(Probably) Concave Graph Matching

Haggai Maron **Yaron Lipman**

Weizmann Institute of Science

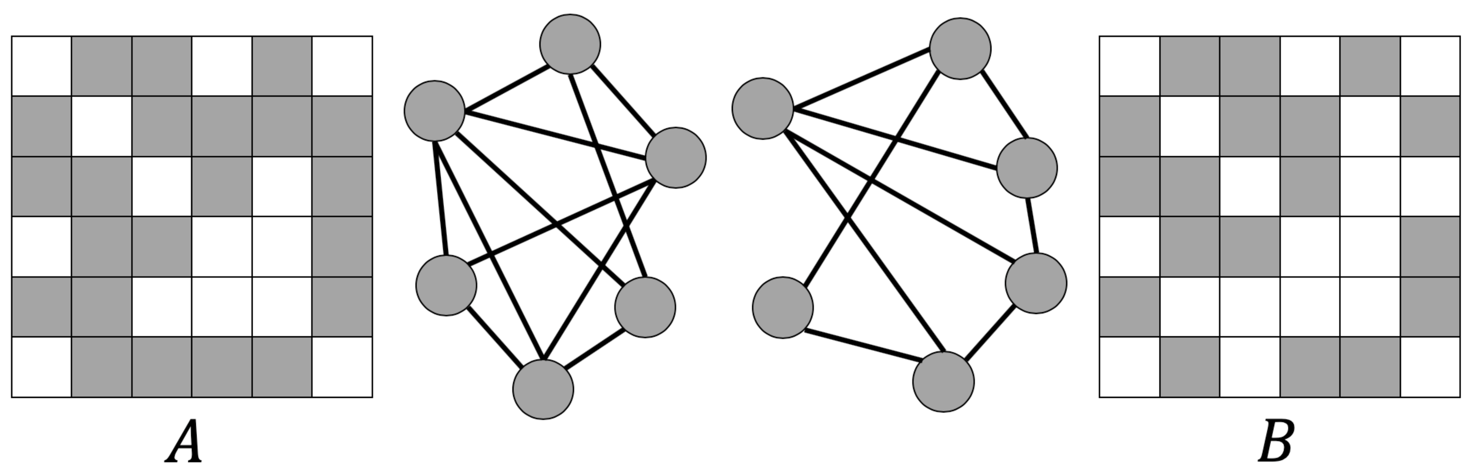
Background

Graph matching is a generic and popular modeling tool for problems in computational sciences such as computer vision, computer graphics, medical imaging and machine learning.

In general, graph matching refers to several different optimization problems of the form:

$$\min_X \; E(X) \quad \text{s.t.} \quad X \in \mathcal{F} \tag{1}$$

where $\mathcal{F} \subset \mathbb{R}^{n \times n_0}$ is a collection of *matchings* between vertices of two graphs G_A and G_B. Usually, $E(X) = [X]^T M[X] + a^T[X]$ is a quadratic function in X and $\mathcal{F}$ is the set of permutation matrices. Often, M quantifies the discrepancy between edge affinities exerted by the matching X. Edge affinities are represented by symmetric matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n_0 \times n_0}$.



Multiple works have considered the relaxation:

$$E(X) = -\mathrm{tr}(BX^TAX) \quad \text{s.t.} \quad X \in \mathrm{DS} \tag{2}$$

Where DS is the set of doubly-stochastic matrices, the convex-hull of the permutation matrices. This energy measures the inner product of the relabeled affinity matrix XAX^T with the matrix B. For example if the two graphs are isomorphic the minimum of this functional is obtained by the permutation X that represents this isomorphism.

E is in general indefinite, resulting in a non-convex relaxation. In case E is *concave* the relaxation enjoys three key properties:

- Its solution set is the same as the original problem.
- Its set of local optima are all permutations.
- For every descent direction the maximal step is optimal.

[Vogelstein et al., 2015, Lyzinski et al., 2016] suggest to locally optimize this relaxation with the Frank-Wolfe algorithm and motivate it by proving that for a specific class of adjacency matrices, the optimal solution of the relaxation almost always coincides with the GM optimal solution. [Vestner et al., 2017, Boyarski et al., 2017] make the useful observation that E is *concave* for some important cases of affinities such as the heat kernels and Gaussians.

This work

In this paper we introduce the concepts of *conditionally concave* and *probably conditionally concave* energies $E(X)$ and show that these energies encapsulate many practical instances of the graph matching problem, including matching Euclidean graphs and graphs on surfaces.

Conditionally concave energies

Conditionally concave energy $E(X)$ means that the restriction of the Hessian M of the energy E to the linear space

$$\mathrm{lin}(\mathrm{DS}) = \left\{ X \in \mathbb{R}^{n \times n} \mid X\mathbf{1} = 0, X^T\mathbf{1} = 0 \right\} \tag{3}$$

is negative definite, Where $\mathrm{lin}(\mathrm{DS})$ is the linear part of the affine-hull of the doubly-stochastic matrices.

Our first result shows that there is a large class of affinity matrices resulting in conditionally concave E.

Proposition 1. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$, $\Psi : \mathbb{R}^s \to \mathbb{R}$ be both conditionally positive (or negative) definite functions of order 1. For any pair of graphs with affinity matrices $A, B \in \mathbb{R}^{n \times n}$ so that

$$A_{ij} = \Phi(x_i - x_j), \quad B_{ij} = \Psi(y_i - y_j) \tag{4}$$

for some arbitrary $\{x_i\} \subset \mathbb{R}^d$, $\{y_i\} \subset \mathbb{R}^s$, the energy $E(X)$ is conditionally concave.

Proof idea: we show that the projection onto $\mathrm{lin}(\mathrm{DS})$ can be obtained as the Kronecker product of a projection onto $\{\mathbf{1}\}^\perp$.

Application. Matching graphs with Euclidean affinities:

$$A_{ij} = \|x_i - x_j\|_2, \quad B_{ij} = \|y_i - y_j\|_2 \tag{5}$$

Includes distances that can be isometrically embedded in Euclidean spaces such as diffusion distances, distances induced by deep learning embeddings, Mahalanobis distances and the spherical distances.

Probably conditionally concave energies

An energy E is called *probably conditionally concave* if it is rare to find a linear subspace D of $\mathrm{lin}(\mathrm{DS})$ so that the restriction of E to it is convex, that is $M|_D \succeq 0$. The following theorem bounds the probability of finding uniformly at random a linear subspace D such that the restriction of $M \in \mathbb{R}^{m \times m}$ to D is convex.

Proposition 2. Let $M \in \mathbb{R}^{m \times m}$ be a symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_m$. Then, for all $t \in (0, \frac{1}{2\lambda_{\max}})$:

$$P_r(M|_D \succeq 0) \le \prod_{i=1}^{m} (1 - 2t\lambda_i)^{-\frac{d}{2}}, \tag{6}$$

Example. Let U be a unitary matrix and consider the matrix:

$$U \cdot \mathrm{diag}\Big(\overbrace{-1, -1, \ldots, -1}^{0.51m}, \overbrace{1, 1, \ldots, 1}^{0.49m}\Big) \cdot U^T$$

It is extremely unlikely to sample a convex direction in dimension $m \approx 300^2$, *i.e.*, the probability will be $\approx 4 \cdot 10^{-5}$.

Application. For example, matching problems in which the affinities A, B describe geodesic distances on surfaces result in probably conditionally concave E.

We further use equation (6) to prove the following:

Proposition 3. There exists a rather general probability space of Hessians M so that the probability of local minima of $E(X)$ to be outside $\mathcal{F}$ is very small.

This probability space is of the form

$$\Omega_m = \left\{ U\Lambda_{m,p}U^T \mid U \in O(m) \right\}, \quad \Lambda_{m,p} = \Big(\overbrace{\lambda_1, \lambda_2, \ldots}^{(1-p)m}, \overbrace{\mu_1, \mu_2, \ldots}^{pm}\Big)$$

where $\lambda_i \le -b$, $b > 0$ are the negative eigenvalues; $0 \le \mu_i \le a$, $a > 0$ are the positive eigenvalues, the ratio of positive to negative eigenvalues is a constant $p \in (0, 1/2)$ and $a \le b$.

Graph matching with one sided permutations

In case we want to optimize over the one-sided permutations, namely $\mathcal{F} = \{X \in \{0,1\}^{n \times n_0} \mid X\mathbf{1} = \mathbf{1}\}$, some of the previous energies might not yield concave relaxations. In this case we devise a variation of the Frank-Wolfe algorithm using a *concave search* procedure. That is, in each iteration, instead of standard line search we subtract a convex energy from $E(X)$ that is constant on $\mathcal{F}$ until we find a descent step.

Experiments

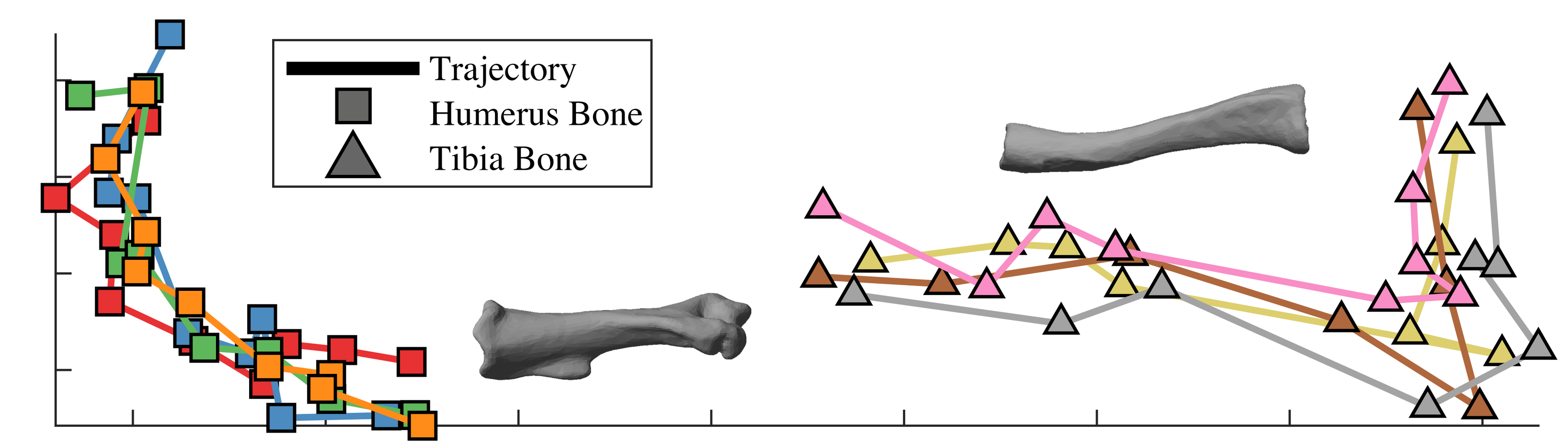


Figure 1: Anatomical dataset embedding in the plane. Squares and triangles represent different bone types, lines represent temporal trajectories.

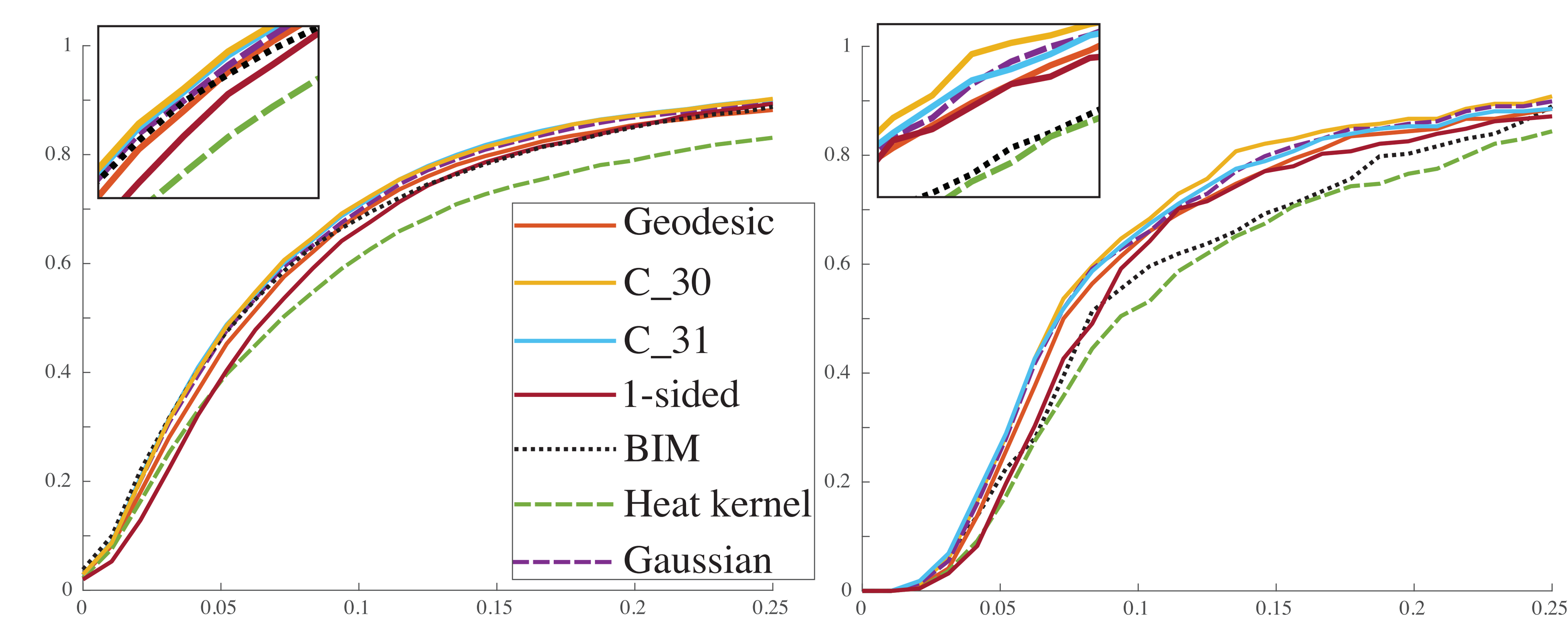


Figure 2: (a) SHREC07 benchmark: Cumulative distribution functions of all errors (left) and mean error per shape (right).

References

[Boyarski et al., 2017] Boyarski, A., Bronstein, A., Bronstein, M., Cremers, D., Kimmel, R., Lähner, Z., Litany, O., Remez, T., Rodola, E., Slossberg, R., et al. (2017). Efficient deformable shape correspondence via kernel matching. *arXiv preprint arXiv:1707.08991*.

[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*.

[Vestner et al., 2017] Vestner, M., Litman, R., Rodola, E., Bronstein, A., and Cremers, D. (2017). Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6681--6690. IEEE.

[Vogelstein et al., 2015] Vogelstein, J. T., Conroy, J. M., Lyzinski, V., Podrazik, L. J., Kratzer, S. G., Harley, E. T., Fishkind, D. E., Vogelstein, R. J., and Priebe, C. E. (2015). Fast approximate quadratic programming for graph matching. *PLOS one*, 10(4):e0121002.