GEODESIC SINKHORN FOR FAST AND ACCURATE OPTIMAL TRANSPORT ON MANIFOLDS

Guillaume Huguet¹,⋆ Alexander Tong¹,⋆ María Ramos Zapatero²
Christopher J. Tape² Guy Wolf¹ Smita Krishnaswamy³

¹ Université de Montréal; Mila - Quebec AI Institute
² University College London Cancer Institute ³ Yale University

ABSTRACT

Efficient computation of optimal transport distance between distributions is of growing importance in data science. Sinkhorn-based methods are currently the state-of-the-art for such computations, but require \(O(n^2)\) computations. In addition, Sinkhorn-based methods commonly use an Euclidean ground distance between datapoints. However, with the prevalence of manifold structured scientific data, it is often desirable to consider geodesic ground distance. Here, we tackle both issues by proposing Geodesic Sinkhorn—based on diffusing a heat kernel on a manifold graph. Notably, Geodesic Sinkhorn requires only \(O(n \log n)\) computation, as we approximate the heat kernel with Chebyshev polynomials based on the sparse graph Laplacian. We apply our method to the computation of barycenters of several distributions of high dimensional single cell data from patient samples undergoing chemotherapy. In particular, we define the barycentric distance as the distance between two such barycenters. Using this definition, we identify an optimal transport distance and path associated with the effect of treatment on cellular data.

1. INTRODUCTION

Optimal Transport (OT) distances or Wasserstein distances are computed by lifting ground distances between points to distances between measures. This distance is computed relative to a ground distance on the support of the distributions, making it more informative than distances based only on a pointwise comparison of the densities. However, to compute the Wasserstein, one needs to find the optimal transport plan from the source distribution to a target distribution; this is a linear programming problem requiring \(O(n^3)\) computations. An important drawback of the Sinkhorn algorithm is that it re-scales a Gaussian kernel based on the distance matrix. It is equivalent to the Schrödinger Bridge problem, for which similar algorithms were developed [3, 4, 5]. In the discrete case, it requires \(O(n^2)\) for distributions of size \(n\), since it relies on matrix-vector products. Furthermore, this formulation allows for fast computation of the discrete barycenter with fixed support (the average distributions w.r.t. the Sinkhorn distance). An important drawback of the Sinkhorn algorithm is the necessity of storing and multiplying the pairwise distance matrix with a vector.

Additionally, the ground distance is commonly chosen as the Euclidean distance. The Euclidean distance is often sub-optimal for high-dimensional datasets over larger distances according to the manifold hypothesis, which says observations lie near a low dimensional (curved) manifold in high dimensional space [6]. For higher dimensional datasets assumed to be sampled from a lower dimensional manifold, using a distance closer to the manifold for OT has shown promising results [7, 8, 9, 10].

In this work, we present Geodesic Sinkhorn¹; a Sinkhorn-based method for fast optimal transport with a heat-geodesic ground distance. Our method is based on the geometry of the dataset constructed from a common graph and uses the heat kernel on the graph to define a heat-geodesic distance. Key to this approach, we will never need to construct or operate on an \(n \times n\) distance matrix, and we will only use the sparse Laplacian matrix and sparse matrix-vector products. For sparse graphs, this can be used for \(O(n \log n)\) computation of the Sinkhorn distance with a manifold ground distance, improving on the \(O(n^2)\) implementations based on dense matrices.

Increasing the state-of-the-art efficiency in Sinkhorn computation opens us up to being able to perform complex operations on large groups of datasets. In particular, we consider interpolating between datasets and show that using our heat-geodesic distance improves the interpolation accuracy.

*Equal contribution. This research was partially funded by ESP Mérité [G.H.], NSERC Discovery grant 03267 [G.W.], Canada CIFAR AI Chair [G.W.], NIH grant R01GM135929 [G.W., S.K.], Cancer Research UK (C60693 / A23783) [C.J.T.], the Rosetrees Trust (M872 / A2292) [C.J.T.], and the Yale-UCL Collaborative Student Exchange Programme. Correspondence to: guy.wolf@umontreal.ca, smita.krishnaswamy@yale.edu

1https://github.com/KrishnaswamyLab/GeoSinkhorn
compared to OT with Euclidean distance. The barycenter corresponds to the average distribution of a set of distributions. Our method allows for finer-grained barycenters on a data manifold, which motivates us to define a novel notion of dissimilarity between families of distributions called barycentric distance.

We apply the barycentric distance to single cell data from patient-derived cancer organoids (PDOs) to assess the effect of treatments (such as drugs and chemotherapy). Here we have one set of PDOs from control conditions, and another set that are treated. The treatment effect is thus the distance between these barycenters. In addition, we use Geodesic Sinkhorn’s barycenter to compare the effect from one family of distributions to another.

Our main contributions include: (1) A new method for computing optimal transport distances on a manifold called Geodesic Sinkhorn, which is highly efficient in time and memory. (2) Defining the barycentric distance; a novel distance between families of distributions, and showing its utility in deriving treatment effect from control and treated patient samples.

2. RELATED WORK

Geodesic Sinkhorn is related to prior work linking the entropy-regularized optimal transport problem triangular mesh with the heat operator [11, 8], but using different graph filtering techniques. These approaches approximate the application of the heat kernel to a vector by discretizing the heat equation and solving systems of linear equations. This technique was used in different contexts, either with the cotangent Laplacian [8] or to learn a ground metric [12]. Solving these systems for each Sinkhorn iteration can be done efficiently with a sparse Cholesky decomposition. However, this method’s efficiency depends mainly on the efficiency of the Cholesky decomposition which can be slow depending on the sparsity pattern is \(O(n^3)\) for an \(n \times n\) matrix, and necessitates solving 2K systems of linear equations per Sinkhorn iteration, where \(K\) is the number of sub-steps in the backward Euler discretization.

3. PRELIMINARIES

In this section, we start by reviewing the basics of OT and the Wasserstein distance, as well the Sinkhorn distance. Then we review two notions fundamental to our method; the heat equation on a graph and the Chebyshev approximation of the heat kernel.

3.1. Wasserstein Distance

In the following, we assume that all distributions admit a density or a probability mass function, and we use the same notation for both. Let \(\mu, \nu\) be two probability distributions on a measurable space \(\mathcal{X} \subseteq \mathbb{R}^d\) with metric \(d(\cdot, \cdot)\), let \(\Pi(\mu, \nu)\) be the set of joint probability distributions \(\pi\) on the space \(\mathcal{X} \times \mathcal{X}\) where, for any measurable subset \(\omega \subset \mathcal{X}\), \(\pi(\omega \times \mathcal{X}) = \mu(\omega)\) and \(\pi(\mathcal{X} \times \omega) = \nu(\omega)\). The p-Wasserstein distance is defined as:

\[
W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}^2} d(x, y)^p \, d\pi(x, y)\right)^{1/p}.
\]

(1)

In the following, we consider \(p = 2\). An exact algorithm based on linear programming can solve this problem in \(O(n^3 \log n)\) time for discrete distributions of size \(n\).

3.2. Sinkhorn Distances

The Kullback-Leibler (KL) divergence between \(\pi\) and some strictly positive \(K\) on \(\mathcal{X} \times \mathcal{X}\) is defined as

\[
D_{KL}(\pi|K) := \int_{\mathcal{X}^2} \left(\ln \frac{\pi(x, y)}{K(x, y)} - 1\right) \, d\pi(x, y).
\]

(2)

The Sinkhorn distance\(^2\) is a relaxation of equation 1 where the infimum is over all coupling in \(\{\pi \in \Pi(\mu, \nu) : D_{KL}(\pi|\mu \times \nu) \leq \xi\}\) for \(\xi > -1\). Introduced in [13], the optimization of this distance can be solved by considering the entropy-regularized transport

\[
W^2_{d, \lambda}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}^2} d(x, y)^2 \, d\pi(x, y) - \lambda H(\pi)\right)^{1/2},
\]

(3)

where we define the entropy of a coupling \(\pi\) as \(H(\pi) := -\int \ln \pi(x, y) \, d\pi(x, y)\), and \(\lambda > 0\). This formulation converges to the Wasserstein distance as \(\lambda \to 0\), and can be solved with the Sinkhorn algorithm with complexity of the order \(O(n^2/\epsilon)\) for discrete distributions of size \(n\) [13].

In the discrete case, the transport matrix \(\pi\) admits the form \(\text{diag}(v)K\text{diag}(w)\), where \(v, w\) are vectors of size \(n\). The Sinkhorn algorithm iteratively updates the vectors as \((v, w) \leftarrow (\mu/K(w)\pi(v), \nu/K'\pi(v))\), where \(K_\lambda := e^{-d(x, y)2/\lambda}\).

Following [8], using the kernel \(K_\lambda\) gives an alternative interpretation of the Sinkhorn distance as

\[
W^2_{d, \lambda}(\mu, \nu) = \lambda^{1/2} \left(1 + \min_{\pi \in \Pi(\mu, \nu)} D_{KL}(\pi|K_\lambda)\right)^{1/2}.
\]

(4)

The problem in equation 3 is strictly convex and continuous yielding a unique minimizer. In the discrete case, this leads to an algorithm for the entropy-regularized Wasserstein distance based on the Sinkhorn algorithm enforcing the marginal constraints on the kernel \(K_\lambda\) while minimizing the distance as quantified by \(D_{KL}\).

The underlying metric \(d(\cdot, \cdot)\) is generally unknown, thus the kernel \(K_\lambda\) cannot be evaluated. The authors of [8] proposed to approximate \(K_\lambda\) with the heat kernel \(H_t(x, y)\) on \(\mathcal{X}\).

\(^2\)With a slight abuse of language we use the term distance, although the entropy-regularized formulation does not respect the identity of indiscernibles.
According to Varadhan’s formula [14], the geodesic distance on a manifold can be recovered from the heat transfer at small timescales as

$$d(x, y)^2 = \lim_{t \to 0^+} -4t \ln H_t(x, y). \quad (5)$$

Hence, motivating the use of the heat-geodesic distance $d_H^2(x, y) := -4t \ln H_t(x, y)$, with associated kernel $K(x, y) = H_t(x, y).$ Interestingly, Sinkhorn-based methods admit an efficient algorithm to solve the barycenter problem which we present next.

### 3.3. Interpolation with discrete support

By constraining the support to a set $\mathcal{X}$ (or a graph), we can efficiently interpolate between more than two distributions. The barycenter problem [8, 1, 15] generalizes the notion of average between points to an average between distributions. For a set of $m$ distributions $\{\mu_1, \ldots, \mu_m\}$ supported on $\mathcal{X}$, the objective is to find a distribution minimizing the average distance

$$\mu^* := \arg \min_{\mu \in \mathcal{P}(\mathcal{X})} \sum_{i=1}^m \alpha_i W^p_d(\mu, \mu_i)^2,$$

where $\mathcal{P}(\mathcal{X})$ denotes the space of probability distributions supported on $\mathcal{X}$, and $\{\alpha_1, \ldots, \alpha_m\}$ are non-negative weights. Finding the barycenter is a challenging optimization problem, however the barycenter for Sinkhorn-based methods admits an efficient computation. It involves updating $m$ vectors $v_i, w_i$, which define a transport plan from $\mu_i$ to the barycenter $\mu^*$. The support of the barycenter is constrained to $\mathcal{X}$, for most Sinkhorn-based methods the size of $\mathcal{X}$ needs to be small for computational reason. Our method does not suffer from such a limitation. Hence, we can consider barycenter with greater expressivity, and interpolate between large sets of distributions.

### 3.4. Heat Diffusion on a Graph

Consider an undirected graph $G = (V, E)$ with a set $V$ of $n$ vertices and a set of edges $E$, and its weighted adjacency matrix $A$ with non-negative edge weights, and the diagonal degree matrix $D$, where $D_{ii} := \sum_k A_{ik}$. We define the combinatorial Laplacian as $L := D - A$, for any function $f : V \to \mathbb{R}$ we have $(Lf)(v) = \sum_u a_{u,v}(f(v) - f(u))$. The combinatorial Laplacian is a symmetric positive semi-definite matrix, and has an eigendecomposition $L = \Psi \Lambda \Psi^T$ with orthonormal eigenvectors $\Psi$ and diagonal eigenvalue matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, such that $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. The combinatorial Laplacian is a natural extension of the negative of the Laplacian operator to a graph. For a signal $f_0 \in \mathbb{R}^n$ on $G$, the diffusion of $f_0$ on the graph evolves according to the heat equation

$$\frac{d}{dt}f(t) + Lf(t) = 0, \text{ s.t. } f(0) = f_0 \quad t \in \mathbb{R}^+.$$  

The heat kernel solves this ODE, it is defined by the matrix exponential $H_t := e^{-tL}$. By orthogonality of the eigenvectors of $L$, we can write $H_t = \Psi e^{-t\Lambda} \Psi^T$ and $f(t) = H_tf_0$. Computing $H_t$ by eigendecomposition would require $O(n^3)$ operations. Recall that, for the Sinkhorn algorithm, we are only concerned with the application of the heat operator $H_t$ on a signal $f \in \mathbb{R}^n$. For larger diffusion time, the heat kernel converges to its eigenvector associated to the lowest eigenvalues of the Laplacian, hence, intuitively, the heat kernel corresponds to a low-pass filter. In Geodesic Sinkhorn, we use Chebyshev polynomials [16, 17] to approximate the application of the heat operator to a signal. For a short timescale $t$, using the heat kernel accounts for using the geodesic distance as ground distance in the entropy-regularized OT formulation equation 3.

### 3.5. Chebyshev Polynomials

Polynomial sequences are often used to approximate functions or operator. With Chebyshev polynomials, we can approximate the application of the matrix exponential $H_t := e^{-tL}$ to a signal $f$ on the graph. An attractive property of Chebyshev polynomials is that the approximation error decays exponentially with the maximum degree $K$. They are defined by the recursive relation $\{T_k\}_{k \in \mathbb{N}}$ with $T_0(y) = 0, T_1(y) = y$ and $T_k(y) = 2yT_{k-1}(y) - T_{k-2}(y)$ for $k \geq 2$. On $[-1, 1]$ these polynomials are orthogonal w.r.t. the weight $(1 - y^2)^{-1/2}$, and can be used to express the operator $H_t$. Assuming the largest eigenvalue $\lambda_n \leq 2$, we can write

$$H_t = \frac{b_0}{2} + \sum_{k=1}^{\infty} b_k T_k(L - Id),$$

where the $K + 1$ scalar coefficient $\{b_k\}$ depend on time and can be evaluated with the Bessel function. The approximation of $H_t$ is based on the first $K$ term of the series which we note $p_K(L, t)$. It results in $K$ matrix-vector products which can be efficient since, in general, $L$ is a sparse matrix. On a $m$-nearest neighbor graph, this can be $O(Kmm/\lambda)$, where $\lambda$ is a regularization parameter. Chebyshev polynomials admits interesting theoretical properties and are known to converge faster than other polynomials [17, 18]. The choice of the parameter $K$ is related to the number of neighbors or the connectivity of the graph. For small diffusion time, hence only diffusing in a local neighborhood, the approximation is accurate even with a small $K$. As the diffusion time increases, $K$ has to increase in order to consider a larger neighborhood around a node. For OT, we consider small diffusion time, and we found that our results were stable for all $K$ greater than 10.

### 4. Geodesic Sinkhorn Distances

We define the Geodesic Sinkhorn distance between any signals or distributions on a graph $G$ by the entropy-regularized OT with the heat kernel $H_t$ on the graph. This construction is
Algorithm 1 Geodesic Sinkhorn

Input: \( n \times n \) Laplacian \( L \), distributions (signals) \( \mu, \nu \) on \( G \), maximum Chebyshev degree \( K \), regularization \( \lambda \), \( a \) vertices weights.

Output: \( W^2_{\alpha,\lambda}(\mu, \nu) \)

// Initialization
\( w \leftarrow 1 \)

// Sinkhorn iterations
for \( j = 1, 2, \ldots \) do

\( v \leftarrow \mu \odot p_K(L, t)(\alpha \odot w) \)

\( w \leftarrow \nu \odot p_K(L, t)(\alpha \odot v) \)

end for

Return: \( 4t \sum [\alpha \odot (\mu \odot \ln v + \nu \odot \ln w)] \)

also valid between any point cloud datasets. In that case, for \( m \) datasets \( \{X_1, \ldots, X_m\} \) sampled from a set of distributions \( \{\mu_1, \ldots, \mu_m\} \), we construct a common graph using an affinity kernel on the \( m \) datasets and compare two samples by taking the distance between two indicator functions on the graph. We approximate the heat kernel \( H_t \), with Chebyshev polynomials \( p_K(L, l) \) of order \( K \). In Algorithm 1, we present the main steps to evaluate the Geodesic Sinkhorn. It is based upon Sinkhorn iterations [2, 13], where \( \odot \) and \( \circ \) denote respectively the elementwise division and multiplication. Note that, as opposed to the usual Sinkhorn algorithm, we never have to store a dense \( n \times n \) distance matrix, but only the usually sparse graph Laplacian.

Definition 4.1. The Geodesic Sinkhorn distance between two distributions \( \mu, \nu \) on a graph \( G \) is

\[
W_{H_t}(\mu, \nu) := 4t^{1/2} \left( 1 + \min_{\pi \in \Pi(\mu, \nu)} D_{KL}(\pi | H_t) \right)^{1/2}.
\]

In the following proposition, we find the ground distance implicitly used in the optimal transport defined by Geodesic Sinkhorn. We use \( \simeq \) for the equivalence relation between distances.

Proposition 4.2. There exists a maximum Chebyshev polynomial degree \( K \) such that the ground distance in Geodesic Sinkhorn is equivalent to the one based \( H_t \)

\[-4t \log(p_K(L, t))_{ij} \simeq -4t \log(H_t)_{ij}\]

In particular, the Wasserstein distances with these ground distances are equivalent.

Proof. Because the approximation error decreases exponentially in \( K \) [17], we have that for any \( \epsilon > 0 \) sufficiently small there exist \( K_t \) such that \((H_t)_{ij} - \epsilon < (p_K)_{ij} < (H_t)_{ij} + \epsilon\). Choose \( K \) such that this is true for all vertices \( K := \max\{K_1, \ldots, K_t\} \). We define

\[
c := \min_{ij} \frac{(H_t)_{ij} - \epsilon}{(H_t)_{ij}} \quad \text{and} \quad C := \max_{ij} \frac{(H_t)_{ij} + \epsilon}{(H_t)_{ij}}
\]

and we have the equivalence between the distances since

\[
c(H_t)_{ij} \leq (p_K(L, t))_{ij} \leq C(H_t)_{ij} \quad \text{for all} \quad ij \in [n],
\]

and since the logarithm is a monotonic function.

In [8, 12], using the Euler implicit discretization results in a ground cost of the form \( -\epsilon \ln((\text{Id} - \frac{1}{\epsilon} L)^{-K}) \), where \( \text{Id} \) is the identity matrix, and can be seen as another approximation for the matrix exponential.

The efficiency of Geodesic Sinkhorn improves the notion of barycenter as it is possible to consider much larger graph \( G \), thus a finer grained support of the barycenter. This leads us to define a novel distance between families of distributions.

Definition 4.3. For two finite families of distributions \( T \) and \( C \) supported on \( G \), we define the barycentric distance between the families \( T, C \) as

\[
\gamma(T, C) := W_{H_t}(\mu_T^\gamma, \mu_C^\gamma)
\]

where \( \mu_T^\gamma, \mu_C^\gamma \) are respectively the barycenters of \( T \) and \( C \).

The previous definition is valid for any distances between distributions or barycenters. However, OT barycenters are known to be more informative than others [15]. We will further explore this comparison in our experiments. We use it to distinguish between two groups in a medical setting where a set of patients received a treatment (defining the family \( T \)), and another set acts as a control family \( C \), Following this idea, we define a notion of effect between two families.

Definition 4.4. For two family of distributions \( T \) and \( C \) supported on \( G \), define the Expected Barycenter Effect of \( T \) as

\[
\tau(T) := \mathbb{E}_{\mu_T^\star}(Y_t) - \mathbb{E}_{\mu_C^\star}(Y_c),
\]

where \( \mu_T^\star, \mu_C^\star \) are respectively the barycenters of \( T \) and \( C \), and the features \( Y_c \sim \mu_C^\star \) and \( Y_t \sim \mu_T^\star \).

Note that we compute the average on the family of distributions instead of the average on their support, hence we evaluate their expectations in a closed form. This definition also extends to a conditional equivalent where families of distribution can be subdivided with discrete covariate variables. When the barycenters are computed with the total variation, this definition is equivalent to the naïve Average Treatment Effect (ATE) [19]; i.e. difference of empirical means.

5. Results

We demonstrate the accuracy and efficiency of the Geodesic Sinkhorn distance on two tasks: (1) Nearest-Wasserstein-neighbor calculation on simulated data with manifold structure similar to the setup of [10]; (2) A newly defined Barycentric distance between families of distributions computed to quantify the effect of a treatment on patient-derived organoids.
We test if we can identify a linear treatment effect with the Ex-
pected Barycenter Effect (EBE) with one outlier distribution centered at -60, -30, 0, or 5. Comparison using the barycenter from Sinkhorn, total variation, or Geodesic Sinkhorn. Values closer to the real treatment effect of 5 are better.

### Table 2

| Outlier | EBE Geo Sinkhorn | EBE Sinkhorn | EBE TV |
|---------|-----------------|--------------|--------|
| -60     | 5.016 ± 0.226   | -0.103 ± 0.005 | -1.429 ± 0.144 |
| -30     | 5.053 ± 0.196   | 0.355 ± 0.049  | 1.571 ± 0.144  |
| 0       | 4.917 ± 0.315   | 4.954 ± 0.157  | 4.571 ± 0.144  |
| No outlier | 5.059 ± 0.159   | 5.054 ± 0.16   | 5.071 ± 0.144  |

5.1. Nearest-Wasserstein-neighbor distributions

In this experiment, we compare our method with Sinkhorn [13], and LR Sinkhorn [20], both algorithms with Euclidean and squared Euclidean ground distance, with DiffusionEMD [21], and Sinkhorn with Euler approximation of the heat filter. We created 15 Gaussian distributions sampled randomly on a swiss roll dataset, and sampled 10k observations from each distribution. We rotated the observations in 10 dimensions. We consider a k-nearest neighbors task on these distributions. We evaluate the methods with the ground truth, since we know the exact geodesic distance on the manifold. In Tab. 1, we report the average and standard deviation over 10 seeds of the Spearman and Pearson correlations to the ground truth, and the runtime in seconds with and without the computation of the graph. Our method is the most accurate while being much faster than other Sinkhorn-based methods.

5.2. Barycentric distance

We test if we can identify a linear treatment effect with the Expected Barycenter Effect (EBE). In this experiment, we create a control family of distributions \( C \) of ten standard Gaussian distributions. The treatment group consists of nine Gaussian distributions \( \mathcal{N}(5, 1) \), and one outlier centered at different means. For each distribution, we sample 500 observations, and reproduce the experiment over ten seeds. In Table 2, we report the EBE and its standard deviation with the Geodesic Sinkhorn, the Total Variation (TV) distance, and Sinkhorn. Since the TV only compares the mean, it is sensitive to the outlier, whereas our method can identify the true treatment effect.

5.3. Single-cell signaling data

We use single-cell signaling data produced by mass cytometry (MC) for a screening study to compare the treatment effect of different chemotherapies on 10 colorectal cancer (CRC) patient-derived organoids (PDOs) [22]. These PDOs can be grouped into chemoresistant PDOs, that show little-to-no effect when treated with chemotherapies; and chemosensitive PDOs, that present strong shifts in their phenotypes upon treatment. The observations include single-cell data information on the cell cycle and signaling changes upon treatment of PDOs with different CRC treatments at a range of concentrations. In Fig. 1, we present the barycentric distances matrices between treatments a) and between four concentrations of treatment SN-38 (S) b). In both cases, the control groups corresponds to AH and DMSO, the two rightmost columns. We compare the distance matrices between Sinkhorn (left) and our method (right). Our method provide a finer distinction between treatments (Fig. 1 top) and concentrations (Fig. 1 bottom), especially for the chemosensitive group. As observed in [22], chemosensitive PDOs show little-to-no response to lower concentrations of SN-38 (S1), but their phenotype shifts very strongly upon treatment with higher concentrations (S2, S3, and S4) (Fig. 1 b). When comparing combinations of different treatments (Fig. 1 a), Geodesic Sinkhorn better resolves the difference between SN-38 (S) alone and in combination with Cetuximab (C), showing that S is the main agent creating the treatment effect and the combination with C does not resolve in a synergistic effect [22]. Note that we only consider the relative magnitude of the distances, since the two algorithm use different ground distances.

6. CONCLUSION

In this work, we considered the use of OT for graphs and large datasets in high dimensions potentially sampled from a lower
dimensional manifold. We proposed Geodesic Sinkhorn, a fast implementation of the Sinkhorn algorithm using the graph Laplacian and Chebyshev polynomials. Our method is well adapted for large and high dimensional datasets as it is defined with a geodesic ground distance, which takes into account the underlying geometry of the data, and requires less computation time and less memory. On a synthetic dataset, we showed that Geodesic Sinkhorn is much faster than other Sinkhorn-based methods while being more accurate. With the Wasserstein barycenter, we defined the barycentric distance to compare entire families of distributions, and the expected barycenter effect, then applied both methods to a large PDO drug screen dataset.

7. REFERENCES

[1] Gabriel Peyré and Marco Cuturi, “Computational Optimal Transport,” arXiv, 2020.

[2] Richard Sinkhorn and Paul Knopp, “Concerning nonnegative matrices and doubly stochastic matrices,” Pacific Journal of Mathematics, 1967.

[3] Robert Fortet, “Résolution d’un système d’équations de m. schrödinger,” J. Math. Pure Appl. IX, 1940.

[4] Solomon Kullback, “Probability densities with given marginals,” The Annals of Mathematical Statistics, 1968.

[5] Philip A Knight and Daniel Ruiz, “A fast algorithm for matrix balancing,” IMA Journal of Numerical Analysis, 2013.

[6] Kevin R. Moon, Jay S. Stanley, Daniel Burkhardt, David van Dijk, Guy Wolf, and Smita Krishnaswamy, “Manifold learning-based methods for analyzing single-cell RNA-sequencing data,” Current Opinion in Systems Biology, 2018.

[7] Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy, “Manifold interpolating optimal-transport flows for trajectory inference,” NeurIPS, 2022.

[8] Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas, “Convolutional wasserstein distances: Efficient optimal transportation on geometric domains,” ACM Transactions on Graphics (ToG), 2015.

[9] Alexander Tong, Guillaume Huguet, Dennis Shung, Amine Natik, Manik Kuchroo, Guillaume Lajoie, Guy Wolf, and Smita Krishnaswamy, “Embedding signals on graphs with unbalanced diffusion earth mover’s distance,” in ICASSP. IEEE, 2022.

[10] Alexander Tong, Guillaume Huguet, Amine Natik, Kincaid MacDonald, Manik Kuchroo, Ronald Coifman, Guy Wolf, and Smita Krishnaswamy, “Diffusion earth mover’s distance and distribution embeddings,” in ICML, 2021.

[11] Keenan Crane, Clarisse Weischedel, and Max Wardetzky, “Geodesics in heat: A new approach to computing distance based on heat flow,” ACM Transactions on Graphics, 2013.

[12] Matthieu Heitz, Nicolas Bonneel, David Coeurjolly, Marco Cuturi, and Gabriel Peyré, “Ground metric learning on graphs,” J. of Math. Imaging and Vision, 2021.

[13] Marco Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” NeurIPS, 2013.

[14] S. R. S. Varadhan, “On the behavior of the fundamental solution of the heat equation with variable coefficients,” Communications on Pure and Applied Mathematics, 1967.

[15] Marco Cuturi and Arnaud Doucet, “Fast computation of wasserstein barycenters,” in ICML, 2014.

[16] David I Shuman, Pierre Vandergheynst, and Pascal Frossard, “Chebyshev polynomial approximation for distributed signal processing,” in DOCSS. 2011, IEEE.

[17] Sibylle Marcotte, Amélie Barbe, Rémi Gribonval, Titouan Vayer, Marc Sebban, Pierre Borgnat, and Paulo Gonçalves, “Fast multiscale diffusion on graphs,” in ICASSP. IEEE, 2022.

[18] Shih-Gu Huang, Ilwoo Lyu, Anqi Qiu, and Moo K Chung, “Fast polynomial approximation of heat kernel convolution on manifolds and its application to brain sulcal and gyral graph pattern analysis,” IEEE transactions on medical imaging, 2020.

[19] Guido W Imbens and Donald B Rubin, Causal inference in statistics, social, and biomedical sciences, Cambridge University Press, 2015.

[20] Meyer Scetbon, Marco Cuturi, and Gabriel Peyré, “Low-rank sinkhorn factorization,” in ICML, 2021.

[21] Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy, “TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics,” in ICML, 2020.

[22] María Ramos Zapatero, Alexander Tong, Jahangir Sufi, Petra Vlckova, Ferran Cardoso Rodriguez, Callum Nattress, Xiao Qin, Daniel Hochhauser, Smita Krishnaswamy, and Christopher J. Tape, “Trellis single-cell screening reveals stromal regulation of patient-derived organoid drug responses,” bioRxiv, 2023.