomplish by finding a member $\bar{A} \in \mathcal{V}(G)$ for which $s(\bar{A})$ is discriminative.

To construct \bar{A} let I_1, I_2, \dots, I_k be the partition of the vertices \mathbf{a} induced by the action of G . For all $r \leq k$ and $\mathbf{a}_i, \mathbf{a}_j \in I_r$ we set

$$\bar{A}_{ij} = r|I_r|^{-1}.$$

If $\mathbf{a}_j \notin [\mathbf{a}_i]$ we set $\bar{A}_{ij} = 0$. The constructed graph \bar{A} is a member of $\mathcal{V}(G)$, and the vector $s(\bar{A})$ is discriminative since for all $r \leq k$ and $i \in I_r$ we have $s_i(\bar{A}) = r$. This concludes the proof of Theorem 3.

Counter example While weak exactness is guaranteed almost everywhere, it can still fail in very simple examples. Such examples can be constructed using the fact that if $\mathbf{1}$ is an eigenvector of A then $\frac{1}{n}\mathbf{1}^T\mathbf{1}$ is always a valid convex automorphism. For example the graph

$$A_0 = \begin{bmatrix} 6 & 1 & 2 \\ 1 & 5 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

is asymmetric, but satisfies $A_0\mathbf{1} = \lambda\mathbf{1}$ for $\lambda = 9$ and so $\frac{1}{n}\mathbf{1}^T\mathbf{1}$ is a convex isomorphism, and so weak exactness, and certainly exactness, does not hold.

One method for overcoming such counter examples is adding a linear term to the graph matching energy penalizing for correspondences which do not respect isomorphism-invariants. For example, For each vertex \mathbf{a}_i we can define $L(i)$ to be the sorted values of the i -th row of the graph. Clearly if $\mathbf{a}_j \in [\mathbf{a}_i]$ then $L(i) = L(j)$ so L is an isomorphism invariant. Since in our example $L(i), i = 1, 2, 3$ are all distinct, the only zero-energy solution of the modified relaxation

$$\min_{S \in \text{DS}} \|SA_0 - A_0S\|_F^2 + \sum_{ij} S_{ij} \|L(i) - L(j)\| \quad (7)$$

is the identity matrix.

3 Convex exactness for reflective groups

Our goal in this section is proving convex exactness holds generically for reflective groups with a full orbit (Theorem 1). We break up the proof of the theorem into two parts: The first part establishes sufficient conditions which guarantee exact recovery, and the second part proves these sufficient conditions hold generically if G is reflective and has a full orbit.

Fix some G and $A \in \mathcal{A}(G)$. As in the previous section let I_1, I_2, \dots, I_k be the partition of the vertices \mathbf{a} induced by the action of G , and denote $n_j = |I_j|$. Let S be a convex automorphism of A . Denote by S_{ij} and A_{ij} the submatrices of S, A corresponding to the indices $I_i \times I_j$. If A is weakly exact then $S_{ij} = 0$ whenever $i \neq j$ and so the equation $AS = SA$ takes the form

$$A_{ij}S_{jj} = S_{ii}A_{ij}, \quad \text{for all } 1 \leq i, j \leq k. \quad (8)$$

3.1 Sufficient conditions for convex exactness

Proposition 3.1. *Let A be a graph. If $\exists r, 1 \leq r \leq k$ such that*

1. *The vector $s(A)$ is discriminative.*
2. *$\text{rank}(A_{rj}) = n_j$ for all $j \neq r$.*
3. *A_{rr} has simple spectrum.*

then $DS(A)$ is convex exact.

Proof of Proposition 3.1. The first condition guarantees weak exactness, and thus that all convex isomorphisms will satisfy (8). Setting $i = r$ in this equation we obtain

$$A_{rj}S_{jj} = S_{rr}A_{rj}, \quad \text{for all } j$$

and so by the second condition S_{jj} is determined uniquely by S_{rr} . It follows that the restriction of the linear map

$$S \mapsto S_{rr}$$

to $\text{Aut}_{\text{conv}}(A)$ is injective. Therefore it is sufficient to show that S_{rr} is a convex combination of the permutation matrices P_{rr} obtained by restricting the automorphisms $P \in \text{Aut}(A)$ to $I_r \times I_r$. By taking $i = r, j = r$ in (8) we see that S_{rr} is a convex automorphism of the subgraph A_{rr} , and the group

$$H = \{P_{rr} \mid P \in \text{Aut}(A)\}$$

is a subgroup of $\text{Aut}(A_{rr})$ which acts transitively on I_r . By the third assumption A_{rr} has simple spectrum. Thus to show A_{rr} is a convex combination of elements of H it is sufficient to prove

Lemma 1. *If $A \in \mathcal{S}^n$ is a graph with simple spectrum, and $H \leq \text{Aut}(A)$ acts transitively on the vertices \mathbf{a} , then $H = \text{Aut}(A)$ and $DS(A)$ is convex exact.*

We now conclude the proof of the proposition by proving the lemma. In this proof B_j denotes the j -th column of the matrix B , and $B_{i\star}$ denotes the i -th row of B .

To prove the lemma it is sufficient to show that

$$\text{Aut}_{\text{conv}}(A) \subseteq \text{conv } H \quad (9)$$

because this implies that

$$\text{conv Aut}(A) \subseteq \text{Aut}_{\text{conv}}(A) \subseteq \text{conv } H \subseteq \text{conv Aut}(A)$$

which proves that $DS(A)$ is convex exact. Additionally all automorphisms $P \in \text{Aut}(A)$ are in $\text{conv } H$, and since permutations are extreme points of DS this can only occur if $P \in H$, and so $H = \text{Aut}(A)$.

We prove (9) using an argument from [Dym and Lipman, 2016]; If S is a convex automorphism of A and v is an eigenvector of A with eigenvalue λ , Then

$$ASv = SA v = \lambda S v$$

so either $Sv = 0$ or Sv is an eigenvector of A with eigenvalue λ . Since A has simple spectrum it follows that $Sv = \alpha v$ for some $\alpha \in \mathbb{R}$. If $S = P$ is an automorphism of A then $\alpha \in \{-1, 1\}$.

Let S be a convex automorphism of A . We want to show that $S \in \text{conv } H$. Since H acts transitively on \mathbf{a} , there are permutation matrices $P(1), \dots, P(n) \in H$ such that for any vector $w \in \mathbb{R}^n$

$$w_k = (P(k)w)_1.$$

Denote by V the matrix whose columns are the eigenvectors of A . Then there is a diagonal matrix D and diagonal matrices $D(1), \dots, D(n)$ such that

$$D = V^T SV \text{ and } D(k) = V^T P(k)V, \text{ for all } 1 \leq k \leq n. \quad (10)$$

Note that S is a convex combination of $P(1), \dots, P(n)$ if and only if D is a convex combination of $D(1), \dots, D(n)$. If v is the j -th eigenvector of A then

$$|v_k| = |(P(k)v)_1| = |D_{jj}(k)v_1| = |v_1|$$

so in particular v has no zero coordinates.

From (10) we obtain

$$(VD)_{1*} = (SV)_{1*} = \sum_{k=1}^n S_{1k} V_{k*} = \sum_{k=1}^n S_{1k} (P(k)V)_{1*} = \sum_{k=1}^n S_{1k} (VD(k))_{1*}$$

and therefore

$$V_{1*} \left(D - \sum_{k=1}^n S_{1k} (D(k)) \right) = 0.$$

Since all entries of V are non-zero the only diagonal matrix solving the equation above is the zero matrix. Thus we obtain D as a convex combination of $D(k)$:

$$D = \sum_{k=1}^n S_{1k} (D(k)).$$

□

3.2 Genericity of the sufficient conditions

In this subsection we prove the sufficient conditions of Proposition 3.1 hold generically if G is reflective and has a full orbit. The first condition was proved to hold generically for any symmetry group G in Theorem 3. We choose the r appearing in the last two conditions of of Proposition 3.1 such that I_r is a full orbit, or equivalently $|I_r| = |G|$. We begin with some preliminaries.

Preliminaries Recall that for a symmetry group G of dimension d , the measure μ_G can be defined as the restriction to $\mathcal{V}(G)$ of the d -dimensional Hausdorff measure H^d on $\mathbb{R}^{n \times n}$. We cite some basic properties of the Hausdorff measure and dimension from chapter 2 in [Falconer, 2004] which will be helpful for the proof of Lemma 3.

1. If C has Hausdorff dimension k and $s > k$, then $H^s(C) = 0$.
2. If $M \subseteq \mathbb{R}^n$ is a submanifold of dimension d , then its Hausdorff dimension is d as well.
3. If $B = \cup_{i \in I} B_i$ and I is countable, then $\dim B = \sup_{i \in I} \dim B_i$.
4. If $B \subseteq C$ then $\dim(B) \leq \dim(C)$.

5. If $B \subseteq \mathbb{R}^m$ and $f : B \rightarrow \mathbb{R}^n$ is Lipschitz, then $\dim f(B) \leq \dim(B)$.

An immediate consequence is that the latter inequality holds if f is a C^1 function defined on all of \mathbb{R}^m . To see this denote $B_k = B \cap \{x \mid \|x\| \leq k\}$ and note that the restriction of f to B_k is Lipschitz. Therefore

$$\dim f(B) = \dim \cup_k f(B_k) = \sup_k \dim f(B_k) \leq \sup_k \dim B_k \leq \dim B \quad (11)$$

For Lemma 4 we will need the following simple lemma. We include a proof for completeness:

Lemma 2. *If $p(x)$ is a non-zero multivariate polynomial $p : \mathbb{R}^d \rightarrow \mathbb{R}$, then the set $\{x \mid p(x) = 0\}$ has Lebesgue measure zero.*

Proof. By induction. For $d = 1$ the claim is obvious. We assume the claim holds for $d - 1$ and show it holds for d . Rewrite p as

$$p(x) = \sum_{j=1}^n q_j(x_1, \dots, x_{d-1}) x_d^j.$$

By the induction hypothesis the set

$$C = \{(x_1, \dots, x_{d-1}) \mid \forall j, q_j(x_1, \dots, x_{d-1}) = 0\}$$

has measure zero in \mathbb{R}^{d-1} . For any fixed (x_1, \dots, x_{d-1}) in the complement of C , $p(x)$ is a univariate non-zero polynomial and has zeros in a (finite) subset of \mathbb{R} of measure zero. Using Fubini's theorem this implies that the set $\{x \mid p(x) = 0\}$ has Lebesgue measure zero. \square

Proof of genericity A graph A is in $\mathcal{V}(G)$ if it is symmetric and (8) is satisfied when S is replaced with all permutations $P \in G$. This means that $\mathcal{V}(G) = \oplus_{i \geq j} \mathcal{V}_{ij}(G)$ where

$$\mathcal{V}_{jj}(G) = \{A_{jj} \in \mathcal{S}(n_j) \mid A_{jj} P_{jj} = P_{jj} A_{jj} \text{ for all } P \in G\}.$$

and for $i > j$

$$\mathcal{V}_{ij}(G) = \{A_{ij} \in \mathbb{R}^{n_i \times n_j} \mid A_{ij} P_{jj} = P_{ii} A_{ij}, \text{ and } A_{ij}^T P_{ii} = P_{jj} A_{ij}^T\}.$$

Thus to prove the second condition is generic it is sufficient to show that almost every $A_{rr} \in \mathcal{V}_{rr}(G)$ has simple spectrum, and that to prove the third condition is generic we need to show that almost every $A_{rj} \in \mathcal{V}_{rj}(G)$ has full rank. Thus the second condition follows by setting $A = A_{rj}$ and $\mathcal{V} = \mathcal{V}_{rj}$ in the following Lemma:

Lemma 3. *If G is reflective then almost all $A \in \mathcal{V}(G)$ has simple spectrum.*

Proof. Since G is reflective all $P \in G$ satisfy

$$P = P^{-1} = P^T.$$

Members $P, Q \in G$ commute because

$$PQ = (PQ)^T = Q^T P^T = QP.$$

Thus G can be diagonalized simultaneously, and so we can partition \mathbb{R}^n into a direct sum of eigenspaces $\mathbb{R}^n = \oplus_{i=1}^\ell W_i$. We denote the dimension of each eigenspace by d_i . Select for each subspace W_i a matrix $V_i \in \mathbb{R}^{n \times d_i}$ whose columns form an orthogonal eigenbasis of W_i , and denote

$$V = [V_1, \dots, V_\ell].$$

A graph A is in $\mathcal{V}(G)$ if and only if it is symmetric and it commutes with the members of G . This in turn occurs if and only if A and all members of G can be diagonalized simultaneously, and so there are symmetric matrices \bar{A}_i , $i = 1, \dots, \ell$ such that

$$V^T A V = \begin{bmatrix} \bar{A}_1 & & & \\ & \bar{A}_2 & & \\ & & \ddots & \\ & & & \bar{A}_\ell \end{bmatrix}$$

It follows that $\mathcal{V}(G)$ can be identified with $\bigoplus_{i=1}^\ell S(d_i)$, and is thus of dimension

$$\dim(\mathcal{V}(G)) = \sum_{i=1}^\ell \frac{d_i^2 + d_i}{2}$$

For $(U_i, \lambda_i) \in \mathcal{O}(d_i) \times \mathbb{R}^{d_i}$ we define $D(\lambda_i)$ to be the diagonal matrix whose diagonal entries are λ_i , and define

$$\bar{A}_i(U_i, \lambda_i) = U_i D(\lambda_i) U_i^T.$$

Now consider the function $f : \prod_{i=1}^\ell (\mathcal{O}(d_i) \times \mathbb{R}^{d_i}) \rightarrow \mathbb{R}^{n \times n}$ defined by

$$f(U_1, \lambda_1, \dots, U_\ell, \lambda_\ell) = V \begin{bmatrix} \bar{A}_1(U_1, \lambda_1) & & & \\ & \bar{A}_2(U_2, \lambda_2) & & \\ & & \ddots & \\ & & & \bar{A}_\ell(U_\ell, \lambda_\ell) \end{bmatrix} V^T$$

The image of f is precisely $\mathcal{V}(G)$. Moreover the dimension of the domain of f is

$$\dim \text{dom}(f) = \sum_{i=1}^\ell \frac{d_i^2 - d_i}{2} + d_i = \dim \mathcal{V}(G).$$

The complement of the set of graphs $A \in \mathcal{V}(G)$ with simple spectrum is a union of sets of the form $f(E_{qr})$ where

$$E_{qr} = \{(U_i, \lambda_i)_{i=1}^\ell \mid \text{for } \bar{\lambda} = (\lambda_1, \dots, \lambda_\ell), \bar{\lambda}_q = \bar{\lambda}_r\}.$$

Since the dimension of each such set is strictly smaller than the dimension of the domain we obtain:

$$\dim f(E_{qr}) \leq^{(11)} \dim E_{qr} \leq \dim \text{dom}(f) - 1 = \dim \mathcal{V}(G) - 1$$

and so the complement of the set of graphs $A \in \mathcal{V}(G)$ with simple spectrum is dimension deficient and thus has zero Hausdorff measure. \square

We now prove the third condition holds generically.

Lemma 4. *If G has a full orbit I_r then almost every $A_{rj} \in \mathcal{V}_{rj}(G)$ has full rank.*

Proof. In this proof we denote members of the vector space $\mathcal{V}_{rj}(G)$ by \bar{A} . We identify this vector space with \mathbb{R}^ℓ for some ℓ via a linear isomorphism $\bar{A} : \mathbb{R}^\ell \rightarrow \mathcal{V}_{rj}(G)$, and define a multivariate polynomial $p : \mathbb{R}^\ell \rightarrow \mathbb{R}$ by

$$p(x) = \det(\bar{A}^T(x) \bar{A}(x))$$

Note that $p(x) = 0$ if and only if $\bar{A}(x)$ has full rank. Thus due to Lemma 2 it is sufficient to show that p isn't identically zero, or equivalently, establish the existence of a full rank matrix in $\mathcal{V}_{rj}(G)$. We now construct such a matrix which we will denote by \hat{A} .

Note that $\hat{A} \in \mathcal{V}_{rj}$ if and only if

$$\hat{A}_{sq} = \hat{A}_{\sigma(s)\sigma(q)}, \text{ for all } s \in I_r, q \in I_j, \sigma \in G. \quad (12)$$

This means that the values of \hat{A} are required to be constant along the orbits of the action of the group G on $I_r \times I_j$ defined by

$$(\sigma, (s, q)) \mapsto (\sigma(s), \sigma(q)).$$

We choose some orbit $[(s, q)]$ of this action and define \hat{A} by the requirement that $\hat{A}_{ij} = 1$ if (i, j) are member of this orbit, and otherwise $\hat{A}_{ij} = 0$. By construction $\hat{A} \in \mathcal{V}_{rj}(G)$ and it remains to verify that it has full rank. The orbit $[(s, q)]$ has $n_r = |G|$ elements $(s_1, q_1), \dots, (s_{n_r}, q_{n_r})$ where the s_i are all distinct, and we can order the orbit so that the first n_j elements of the sequence q_i are distinct as well. Thus

$$(\hat{A}_{s_i, q_\ell})_{i, \ell=1}^{n_j} = I_{n_j}$$

and so $\text{rank } \hat{A} = n_j$. □

4 Almost all or nothing

In this section we prove Theorem 2; we show that for any symmetry group G , either the DS relaxation is never convex exact for any $A \in \mathcal{A}(G)$, or the DS relaxation is convex exact for almost every $A \in \mathcal{A}(G)$. We then explain how generic convex exactness can be established/refuted for a given group G .

Our proof uses another notion of exactness which we will call *affine exactness*: The affine automorphisms of a graph A are the members of the affine set

$$\text{Aut}_{\text{aff}}(A) = \{S \mid S\mathbf{1} = \mathbf{1}, \mathbf{1}^T S = \mathbf{1}^T, S \in \mathcal{N}(\text{Aut}(A)), AS = SA\}.$$

We note that affine automorphisms and convex automorphisms differ in two aspects: On the one hand the entries of convex automorphisms are required to be non-negative while the entries of affine automorphisms are not. On the other hand, affine automorphisms must be members of $\mathcal{N}(\text{Aut}(A))$, a requirement we do not impose on convex automorphisms (although by Theorem 3 convex automorphisms will "usually" satisfy this property).

We say that affine exactness holds at A if

$$\text{Aut}_{\text{aff}}(A) = \text{affAut}(A).$$

We begin by establishing a connection between affine exactness and convex exactness:

Proposition 4.1. *For any graph $A \in \mathcal{A}(G)$, convex exactness holds at A if and only if affine exactness holds at A and*

$$\text{affAut}(A) \cap \{S \geq 0\} = \text{conv Aut}(A). \quad (13)$$

Note that the RHS of (13) is always contained in the LHS.

Proof. If affine exactness and (13) hold then

$$\text{conv Aut}(A) \subseteq \text{Aut}_{\text{conv}}(A) \subseteq \text{Aut}_{\text{aff}}(A) \cap \{S \geq 0\} = \text{affAut}(A) \cap \{S \geq 0\} = \text{conv Aut}(A).$$

so convex exactness holds as well.

Now assume convex exactness holds. Then (13) holds since

$$\text{conv Aut}(A) \subseteq \text{affAut}(A) \cap \{S \geq 0\} \subseteq \text{Aut}_{\text{conv}}(A) = \text{conv Aut}(A).$$

To show affine exactness holds we choose some $S \in \text{Aut}_{\text{aff}}(A)$ and show that $S \in \text{affAut}(A)$. The "centroid" convex automorphism S_c is non-zero in all coordinates except for coordinates (i, j) which satisfy $P_{ij} = 0$ for all automorphisms P . Since at such coordinates S is also zero, there is some $\epsilon > 0$ such that

$$S_1 = (1 - \epsilon)S_c + \epsilon S$$

is doubly stochastic. S_1 is also an affine automorphism, and thus is a convex automorphism. By assumption S_1 is a convex combination of members of $\text{Aut}(A)$. In particular $S_1, S_c \in \text{affAut}(A)$ and therefore so is

$$S = \frac{1}{\epsilon}S_c + (1 - \frac{1}{\epsilon})S_1.$$

□

Note that the condition (13) depends only on G and not on a specific choice of $A \in \mathcal{A}(G)$. Thus if the condition does not hold then $\text{DS}(A)$ will not be convex exact for any $A \in \mathcal{A}(G)$. If G is such that the condition does hold then convex exactness for specific $A \in \mathcal{A}(G)$ is equivalent to affine exactness. Thus Theorem 2 follows from the following proposition:

Proposition 4.2. *Let G be a symmetry group. Then either affine recovery holds for almost every $A \in \mathcal{A}(G)$, Or affine recovery fails for all $A \in \mathcal{A}(G)$.*

Proof of Proposition 4.2. As in the proof of Lemma 4 we show the set of $A \in \mathcal{A}(G)$ for which affine recovery fails is a null set of a suitable multivariate polynomial.

For given $A \in \mathcal{A}(G)$, the affine automorphisms of A are the matrices S satisfying the affine equations defining $\text{Aut}_{\text{aff}}(A)$. Denoting the map which identifies $n \times n$ matrices S with $n^2 \times 1$ vectors by $\text{vec}(\cdot)$, these equations can be written in the form

$$F(A)\text{vec}(S) = b \tag{14}$$

where $F(A)$ depends linearly on A . Since $A \in \mathcal{A}(G)$, all members of $\text{vec}(G)$ are solutions of (14). Thus the kernel of $F(A)$ always includes

$$W = \text{span}(\text{vec}(G - I_n))$$

and affine exactness holds iff $\text{Ker}(F(A)) = W$. Let

$$U = [U_0, U_1]$$

be a unitary matrix, such that U_0 forms an orthonormal basis of W . Then affine exactness holds iff $F(A)U_1$ has full rank. Now pick some linear isometry $x \mapsto \bar{A}(x)$ from \mathbb{R}^ℓ to the vector space $\mathcal{V}(G)$. affine exactness holds at $\bar{A}(x)$ iff x is not a zero of the multivariate polynomial

$$p(x) = \det(U_1^T F^T(\bar{A}(x)) F(\bar{A}(x)U_1)).$$

This concludes the proof of the proposition, due to Lemma 2. □

Checking exactness for given groups We now explain how we check whether convex exactness holds generically for the groups $G_i, i = 1, \dots, 9$ defined by the shapes in Figure 2. We first note that condition (13) holds for these groups. This is because G_i all contain a full orbit. If G has a full orbit $[\mathbf{a}_k]$ and S is an affine combination of members of G , then the coefficients of the affine combination are just the values of the column S_k . In particular if S is doubly stochastic then the affine combination is in fact a convex combination since S_k is a probability vector.

Since (13) holds we have generic convex exactness for G_i if and only if the polynomial $p = p(G_i)$ is a non-zero polynomial. This can be checked either by computing the polynomial

symbolically or by evaluating it on random input. We used the latter method: For each of the groups G_i we generated 100 random graphs in $\mathcal{V}(G_i)$ and evaluated the polynomial on these graphs. For groups G_1, \dots, G_3 all 100 evaluations of the polynomial were zero, and for the remaining graphs the polynomial was found to be non-zero at all 100 evaluated points. These results are summarized in the first column of Table 1.

5 Centroid

Recall that the centroid solution S_c is the matrix obtained by averaging over all members of G as defined in 6. In this section we show that for any symmetry group G and almost every $A \in \mathcal{A}(G)$ the centroid solution S_c can be recovered efficiently. We also show that S_c is the solution which will be obtained by interior-point methods when solving $\text{DS}(A)$.

We begin by giving an explicit construction of S_c . Let us denote as before the equivalence classes of the action of G on \mathbf{a} by I_1, \dots, I_k . Assume that the vertices are arranged so that

$$I_1 = \{\mathbf{a}_1, \dots, \mathbf{a}_{m_1}\}, \quad I_2 = \{\mathbf{a}_{m_1+1}, \dots, \mathbf{a}_{m_2}\}, \quad \dots, \quad I_k = \{\mathbf{a}_{m_{k-1}+1}, \dots, \mathbf{a}_{m_k}\}.$$

Set $n_j = |I_j|, j = 1, \dots, k$, and for any integer p let $J_p \in \mathbb{R}^{p \times p}$ be the constant matrix whose entries are all p^{-1} . Note that S_c is in $\mathcal{N}(G)$ and is invariant under multiplication from the left or right by elements of G . The only doubly stochastic matrix satisfying these properties is

$$S_0 = \begin{bmatrix} J_{n_1} & & & \\ & J_{n_2} & & \\ & & \ddots & \\ & & & J_{n_k} \end{bmatrix}$$

and therefore $S_c = S_0$. Recall that for any symmetry group G and almost every $A \in \mathcal{A}(G)$ the vector $s(A)$ is discriminative. In this case the centroid solution $S_c = S_0$ can be easily computed without even solving the DS relaxation: We first construct a matrix S by setting $S_{ij} = 1$ if $s_i(A) = s_j(A)$ and $S_{ij} = 0$ otherwise. We can then obtain S_c by normalizing the rows of S .

Interior point algorithms Interior point algorithms solve (2) by solving problems of the form

$$\min_{S \in \text{DS}} E_\alpha \|AS - SA\|_F^2 + \alpha F(S) \tag{15}$$

and taking $\alpha \rightarrow 0$ to obtain a solution for (2). The function F is chosen so that it explodes at the boundary, and so the constraints $S \geq 0$ will never be active in (15). A common choice [Wright and Nocedal, 1999] for F is $F(S) = -\sum_{ij} \log(S_{ij})$. Specialized solvers for (2) such as [Rangarajan et al., 1996, Solomon et al., 2016] often use $F(S) = \sum_{ij} S_{ij} \log S_{ij}$. Note that while this F does not explode at the boundary, its derivatives do.

To include both choices of F , as well as other possible choices, we will deal with general F which are of the form

$$F(S) = \sum_{ij} f(S_{ij})$$

where $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R} \cup \{\infty\}$ is continuous and strictly convex and $f(t) < \infty$ if $t > 0$.

Theorem 4. *Let G be a symmetry group, and F be a function satisfying the conditions described previously. Then for almost every $A \in \mathcal{A}(G)$ the unique minimizers S_α^* of (15) converge to S_c as α tends to zero.*

Proof. We assume that $\text{DS}(A)$ is weakly exact. This assumption holds for almost every $A \in \mathcal{A}(G)$. By passing to a subsequence we can assume that S_α^* converges to some S^* in the compact set DS . We need to show that $S^* = S_c$.

We note that for any $S \in \text{DS}$, $P \in G$, $\alpha > 0$

$$E_\alpha(SP) = E_\alpha(S) = E_\alpha(PS).$$

Since S_α^* is the unique minimizer of E_α this equality implies that S_α^* is invariant under multiplication by elements of G from the right and the left. Thus this is true for S^* as well. Due to continuity F is bounded from below and so it can be shown that

$$\|AS^* - S^*A\|_F^2 = \lim_{\alpha \rightarrow 0} \|AS_\alpha^* - S_\alpha^*A\|_F^2 = 0.$$

It follows that S^* is a convex automorphism and since $\text{DS}(A)$ is weakly exact $S^* \in \mathcal{N}(G)$. Since we also showed S^* to be invariant under multiplication by G from the left and right it follows that $S^* = S_c$.

□

6 Retrieving isomorphisms

In this section we discuss how convex exactness can be used to retrieve isomorphisms. We will discuss two classes of methods. The first class searches for extreme points of the convex set of convex isomorphisms. We will show that under the assumptions of Proposition 3.1 all isomorphisms of the graphs can be retrieved quite efficiently. However finding extreme points is not a stable method for retrieving isomorphisms once noise is introduced. This leads to the second class of methods, which we call projection methods. Projection methods are the methods typically used in practice to achieve a permutation solution from the original solution of the DS relaxation. We show theoretically that the popular "convex to concave" projection method is able to retrieve a correct isomorphism, and explore experimentally the behavior of this method as well as the L_2 projection method when noise is introduced.

6.1 Finding extreme points

When convex exactness holds, finding an isomorphism is reduced to the problem of finding an extreme point of the optimal set $\text{ISO}_{\text{conv}}(A, B)$ defined by the linear constraints

$$\begin{aligned} AS &= SB \\ S &\in \text{DS}. \end{aligned}$$

An extreme point (=basic feasible solution) of this linear feasibility program can be found using the simplex algorithm. Extreme points can also be found using interior point algorithms by optimizing a random linear energy over $\text{ISO}_{\text{conv}}(A, B)$. In [Dym and Lipman, 2016] a similar problem is discussed, and it is shown that if the linear energies are randomly drawn from the uniform distribution on S^{n^2-1} , then with probability one the obtained linear program will have a unique solution, which will be an extreme point. Moreover all extreme points will be obtained with equal probability.

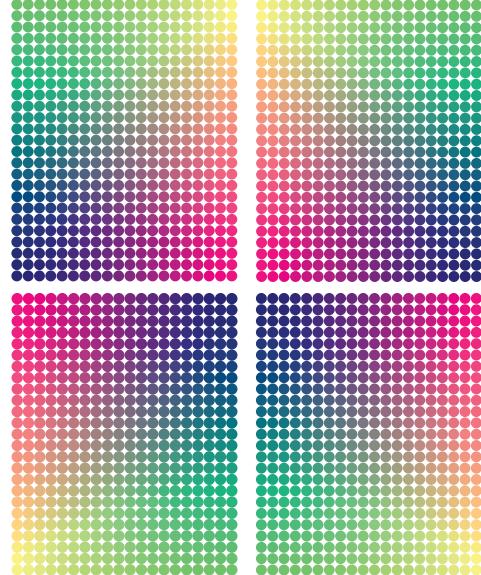


Figure 4: Efficient retrieval of all symmetries of a grid.

Table 1 shows the successfulness of the latter method in returning isomorphisms for symmetric problems in which convex exactness holds. For each of the nine symmetry groups G_i defined by the shapes in Figure 2 we generated 100 random graphs in $\mathcal{V}(G_i)$ according to the distribution μ_{G_i} . For each such graph we then found an extreme point by maximizing a random linear energy over the optimal set. As shown in Table 1 for the graphs $G_i, i > 3$ for which convex exactness holds generically, this algorithm succeeded in returning a permutation in all 100 experiments. For the groups $G_i, i = 1, 2, 3$ for which convex exactness does not hold, this algorithm returned permutations in more than half of the experiments, but non-integer solutions were also obtained. This is due to the fact that the optimal set contains non-integer extreme points in this case.

Next we suggest a more efficient method for obtaining all extreme points of the set of convex isomorphisms, under the assumption that the assumptions of Proposition 3.1 hold and $k = |\text{ISO}(A)|$ is not too large $k \ll n$.

If $s(A)$ is discriminative, then the centroid solution S_c can be found directly as described in the previous section.

Once a convex isomorphism $S = S_c$ was found, we use the technique of [Pataki, 1996] to find an extreme point. We now describe this technique:

We begin with some preliminaries: For $S, T \in \mathbb{R}^{n \times n}$, we say that $S \preceq T$ if $S_{ij} = 0$ whenever $T_{ij} = 0$. We say that $S \prec T$ if $S \preceq T$ but the converse inequality $T \preceq S$ does not hold.

A *face* of a convex set K is a subset $F \subseteq K$ such that for all $x, y \in F$ and $t \in (0, 1)$ satisfying

$$(1 - t)x + ty \in F$$

necessarily $x, y \in F$. An *extreme point* is a face which is a singleton. If K is a convex compact set then it is the convex hull of its extreme points E . Moreover, for each face $F \subseteq K$,

$$F = \text{conv}(E \cap F). \quad (16)$$

Every $S \in \text{ISO}_{\text{conv}}(A, B)$ defines a face

$$F(S) = \{Q \in \text{ISO}_{\text{conv}}(A, B) \mid Q \preceq S\}$$

and an affine space obtained from $F(S)$ by removing the positivity constraints, i.e.,

$$V(S) = \{R \mid R\mathbf{1} = \mathbf{1}, R^T\mathbf{1} = \mathbf{1}, AR = RB, R \preceq S\}.$$

We note that S is in the relative interior of $F(S) \subseteq V(S)$. This means that for all $R \in V(S)$ there is a sufficiently small $t > 0$ such that $(1 - t)S + tR \in F(S)$. The boundary of $F(S)$ in $V(S)$ is the set:

$$\partial F(S) = \{Q \in F(S) \mid Q \prec S\}.$$

We can now describe the algorithm of [Pataki, 1996]:

1. We are given as input some $S \in \text{ISO}_{\text{conv}}(A, B)$ and set $r = 0$ and $S_r = S$.
2. We compute a spanning subset to the affine space $V(S_r)$. If $V(S_r) = \{S_r\}$ then S_r is an extreme point and we are done.
3. Otherwise we choose some $R_r \neq S_r$ in $V(S_r)$. We then find the unique $t > 0$ such that

$$(1 - t)S_r + tR_r \in \partial F(S)$$

and set S_{r+1} to be the matrix on the left hand side. We then return to the previous step.

	exact?	% permutations
(a)	no	0.64
(b)	no	0.76
(c)	no	0.67
(d)	yes	1
(e)	yes	1
(f)	yes	1
(g)	yes	1
(h)	yes	1
(i)	yes	1

Table 1

The iterative process can only terminate when $V(S_r) = \{S_r\}$. This will necessarily occur after a finite number of steps since S_{r+1} always has more zeros than S_r . In the convex exact case, a permutation will be attained within $k = |\text{ISO}(A, B)|$ steps. This is because each face $F(S_{r+1})$ is strictly contained in the former face $F(S_r)$ and therefore according to (16) the number of extreme points=permutations in $F(S_{r+1})$ is strictly smaller than the number of extreme points in $F(S_r)$.

Once a permutation $P(1)$ is obtained, an additional permutation can be sought for by repeating the process above, but beginning with $S_0^1 = (1-t)S_0 + tP(1)$ where $t < 0$ is the smallest possible so that $S_0(t)$ is doubly stochastic. This choice gives an initial convex isomorphism such that $P(1) \notin F(S_0)$, guaranteeing that the algorithm will return a new permutation $P(2)$. In the next step we can set $S_0^2 = (1-t)S_0^1 + tP(2)$ and continue in this manner until we obtain a collection of isomorphisms $P(1), \dots, P(L)$, and S_c is a convex combination of these isomorphisms. In fact under the full orbit assumption $P(1), \dots, P(L)$ will be all the isomorphisms. This is because S_c can be written as a positive convex combination of all members of $\text{ISO}(A, B)$, and the members of $\text{ISO}(A, B)$ are linearly independent, implying that this is the only possible convex combination giving S , so that all isomorphisms were obtained. The linear independence of $\text{ISO}(A, B)$ follows from the fact that it has full orbit, and so each isomorphism has a non-zero coordinate i, j on which all other isomorphisms vanish.

From a computational perspective, under the conditions of Proposition 3.1, The algorithm above will return an isomorphism within k steps, and all isomorphisms within $O(k^2)$ iterations. Computing the first affine space $V(S_c)$ is basically the problem of finding a linear basis to the solution set of the linear equations defining $V(S_c)$. Since S_c has at most nk non-zero entries, this is a linear equation in $O(n)$ variables instead of n^2 variables. For finding the subsequent affine spaces $V(S_r)$ additional computational saving can be obtained due to the fact that $V(S_r) \subseteq V(S_c)$. Thus all elements in $V(S_r)$ are affine combinations of $k+1$ spanning element of the affine space $V(S_c)$, so $V(S_r)$ is obtained by solving a linear equation in only $k+1$ variables.

Figure 4 shows the results of applying the algorithm described above to find the symmetries of a 20×25 grid. The grid has a reflective symmetry group with full orbit and thus fulfills the conditions of Theorem 1. We took $A = B$ to be the Euclidean distance matrix of the grid (here $n = 500$) and used the algorithm described above to obtain all symmetries of the grid. In our implementation in Matlab this calculation took around ten seconds.

6.2 Projection methods

The classical approach [Aflalo et al., 2015] for projecting a permutation solution from the doubly stochastic relaxation is using the standard L_2 projection, which can be implemented as a linear program and solved efficiently using the Hungarian algorithm. See [Zaslavskiy et al., 2009] for more details. A more accurate and more computationally demanding method is the "convex to concave" method. We will explain this method in the formulation used in the DS++ algorithm [Dym et al., 2017]. Similar suggestions appear in [Zaslavskiy et al., 2009, Ogier and Beyer, 1990]. We then prove DS++ obtains a permutation solution in the convex exact case (up to some technicalities which will be explained), and examine the behavior of both projection methods when noise is added.

Convex to concave projection The convex to concave method sequentially solves optimization problems of the form

$$\min_{S \in DS} E(S, a) = \|AS - SB\|_F^2 + a(n - \|S\|_F^2), \quad a \geq 0. \quad (17)$$

The strictly concave function

$$g(S) = n - \|S\|_F^2$$

is non-negative on DS, and $g(S) = 0$ if and only if S is a permutation. Additionally if a is sufficiently large so that $E(S, a)$ is strictly concave, then the (global and local) minima of (17) will necessarily be permutations since the minima of a strictly concave function on a convex compact set are always extreme points. Thus the global minimum of the relaxed (17) and the original quadratic assignment problem are identical. Note however that since (17) is not convex computing the global minimum is no longer tractable.

Building on this observation, the convex to concave method minimizes (locally) a sequence of optimization problems of the form (17) on a sequence of choices

$$a_0 < a_1 < \dots < a_N$$

and in each step uses the obtained solution S_i as a warm start to the optimization of $E(S, a_{i+1})$. The first point a_0 is selected so that $E(S, a_0)$ is convex, and the last point is selected so that $E(S, a_N)$ is strictly concave and thus the obtained local minima S_N is guaranteed to be a permutation.

The first point a_0 can be selected to be zero to ensure that $E(S, a_0)$ is convex. However a better selection is $a_0 = \lambda_{\min}$ where $\lambda_{\min} \geq 0$ is the minimal eigenvalue of the quadratic form

$$S \mapsto \|AS - SB\|_F^2$$

when restricted to the subspace

$$\{S \mid S\mathbf{1} = 0, S^T\mathbf{1} = 0\}.$$

Similarly the last point a_N is selected to be (slightly larger than) the maximal eigenvalue λ_{\max} of the same quadratic form over the same subspace. This choice ensures that $E(S, a_N)$ is (strictly) concave. The remaining points a_i can be uniformly sampled in the interval $[a_0, a_N]$ (for lack of a better strategy).

Note that if A and B are isomorphic, then for any $a > 0$ the global minimizers of $E(S, a)$ are precisely $\text{ISO}(A, B)$ (while for $a = 0$ the global minimizers are $\text{ISO}_{\text{conv}}(A, B)$). This observation suggests the "convex to concave" method may be successful in retrieving isomorphisms even for symmetric problems, and possibly could return integer solutions S_i even for $i < N$. We now give a theoretical justification for these observations.

We assume that we obtain each S_i^* from a local minimization algorithm with the following properties:

1. Monotonicity: $E(S_i^*, a_i) \leq E(S_{i-1}^*, a_i)$.
2. The first-order necessary condition (KKT conditions) for local minima is satisfied at S_i^* .
3. The second-order necessary condition for local minima of $E(\cdot, a_i)$ is satisfied at S_i^* . That is

$$\text{vec}(S - T)^T H \text{vec}(S - T) \geq 0, \text{ for all } S, T \in F(S_i^*). \quad (18)$$

Here H is the Hessian of the quadratic form $E(\cdot, a_i)$.

Under these assumptions we prove

Theorem 5. *Assume A and B are isomorphic and the DS relaxation is convex exact at A . Assume $S_i^*, i = 0, \dots, N$ satisfy conditions (1)-(3). Then if a_1 is sufficiently close to a_0*

$$S_i^* \in \text{ISO}(A, B), \text{ for all } i \geq 1.$$

The theorem is proved in Appendix A.

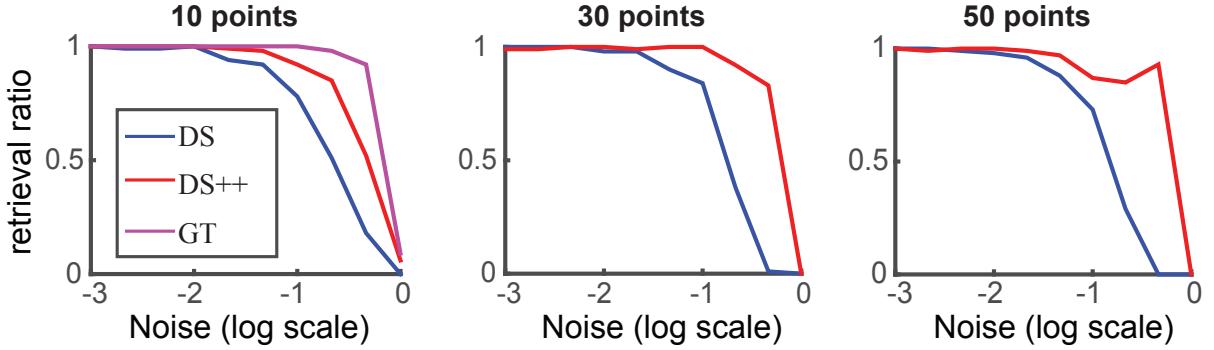


Figure 5: Evaluation of DS and DS++ accuracy as a function of noise for symmetric graphs. The details are explained in the text.

Isomorphism retrieval for noisy problems We examine the behavior of the DS relaxation coupled with the projections described above for noisy symmetric problems by conducting the following experiment:

We construct a random bilaterally symmetric graphs $A \in \mathbb{R}^{n \times n}$ and choose $B = A$. We then perturb these graphs by two randomly selected symmetric matrices $\Delta A, \Delta B$, and solve the DS relaxation using both projection methods. We do this for $n = 10, 30, 50$ and for matrices $\Delta A, \Delta B$ with Frobenius norm $\epsilon = 10^\alpha$ where we use ten values of α uniformly chosen from the interval $[-3, 0]$. The graph A is chosen by computing an isometry $L : \mathbb{R}^k \rightarrow \mathcal{V}(G)$ where G is a permutation subgroup with two elements, and then sampling a vector $x \in \mathbb{R}^k$ uniformly from the unit sphere to obtain $A = L(x)$. For each fixed value of n, α we repeat 100 different instances of the experiment, and compute the retrieval ratio of both methods, which we define as the number of times the method returned a permutation from G divided by the number of experiments (100). The results are shown in Figure 5.

It can be seen that both methods succeed in retrieving a correct permutation at low noise levels, but the convex-to-concave method (denoted by DS++) is more successful than the L_2 projection method (denoted by DS) at higher noise levels. In the case $n = 10$ we also add the "ground truth retrieval ratio", that is the number of instances in which the global minimizer of the graph matching energy was indeed in G divided by the number of experiments. It can be seen that as the noise level approaches $10^0 = 1$ the noise "takes over the problem" and the members of G are no longer the global minimizers. The ground truth solutions were obtained by the semi-definite relaxation of [Kezurer et al., 2015] which is known to be very tight, though computationally expensive. We verify that the solution obtained from the semi-definite relaxation is indeed the correct solution by checking that the difference between the lower bound provided by the relaxation and the upper bound provided by projecting the solution of the relaxation are negligible.

As a side note, we observe that at low noise levels DS++ obtains a solution in G after two iterations in accordance with Theorem 5, and that even at high noise levels a permutation solution is usually attained after four iterations. This indicates that it might be worthwhile to choose a smaller a_N , or alternatively to consider less steps in the convex-to-concave process.

References

- [Aflalo et al., 2015] Aflalo, Y., Bronstein, A., and Kimmel, R. (2015). On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*, 112(10):2942–2947.
- [Babai, 2016] Babai, L. (2016). Graph isomorphism in quasipolynomial time [extended abstract]. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*.

ing, pages 684–697. ACM.

[Babai et al., 1982] Babai, L., Grigoryev, D. Y., and Mount, D. M. (1982). Isomorphism of graphs with bounded eigenvalue multiplicity. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 310–324. ACM.

[Conte et al., 2004] Conte, D., Foggia, P., Sansone, C., and Vento, M. (2004). Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298.

[Dym and Lipman, 2016] Dym, N. and Lipman, Y. (2016). Exact recovery with symmetries for procrustes matching. *arXiv:1606.01548*.

[Dym et al., 2017] Dym, N., Maron, H., and Lipman, Y. (2017). Ds++: A flexible, scalable, and provably tight relaxation for matching problems. *arXiv:0902.0885*.

[Falconer, 2004] Falconer, K. (2004). *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons.

[Fiori and Sapiro, 2015] Fiori, M. and Sapiro, G. (2015). On spectral properties for graph matching and graph isomorphism problems. *Information and Inference*, 4(1):63–76.

[Giorgi et al., 2007] Giorgi, D., Biasotti, S., and Paraboschi, L. (2007). Shape retrieval contest 2007: Watertight models track. *SHREC competition*, 8.

[Kezurer et al., 2015] Kezurer, I., Kovalsky, S. Z., Basri, R., and Lipman, Y. (2015). Tight relaxation of quadratic matching. *Comput. Graph. Forum*, 34(5):115–128.

[Loiola et al., 2007] Loiola, E. M., de Abreu, N. M. M., Boaventura-Netto, P. O., Hahn, P., and Querido, T. (2007). A survey for the quadratic assignment problem. *European journal of operational research*, 176(2):657–690.

[Lyzinski et al., 2016] Lyzinski, V., Fishkind, D. E., Fiori, M., Vogelstein, J. T., Priebe, C. E., and Sapiro, G. (2016). Graph matching: Relax at your own risk. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):60–73.

[Mémoli, 2007] Mémoli, F. (2007). On the use of gromov-hausdorff distances for shape comparison.

[Mémoli, 2011] Mémoli, F. (2011). Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487.

[Ogier and Beyer, 1990] Ogier, R. G. and Beyer, D. (1990). Neural network solution to the link scheduling problem using convex relaxation. In *Global Telecommunications Conference, 1990, and Exhibition. ‘Communications: Connecting the Future’, GLOBECOM’90.*, IEEE, pages 1371–1376. IEEE.

[Pataki, 1996] Pataki, G. (1996). Cone-lp’s and semidefinite programs: Geometry and a simplex-type method. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 162–174. Springer.

[Rangarajan et al., 1996] Rangarajan, A., Gold, S., and Mjolsness, E. (1996). A novel optimizing network architecture with applications. *Neural Computation*, 8(5):1041–1060.

[Sahni and Gonzalez, 1976] Sahni, S. and Gonzalez, T. (1976). P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565.

- [Scheinerman and Ullman, 2011] Scheinerman, E. R. and Ullman, D. H. (2011). *Fractional graph theory: a rational approach to the theory of graphs*. Courier Corporation.
- [Solomon et al., 2012] Solomon, J., Nguyen, A., Butscher, A., Ben-Chen, M., and Guibas, L. (2012). Soft maps between surfaces. In *Computer Graphics Forum*, volume 31, pages 1617–1626. Wiley Online Library.
- [Solomon et al., 2016] Solomon, J., Peyré, G., Kim, V. G., and Sra, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans. Graph.*, 35(4):72:1–72:13.
- [Wright and Nocedal, 1999] Wright, S. and Nocedal, J. (1999). Numerical optimization. *Springer Science*, 35:67–68.
- [Zaslavskiy et al., 2009] Zaslavskiy, M., Bach, F., and Vert, J.-P. (2009). A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2227–2242.

A Convex to concave

Proof of Theorem 5. We begin with some preliminaries. First note that if S_{i-1}^* is an isomorphism, then $E(S_i^*, a_i) = 0$ due to the monotonicity condition and thus S_i^* is an isomorphism.

In the asymmetric case the claim is trivial: Since $E(S, a_0)$ is convex its local minimizers are also global minimizers. Since in the asymmetric case is the unique isomorphism between A and B is the only global minimizer for any $a_0 \geq 0$ it follows that S_0^* is that unique minimizer. Therefore in this proof we will focus on the symmetric case only.

In the symmetric case there are at least two isomorphisms P_0, P_1 . Thus $P_1 - P_0$ is an eigenvector of the energy $E(S)$ with eigenvalue $\lambda_{\min} = 0$ and so $a_0 = 0$.

Our claim follows easily from the following lemma:

Lemma 5. *There exists an open set U containing $\text{ISO}_{\text{conv}}(A, B)$ such that for all $a > 0$, The only points satisfying the first and second order conditions for local minimization of $E(S, a)$ are the members of $\text{ISO}(A, B)$.*

To obtain the theorem from the lemma, let $m > 0$ be the minimum of $E(S)$ on the compact set $DS \setminus U$. For any $a_1 > 0$ sufficiently small so that $a_1 \max_{S \in DS} g(S) < m$ we obtain

$$E(S_1^*, a_1) \leq E(S_0^*, a_1) = a_1 g(S_0^*) < m \leq \min_{S \in DS \setminus U} E(S, a_1).$$

It follows that $S_1^* \in U$ and since it satisfies the first and second order conditions for local minimization of $E(\cdot, a_1)$ it follows that $S_1^* \in \text{ISO}(A, B)$.

Proof of Lemma 5. We construct for each $S \in \text{ISO}_{\text{conv}}(A, B)$ an open set U_S satisfying the properties required from U and then choose

$$U = \bigcup_{S \in \text{ISO}_{\text{conv}}(A, B)} U_S.$$

If S is a convex isomorphism but not a permutation we choose

$$U_S = DS \cap \{Q \mid Q_{ij} > 0 \text{ if } S_{ij} > 0\} = \{Q \mid S \in F(Q)\}.$$

Fix some $Q \in U_S$, we claim that the second-order necessary condition for minimizing $E(\cdot, a)$ is not satisfied at Q for any $a > 0$. Since S is a convex combination of isomorphisms we can choose an isomorphism P such that $S \succeq P$, and so $P, S \in F(Q)$. Since P, S are both zeros of

the convex quadratic form $E(\cdot, 0)$, it follows that the second-order condition does not hold since (denoting by H_g the Hessian of g)

$$\text{vec}(P - S)^T H \text{vec}(P - S) = a \text{ vec}(P - S)^T H_g \text{vec}(P - S) = -2a \|P - S\|_F^2 < 0.$$

For isomorphisms P we choose U_P as follows:

For any $S \in DS \setminus \Pi_n$ the concavity of g implies

$$\nabla_P^T g(S - P) \geq g(S) - g(P) > 0.$$

In particular this is true for any S in the compact set

$$K = \{S \in DS \mid \sup_{i,j} |P_{ij} - S_{ij}| = 0.5\}.$$

Since g is C^1 the function

$$F(Q, S) = \nabla_Q^T g(S - Q)$$

is continuous. Thus, there is a neighborhood $U_P \subseteq \{S \in DS \mid \sup_{i,j} |P_{ij} - S_{ij}| \leq \frac{1}{4}\}$ of P on which

$$\nabla_Q^T g(S - Q) > 0, \quad \forall (Q, S) \in U_P \times K. \quad (19)$$

Fix some $Q \in U_P$. Define

$$t_M = \sup\{t \geq 0 \mid Q(t) = (1-t)P + tQ \in DS\}.$$

Note that $\sup_{i,j} |P_{ij} - Q_{ij}(t_M)| = 1$ and therefore there is some $1 < t_0 < t_M$ such that $S = Q(t_0) \in K$. It follows from (19) that

$$0 < \nabla_Q^T g(Q(t_0) - Q) = (1 - t_0) \nabla_Q^T g(P - Q)$$

and therefore

$$\nabla_Q^T g(P - Q) < 0$$

The convexity of E implies that for all $Q \in U_P$,

$$\nabla_Q^T E(P - Q) \leq E(P) - E(Q) \leq 0.$$

From the last two equations it follows that for any $a > 0$ the energy $E(\cdot, a)$ has a descent direction $P - Q$ at any point $Q \in U_P \setminus \{P\}$. Note that this direction is orthogonal to the gradients of the constraints defining DS , since $Q + t(P - Q)$ is feasible if $|t|$ is small enough. Thus the first-order condition does not hold at Q . \square