Entropic Metric Alignment for Correspondence Problems

Justin Solomon* MIT
Gabriel Peyré CNRS & Univ. Paris-Dauphine
Vladimir G. Kim Adobe Research
Suvrit Sra MIT

Abstract

Many shape and image processing tools rely on computation of correspondences between geometric domains. Efficient methods that stably extract "soft" matches in the presence of diverse geometric structures have proven to be valuable for shape retrieval and transfer of labels or semantic information. With these applications in mind, we present an algorithm for probabilistic correspondence that optimizes an entropy-regularized Gromov-Wasserstein (GW) objective. Built upon recent developments in numerical optimal transportation, our algorithm is compact, provably convergent, and applicable to any geometric domain expressible as a metric measure matrix. We provide comprehensive experiments illustrating the convergence and applicability of our algorithm to a variety of graphics tasks. Furthermore, we expand entropic GW correspondence to a framework for other matching problems, incorporating partial distance matrices, user guidance, shape exploration, symmetry detection, and joint analysis of more than two domains. These applications expand the scope of entropic GW correspondence to major shape analysis problems and are stable to distortion and noise.

Keywords: Gromov-Wasserstein, matching, entropy

Concepts: •**Computing methodologies** → **Shape analysis;**

1 Introduction

A basic component of the geometry processing toolbox is a tool for *mapping* or *correspondence*, the problem of finding which points on a target domain correspond to points on a source. Many variations of this problem have been considered in the graphics literature, e.g. with some sparse correspondences provided by the user. Regardless, the basic task of geometric correspondence facilitates the transfer of properties and edits from one shape to another.

The primary factor that distinguishes correspondence algorithms is the choice of objective functions. Different choices of objective functions express contrasting notions of which correspondences are "desirable." Classical theorems from differential geometry and most modern algorithms consider *local* distortion, producing maps that take tangent planes to tangent planes with as little stretch as possible; slightly larger neighborhoods might be taken into account by e.g. aligning heat kernels. These approaches are justified by classical differential geometry when the matched domains satisfy conditions like near-isometry or near-conformality, but when these conditions are violated these algorithms suffer from having to patch together local elastic terms into a single global map.

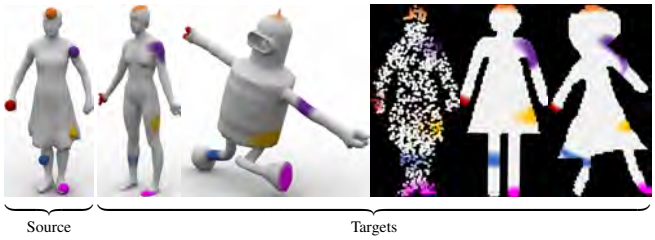


Figure 1: *Entropic GW can find correspondences between a source surface (left) and a surface with similar structure, a surface with shared semantic structure, a noisy 3D point cloud, an icon, and a hand drawing. Each fuzzy map was computed using the same code.*

In this paper, we propose a new correspondence algorithm that minimizes distortion of long- and short-range distances alike. We study an entropically-regularized version of the *Gromov-Wasserstein* (GW) mapping objective function from [Mémoli 2011] measuring the distortion of geodesic distances. The optimizer is a probabilistic matching expressed as a "fuzzy" correspondence matrix in the style of [Kim et al. 2012; Solomon et al. 2012]; we control sharpness of the correspondence via the weight of an entropic regularizer.

Although [Mémoli 2011] and subsequent work identified the possibility of using GW distances for geometric correspondence, computational challenges hampered their practical application. To overcome these challenges, we build upon recent methods for regularized optimal transportation introduced in [Benamou et al. 2015; Solomon et al. 2015]. While optimal transportation is a fundamentally different optimization problem from regularized GW computation (linear versus quadratic matching), the core of our method relies upon solving a sequence of regularized optimal transport problems.

Our remarkably compact algorithm (see Algorithm 1) exhibits global convergence, i.e., it *provably* reaches a local minimum of the regularized GW objective function regardless of the initial guess. Our algorithm can be applied to any domain expressible as a metric measure space (see §2). Concretely, only distance matrices are required as input, and hence the method can be applied to many classes of domains including meshes, point clouds, graphs, and even more abstract structures.

A major advantage of our framework is its extensibility. In addition to the conventional correspondence problem, we apply our method to organizing shape collections and show how to find correspondences given user guidance or incomplete pairwise distances. We also provide algorithms to extract multiple maps in the presence of symmetry and to compute consistent maps within a collection.

Contributions. We present a fuzzy mapping algorithm minimizing the Gromov-Wasserstein (GW) objective with entropic regularization. In summary, our key contributions are the following:

- discretization of an entropically-regularized GW objective suitable for domains in graphics and geometry processing;
- a simple-to-implement algorithm for minimizing this objective that relies only upon scalable low-level linear algebra;

*E-mail: jsolomon@mit.edu

SIGGRAPH '16 Technical Paper, July 24 - 28, 2016, Anaheim, CA
ISBN: 978-1-4503-4279-7/16/07
DOI: http://dx.doi.org/10.1145/2897824.2925903

- a convergence proof for the iterative optimization algorithm;
- comprehensive experiments establishing reliability, efficiency, and versatility of the regularized GW algorithm; and
- proof-of-concept extensions of regularized GW correspondence to a large variety of problems, demonstrating its versatility within the geometry processing and graphics toolboxes.

2 Related Work

Our algorithm is a general technique for mapping, a problem that has a long history in graphics. In this section, we focus on works relevant to our discussion; see [van Kaick et al. 2011] for a thorough survey of correspondence algorithms.

Gromov-Wasserstein (GW) distances. We focus on the GW mapping objective [Mémoli 2007], a relaxation of the Gromov-Hausdorff (GH) distance between metric spaces [Gromov 2001]. GH measures the single distance most distorted by a map. While it has been applied to geometry processing [Bronstein et al. 2010], GH is costly to optimize and prioritizes one distance value at a time.

GW is a distance between *metric measure spaces*, i.e., metric spaces equipped with a probability distribution (see [Mémoli 2011; Mémoli 2014; Sturm 2012] for more details). This additional feature is crucial, especially because it allows one to measure the *expected* distortion of distances. Matchings considered by GW distances thus differ from the point-to-point maps considered by Gromov-Hausdorff distances. GW distances optimize over probabilistic "measure couplings" (see §3), the continuous counterpart of "fuzzy" mapping matrices introduced in [Kim et al. 2012; Solomon et al. 2012]. Using an additional probability distribution on a metric space is also beneficial from an application point of view, because one can encode additional information (such as spatially varying confidence or noise level) in the probability distribution; see §5.3 for an example.

Mapping objectives. Most mathematical methods for correspondence are built around criteria for a desirable map, expressed in algorithmic design choices or as terms in an objective function. Here, we attempt to place GW distances in this larger context.

Many mapping methods attempt to minimize *local* distortion, measuring the stretch of the source onto the target. In differential geometry, harmonic maps minimize a local measure of stretch and shear [Urakawa 2013]. Heat kernel-based methods like [Ovsjanikov et al. 2010] prefer maps that distort local distances and curvatures; Kezurer et al. [2015] use a similar objective. "Kernelized sorting" [Quadrianto et al. 2009] optimizes the Hilbert Schmidt Independence Criterion [Smola et al. 2007], which coincides with the GW objective function after replacing geodesic distances with diffusion kernels. Fried et al. [2015] apply this machinery to image layout problems. These methods are subject to the challenge of assembling many local constraints into one map.

Other mapping methods seek to preserve global structure. Bronstein et al. [2006] showed that one such objective is effective for nonrigid mesh registration, optimized using simple gradient-based steps; Aflalo and Kimmel [2013] introduce a spectral approximation in the presence of a Laplacian operator for machine learning applications. These methods can make it difficult to localize points in small neighborhoods but are well-suited to fuzzy mapping applications. Our regularized GW mapping technique belongs to this class but is accompanied by a more stable optimization routine than its peers.

If more is known about the mapped geometry, specialized algorithms may apply. For instance, Kim et al. [2011] seek nearly conformal maps between surfaces and We et al. [2015] train for maps between human body models. Chen and Koltun [2015] rely on extrinsic alignment to reduce matching ambiguity for triangulated surfaces. We aim to devise a versatile method that does not depend on these assumptions, making it applicable to many domains (see §4).

Our measure couplings bear some rough similarity to the functional map matrices of Ovsjanikov et al. [2012], but the underlying variables are different. We minimize geometric distortion, while they use weaker "commutativity" regularizers and descriptors. Functional map matrices in the Laplace-Beltrami basis also can take on negative values, which are explicitly avoided in our probabilistic formulation.

Optimal transportation. As we will see in §3, the GW matching objective function inherits part of its name from the Wasserstein distance between probability distributions on a geometric domain. These latter distances have recently been applied to several problems in graphics; a small sampling includes [Bonneel et al. 2011; De Goes et al. 2011; Mérigot 2011; de Goes et al. 2012; Schwartzburg et al. 2014; Solomon et al. 2014; de Goes et al. 2015; Solomon et al. 2015]. Despite this relationship, the Wasserstein and GW distances serve contrasting purposes and require different computational machinery.

GW and Wasserstein distances apply to different problems. Wasserstein distances are between distributions on the same geometric domain. GW distances are between different geometric domains. From an optimization standpoint, Wasserstein distances are defined by a convex linear program [Villani 2003; Rubner et al. 2000], while GW distances require solving a nonconvex quadratic program.

Nonetheless, we leverage algorithms for Wasserstein distances as building blocks. We apply entropic regularization, proposed for optimal transportation in [Cuturi 2013; Benamou et al. 2015] and introduced to graphics in [Solomon et al. 2015]; the entropic Sinkhorn algorithm comprises our inner loop. Our algorithm resembles "soft-assign" in machine learning [Rangarajan et al. 1997] (see §3.3). The entropic term controls the fuzziness of our coupling; highly-regularized problems can generate meaningful rough maps in just a few iterations, while decreasing the regularizer better approximates the GW problem at the cost of more expensive optimization.

Quadratic assignment and correspondence. GW computation is an instance of the quadratic assignment problem [Pardalos and Wolkowicz 1994; Loiola et al. 2007; Çela 2013]. Solving this problem with global optimality and point-to-point constraints likely is algorithmically intractable—even within a constant approximation factor [Sahni and Gonzalez 1976]. After relaxing integer constraints the problem remains similarly intractable [Sahni 1974].

Optimization. Limited attention has been dedicated to the problem of computing GW distances efficiently. Initial work used general-purpose solvers. Mémoli [2009] introduces a spectral approximation bounded by solving a sequence of linear programs.

An alternative approach with theoretical guarantees is convex relaxation, removing constraints from nonconvex problems until they become convex programs. Some theoretical results provide conditions under which these methods recover a global optimum for the original problem, that is, when the relaxations are *tight*. While the potential for global optimality is attractive, the number of relaxed variables can be huge. For instance, Kezurer et al. [2015] optimize over $n^2 \times n^2$ semidefinite matrices, where n is the number of mapped points, and [Solomon et al. 2012; Solomon et al. 2013] include many large-scale transportation subproblems. Furthermore, relaxed problems with non-unique solutions, e.g. mapping in the presence of symmetries, admit large spaces of unusable outputs. See [Aflalo et al. 2014; Aflalo et al. 2015; Lyzinski et al. 2015] for similar trade-offs in graph matching. Eigenvalue relaxations, e.g. [Leordeanu and Hebert 2005], also provide some notion of

global optimality, with much less favorable conditions for tightness and after removing many constraints.

Our algorithm instead inherits convergence and *local* optimality from recent work on nonconvex optimization [Boţ et al. 2015]. It resembles a discretized gradient flow in the Kullback-Leibler (KL) metric, similar to the forward–backward proximal gradient algorithm [Bauschke and Combettes 2011; Combettes and Pesquet 2011] extended to Bregman divergences [Bauschke et al. 2006]. Guarantees for nonconvex objectives come from the Kurdyka-Łojasiewicz (also KL) property [Kurdyka 1998; Łojasiewicz 1963; Łojasiewicz 1993], applied to nonconvex optimization by Attouch et al. [2010].

3 Matching a Pair of Domains

3.1 Gromov-Wasserstein Matching

Suppose we are given two geometric domains Σ_0 and Σ accompanied with nonnegative unit measures μ_0 and μ, resp. Following [Mémoli 2011], define the space $\mathcal{M}(\mu_0,\mu)$ of *measure couplings* as the set of measures $\gamma \in \mathrm{Prob}(\Sigma_0 \times \Sigma)$ satisfying

$$\gamma(S_0 \times \Sigma) = \mu_0(S_0) \quad \text{and} \quad \gamma(\Sigma_0 \times S) = \mu(S)$$

for all measurable $S_0 \subseteq \Sigma_0$ and $S \subseteq \Sigma$. We will treat γ as a "soft correspondence" in the language of [Solomon et al. 2012], that is, a high probability assigned by γ to $(p_0, p) \in \Sigma_0 \times \Sigma$ indicates that p_0 and p should be matched.

To add geometric structure, assume Σ_0 and Σ are additionally accompanied with pairwise distance functions $d_0 : \Sigma_0 \times \Sigma_0 \to \mathbb{R}_+$ and $d : \Sigma \times \Sigma \to \mathbb{R}_+$. The *2-Gromov-Wasserstein* (GW) distance between (μ_0, d_0) and (μ, d) measures the minimal distortion induced by a measure coupling between the two domains:

$$\mathrm{GW}_2^2((\mu_0,d_0),(\mu,d)) := \min_{\gamma\in\mathcal{M}(\mu_0,\mu)} \iint_{\Sigma_0\times\Sigma} \left[d_0(x,x') - d(y,y')\right]^2 d\gamma(x,y)\, d\gamma(x',y'). \tag{1}$$

In contrast to optimal transportation, this optimization is a potentially nonconvex *quadratic program* rather than a linear program due to the product of two $d\gamma$'s.

The GW objective is constructed from the assumption that if a map pairs $x \mapsto y$ and $x' \mapsto y'$, then the distance between x and x' on Σ_0 should be similar to the distance between y and y' on Σ. If d is the geodesic distance function, then the objective is zero when γ encodes an isometry between Σ_0 and Σ. That said, γ is meaningful even when Σ_0 and Σ are not isometric, measuring the optimal deviation from preserving the distance structure of a surface.

While our subsequent derivation could be written in continuous language, for clarity and to focus on computational applications we transition to discrete notation. Assume Σ_0 is discretized using n_0 points and that Σ is discretized using n points; we accompany these samplings with discrete measures $\boldsymbol{\mu}_0 \in \mathbb{R}_+^{n_0}$ and $\boldsymbol{\mu} \in \mathbb{R}_+^{n}$ such that $\mathbf{1}^\top\boldsymbol{\mu}_0 = \mathbf{1}^\top\boldsymbol{\mu} = 1$. In our experiments, we take $\boldsymbol{\mu}_0, \boldsymbol{\mu}$ either to be constant vectors or—for triangle meshes—vectors of per-vertex barycentric area weights; future work might consider more general choices of these vectors. We use symmetric matrices $\mathbf{D}_0 \in \mathbb{R}_+^{n_0\times n_0}$ and $\mathbf{D} \in \mathbb{R}_+^{n\times n}$ to denote pairwise distances on Σ_0 and Σ.

In this language, the set of measure couplings is

$$\overline{\mathcal{M}}(\boldsymbol{\mu}_0,\boldsymbol{\mu}) := \{\boldsymbol{\Gamma} \in \mathbb{R}_+^{n_0\times n} : \boldsymbol{\Gamma}\boldsymbol{\mu} = \mathbf{1}, \boldsymbol{\Gamma}^\top\boldsymbol{\mu}_0 = \mathbf{1}\}, \tag{2}$$

and the 2-Gromov-Wasserstein distance is

$$\overline{\mathrm{GW}}_2^2(\mathbf{D}_0,\mathbf{D}) := \min_{\boldsymbol{\Gamma}\in\overline{\mathcal{M}}} \sum_{ijk\ell} (\mathbf{D}_{0ij} - \mathbf{D}_{k\ell})^2 \boldsymbol{\Gamma}_{ik}\boldsymbol{\Gamma}_{j\ell}\boldsymbol{\mu}_{0i}\boldsymbol{\mu}_{0j}\boldsymbol{\mu}_k\boldsymbol{\mu}_\ell. \tag{3}$$

We think of $\boldsymbol{\Gamma}$ as a function sampled from $\Sigma_0 \times \Sigma$, so the products $\boldsymbol{\Gamma}^\top\boldsymbol{\mu}_0$ and $\boldsymbol{\Gamma}\boldsymbol{\mu}$ integrate out Σ_0 and Σ, resp. The linear constraints defining $\overline{\mathcal{M}}$ reflect the fact that $\boldsymbol{\Gamma}$ should marginalize to the constant probability measure on Σ_0 and Σ.

Define an inner product of measure couplings as

$$\langle \boldsymbol{\Gamma}, \boldsymbol{\Gamma}' \rangle := \sum_{ik} \boldsymbol{\Gamma}_{ik}\boldsymbol{\Gamma}'_{ik}\boldsymbol{\mu}_{0i}\boldsymbol{\mu}_k.$$

After expanding the square, we can write

$$\overline{\mathrm{GW}}_2^2(\mathbf{D}_0,\mathbf{D}) = 2 \min_{\boldsymbol{\Gamma}\in\overline{\mathcal{M}}} \langle \boldsymbol{\Gamma}, \boldsymbol{\Lambda}(\boldsymbol{\Gamma})\rangle, \tag{4}$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\Gamma}) := \frac{1}{2}\mathbf{D}_0^{\wedge 2}[\![\boldsymbol{\mu}_0]\!]\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{1}\mathbf{1}^\top - \mathbf{D}_0[\![\boldsymbol{\mu}_0]\!]\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{D} + \frac{1}{2}\mathbf{1}\boldsymbol{\mu}_0^\top\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{D}^{\wedge 2}.$$

The superscript $\wedge 2$ denotes the elementwise square of a matrix, and $[\![\mathbf{v}]\!]$ is the diagonal matrix constructed from vector $\mathbf{v}$.

Applying the marginalization constraints (2) shows

$$\overline{\mathrm{GW}}_2^2(\cdot) = C(\mathbf{D}_0,\mathbf{D}) - 2 \max_{\boldsymbol{\Gamma}\in\overline{\mathcal{M}}} \langle \boldsymbol{\Gamma}, \mathbf{D}_0[\![\boldsymbol{\mu}_0]\!]\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{D}\rangle, \tag{5}$$

where

$$C(\mathbf{D}_0,\mathbf{D}) := \sum_{ij} \mathbf{D}_{0ij}^2\boldsymbol{\mu}_{0i}\boldsymbol{\mu}_{0j} + \sum_{k\ell} \mathbf{D}_{k\ell}^2\boldsymbol{\mu}_k\boldsymbol{\mu}_\ell.$$

This maximization problem for $\boldsymbol{\Gamma}$ is a nonconvex quadratic program for $\boldsymbol{\Gamma}$, removing the possibility of using convex optimization tools. Any time Σ_0 or Σ admit a symmetry, the objective has multiple optima, one for each symmetry. For example, when mapping human models, the two optima will be the orientation-preserving map and a left-to-right flipped map. Unlike convex relaxations, however, a convex combination of the orientation-preserving and orientation-reversing maps will *not* necessarily be optimal.

3.2 Entropic Regularization

Following [Solomon et al. 2015], we define the *entropy* of a measure coupling as

$$H(\boldsymbol{\Gamma}) := -\langle \boldsymbol{\Gamma}, \ln \boldsymbol{\Gamma}\rangle,$$

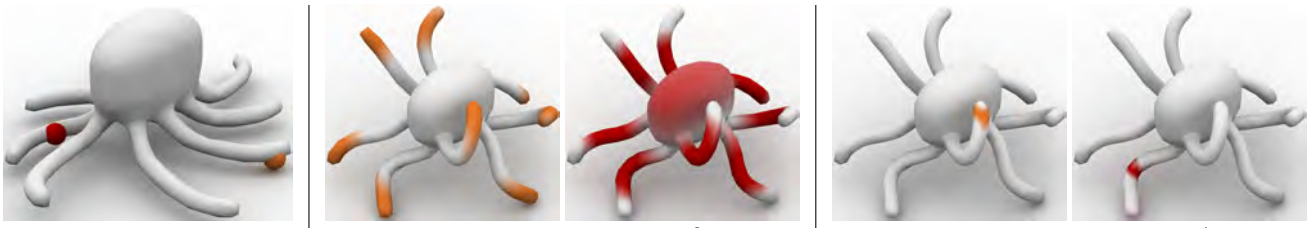
$H(\boldsymbol{\Gamma})$ is large when $\boldsymbol{\Gamma}$ has many nonzero entries and is small when $\boldsymbol{\Gamma}$ encodes a map close to a permutation matrix.

Suppose we regularize the objective (5) by adding $-\alpha H(\boldsymbol{\Gamma})$ for some $\alpha > 0$. As illustrated in Figure 2, larger values of α increase the softness of $\boldsymbol{\Gamma}$, effectively smoothing the base GW objective function. As $\alpha \to 0$, optimal matchings $\boldsymbol{\Gamma}$ begin to resemble permutation matrices (see [Quadrianto et al. 2009, eq. (10)]) and no longer superpose symmetric matches.

Now, define the *KL-divergence* between $\boldsymbol{\Gamma}, \overline{\boldsymbol{\Gamma}} \in \mathbb{R}_+^{n_0\times n}$ as

$$\mathrm{KL}(\boldsymbol{\Gamma}|\overline{\boldsymbol{\Gamma}}) := \sum_{ik} \boldsymbol{\mu}_{0i}\boldsymbol{\mu}_k \left[\left(\boldsymbol{\Gamma}_{ik} \ln \frac{\boldsymbol{\Gamma}_{ik}}{\overline{\boldsymbol{\Gamma}}_{ik}}\right) - \boldsymbol{\Gamma}_{ik} + \overline{\boldsymbol{\Gamma}}_{ik}\right].$$

This KL divergence contains an additional linear term from the one in [Solomon et al. 2015], but since it only involves $\overline{\boldsymbol{\Gamma}}$ it is a constant in their objective. This use of "generalized" KL (also called



Source surface | Target surface ($\alpha = 8 \times 10^{-3}$) | Target surface ($\alpha = 7 \times 10^{-4}$)

Figure 2: *Effect of regularization parameter α. Colored points (left) are mapped to colored distributions (right). As α decreases, the map sharpens; eventually, eight-way symmetry is broken in favor of a more bijective map matching tentacles based on asymmetries in their lengths.*

function GROMOV-WASSERSTEIN($\boldsymbol{\mu}_0, \mathbf{D}_0, \boldsymbol{\mu}, \mathbf{D}, \alpha, \eta$)
// Computes a local minimizer $\boldsymbol{\Gamma}$ of (6)
$\boldsymbol{\Gamma} \leftarrow$ ONES($n_0 \times n$)
for $i = 1, 2, 3, \ldots$
$\mathbf{K} \leftarrow \exp(\mathbf{D}_0[\![\boldsymbol{\mu}_0]\!]\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{D}^\top/\alpha)$
$\boldsymbol{\Gamma} \leftarrow$ SINKHORN-PROJECTION($\mathbf{K}^{\wedge\eta} \otimes \boldsymbol{\Gamma}^{\wedge(1-\eta)}; \boldsymbol{\mu}_0, \boldsymbol{\mu}$)
return $\boldsymbol{\Gamma}$

function SINKHORN-PROJECTION($\mathbf{K}; \boldsymbol{\mu}_0, \boldsymbol{\mu}$)
// Finds $\boldsymbol{\Gamma}$ minimizing $\mathrm{KL}(\boldsymbol{\Gamma}|\mathbf{K})$ *subject to* $\boldsymbol{\Gamma} \in \overline{\mathcal{M}}(\boldsymbol{\mu}_0, \boldsymbol{\mu})$
$\mathbf{v}, \mathbf{w} \leftarrow \mathbf{1}$
for $j = 1, 2, 3, \ldots$
$\mathbf{v} \leftarrow \mathbf{1} \oslash \mathbf{K}(\mathbf{w} \otimes \boldsymbol{\mu})$
$\mathbf{w} \leftarrow \mathbf{1} \oslash \mathbf{K}^\top(\mathbf{v} \otimes \boldsymbol{\mu}_0)$
return $[\![\mathbf{v}]\!]\mathbf{K}[\![\mathbf{w}]\!]$

Algorithm 1: *Iteration for finding regularized Gromov-Wasserstein distances. $\otimes, \oslash$ denote elementwise multiplication and division.*

I-divergence) will figure into our method for joint domain analysis (§5.5) and identifies KL divergence as a Bregman divergence [Bregman 1967].

Finally, define

$$\mathbf{f}_\alpha(\boldsymbol{\Gamma}) := \exp(\mathbf{D}_0[\![\boldsymbol{\mu}_0]\!]\boldsymbol{\Gamma}[\![\boldsymbol{\mu}]\!]\mathbf{D}/\alpha).$$

After regularizing by adding $-\alpha H(\boldsymbol{\Gamma})$ to the objective (5), an identical argument to the one in [Benamou et al. 2015] shows that we can recover $\boldsymbol{\Gamma}$ by solving the following minimization problem:

$$\underset{\boldsymbol{\Gamma} \in \overline{\mathcal{M}}(\boldsymbol{\mu}_0, \boldsymbol{\mu})}{\arg\min} \quad \alpha \mathrm{KL}(\boldsymbol{\Gamma}|\mathbf{f}_\alpha(\boldsymbol{\Gamma})). \tag{6}$$

The fact that $\boldsymbol{\Gamma}$ appears twice makes this objective nonconvex. As derived in Appendix A, we compute $\boldsymbol{\Gamma}$ using KL-projected gradient descent:

$$\boxed{\boldsymbol{\Gamma}^{(k+1)} \leftarrow \underset{\boldsymbol{\Gamma} \in \overline{\mathcal{M}}}{\arg\min}\, \mathrm{KL}\left(\boldsymbol{\Gamma} \middle| \left[\mathbf{f}_\alpha(\boldsymbol{\Gamma}^{(k)})\right]^{\wedge\eta} \otimes \left[\boldsymbol{\Gamma}^{(k)}\right]^{\wedge(1-\eta)}\right)}, \tag{7}$$

where $\otimes$ denotes elementwise (Hadamard) matrix multiplication.

Iteration (7) is a convex program identical to the optimal transportation problems considered in [Benamou et al. 2015]; as derived in that paper, it can be optimized using iterative Sinkhorn projection. The parameter $\eta \in (0, 1]$ controls the *inertia* of the iterative scheme; η acts similarly to the "learning rate" parameter of gradient descent algorithms. Pseudocode is provided in Algorithm 1.

3.3 Convergence

Convergence of (7) as $k \to \infty$ is a subtle matter. While existing work suggests monotonicity of the objective (6) under weak assumptions (see below), this does *not* guarantee convergence of the iterates $\boldsymbol{\Gamma}^{(k)}$. Given the dependence of our subsequent discussion on this algorithm as well as its appearance in many other contexts, we prove the following convergence result:

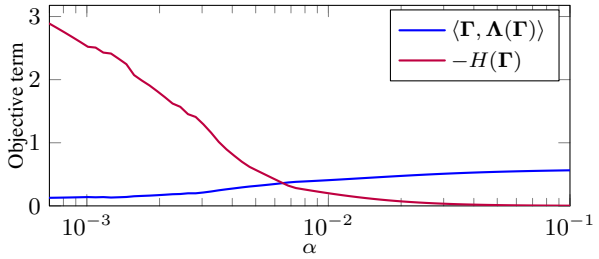


Figure 3: *Dependence of objective terms on regularizer α for the experiment in Figure 2.*

Proposition 1. $\{\boldsymbol{\Gamma}^{(k)}\}_{k=0}^{\infty}$ *converges to a critical point of* (6) *when* $\eta < c\alpha/1+c\alpha$, *where* $c := [\min_{ij} \boldsymbol{\mu}_{0i}\boldsymbol{\mu}_j]/\|\boldsymbol{\mu}_0\|^2\|\boldsymbol{\mu}\|^2\|\mathbf{D}_0\|\|\mathbf{D}\|$.

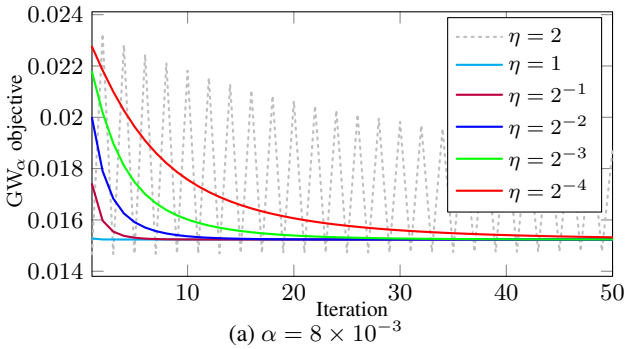
As α increases, the required bound loosens, reflecting the improved conditioning of the optimization problem. We prove this result in the appendix by expressing it as an instance of a modern method for nonconvex optimization [Boţ et al. 2015].

From a practical standpoint, we halt GW_α iteration when the relative change in $\boldsymbol{\Gamma}$ or objective value is less than a fixed threshold (10^{-7} in our experiments).

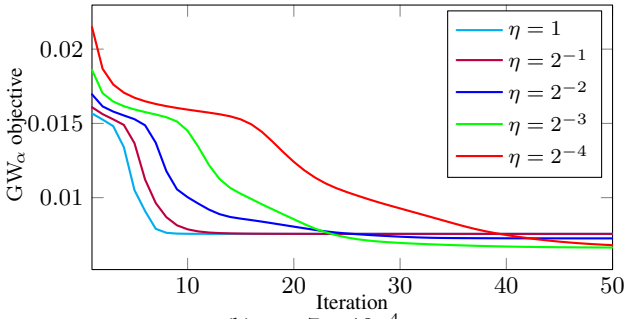
Comparison to softassign. While we derived it for GW_α matching, iteration (7) with $\eta = 1$ is similar to the "softassign" algorithm in machine learning for quadratic assignment [Gold and Rangarajan 1996; Rangarajan et al. 1996]. Rangarajan et al. [1997; 1999] consider the convergence of softassign. They mainly focus on monotonicity of the objective rather than convergence of the iterates. More critically, their analysis assumes convexity; see, e.g., the conclusion of the proof of [Rangarajan et al. 1999, Theorem 1]. Such an assumption is unlikely to be valid for geometric problems with local and global symmetries. Their analysis introduces convexity via a "self-amplification term" along the diagonal of the objective matrix, which modifies the optimization problem and obliterates the possibility of identifying symmetric ambiguity.

3.4 Parameter Selection

Figure 3 illustrates the trade-off between the two objective terms for the experiments in Figure 2 as a function of α. As α increases, so does the GW objective (3) while negative entropy decreases. A phase transition in the GW objective indicates a change in the nature of the map, in this case from mapping individual tentacles to superposing all tentacle targets. This example illustrates a more general pattern: large choices of α lead to higher-entropy correspondence matrices that still contain meaningful match information, while lower α's make for sharper map matrices that express a single coherent match-



(a) $\alpha = 8 \times 10^{-3}$



(b) $\alpha = 7 \times 10^{-4}$

Figure 4: *Convergence for the examples in Figure 2.*

ing. When α is small but there exist multiple symmetric maps, they can be extracted using the method in §5.4.

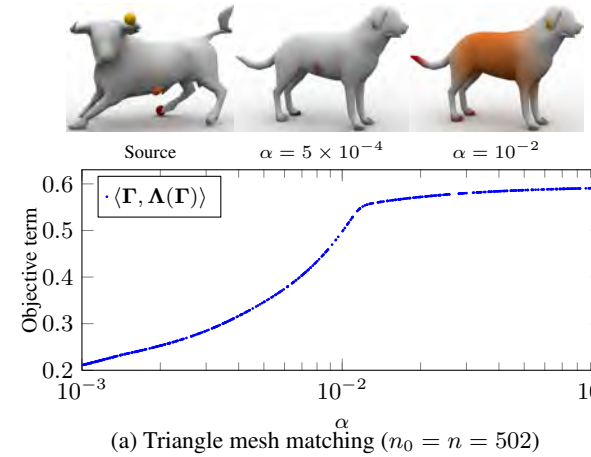
Our algorithm converges faster with a warm start, and hence we can generate plots like the one in Figure 3 efficiently by incrementally adjusting α and updating $\mathbf{\Gamma}$. User-guided selection of α can be guided using this plot to find significant changes in the map as it depends on regularization.

4 Experiments

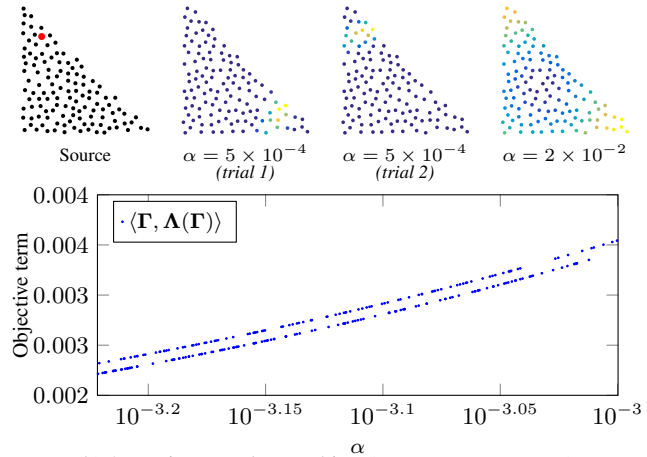
Algorithm 1 provides an extremely simple technique for optimizing the GW$_\alpha$ objective function. We view this simplicity as an advantage; our algorithm is implementable in nearly any framework and generalizable to many classes of domains. In this section, we verify that simplicity does not come at the cost of performance through experiments designed to reveal properties of our iterative technique.

Convergence. Figure 4 illustrates convergence of our algorithm on the experiments in Figure 2, with several choices of η. The experiments illustrate that the bound on η in Proposition 1 is loose, that is, some choices of $\eta \geq {}^{c\alpha}/_{1+c\alpha}$ still exhibit convergence. In our experiments, the algorithm appears to converge monotonically even when $\eta = 1$. While we are unable to provide rigorous proof of convergence in this regime, from an engineering standpoint larger η values can hasten completion time. Figure 4a shows an example with $\eta > 1$ that does not converge. As a trade-off, Figure 4b illustrates that the larger steps can skip past optima with better objective values; in this case, the different local optima correspond to different matchings of the octopus's tentacles.

Sensitivity. Since Algorithm 1 optimizes a non-convex objective, the output may depend on the initial $\mathbf{\Gamma}$. Figure 5 tests sensitivity to the initial guess. Like the previous test, for a given matching task we plot the optimized GW objective $\langle \mathbf{\Gamma}, \mathbf{\Lambda}(\mathbf{\Gamma}) \rangle$ as a function of α; this roughly should increase as α increases. In this test, however, we randomly generate an initial $\mathbf{\Gamma}$ in each trial.



(a) Triangle mesh matching ($n_0 = n = 502$)



(b) 2D point sample matching ($n_0 = 119$, $n = 121$)

Figure 5: *Sensitivity to initial guess of $\mathbf{\Gamma}$ for matching (a) surfaces and (b) planar point samples.*

We evaluate surface matching stability in Figure 5a; distance matrices $\mathbf{D}_0, \mathbf{D}$ contain geodesic distances. This test, representative of many similar experiments involving surfaces, shows little dependence on the initial $\mathbf{\Gamma}$ even for sharp matchings (low α), as reflected by the single coherent curve of objective values.

To reveal a case where multiple optima appear, we consider the task of matching two Poisson disk samples from a right triangle in Figure 5b; here, $\mathbf{D}_0, \mathbf{D}$ contain planar Euclidean distances. Since the triangle has a reflectional symmetry, we expect there to be two nearly-optimal maps from the source to the target. When α is sufficiently small, we do observe two different optimized objectives for the same α depending on which symmetric map is closer to the initial estimate of $\mathbf{\Gamma}$. For large enough α (displayed above the plot), the optimal coupling superposes the two optimal maps and the two parallel curves merge.

Efficiency. Figure 6 tests efficiency of our technique. Here, we plot the GW$_\alpha$ objective as a function of time, measured using a single-threaded implementation in Matlab on a 2.1 GHz Intel i7 CPU with 8GB memory. [Mémoli 2011] and subsequent works mention two optimization algorithms (without regularization): gradient descent and alternation. The latter solves an optimal transportation-style linear program in each step. Our method directly improves that technique, so to compare to a more distant alternative we consider gradient-based routines popular in graphics. In particular, we employ a standard implementation of the interior point method [Waltz et al.

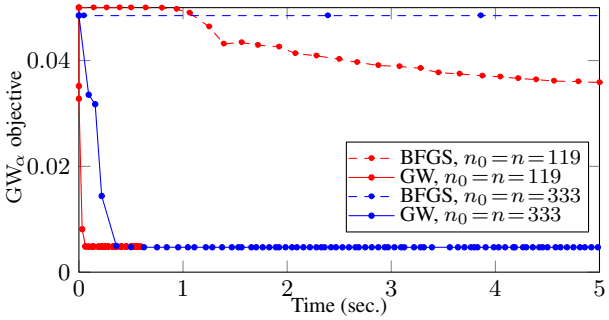


Figure 6: *Objective value vs. time for triangular samples as in Figure 5b ($\alpha \equiv 10^{-3}$); markers indicate iterations.*

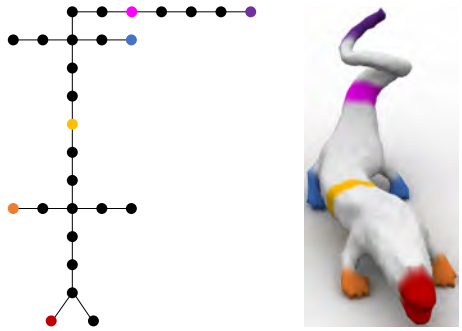


Figure 8: *Mapping a graph onto a shape ($n_0 = 27, n = 2502, \alpha = 1.8 \times 10^{-4}$).*

2006] with "limited-memory" BFGS Hessian estimates. Both in terms of elapsed time and number of iterations, our method reaches a critical point far faster than the interior point alternative. The difference widens as the size of the problem increases.

Robustness. To demonstrate the robustness of GW$_\alpha$-based matching, Figure 7 illustrates couplings between pairs of triangulated surfaces computed using our algorithm. Even when the surfaces undergo significant geometric and topological changes (e.g. adding slats to the chair backs and matching a cup to a mug twice its height) that deviate considerably from isometry, conformality, and other assumptions imposed by surface matching algorithms, our method stably recovers a reasonable correspondence. We visualize maps by showing rows of the measure coupling color coded on a target from corresponding points on a source.

Generality. Throughout this paper, we attempt to demonstrate our algorithm on a variety of matching tasks. In this section, we highlight a few additional examples to underscore its wide applicability.

Figure 1 (page 1) tests our algorithm's flexibility to domain class. Here, we map a triangle mesh to five different classes of target domains. Despite the contrasting structures and even dimensionalities of the targets, our algorithm extracts a meaningful correspondence.

Similarly, Figure 8 shows an example of mapping a graph onto a surface. Here, distances $\mathbf{D}_0$ on the graph are all-pairs shortest-path distances with unit edge weights, and $\mathbf{D}$ contains geodesic distances on the target surface. Convergence for this example was particularly fast given the small size of the source domain.

Figure 9 applies our machinery to mapping a set of images onto a grid layout; as a point of comparison, Fried et al. [2015] recently proposed a more complicated optimization routine for the same problem. Here, $\mathbf{D}_0$ contains distances between average LAB color

Figure 9: *Mapping 196 images onto a 14×14 image grid while preserving color similarity. (Images from Flickr public domain collection.)*

over an image, and $\mathbf{D}$ contains Euclidean distances between grid cells. For this application, we need to extract a matching rather than a soft measure coupling $\mathbf{\Gamma}$. To do so, after optimizing for $\mathbf{\Gamma}$, we solve a linear assignment problem

$$\begin{array}{ll} \max_{\tilde{\mathbf{\Gamma}}} & \langle \tilde{\mathbf{\Gamma}}, \mathbf{\Gamma} \rangle \\ \text{s.t.} & \tilde{\mathbf{\Gamma}} \geq 0 \\ & \tilde{\mathbf{\Gamma}}\mathbf{1} = \mathbf{1}n \qquad \tilde{\mathbf{\Gamma}}^\top \mathbf{1} = \mathbf{1}n_0. \end{array}$$

When $n_0 = n$, this linear program admits a permutation matrix solution computable using a standard convex optimization tool or any of the many fast solvers for assignment problems.

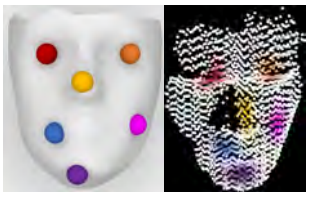


Figure 10: *Point clouds.*

Figure 10 shows an example matching a triangle mesh to a point cloud. Here, we match a noisy and sparse point cloud face scan to a generic mesh of a face. Since the two shapes are in approximately the same pose, we take $\mathbf{D}_0, \mathbf{D}$ to contain Euclidean (rather than geodesic) distances. We use mesh-based area weights on the source and uniform area weights on the target. The match is successful despite the nonuniform sampling of the point cloud and a missing patch on the nose.

Figure 11

The supplemental document includes a 2D shape matching test (data from [Thakoor et al. 2007]). The shapes are point clouds without topology. We use the rounding procedure outlined for image layout to display the maps, transferring colors from the matched point on a source shape (boxed). Our algorithm recovers smooth maps for most 2D test cases. The most challenging test in the dataset is the "fork" class; these shapes admit symmetries *and* thin structures that make it difficult to extract a point-to-point map after applying the entropic regularizer. Figure 11 shows an example before rounding.

Figure 7: *Additional surface correspondences computed using Algorithm 1 ($\alpha \equiv 5\times10^{-4}, \eta \equiv 1$). As in [Solomon et al. 2015], colored points on the left are mapped to colored distributions on the right. §5.4 details a method for addressing the symmetry reversal in the third map.*

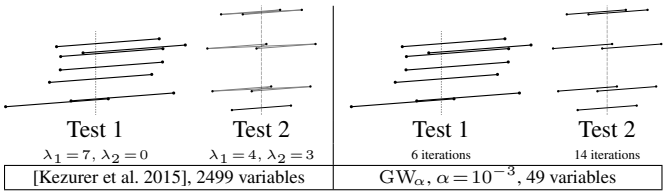


Figure 12: *Comparison to SDP matching. The points on the left of the vertical dotted lines are mapped to the right; edge darkness indicates value of $\mathbf{\Gamma}$. The two largest eigenvalues are shown for [Kezurer et al. 2015]; the second test case is not tight. The number of iterations for six digits of precision is shown for GW_α.*

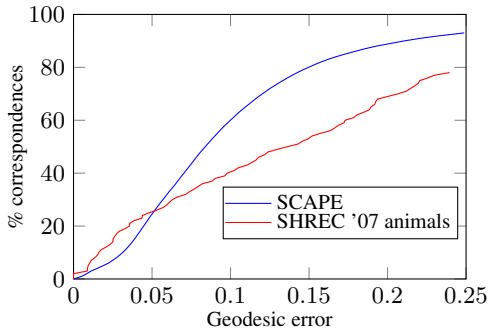


Figure 14: *Benchmark tests from [Kim et al. 2011].*

Comparison. Here we provide a few examples qualitatively demonstrating how our method differs from existing work.

Figure 12 compares to [Kezurer et al. 2015] for a 2D matching problem using pairwise Euclidean distances. We show a symmetric test case and an asymmetric test case. For fair comparison, we use their SDP relaxation on the GW objective. Our unoptimized implementation instantiated 2499 variables to match seven points using their method, which has $O(n^4)$ variables for n matched points. Beyond an obvious difference in timings, their relaxation was not tight in the symmetric test. While optimization via the interior point method recovered a map that arbitrarily superposed the two equally optimal maps, the weights of the superposition were not isotropic or controllable. Our method recovered a sharp map; the symmetric map can be recovered as described in §5.4.

The supplemental document includes a comparison to [Aflalo et al. 2015] in which we minimize $\|\mathbf{D}_0\mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{D}\|_{\mathrm{Fro}}$ over doubly stochastic $\mathbf{\Gamma}$. To match their setup, we assume $n_0 = n$ with uniform area weights. Although their method does not require an SDP, the convexity of their objective is problematic for (nearly-)symmetric domains; see the supplemental document for discussion.

Figure 13 compares to the Blended Intrinsic Maps (BIM) method of [Kim et al. 2011] for triangle mesh correspondence. BIM constructs point-to-point maps between triangulated surfaces by patching together conformal (angle-preserving) maps. For comparison, we compute a sharp GW_α map ($\alpha = 5\times10^{-4}$) and show the highest-probability match for each source point. Because BIM restricts to conformal maps rather than minimizing a measure of stretch like GW_α, it cannot capture the stretch of the neck and legs from the bull source model to the giraffe target. GW_α more evenly distributes the distortion.

Figure 13: *Blended intrinsic maps.*

Figure 14 illustrates performance on benchmarks from [Kim et al. 2011]; the units of the plots and experiments are identical to the ones used in their paper to ease comparison to [Kim et al. 2011] and follow-on work. For these experiments, we round measure couplings to sharp maps by mapping source points to the maximum of their corresponding rows in the coupling. Despite our method's generality and design for rough rather than dense correspondence, it behaves comparably to algorithms for dense, isometry-invariant mapping on the near-isometric SCAPE dataset. Error is higher the non-isometric SHREC '07 dataset; for some geometrically diverse surface pairs, GW_α takes the front of one animal to the back of another. While the plots show results without user guidance, these negative results could easily be addressed using the supervised matching pipeline outlined in §5.2.

5 Applications and Extensions

The basic GW_α optimization can be applied to many graphics tasks and admits countless extensions to variations of the original problem. To emphasize how GW_α fits into different pipelines, here we provide several examples of adapting GW_α to diverse matching problems, accompanied by proof-of-concept examples validating stability to these changes.

5.1 Organizing Shape Collections

The GW_α objective value after minimization is a *distance* between metric spaces. These distances can be used for shape retrieval, search, exploration, or organization of shape databases. GW_α distances between geodesic distance matrices are isometry-invariant, popular for deformable shape retrieval [Bronstein et al. 2011; Rustamov et al. 2013]. For example, Figure 15 shows a multidimensional scaling (MDS) embedding of 80 models from four different categories in the SHREC dataset [Giorgi et al. 2007] using pairwise GW_α distances. Note that models from the same class form distinct clusters.

Computing pairwise GW_α distances can be prohibitively expensive for large datasets, so we leverage the observation of Rustamov et al. [2013] that collections of similar 3D models can be compared to a single "base shape." In this case, we use the GW_α maps $\mathbf{\Gamma}_{i,0}$ (where

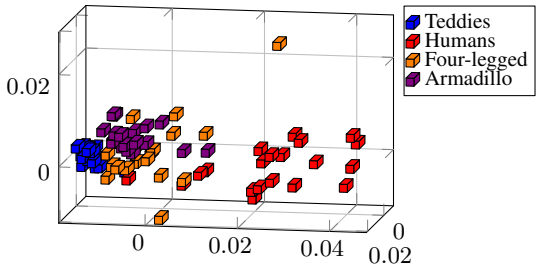


Figure 15: *MDS embedding of four classes from SHREC dataset.*

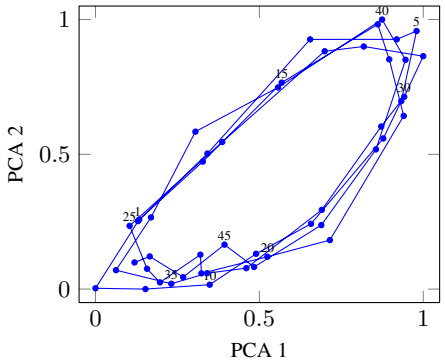


Figure 16: *Recovery of galloping horse sequence.*

0 is the base shape) as a feature vector for shape i. We reproduce the result presented in the work of Rustamov et al., recovering the circular structure of meshes from a galloping horse animation sequence (Figure 16). Unlike Rustamov et al., however, our method does not require ground truth maps between shapes as input.

5.2 Supervised Matching

An important feature of a matching tool is the ability to incorporate user input, e.g. ground truth matches of points or regions. In the GW_α framework, one way to enforce these constraints is to provide a *stencil* $\mathbf{S}$ specifying a sparsity pattern for the map $\mathbf{\Gamma}$. Incorporating constraints in this form is as simple as replacing $\mathbf{K} \leftarrow \mathbf{K} \otimes \mathbf{S}$ in Algorithm 1 before Sinkhorn projection.

Figure 17 illustrates a prototype "user session" in interactive map design. Initially, we optimize with high regularization and no constraints, yielding a superposition of symmetric maps. The user is prompted with the highest-entropy row (leftmost target; source point marked in black) and inputs a ground truth match, marked in black in the remaining target images. The map is recomputed with a lower regularizer and the ground truth match in $\mathbf{S}$, disambiguating the rotational symmetry (center target). GW_α is still unable to disambiguate the top-bottom reflectional symmetry, so the user further constrains the upper half of the source to map to the upper half of the target by using § to zero out undesired matches. With this additional change and after decreasing regularization, the algorithm produces a near point-to-point map (rightmost target). Each map update takes a few seconds.

5.3 Weighted Distance Matrix

In some scenarios, only distances in a particular range are relevant to matching, e.g. keeping certain points close to one another while pushing others far apart. In other contexts, distance values may be known with varying confidence.

More generally, suppose in addition to distance matrices $\mathbf{D}_0 \in \mathbb{R}_+^{n_0 \times n_0}$ and $\mathbf{D} \in \mathbb{R}_+^{n \times n}$ we are given symmetric weight matrices $\mathbf{W}_0 \in \mathbb{R}_+^{n_0 \times n_0}$ and $\mathbf{W} \in \mathbb{R}_+^{n \times n}$. We could solve a *weighted* version of the GW_α matching problem (3) that prioritizes maps preserving distances corresponding to large $\mathbf{W}$ values:

$$\min_{\mathbf{\Gamma} \in \mathcal{M}} \sum_{ijk\ell} (\mathbf{D}_{0ij} - \mathbf{D}_{k\ell})^2 \mathbf{\Gamma}_{ik} \mathbf{\Gamma}_{j\ell} \mathbf{W}_{0ij} \mathbf{W}_{j\ell} \boldsymbol{\mu}_{0i} \boldsymbol{\mu}_{0j} \boldsymbol{\mu}_k \boldsymbol{\mu}_\ell. \quad (8)$$

For instance, $(\mathbf{W}_0, \mathbf{W})$ might contain confidence values expressing the quality of the entries of $(\mathbf{D}_0, \mathbf{D})$. Or, $\mathbf{W}_0, \mathbf{W}$ could take values in $\{\varepsilon, 1\}$ reducing the weight of distances that are unknown or do not need to be preserved by $\mathbf{\Gamma}$.

Following the same simplifications as §3.1, we can optimize this objective by minimizing $\langle \mathbf{\Gamma}, \mathbf{\Lambda}_{\mathbf{W}}(\mathbf{\Gamma}) \rangle$, where

$$\begin{aligned} \mathbf{\Lambda}_{\mathbf{W}}(\mathbf{\Gamma}) := & \frac{1}{2} [\mathbf{D}_0^{\wedge 2} \otimes \mathbf{W}_0] [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma} [\![\boldsymbol{\mu}]\!] \mathbf{W} \\ & - [\mathbf{D}_0 \otimes \mathbf{W}_0] [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma} [\![\boldsymbol{\mu}]\!] [\mathbf{D} \otimes \mathbf{W}] \\ & + \frac{1}{2} \mathbf{W}_0 [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma} [\![\boldsymbol{\mu}]\!] [\mathbf{D}^{\wedge 2} \otimes \mathbf{W}] \end{aligned}$$

The remainder of the derivation in §3.1 remains unchanged, and hence we can optimize (8) using a small modification of Algorithm 1 corresponding to the new update rule

$$\mathbf{\Gamma}_{\mathbf{W}}^{(k+1)} \leftarrow \arg\min_{\mathbf{\Gamma} \in \mathcal{M}} \mathrm{KL}\left(\mathbf{\Gamma} \middle\| \left[\exp\left(-\frac{\mathbf{\Lambda}_{\mathbf{W}}(\mathbf{\Gamma}_{\mathbf{W}}^{(k)})}{\alpha}\right)\right]^{\wedge \eta} \otimes \left[\mathbf{\Gamma}_{\mathbf{W}}^{(k)}\right]^{\wedge (1-\eta)}\right).$$

In practice, giving $\mathbf{W}$ exactly zero entries can cause numerical problems during Sinkhorn rescaling, so we bound it below by a small $\varepsilon > 0$.

Figure 18: *Mapping a set of 185 images onto a two shapes while preserving color similarity.* *(Images from Flickr public domain collection.)*

Figure 18 illustrates an application of this technique ($\alpha = 3 \times 10^{-3}, \eta = 2.5 \times 10^{-2}$). Here, we extend the image example from §4 to map images onto shapes on a shared grid. We take $\mathbf{W}_{ij} = 1$ if i and j are in the same connected component and $\mathbf{W}_{ij} = 0$ otherwise; this objective does not enforce color continuity between the connected components. As a result, the components exhibit different (but still smooth) color variations, e.g. the clusters of blue images on the two components are distant from each other. Note the optimization decides which image appears in which cluster.

5.4 Symmetry Detection

[Pauly 2015] recently posed the question of whether optimal transportation can be used for symmetry analysis. We outline one strategy

Source | Unconstrained (black point image) $\alpha = 8\times10^{-3}$ | Point constraint $\alpha = 5\times10^{-4}$ | Point & plane constraints $\alpha = 2.5\times10^{-4}$

Figure 17: *Steps in supervised map design user session, outlined in §5.2.*

here for dealing with symmetries intrinsic to the distance matrices $\mathbf{D}_0, \mathbf{D}$ leveraging the *non*-convexity of the GW_α objective function.

Suppose Algorithm 1 yields a measure coupling $\mathbf{\Gamma}$ corresponding to a map $\Sigma_0 \mapsto \Sigma$. If the distance matrix $\mathbf{D}$ on Σ has a symmetry, then there exists a second coupling $\mathbf{\Gamma}'$ with a similar GW_α objective value encoding the symmetric map. We optimize for this map by biasing the GW_α objective slightly against $\mathbf{\Gamma}$ in (6):

$$\mathbf{f}_\alpha^{\mathrm{symm}}(\mathbf{\Gamma}';\mathbf{\Gamma}) := \exp([\mathbf{D}_0[\![\boldsymbol{\mu}_0]\!]\mathbf{\Gamma}'[\![\boldsymbol{\mu}]\!]\mathbf{D} - \varepsilon\mathbf{\Gamma}]/\alpha). \qquad (9)$$

Now, the symmetric map should have a slightly lower objective value than $\mathbf{\Gamma}$, which we can compute using the same algorithm.

Figure 19 illustrates an example of symmetric map computation. Algorithm 1 yielded the left-right reversing map between the meshes of humans (same as Figure 7), which is equally optimal to the orientation-preserving map. Fixing this map as $\mathbf{\Gamma}$ and optimizing for $\mathbf{\Gamma}'$ using (9) ($\varepsilon = 10^{-6}$) yields the orientation-preserving map as desired.

Σ_0 $\quad$ $\mathbf{\Gamma}$ $\quad$ $\mathbf{\Gamma}'$

Figure 19: *Symmetric maps.*

Figure 20 shows the more conventional case of intrinsic "self"-symmetry detection. In this case, we seek maps from the mouse model onto itself that preserve distance structure ($\alpha = 10^{-3}, \varepsilon = 10^{-3}$). To avoid the obvious identity map, we optimize for $\mathbf{\Gamma}'$ after taking $\mathbf{\Gamma}$ to be the identity matrix. This yields the left-right symmetric map of the model.

Σ_0 $\quad$ $\mathbf{\Gamma}'$

Figure 20: *Self symmetry.*

Notice the construction of this technique relies upon the non-convexity of the GW_α objective. In particular, averaging the orientation-preserving and orientation-reversing maps *increases* the GW_α objective, a property that cannot hold for convex functions.

5.5 Joint Domain Analysis

A valuable property of mapping within a collection is *consistency*. If we map $\mathcal{A}$ to $\mathcal{B}$, $\mathcal{B}$ to $\mathcal{C}$, and $\mathcal{C}$ back to $\mathcal{A}$, we may wish to constrain the composition to approximate the identity. Consistency can help improve a collection of maps by reinforcing correct matches.

Suppose we are given domains $\{\mathcal{D}_1, \ldots, \mathcal{D}_p\}$ with pairwise couplings $\mathbf{\Gamma}_{ij} \in \overline{\mathcal{M}}(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j)$; $\mathbf{\Gamma}_{ij}$ and $\mathbf{\Gamma}_{jk}$ can be composed as $\mathbf{\Gamma}_{ij}[\![\boldsymbol{\mu}_j]\!]\mathbf{\Gamma}_{jk}$ (see appendix). Following [Huang and Guibas 2013], if the $\mathbf{\Gamma}_{ij}$'s are consistent, they factor through an "urshape," that is, we could map every $\mathcal{D}_i$ to a single target and back out. If the urshape approximately consists of m uniformly-sampled points, we

function SYMMETRIC-NNMF($\mathbf{B}$)
  $\mathbf{A} \leftarrow$ RANDOM($pm \times m$) *// Initialize with nontrivial* $\mathbf{A}$
  for $\ell = 1, 2, 3, \ldots$
    $\mathbf{U} \leftarrow \mathbf{A} \otimes [(\mathbf{B} \oslash \mathbf{A}\mathbf{A}^\top)\mathbf{A}]$
    $\mathbf{A} \leftarrow \mathbf{U}[\![\mathbf{1} \oslash \sqrt{\mathbf{U}^\top\mathbf{1}}]\!]$
  return $\mathbf{A}$ *//* $\mathbf{B} \approx \mathbf{A}\mathbf{A}^\top$

Algorithm 2: *Nonnegative matrix factorization minimizing* $\mathrm{KL}(\mathbf{B}|\mathbf{A}\mathbf{A}^\top)$ *with respect to* $\mathbf{A}$.

can write $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_p \in \overline{\mathcal{M}}(\boldsymbol{\mu}_i, \mathbf{1}/m)$ such that $\mathbf{\Gamma}_{ij} = \mathbf{\Gamma}_i\mathbf{\Gamma}_j^\top/m$:

$$\underbrace{\begin{pmatrix} \mathbf{\Gamma}_{11} \cdots \mathbf{\Gamma}_{1p} \\ \vdots \ddots \vdots \\ \mathbf{\Gamma}_{p1} \cdots \mathbf{\Gamma}_{pp} \end{pmatrix}}_{\mathbf{G}} = \frac{1}{m} \underbrace{\begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_p \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_p \end{pmatrix}^\top}_{\mathbf{A}^\top}. \qquad (10)$$

This shows that $\mathbf{G}$ admits a *low-rank, symmetric nonnegative factorization* proportional to $\mathbf{A}\mathbf{A}^\top$.

Suppose we are given metric matrices $\mathbf{D}_1, \ldots, \mathbf{D}_p$. For simplicity, we assume $\mathbf{D}_i \in \mathbb{R}_+^{m\times m}\ \forall i$ with constant area weights. From (10), we optimize for a set of nearly-consistent maps $\mathbf{\Gamma}_{ij}$ as follows:

$$\begin{aligned} \min_{\mathbf{G},\mathbf{A}}\ & \beta\mathrm{KL}(\mathbf{G}|\mathbf{A}\mathbf{A}^\top) + \textstyle\sum_{ij} \mathrm{KL}(\mathbf{\Gamma}_{ij}|\mathbf{f}_\alpha(\mathbf{\Gamma}_{ij};\mathbf{D}_i,\mathbf{D}_j)) \\ \text{s.t. } & \mathbf{G} \text{ has blocks } \mathbf{\Gamma}_{ij} \text{ and } \mathbf{A} \text{ has blocks } \mathbf{\Gamma}_i \\ & \mathbf{\Gamma}_{ij}, \mathbf{\Gamma}_i \in \overline{\mathcal{M}}(\mathbf{1}/m, \mathbf{1}/m)\ \forall i, j \end{aligned} \qquad (11)$$

The second term is the GW_α objective (6), and the first approximates $\mathbf{G}$ as $\mathbf{A}\mathbf{A}^\top$. This resembles factorizations in probabilistic latent semantic analysis (pLSA) [Hofmann 1999], which approximate $\mathbf{B} \approx \mathbf{W}\mathbf{H}$ by minimizing $\mathrm{KL}(\mathbf{B}|\mathbf{W}\mathbf{H})$ [Gaussier and Goutte 2005].

We minimize (11) by alternating between $\mathbf{G}$ and $\mathbf{A}$. Each alternation has a straightforward algorithm detailed below and in Algorithm 3. Our method ensures that the objective of (11) decreases in each step, but we leave to future work a result like Proposition 1.

First, suppose $\mathbf{A}$ is fixed and that we optimize for $\mathbf{G}$. Denote by $\mathbf{F}_\alpha(\mathbf{G})$ the matrix with blocks $\mathbf{f}_\alpha(\mathbf{\Gamma}_{ij};\mathbf{D}_i,\mathbf{D}_j)$. Algebraic manipulation of (11) with $\mathbf{A}$ fixed yields an equivalent problem

$$\begin{aligned} \min_{\mathbf{G}}\ & \mathrm{KL}\left(\mathbf{G}\,\middle|\,\mathbf{F}_\alpha(\mathbf{G})^{\wedge(1-\delta)} \otimes (\mathbf{A}\mathbf{A}^\top)^{\wedge\delta}\right) \\ \text{s.t. } & \mathbf{G} \text{ has blocks } \mathbf{\Gamma}_{ij} \text{ and } \mathbf{A} \text{ has blocks } \mathbf{\Gamma}_i \\ & \mathbf{\Gamma}_{ij}, \mathbf{\Gamma}_i \in \overline{\mathcal{M}}(\mathbf{1}/m, \mathbf{1}/m)\ \forall i, j, \end{aligned} \qquad (12)$$

where $\delta := 1/1+\beta$. Problem (12) decouples over the $m\times m$ blocks of $\mathbf{G}$, and each individual problem is an instance of the GW_α objective with elementwise scaling from $\mathbf{A}\mathbf{A}^\top$.

Now, suppose $\mathbf{G}$ is fixed, leaving minimization of $\mathrm{KL}(\mathbf{G}|\mathbf{A}\mathbf{A}^\top)$ with respect to $\mathbf{A}$. Unlike pLSA, we factor $\mathbf{G} \approx \mathbf{A}\mathbf{A}^\top$ rather than

function JOINT-GW$_\alpha(\mathbf{D}_1, \ldots, \mathbf{D}_p; \delta, \eta)$
// Computes consistent maps via (12).
$\mathbf{A} \leftarrow$ ONES$(pm \times m)$
$\mathbf{G} \leftarrow$ ONES$(pm \times pm)$
for $k = 1, 2, 3, \ldots$
$\mathbf{G}_0 \leftarrow \frac{1}{m}\mathbf{A}\mathbf{A}^\top$ *// Consistent pairwise maps*
for $i, j = 1, 2, \ldots, p$ *// Fix* $\mathbf{A}$ *and optimize* $\mathbf{G}$
$\mathbf{\Gamma}_{ij} \leftarrow$ BLOCK-PROJECTION$(\mathbf{D}_i, \mathbf{D}_j, (\mathbf{G}_0)_{ij}; \delta, \eta)$
BLOCK$(\mathbf{G}, i, j) \leftarrow \mathbf{\Gamma}_{ij}$ *// Replace block of* $\mathbf{G}$
$\mathbf{A} \leftarrow$ SYMMETRIC-NNMF$(\mathbf{G}^{\wedge(1/\delta)} \otimes \mathbf{F}_\alpha(\mathbf{G})^{\wedge(1-1/\delta)})$
return $(\mathbf{A}, \mathbf{G})$

function BLOCK-PROJECTION$(\mathbf{D}_i, \mathbf{D}_j, \mathbf{V}; \delta, \eta)$
// Blockwise projection for consistent GW
$\mathbf{\Gamma} \leftarrow$ ONES$(m \times m)$
for $\ell = 1, 2, 3, \ldots$
$\mathbf{K} \leftarrow \mathbf{f}_\alpha(\mathbf{\Gamma})^{\wedge(1-\delta)} \otimes \mathbf{V}^{\wedge\delta}$
$\mathbf{\Gamma} \leftarrow$ SINKHORN-PROJECTION$(\mathbf{K}^{\wedge\eta} \otimes \mathbf{\Gamma}^{\wedge(1-\eta)}; \frac{1}{m}, \frac{1}{m})$
return $\mathbf{\Gamma}$

Algorithm 3: *Algorithm for joint GW matching.*

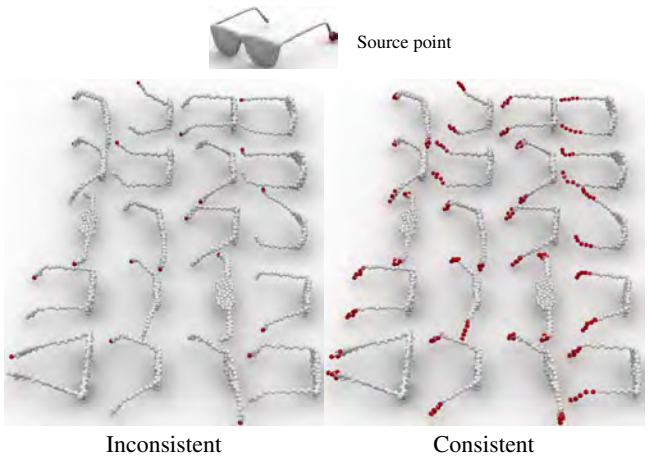


Figure 21: *Consistent mapping via low-rank factorization reveals the left-right symmetric ambiguity of mapping between twenty models of eyeglasses ($\alpha = 10^{-3}, \beta = 1$, $n = 100$, rank 15).*

$\mathbf{A} \approx \mathbf{W}\mathbf{H}$. We minimize with respect to a *single* $\mathbf{A}$ rather than minimizing $\mathrm{KL}(\mathbf{G}|\mathbf{W}\mathbf{H})$ over $\mathbf{W}$ and $\mathbf{H}$ independently. Algorithm 2 provides a simple iteration for this factorization, using majorization-minimization [Hunter and Lange 2000]. We prove the following monotonicity property in the appendix:

Proposition 2. *Denote the iterates of Algorithm 2 as* $\mathbf{A}^{(\ell)}$*. Then,* $\mathrm{KL}(\mathbf{G}|\mathbf{A}^{(\ell)}\mathbf{A}^{(\ell)\top})$ *decreases as a function of* ℓ*, with monotonicity unless* $\mathbf{A}^{(\ell)}$ *is a critical point of* $\mathrm{KL}(\mathbf{G}|\mathbf{A}\mathbf{A}^\top)$.

The work of [Lanckriet and Sriperumbudur 2009] might be used to obtain more refined convergence results.

Figure 21 illustrates an example of consistent mapping using the technique outlined above. In this case, we compute consistent maps between 20 models of eyeglasses, each sampled with 100 points. Pairwise mapping using the GW_α objective yields maps with arbitrary, inconsistent left-right flips. Adding the low-rank term (11) produces maps that superpose the ambiguous symmetry. For this experiment, we take A to be rank 15; this choice of dimensionality accelerates convergence of the alternating algorithm and is reasonable given the symmetry we expect in the final maps.

6 Discussion and Conclusion

Empirically, it appears our entropic regularizer transforms a quadratic matching energy and constraint landscape—well-known to be difficult to navigate—to a smooth one whose basins can be reached via a variant of projected gradient descent. In this sense, despite the presence of multiple local optima for some mapping tasks, it is likely outside the class of "scary" optimization objectives avoided in tractable numerical tools [Sun et al. 2015] since these local optima all represent meaningful correspondences.

Our results suggest many potential extensions and challenging avenues for future research. Principally, what remains is extension to *partial* matching, e.g. taking part of a surface to a full target or vice versa. An obvious approach is to drop either prescribed row or column sums for $\mathbf{\Gamma}$, but this leads to a larger space in which GW_α appears to have more local optima; this is reasonable in that there often exist many ways to align a patch of a domain with a larger space. Application-specific regularization may help in this case. An additional extension might be to improve the efficiency of our technique by alleviating the need for distance matrices on structured domains, similar to [Solomon et al. 2015].

While a theme of this work is that regularization can make challenging matching problems tractable, the longest-standing challenge is to optimize GW and related objectives without smoothing. General quadratic matching is NP-hard, but algorithms for minimizing quadratic objectives derived from smooth geometric domains may have greater hope of succeeding. Recent progress for tasks in graphics including [Chen and Koltun 2015; Wei et al. 2015] suggests that there is much to do in this direction.

There is a formidable spectrum of tools for geometric correspondence. Methods tuned to individual domain classes are well-explored for detailed, application-specific analysis. Contrastingly, regularized GW optimization is a widely-applicable, straightforward, and easily-extended tool for rough correspondence given minimal guidance or assumptions. Our examples show how an identical short piece of code can be used to understand shapes, surfaces, image collections, graphs, icons, and so on. This practical contribution is paired with an unconditional convergence guarantee, a reasonable compromise between slow globally optimal and fast heuristic approaches. As it stands, metric alignment via entropic GW distances provides an easy, stable approach to soft correspondence suitable across a wide collection of matching tasks.

Acknowledgments

T. Funhouser and W. Leeb kindly edited early drafts of this paper. J. Solomon and S. Sra were connected on the MathOverflow platform. C. Lampert suggested the connection to [Quadrianto et al. 2009].

J. Solomon acknowledges the support of the NSF Mathematical Sciences Postdoctoral Research Fellowship (award number 1502435). The work of G. Peyré has been supported by the European Research Council (ERC project SIGMA-Vision). 3D face point cloud courtesy Singularity University.

A Proof of Proposition 1

We appeal to [Boţ et al. 2015, Theorem 12], which establishes convergence of a general nonconvex optimization algorithm. They

consider the iteration (in their notation, fix $\beta_n \equiv 0$ and $\alpha_n \equiv \xi$)

$$x^{(k+1)} \leftarrow \arg\min_u \left\{ D_F(u, x^{(k)}) + \xi \left[u \cdot \nabla g(x^{(k)}) + f(u) \right] \right\}, \quad (13)$$

where F is a Bregman divergence parameterized by convex $F(x)$:

$$D_F(x, y) := F(x) - F(y) - \nabla F(y) \cdot (x - y).$$

Make the following substitutions:

$$\begin{aligned} g(\mathbf{\Gamma}) &:= -1/2 \langle \mathbf{\Gamma}, \mathbf{D}_0 [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma} [\![\boldsymbol{\mu}]\!] \mathbf{D} \rangle \\ &\implies \nabla g(\mathbf{\Gamma}) = -[\![\boldsymbol{\mu}_0]\!] \mathbf{D}_0 [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma} [\![\boldsymbol{\mu}]\!] \mathbf{D} [\![\boldsymbol{\mu}]\!] \\ f(\mathbf{\Gamma}) &:= -\alpha H(\mathbf{\Gamma}) + \chi[\mathbf{\Gamma} \in \overline{\mathcal{M}}] \\ F(\mathbf{\Gamma}) &:= -H(\mathbf{\Gamma}) \implies D_F(\mathbf{\Gamma}, \mathbf{\Gamma}') = \mathrm{KL}(\mathbf{\Gamma} | \mathbf{\Gamma}') \end{aligned}$$

Here, $\chi[\cdot]$ is the $\{0, \infty\}$ indicator of a set. Then, the algorithm becomes

$$\mathbf{\Gamma}^{(k+1)} \leftarrow \arg\min_{\mathbf{\Gamma} \in \overline{\mathcal{M}}} \left\{ \mathrm{KL}(\mathbf{\Gamma} | \mathbf{\Gamma}^{(k)}) - \xi \left[\langle \mathbf{\Gamma}, \mathbf{D}_0 [\![\boldsymbol{\mu}_0]\!] \mathbf{\Gamma}^{(k)} [\![\boldsymbol{\mu}]\!] \mathbf{D} \rangle + \alpha H(\mathbf{\Gamma}) \right] \right\}.$$

After algebraic simplification, this is exactly the iteration (7) with

$$\eta := \frac{\xi \alpha}{1 + \xi \alpha} \iff \xi := \frac{\eta}{\alpha(1 - \eta)}.$$

Notice when $\eta = 0$, we must have $\xi = 0$ and the $\mathbf{\Gamma}^{(k)}$'s do not change. Contrastingly, $\eta \to 1$ as $\xi \to \infty$.

The choice of f is proper, lower semicontinuous, and bounded below, and g is differentiable with Lipschitz constant $L_{\nabla g}$:

$$\begin{aligned} \|\nabla g(\mathbf{\Gamma}) - \nabla g(\mathbf{\Gamma}')\| &= \|[\![\boldsymbol{\mu}_0]\!] \mathbf{D}_0 [\![\boldsymbol{\mu}_0]\!] (\mathbf{\Gamma} - \mathbf{\Gamma}') [\![\boldsymbol{\mu}]\!] \mathbf{D} [\![\boldsymbol{\mu}]\!]\| \\ &\le \underbrace{\|\boldsymbol{\mu}_0\|^2 \|\boldsymbol{\mu}\|^2 \|\mathbf{D}_0\| \|\mathbf{D}\|}_{L_{\nabla g}} \cdot \|\mathbf{\Gamma} - \mathbf{\Gamma}'\|. \end{aligned}$$

The function $x \mapsto x \ln x$ is 1–strongly convex. Hence, $\mathbf{\Gamma} \mapsto -H(\mathbf{\Gamma})$ is σ-strongly convex, where $\sigma := \min_{ij} \boldsymbol{\mu}_{0i} \boldsymbol{\mu}_j$. Finally, $f + g$ is coercive since it is non-infinite on a compact set, and it satisfies the requirements of the Kurdyka-Łojasiewicz inequality because it is definable in an o-minimal structure over the reals (see [Coste 1999]).

Next, we check the requirements of [Boţ et al. 2015, Lemma 5] for convergence; we refer the reader to their paper for notation. Our iteration (13) does not include a linear term and hence we can take $M_2 = 0$. To ensure $M_1 > 0$, we need $\sigma - \xi L_{\nabla g} > 0$, or

$$\xi < \frac{\min_{ij} \boldsymbol{\mu}_{0i} \boldsymbol{\mu}_j}{\|\boldsymbol{\mu}_0\|^2 \|\boldsymbol{\mu}\|^2 \|\mathbf{D}_0\| \|\mathbf{D}\|}.$$

If we define c as the right-hand side, then we need $\eta < c\alpha / 1 + c\alpha$. Regardless, under this condition we satisfy all the requirements for convergence, as desired.

B Composition of Couplings

To check that $\mathbf{\Gamma}_{ij} [\![\boldsymbol{\mu}_j]\!] \mathbf{\Gamma}_{jk}$ is a valid composition of $\mathbf{\Gamma}_{ij}$ and $\mathbf{\Gamma}_{jk}$, we simply verify that this product is a measure coupling in $\overline{\mathcal{M}}(\boldsymbol{\mu}_i, \boldsymbol{\mu}_k)$:

$$\begin{aligned} (\mathbf{\Gamma}_{ij} [\![\boldsymbol{\mu}_j]\!] \mathbf{\Gamma}_{jk}) \boldsymbol{\mu}_k &= \mathbf{\Gamma}_{ij} [\![\boldsymbol{\mu}_j]\!] \mathbf{1} = \mathbf{\Gamma}_{ij} \boldsymbol{\mu}_j = \mathbf{1} \\ (\mathbf{\Gamma}_{ij} [\![\boldsymbol{\mu}_j]\!] \mathbf{\Gamma}_{jk})^\top \boldsymbol{\mu}_i &= \mathbf{\Gamma}_{jk}^\top [\![\boldsymbol{\mu}_j]\!] \mathbf{\Gamma}_{ij}^\top \boldsymbol{\mu}_i = \mathbf{\Gamma}_{jk}^\top [\![\boldsymbol{\mu}_j]\!] \mathbf{1} = \mathbf{\Gamma}_{jk}^\top \boldsymbol{\mu}_j = \mathbf{1} \end{aligned}$$

Intuitively, this composition formula formalizes the notion

$$P(x \in \mathcal{D}_i \mapsto z \in \mathcal{D}_k) = \int_{y \in \mathcal{D}_j} P(x \in \mathcal{D}_i \mapsto y \in \mathcal{D}_j) P(y \in \mathcal{D}_j \mapsto z \in \mathcal{D}_k).$$

C Proof of Proposition 2

We employ a standard result from the theory of majorization-minimization. Suppose functions $F(\mathbf{A})$ and $G(\mathbf{A}, \mathbf{A}')$ satisfy

$$\begin{aligned} F(\mathbf{A}) &= G(\mathbf{A}, \mathbf{A}) \; \forall \mathbf{A} \ge 0 \\ F(\mathbf{A}) &\le G(\mathbf{A}, \mathbf{A}') \; \forall \mathbf{A}, \mathbf{A}' \ge 0. \end{aligned}$$

Then, the following iteration monotonically decreases $F(\mathbf{A}^{(\ell)})$:

$$\mathbf{A}^{(\ell+1)} \leftarrow \arg\min_{\mathbf{A} \ge 0} G(\mathbf{A}, \mathbf{A}^{(\ell)}),$$

with strict decrease unless $\mathbf{A}^{(\ell)}$ is a stationary point of F. This follows directly from the definition of $\mathbf{A}^{(\ell)}$ and structure of F, G.

In our case, we take

$$\begin{aligned} F(\mathbf{A}) &= \mathrm{KL}(\mathbf{G} | \mathbf{A}\mathbf{A}^\top) \\ &\sim \sum_{ik} \left[-\mathbf{G}_{ik} \ln (\mathbf{A}\mathbf{A}^\top)_{ik} + (\mathbf{A}\mathbf{A}^\top)_{ik} \right] \text{ after removing constants} \\ &= \|\mathbf{A}^\top \mathbf{1}\|^2 - \sum_{ik} \mathbf{G}_{ik} \ln \sum_m \frac{\lambda_m^{ik} \mathbf{A}_{im} \mathbf{A}_{km}}{\lambda_m^{ik}}, \end{aligned}$$

where λ is any set of constants satisfying $\sum_m \lambda_m^{ik} = 1$ and $\lambda_m^{ik} \ge 0$. By Jensen's inequality,

$$F(\mathbf{A}) \le \|\mathbf{A}^\top \mathbf{1}\|^2 - \sum_{ik} \mathbf{G}_{ik} \sum_m \lambda_m^{ik} \ln \frac{\mathbf{A}_{im} \mathbf{A}_{km}}{\lambda_m^{ik}}.$$

Take

$$\lambda_m^{ik} = \lambda_m^{ik}(\mathbf{G}') := \frac{\mathbf{G}_{im} \mathbf{G}_{km}}{\sum_{m'} \mathbf{G}_{im'} \mathbf{G}_{km'}}.$$

With this choice, by the inequalities above the following G satisfies the criteria from majorization-minimization:

$$G(\mathbf{A}, \mathbf{A}') := \|\mathbf{A}^\top \mathbf{1}\|^2 - \sum_{ik} \mathbf{G}_{ik} \ln \sum_m \frac{\lambda_m^{ik} \mathbf{A}_{im} \mathbf{A}_{km}}{\lambda_m^{ik}}.$$

This function is strictly convex in $\mathbf{A}$. The iterations of our algorithm solve the root-finding problem $\nabla_{\mathbf{A}} G(\mathbf{A}, \mathbf{A}^{(\ell)}) = 0$ in closed form. Note that solutions of $\nabla_{\mathbf{A}} G(\mathbf{A}, \mathbf{A}^{(\ell)}) = 0$ satisfy $\mathbf{A} > 0$ without adding the constraint explicitly thanks to the log term.

References

AFLALO, Y., AND KIMMEL, R. 2013. Spectral multidimensional scaling. *PNAS 110*, 45, 18052–18057.

AFLALO, Y., BRONSTEIN, A. M., AND KIMMEL, R. 2014. Graph matching: relax or not? *CoRR abs/1401.7623*.

AFLALO, Y., BRONSTEIN, A., AND KIMMEL, R. 2015. On convex relaxation of graph isomorphism. *Proc. National Academy of Sci. 112*, 10, 2942–2947.

ATTOUCH, H., BOLTE, J., REDONT, P., AND SOUBEYRAN, A. 2010. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res. 35*, 2 (May), 438–457.

BAUSCHKE, H., AND COMBETTES, P. 2011. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS.

BAUSCHKE, H. H., COMBETTES, P. L., AND NOLL, D. 2006. Joint minimization with alternating Bregman proximity operators. *Pacific J. Optim. 2*, 3, 401–424.

BENAMOU, J.-D., CARLIER, G., CUTURI, M., NENNA, L., AND PEYRÉ, G. 2015. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comp. 37*, 2, A1111–A1138.

BERG, A. C., BERG, T. L., AND MALIK, J. 2005. Shape matching and object recognition using low distortion correspondences. In *Proc. CVPR*, vol. 1, 26–33.

BONNEEL, N., VAN DE PANNE, M., PARIS, S., AND HEIDRICH, W. 2011. Displacement interpolation using Lagrangian mass transport. *ACM Trans. Graph. 30*, 6 (Dec.), 158:1–158:12.

BOŢ, R. I., CSETNEK, E. R., AND LÁSZLÓ, S. C. 2015. An inertial forward–backward algorithm for the minimization of the sum of two nonconvex functions. *EURO J. Comp. Optim.*, 1–23.

BREGMAN, L. M. 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR comp. math. and math. phys. 7*, 3, 200–217.

BRONSTEIN, A. M., BRONSTEIN, M. M., AND KIMMEL, R. 2006. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *PNAS 103*, 5, 1168–1172.

BRONSTEIN, A. M., BRONSTEIN, M. M., KIMMEL, R., MAHMOUDI, M., AND SAPIRO, G. 2010. A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *IJCV 89*, 2-3, 266–286.

BRONSTEIN, A. M., BRONSTEIN, M. M., GUIBAS, L. J., AND OVSJANIKOV, M. 2011. Shape Google: Geometric words and expressions for invariant shape retrieval. *TOG 30*, 1 (Feb.), 1:1–1:20.

ÇELA, E. 2013. *The Quadratic Assignment Problem: Theory and Algorithms*. Combinatorial Optim. Springer.

CHEN, Q., AND KOLTUN, V. 2015. Robust nonrigid registration by convex optimization. In *Proc. ICCV*, 2039–2047.

CHEN, Y., GUIBAS, L., AND HUANG, Q. 2014. Near-optimal joint object matching via convex relaxation. In *Proc. ICML*, JMLR Workshop and Conference Proceedings, 100–108.

COMBETTES, P. L., AND PESQUET, J.-C. 2011. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*. 185–212.

COSTE, M. 1999. An introduction to O-minimal geometry. Tech. rep., Institut de Recherche Mathematiques de Rennes, November.

CUTURI, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transportation. In *Proc. NIPS*, vol. 26. 2292–2300.

DE GOES, F., COHEN-STEINER, D., ALLIEZ, P., AND DESBRUN, M. 2011. An optimal transport approach to robust reconstruction and simplification of 2d shapes. In *CGF*, vol. 30, 1593–1602.

DE GOES, F., BREEDEN, K., OSTROMOUKHOV, V., AND DESBRUN, M. 2012. Blue noise through optimal transport. *ACM Trans. Graph. 31*, 6 (Nov.), 171:1–171:11.

DE GOES, F., WALLEZ, C., HUANG, J., PAVLOV, D., AND DESBRUN, M. 2015. Power particles: An incompressible fluid solver based on power diagrams. *ACM Trans. Graph. 34*, 4 (July), 50:1–50:11.

FRIED, O., DIVERDI, S., HALBER, M., SIZIKOVA, E., AND FINKELSTEIN, A. 2015. IsoMatch: Creating informative grid layouts. In *CGF*, vol. 34, 155–166.

GAUSSIER, E., AND GOUTTE, C. 2005. Relation between pLSA and NMF and implications. In *Proc. SIGIR*, 601–602.

GIORGI, D., BIASOTTI, S., AND PARABOSCHI, L. 2007. Shrec: shape retrieval contest: Watertight models track. *http://watertight.ge.imati.cnr.it*.

GOLD, S., AND RANGARAJAN, A. 1996. A graduated assignment algorithm for graph matching. *PAMI 18*, 4 (Apr.), 377–388.

GROMOV, M. 2001. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Progress in Math. Birkhäuser.

HOFMANN, T. 1999. Probabilistic latent semantic indexing. In *Proc. SIGIR*, 50–57.

HUANG, Q., AND GUIBAS, L. 2013. Consistent shape maps via semidefinite programming. In *Proc. SGP*, 177–186.

HUNTER, D., AND LANGE, K. 2000. Quantile regression via an MM algorithm. *J. Comp. and Graphical Stats. 9*, 1, 60–77.

KEZURER, I., KOVALSKY, S. Z., BASRI, R., AND LIPMAN, Y. 2015. Tight relaxation of quadratic matching. *CGF*.

KIM, V. G., LIPMAN, Y., AND FUNKHOUSER, T. 2011. Blended intrinsic maps. *ACM Trans. Graph. 30*, 4 (July), 79:1–79:12.

KIM, V. G., LI, W., MITRA, N. J., DIVERDI, S., AND FUNKHOUSER, T. 2012. Exploring collections of 3d models using fuzzy correspondences. *ACM Trans. Graph. 31*, 4 (July), 54:1–54:11.

KURDYKA, K. 1998. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier 48*, 3, 769–783.

LANCKRIET, G. R., AND SRIPERUMBUDUR, B. K. 2009. On the convergence of the concave-convex procedure. In *Proc. NIPS*, 1759–1767.

LEORDEANU, M., AND HEBERT, M. 2005. A spectral technique for correspondence problems using pairwise constraints. In *Proc. ICCV*, 1482–1489.

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 J. Operational Research 176*, 2, 657–690.

ŁOJASIEWICZ, S. 1963. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 87–89.

ŁOJASIEWICZ, S. 1993. Sur la géométrie semi-et sous-analytique. In *Annales de l'institut Fourier*, vol. 43, 1575–1595.

LYZINSKI, V., FISHKIND, D., FIORI, M., VOGELSTEIN, J., PRIEBE, C., AND SAPIRO, G. 2015. Graph matching: Relax at your own risk. *PAMI*.

MÉMOLI, F. 2007. On the use of Gromov–Hausdorff distances for shape comparison. In *Symposium on Point Based Graphics*. 81–90.

MÉMOLI, F. 2008. Gromov–Hausdorff distances in Euclidean spaces. In *Proc. CVPR Workshops*, 1–8.

MÉMOLI, F. 2009. Spectral Gromov-Wasserstein distances for shape matching. In *Proc. ICCV Workshops*, 256–263.

MÉMOLI, F. 2011. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Comp. Math. 11*, 4, 417–487.

MÉMOLI, F. 2014. The Gromov–Wasserstein distance: A brief overview. *Axioms 3*, 3.

MÉRIGOT, Q. 2011. A multiscale approach to optimal transport. In *CGF*, vol. 30, 1583–1592.

MEYER, R. R. 1976. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *J. Comput. Syst. Sci. 12*, 1 (Feb.), 108–121.

OVSJANIKOV, M., MÉRIGOT, Q., MÉMOLI, F., AND GUIBAS, L. 2010. One point isometric matching with the heat kernel. In *CGF*, vol. 29, 1555–1564.

OVSJANIKOV, M., BEN-CHEN, M., SOLOMON, J., BUTSCHER, A., AND GUIBAS, L. 2012. Functional maps: A flexible representation of maps between shapes. *ACM Trans. Graph. 31*, 4 (July), 30:1–30:11.

PACHAURI, D., KONDOR, R., AND SINGH, V. 2013. Solving the multi-way matching problem by permutation synchronization. In *Proc. NIPS*. 1860–1868.

PARDALOS, P., AND WOLKOWICZ, H., Eds. 1994. *Quadratic Assignment and Related Problems*. AMS.

PAULY, M., 2015. The beauty of geometry (SGP keynote), July.

QUADRIANTO, N., SONG, L., AND SMOLA, A. J. 2009. Kernelized sorting. In *Proc. NIPS*, vol. 21. 1289–1296.

RANGARAJAN, A., GOLD, S., AND MJOLSNESS, E. 1996. A novel optimizing network architecture with applications. *Neural Comp. 8*, 5, 1041–1060.

RANGARAJAN, A., YUILLE, A. L., GOLD, S., AND MJOLSNESS, E. 1997. A convergence proof for the softassign quadratic assignment algorithm. *Proc. NIPS*, 620–626.

RANGARAJAN, A., YUILLE, A., AND MJOLSNESS, E. 1999. Convergence properties of the softassign quadratic assignment algorithm. *Neural Comput. 11*, 6 (Aug.), 1455–1474.

RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 2000. The earth mover's distance as a metric for image retrieval. *IJCV 40*, 2 (Nov.), 99–121.

RUSTAMOV, R. M., OVSJANIKOV, M., AZENCOT, O., BEN-CHEN, M., CHAZAL, F., AND GUIBAS, L. 2013. Map-based exploration of intrinsic shape differences and variability. *TOG 32*, 4 (July), 72:1–72:12.

SAHNI, S., AND GONZALEZ, T. 1976. P-complete approximation problems. *J. ACM 23*, 3 (July), 555–565.

SAHNI, S. 1974. Computationally related problems. *SIAM J. Comp. 3*, 4, 262–279.

SCHWARTZBURG, Y., TESTUZ, R., TAGLIASACCHI, A., AND PAULY, M. 2014. High-contrast computational caustic design. *ACM Trans. Graph. 33*, 4 (July), 74:1–74:11.

SMOLA, A., GRETTON, A., SONG, L., AND SCHÖLKOPF, B. 2007. A Hilbert space embedding for distributions. In *Algorithmic Learning Theory*, vol. 4754 of *LNCS*. 13–31.

SOLOMON, J., NGUYEN, A., BUTSCHER, A., BEN-CHEN, M., AND GUIBAS, L. 2012. Soft maps between surfaces. *CGF 31*, 5 (Aug.), 1617–1626.

SOLOMON, J., GUIBAS, L., AND BUTSCHER, A. 2013. Dirichlet energy for analysis and synthesis of soft maps. In *CGF*, vol. 32, 197–206.

SOLOMON, J., RUSTAMOV, R., GUIBAS, L., AND BUTSCHER, A. 2014. Earth mover's distances on discrete surfaces. *ACM Trans. Graph. 33*, 4 (July), 67:1–67:12.

SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T., AND GUIBAS, L. 2015. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *TOG 34*, 4 (July), 66:1–66:11.

STURM, K.-T. 2012. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. Preprint 1208.0434, arXiv.

SUN, J., QU, Q., AND WRIGHT, J. 2015. When are nonconvex problems not scary? *arXiv:1510.06096*.

TANGELDER, J. W., AND VELTKAMP, R. C. 2008. A survey of content based 3d shape retrieval methods. *Multimedia Tools Appl. 39*, 3 (Sept.), 441–471.

THAKOOR, N., GAO, J., AND JUNG, S. 2007. Hidden Markov model-based weighted likelihood discriminant for 2-D shape classification. *Trans. Image Proc. 16*, 11, 2707–2719.

TROPP, J. A., 2003. An alternating minimization algorithm for non-negative matrix approximation.

URAKAWA, H. 2013. *Calculus of Variations and Harmonic Maps*. Translations of Math. Monographs. AMS.

VAN KAICK, O., ZHANG, H., HAMARNEH, G., AND COHEN-OR, D. 2011. A survey on shape correspondence. *CGF 30*, 6, 1681–1707.

VILLANI, C. 2003. *Topics in Optimal Transportation*. Graduate studies in Math. AMS.

WALTZ, R. A., MORALES, J. L., NOCEDAL, J., AND ORBAN, D. 2006. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Math. Program. 107*, 3 (July), 391–408.

WEI, L., HUANG, Q., CEYLAN, D., VOUGA, E., AND LI, H. 2015. Dense human body correspondences using convolutional networks. *CoRR abs/1511.05904*.