Stability of entropic Wasserstein barycenters and application to random geometric graphs

Marc THENEVEAU[1] Nicolas KERIVEN[2]

[1]École polytechnique, Palaiseau, France

[2]CNRS, IRISA, Univ. Rennes 1, Rennes, France

Résumé – L'intérêt pour les graphes s'étant accru ces dernières années, le calcul de divers outils géométriques est devenu essentiel. Dans certains domaines tels que le traitement des maillages, ces outils reposent souvent sur le calcul des géodésiques et des plus courts chemins dans les manifolds discrétisés. Un exemple récent d'un tel outil est le calcul des barycentres de Wasserstein (WB), une notion très générale de barycentres dérivée de la théorie du Transport Optimal, et de leur variante entropique-régularisée. Dans cet article, nous examinons comment les WB sur des maillages discrétisés sont liés à la géométrie de la variété sous-jacente. Nous fournissons d'abord un résultat de stabilité générique en ce qui concerne les matrices de coût d'entrée. Nous appliquons ensuite ce résultat aux graphes géométriques aléatoires sur les variétés, dont les plus courts chemins convergent vers des géodésiques, ce qui prouve la cohérence des WB calculés sur des formes discrétisées.

Abstract – As interest in graph data has grown in recent years, the computation of various geometric tools has become essential. In some area such as mesh processing, they often rely on the computation of geodesics and shortest paths in discretized manifolds. A recent example of such a tool is the computation of *Wasserstein barycenters* (WB), a very general notion of barycenters derived from the theory of Optimal Transport, and their entropic-regularized variant. In this paper, we examine how WBs on discretized meshes relate to the geometry of the underlying manifold. We first provide a generic stability result with respect to the input cost matrices. We then apply this result to random geometric graphs on manifolds, whose shortest paths converge to geodesics, hence proving the consistency of WBs computed on discretized shapes.

1 Introduction

Graphs, and their variants, are becoming increasingly popular in machine learning and signal processing to represent many kinds of data [8], from social or computer networks to molecules and proteins, three-dimensional shapes, and so on. In some areas, graphs are usually associated with the representation of an underlying "geometry". For instance, the study of Graph Neural Networks and their variants is at the origin of the very active domain of Geometric Deep Learning [3], and the analysis of such "geometric" (random) graphs and their limit is encountered in many domains of data science [15].

In the same fashion, Optimal Transport (OT) [14] is a powerful theory that defines geometrically-meaningful distances and mappings, that can be applied to graph-structured data [9]. A resurgence in data science has recently been experienced, mostly due to novel, efficient computation methods [14], for instance based on entropic regularization [6]. Among the many applications derived from OT, Wasserstein barycenters (WB) [1] are powerful tools to compute meaningful geometric means between measures that can represent very general objects. They have found applications in imagery, statistics, machine learning [7], signal processing [16] and so on. Moreover, they are also amenable to fast computations [2].

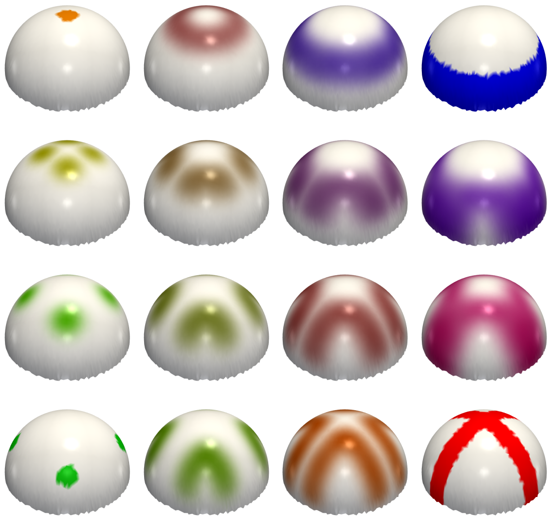
In this paper, we examine some theoretical properties of Wasserstein barycenters on irregular domains such as (random) graphs, where the ground *cost function* may be noisy and converge to some (unknown) limit. We show that WBs are stable to deformations of the cost matrices that represent the distances in the space, more so when entropic regularization is used. We then apply these results on random geometric graphs, where the shortest paths are known to converge toward the geodesic distances on an underlying manifold. As a result, this guarantees for instance that WBs computed on properly discretized 3D shapes with respect to the shortest paths indeed converge toward the "true" WBs (Fig. 1).

Outline. In Sec. 2, we start by preliminary materials on OT and Wasserstein barycenters. In Sec. 3, we give a generic stability results of Wasserstein barycenters to deformation cost, before presenting an application on random geometric graphs in Sec. 4 with some numerical illustrations. The code to reproduce the figures is available at `https://github.com/nkeriven/otrg`. Technical proofs are provided in the Appendix, available at [17].

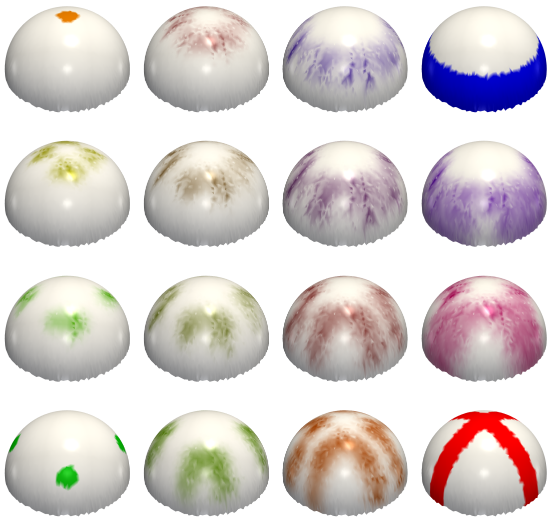
Related Work. Stability of (classical) OT has been mostly studied w.r.t. the input measures, since an important goal is to understand the convergence speed of OT when replacing the measures by a sampled version [11]. There are a few results on the stability w.r.t. cost deformation, with some applications on random graphs [9]. For WBs, stability w.r.t. the input measures has been studied [4], but to our knowledge stability w.r.t. cost deformation is novel. We remark that a recent preprint [5] (published after the preprint version of this paper [17]) includes improved variations of our results with a completely different proof technique.

The relationship between shortest paths on geometric graphs and geodesics on manifolds has been long established, with many applications in shape and graph analysis [13]. OT on shapes has been explored empirically and theoretically [9], and WBs have found applications in imagery, for instance for texture mixing. The theoretical properties of WBs on manifolds has been thoroughly explored e.g. in [10], but results pertaining to the infinite-node limit of discretized manifolds such as the one presented here are, to our knowledge, novel.

Notations. We define the scalar product between two matrices by $< A, B >= tr(A^t B)$. The probability simplex is $\Delta_+^n = \{a \in \mathbb{R}_+^n; \sum_i a_i = 1\}$. The norm $\|\cdot\|_\infty$ refers to the maximal element both for vectors and matrices. Real functions are applied to vectors and matrices element-by-element, for instance e^A or $\log(A)$.



(a) Barycenters with the true geodesics (known for the sphere).



(b) Barycenters with the shortest paths in a random graph.

Figure 1 – Interpolation between $S = 4$ distributions on a half sphere, for varying weights λ_s, with true geodesics (top) or estimated ones on a random graph (bottom). Sec. 3 and 4 prove the convergence of the noisy barycenters to the true ones when the number of nodes goes to infinity.

2 Background on Optimal Transport

Let us start by recalling some background material on discrete Optimal Transport. We consider two finite distributions $a \in \Delta_+^n$ and $b \in \Delta_+^m$ as well as a *cost matrix* $C \in \mathbb{R}_+^{n\times m}$. Usually (but technically not necessarily!), a and b are *weights* associated to two sets of points $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ and C_{ij} indicates how "costly" it is to transport mass from x_i to y_j, often through a metric d elevated to some power $C_{ij} = d(x_i, y_j)^p$. The x_i and y_j can live in different spaces, as long as d is properly defined.

We denote by $U_{a,b} = \{T \in \mathbb{R}_+^{n\times m}, T1_m = a, T^\top 1_n = b\}$ the set of *couplings* between a and b. The OT distance between a and b is defined as:

$$W_C(a, b) := \min_{T\in U(a,b)} < T, C > . \tag{1}$$

When $C_{ij} = d(x_i, y_j)^p$, $W_C(a, b)^{1/p}$ is the so-called p-Wasserstein distance between the measures $\sum_i a_i \delta_{x_i}$ and $\sum_j b_j \delta_{y_j}$. However, note that *only the knowledge of* a, b, C is necessary to compute $W_C(a, b)$.

Computing (1) is a linear problem, which makes it difficult to solve at large scale. This can be handled by adding *entropic regularization* to the cost function [6]: for $\epsilon \geq 0$,

$$W_C^\epsilon(a, b) := \min_{T\in U(a,b)} W_C^\epsilon(a, b, T)$$
$$\text{where} \quad W_C^\epsilon(a, b, T) :=< T, C > -\epsilon H(T) \tag{2}$$

where $H(T) = -\sum_{i,j} T_{i,j} \log T_{i,j}$ with the convention that $0 \log 0 = 0$ by continuity. The resulting problem is strictly convex when $\epsilon > 0$, and can be solved efficiently by a numbers of methods [14], including the celebrated Sinkhorn's algorithm [6]. When $\epsilon \to 0$, the problem converges (in various ways) to the unregularized one (1) [14].

In this paper, we examine so-called *Wasserstein barycenters*. Consider S discrete measures $b_s \in \Delta_+^{m_s}$ of size m_s, along with S cost matrices $C_s \in \mathbb{R}_+^{n\times m_s}$ that indicate the transportation cost from each b_s *to a common space of size* n. The "barycenter" of the b_s is thus a measure $a \in \Delta_+^n$. Given weights $\lambda \in \Delta_+^S$, it is computed by a Fréchet mean w.r.t. the W^ϵ distance: denoting $\Theta = \{\lambda_s, b_s, C_s\}_{s=1}^S$ for short,

$$B^\epsilon(\Theta) := \min_{a\in\Delta_+^n} B^\epsilon(\Theta, a)$$
$$\text{where} \quad B^\epsilon(\Theta, a) := \sum_{s=1}^S \lambda_s W_{C_s}^\epsilon(a, b_s) \tag{3}$$

This is a smooth convex optimization problem [14], with a unique minimizer when $\epsilon > 0$, that we denote by a^Θ. When $\epsilon > 0$, it can be computed by a variant of Sinkhorn's algorithm [14, Chap. 9]. As before, generally (but, again, not necessarily) b_s represent the weights of a discrete measure over positions $\{y_{1s}, \dots, y_{m_s s}\}$, the sought-after barycenter a is over some positions $\{x_1, \dots, x_n\}$, and the cost matrices are defined with metrics $C_{i,j,s} = d_s(x_i, y_{js})^p$. Again, the spaces in which y_{js}, x_i live need *not* be the same, as long as the metrics d_s are over the appropriate domains.

In the next section, we examine the stability of this problem to perturbations of the cost matrices $C_s \in \mathbb{R}^{m_s\times n}$, before presenting an application on random geometric graphs on manifolds in Sec. 4.

3 Stability of Wasserstein barycenters

We study the stability of Wasserstein barycenters (3) to perturbations of the cost matrices C_s. In the rest of the section, we denote $\Theta = \{\lambda_s, b_s, C_s\}_{s=1}^S$ and $\tilde\Theta = \{\lambda_s, b_s, \tilde C_s\}_{s=1}^S$ with the same λ_s, b_s but perturbed cost matrices $\tilde C_s$.

Our first result guarantees closeness of the cost function for any regularization level, including $\epsilon = 0$. It does *not*, however, guarantee proximity of the optimal barycenters. The proof, presented in the Appendix [17], is straightforward.

Proposition 1. *For all $\epsilon \geq 0$, we have*

$$\left| B^\epsilon(\Theta) - B^\epsilon(\tilde\Theta) \right| \leq \sum_s \lambda_s \| C_s - \tilde C_s \|_\infty \tag{4}$$

Hence, if all matrices $\tilde C_s$ converge to C_s in ∞-norm, the cost functions B^ϵ converge to one another. However, this proposition does not provide stability of the barycenter a^Θ itself, which is what interests us in practice. For this we need strict convexity of the problem, which holds only when $\epsilon > 0$. The following theorem is then our main result. Recall that a^Θ is the optimal barycenter in (3).

Theorem 1. *Assume $0 \leq c_{\min} \leq C_{sij}, \tilde C_{sij} \leq c_{\max}$ hold for all s, i, j. For all $\epsilon > 0$ we have*

$$\|a^\Theta - a^{\tilde\Theta}\|_2^2 \lesssim \epsilon e^{3(c_{\max} - c_{\min})/\epsilon} \sum_s \lambda_s \|C_s - \tilde C_s\|_\infty \tag{5}$$

We therefore obtain stability of the optimal barycenters, with a potentially large multiplicative constant in ϵ. We also note that the bound is insensitive to shifting the costs C_s and $\tilde C_s$ by a constant c (which shifts $c_{\min}$ and $c_{\max}$), which is to be expected since this shifts W_C by the same constant and does not affect the transport plans [9].

In the next section, we apply this result to the approximation of Wasserstein barycenters on manifolds. The rest of this section is dedicated to a sketch of proof of this theorem, whose details can be found in Appendix [17].

Sketch of proof. As usual in convex optimization, we work with the dual problem of (3), which can be found e.g. in [14, Chap. 9]. Our proof is available on the arxiv version of the article [17] and consist in three parts. First, we bound the solutions of the dual problem with respect to the parameters of the problem. Then, we bound the error of the optimal solution with the error of $\mathcal{L}$, using strict convexity. Finally, we bound the error of $\mathcal{L}$ with the norm between cost matrices.

4 Random geometric graphs

A classical approach to manipulating manifolds such as 3D shapes is to discretize them, for instance by constructing a random geometric graph [12]. This is done by randomly drawing N points on the manifold and connecting them if their distance in the ambient space is less than a certain h_N which tends to 0 when N tends to infinity. It is then known [9] that the length of the shortest paths in the graph converge, under some conditions, to the geodesic distance of the manifold. More precisely, assume that we have a compact, smooth submanifold $\mathcal{M} \subset \mathbb{R}^d$ of dimension k, without boundary for simplicity. Its geodesic distance is $d(x, y)$, while $\|x - y\|$ refers to the norm in the ambient space $\mathbb{R}^d$. Its diameter is $D_\mathcal{M} := \sup_{u,t\in\mathcal{M}} d(u, t)$.

Consider the following objects: for $1 \leq s \leq S$, distributions $\nu_s \in \Delta_+^{d_s}$, weights λ_s, and m_s supporting points $\{y_{s1}, \dots, y_{sm_s}\} \subset \mathcal{M}$ on the manifold, for each distribution (they need not be distinct). Then, the n supporting points $\{x_1, \dots, x_n\} \subset \mathcal{M}$ on which we are going to compute the barycenter $a \in \Delta_+^n$. Finally, we complete with N additional points $\{z_1, \dots, z_N\} \subset \mathcal{M}$ *drawn i.i.d.* according to some distribution P on $\mathcal{M}$, which we assume to have a density p_z with respect to the uniform measure on $\mathcal{M}$, bounded away from zero: $p_z(z) \geq c_z > 0$. We then construct a random graph with radius h_N on $\mathcal{M}$ using all the points $\mathcal{V} := \{x_i, y_{sj}, z_\ell\}_{ij\ell}$: if any two such points $u, t \in \mathcal{V}$ satisfy $\|u - t\| \leq h_N$, then we add an edge between them. Note that here x_i and y_{sj} are deterministic, while z_ℓ are random. We let $N \to \infty$, and $h_N \to 0$, and aim to prove that the shortest paths length between x_i, y_{sj} converge to the geodesic distance. See Fig. 2.

For $p \geq 1$, we denote by $C_s = [d(x_i, y_{sj})^p]_{ij} \in \mathbb{R}^{n \times m_s}$ the matrices containing the true geodesic distances between our points of interest elevated to some power p, with respect to which we want to compute Wasserstein barycenters. We then denote by $SP(u, t)$ the shortest path (minimal number of edges) in the graph between two vertices u, t, with $SP(u, t) = +\infty$ if they are not connected. We define

$$\tilde C_s = [(h_N SP(x_i, y_{sj}))^p] \tag{6}$$

the matrices containing the shortest paths between x_{is} and y_j, normalized by h_N. Then, the following result is from [9].

Lemma 1 (Theorem 2 in [9])**.** *Consider u, t two vertices among the fixed points $\{x_i, y_{sj}\}$, and $\rho > 0$. For N large enough, with probability $1 - \rho$, we have*

$$|d(u, t) - h_N SP(u, t)| \lesssim h_N + \left(\frac{\log \frac{1}{h_N \rho}}{c_z N h_N^k} \right)^{1/k} \tag{7}$$

where the multiplicative constant depends on $\mathcal{M}$.

This convergence translates into the convergence of the cost matrix of the graph to the cost matrix of the manifold. Hence, using a union bound and the results of Sec. 3, we immediately obtain the following corollary.

Corollary 1. *With probability $1 - \rho$, for all $\epsilon > 0$ we have*

$$\|a^\Theta - a^{\tilde\Theta}\|_2^2 \lesssim p D_\mathcal{M}^{p-1} \epsilon e^{\frac{6 D_\mathcal{M}}{\epsilon}} \left(h_N + \left(\frac{\log \frac{n \sum_s m_s}{h_N \rho}}{c_z N h_N^k} \right)^{\frac{1}{k}} \right)$$

In other words, as long as when $N \to \infty$:

$$h_N \to 0, \quad \frac{N h_N^k}{\log(1/h_N)} \to +\infty$$

then the barycenters computed using the shortest paths in the graph converge to the barycenters that use the true geodesic distance d, see Fig. 2. Note that on a k-manifold the average degree of a random geometric graph is $O(N h_N^k)$, so here the average degree needs to increase to $+\infty$ (the graph is not *sparse*), at least by a logarithmic factor.

Numerical illustration. In Fig. 1 and 2, we illustrate our results on two examples of discretized manifolds: a sphere, where the true geodesics are known and we directly compare the effect of the discretization, and a complicated 2D domain (note that the latter technically has a boundary, while our theoretical results required the absence of boundaries for simplicity.

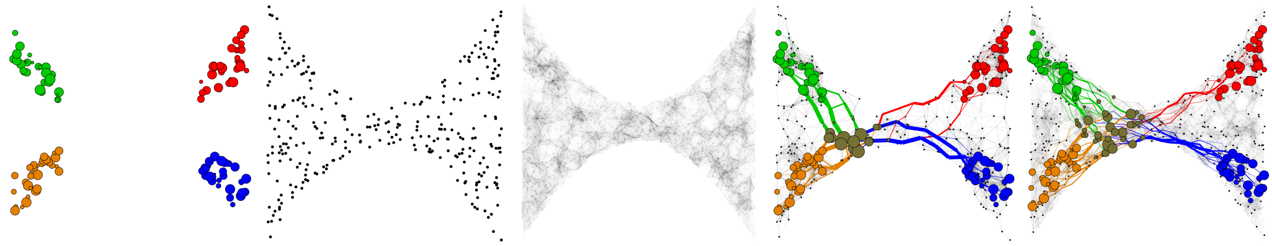


Figure 2 – Example of Wasserstein barycenters in a random graph on a 2D domain. From left to right: distributions ν_s located at points y_{sj}; support points for the barycenter x_i; random geometric graph constructed with x_i, y_{sj}, as well as additional random points $z_1, \dots, z_N$; visualization of the barycenters and the transport plans for two values $N = 10$ and $N = 2000$.

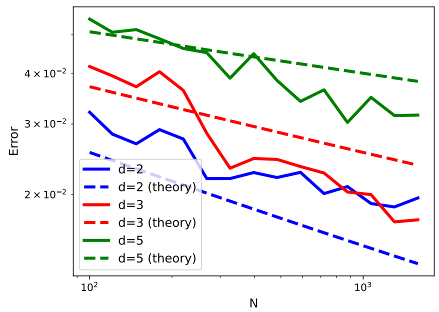


Figure 3 – Error between true WBs on the (hyper-)sphere and barycenters computed with shortest paths, w.r.t. N, for several dimension d, averaged over 30 experiments.

They still seem to be empirically valid). We use $p = 2$, and compute the entropic Wasserstein barycenters with $\epsilon > 0$ using a variant of Sinkhorn's algorithm [14, Chap. 9]. In Fig. 3, we compute $\|a^{\Theta} - a^{\tilde{\Theta}}\|_2^2$ on the sphere, w.r.t. N, and compare with the theoretical rates given by Cor. 1. The bounds appears to be, as expected, quite loose, and the problem quite noisy.

5 Conclusion

In this paper, we have shown the stability of entropic WBs with respect to the cost matrices. We then gave an application to random geometric graphs for which the shortest paths converge to the geodesics of the underlying manifold, guaranteeing for instance the convergence of WBs on discretized 3D shapes. Our theoretical work hints at many potential outlooks. Other models of random graphs could be treated [9], with different applications. Finally, we have assumed fixed the supporting points of the distributions and barycenters, while the stability of WBs to sampling the target measures has recently been shown [4]. Combining the results would finalize the link between continuous [10] and discretized WBs.

Acknowledgement

This work was funded by ANR GRandMa, ANR-21-CE23-0006.

References

[1] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] Jason M. Altschuler and Enric Boix-Adsera. Wasserstein barycenters can be computed in polynomial time in fixed dimension. *Journal of Machine Learning Research*, 22:1–19, 2021.

[3] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. 2021.

[4] Guillaume Carlier, Alex Delalande, and Quentin Merigot. Quantitative Stability of Barycenters in the Wasserstein Space. pages 1–24, 2022.

[5] Lénaïc Chizat. Doubly regularized entropic wasserstein barycenters, 2023.

[6] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances. pages 1–9, 2013.

[7] Nhat Ho, Xuan Long Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, and Dinh Phung. Multilevel clustering via wasserstein means. *34th International Conference on Machine Learning, ICML 2017*, 3:2363–2378, 2017.

[8] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open Graph Benchmark: Datasets for Machine Learning on Graphs. *Neural Information Processing Systems (NeurIPS)*, (NeurIPS):1–34, 2020.

[9] Nicolas Keriven. Entropic optimal transport in random graphs, 2022.

[10] Young Heon Kim and Brendan Pass. Wasserstein barycenters over Riemannian manifolds. *Advances in Mathematics*, 307:640–683, 2017.

[11] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32:1–23, 2019.

[12] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2008.

[13] Gabriel Peyré, Mickael Péchaud, Renaud Keriven, and Laurent D. Cohen. Geodesic methods in computer vision and graphics. *Foundations and Trends in Computer Graphics and Vision*, 5(3-4):197–397, 2010.

[14] Gabriel Peyré and Marco Cuturi. Computational optimal transport. 2018.

[15] Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11:905–934, 2010.

[16] E. Simou, D. Thanou, and P. Frossard. node2coords: Graph representation learning with wasserstein barycenters. *IEEE Transactions on Signal and Information Processing over Networks*, 7:17–29, 2021.

[17] Marc Theveneau and Nicolas Keriven. Stability of Entropic Wasserstein Barycenters and application to random geometric graphs. *Arxiv preprint arXiv:2210.10535*, 2022.