Gromov–Wasserstein Optimal Transport to Align Single-Cell Multi-Omics Data

Pinar Demetci [*1 2] **Rebecca Santorella** [*3] **Björn Sandstede** [3] **William Stafford Noble** [4 5] **Ritambhara Singh** [1 2]

Abstract

Data integration of single-cell measurements is critical for our understanding of cell development and disease, but the lack of correspondence between different types of single-cell measurements makes such efforts challenging. Several unsupervised algorithms are capable of aligning heterogeneous types of single-cell measurements in a shared space, enabling the creation of mappings between single cells in different data modalities. We present Single-Cell alignment using Optimal Transport (SCOT), an unsupervised learning algorithm that uses Gromov Wasserstein-based optimal transport to align single-cell multi-omics datasets. SCOT calculates a probabilistic coupling matrix that matches cells across two datasets based on k-nearest neighbor distances. It uses the resulting coupling matrix to align and project one single-cell dataset onto another. We compare the alignment performance of SCOT with state-of-the-art algorithms on three simulated and two real datasets and demonstrate that SCOT is comparable in quality to competing methods but is significantly faster and requires tuning fewer hyperparameters. The code is available at https://github.com/rsinghlab/SCOT.

1. Introduction

Single-cell measurements provide a fine-grained view of the heterogeneous landscape of cells in a sample, revealing distinct subpopulations and their developmental and regulatory trajectories across time. The availability of measurements capturing various properties of the genome, such as gene expression, chromatin accessibility, histone modifications, and chromatin 3D conformation, has increased the need for data integration methods capable of combining these disparate data types. Due to technical limitations, it is hard to obtain multiple types of measurements from the same individual cell. Furthermore, when we measure different properties of a cell—such as its transcriptional and 3D chromatin profiles—we cannot *a priori* identify correspondences between features in the two domains. Accordingly, integrating two or more single-cell data modalities requires methods that do not rely on either common cells or common features across the data types (Amodio & Krishnaswamy, 2018; Welch et al., 2019; 2017; Stuart et al., 2019).

Liu *et al.* (Liu et al., 2019) proposed a global manifold alignment algorithm based on the maximum mean discrepancy (MMD) measure, called MMD-MA, which can integrate different types of single-cell measurements. Another method, UnionCom (Cao et al., 2020), performs unsupervised topological alignment for single-cell multi-omics data to emphasize both local and global alignment. Neither method requires any correspondence information between the datasets but do require tuning four hyperparameters.

A number of applications across various research areas (Galichon, 2017; Alvarez-Melis & Jaakkola, 2018) use optimal transport to learn a mapping between different data distributions. Optimal transport finds the most cost-effective way to move data points from one domain to another. One way to think about optimal transport is as the problem of moving a pile of sand to fill in a hole through the least amount of work. The optimal transport framework has been used in various biological applications (Schiebinger et al., 2019; Yang et al., 2018; Yang & Uhler, 2019).

The classical optimal transport method requires datasets to be in the same metric space. Mémoli *et al.* (Mémoli, 2011) generalized optimal transport to the Gromov-Wasserstein distance, which compares metric spaces directly instead of comparing samples across spaces. In the natural language processing community, Alvarez *et al.* (Alvarez-Melis & Jaakkola, 2018) used this approach to measure similarities between pairs of words across languages. As far as we are aware, the only biological application of Gromov-Wasserstein optimal transport comes from (Nitzan et al.,

[*]Equal contribution [1]Department of Computer Science, Brown University [2]Center for Computational Molecular Biology, Brown University [3]Division of Applied Mathematics, Brown University [4]Paul G. Allen School of Computer Science and Engineering, University of Washington [5]Department of Genome Sciences, University of Washington. Correspondence to: Ritambhara Singh <ritambhara@brown.edu>.

Presented at the ICML 2020 Workshop on Computational Biology (WCB).

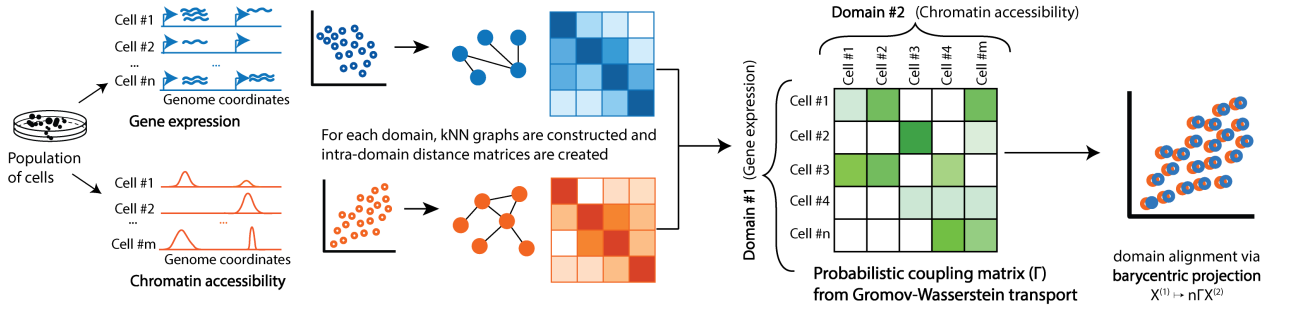

Figure 1. Schematic of SCOT applied to single-cell multi-omics data alignment.

2019), which uses it to reconstruct the spatial organization of cells from transcriptional profiles.

We present Single-Cell alignment using Optimal Transport (SCOT), an unsupervised learning algorithm that employs Gromov Wasserstein optimal transport to align single-cell multi-omics datasets while preserving local geometry. Unlike MMD-MA and UnionCom, our algorithm requires tuning only two hyperparameters and is robust to the choice of one. We compare the alignment performance of SCOT with MMD-MA and UnionCom on three simulated and two real-world datasets to demonstrate that SCOT aligns all the datasets as well as the state-of-the-art methods and converges $\sim$15 and $\sim$50 times faster than MMD-MA and UnionCom, respectively.

2. Methods

SCOT uses Gromov Wasserstein optimal transport, which preserves local neighborhood geometry when moving data points. The output of this transport problem is a matrix of probabilities that represent how likely it is that data points from one space correspond to data points in the other space. These probabilities can then be used to project the data into the same space for alignment. In this section, we first introduce the formulation of optimal transport followed by its extension to Gromov-Wasserstein distance. Finally, we present the details of our SCOT algorithm.

We present these methods for two sets of points: $X = (x_1, x_2, \ldots, x_{n_x})$ from the measure space $(\mathcal{X}, p)$ and $Y = (y_1, y_2, \ldots, y_{n_y})$ from the measure space $(\mathcal{Y}, q)$. We define discrete measures p and q over our data points, representing marginal distributions of X and Y respectively, which we can write as

$$p = \sum_{i=1}^{n_x} p_i \delta_{x_i} \text{ and } q = \sum_{j=1}^{n_y} q_j \delta_{y_j},$$

where δ_{x_i} is the Dirac measure. We do not require any correspondence information across the data sets but assume that there is some underlying shared manifold structure.

2.1. Optimal Transport

The Kantorovich optimal transport problem seeks a minimal cost coupling between two probability distributions to tie them in a meaningful way (Peyré et al., 2019). Referring back to the problem of moving a sand pile to fill in a hole, Kantorovich optimal transport allows us to split the mass of a grain of sand instead of moving the whole grain. For discrete measures, the set of possible couplings are the matrices

$$\Pi(p, q) = \{\Gamma \in \mathbb{R}_+^{n_x \times n_y} : \Gamma \mathbb{1}_{n_y} = p,\ \Gamma^T \mathbb{1}_{n_x} = q\}. \quad (1)$$

Each row Γ_i of a coupling Γ tells us how to split the mass of data point x_i onto the points y_j for $j = 1, \ldots, n_y$, and the condition $\Gamma \mathbb{1}_{n_y} = p$ requires that the sum of each row Γ_i is equal to the probability of sample x_i.

Given discrete measures p and q and a cost matrix $C \in \mathbb{R}^{n_x \times n_y}$ where C_{ij} is the cost of transporting point x_i to point y_j, the discrete optimal transport problem learns a minimal coupling that attains

$$\min_{\Gamma \in \Pi(p,q)} \langle \Gamma, C \rangle. \quad (2)$$

Intuitively, the cost function says how many resources it will take to move x_i to y_j, and the coupling Γ assigns a probability Γ_{ij} that x_i should be moved to y_j for each x_i and y_j in the two spaces.

Although this problem can be solved with minimum cost flow solvers, it is usually regularized with entropy for more efficient optimization and empirically better results (Cuturi, 2013). Thus, the optimal transport problem that is solved numerically is

$$\min_{\Gamma \in \Pi(p,q)} \langle \Gamma, C \rangle - \epsilon H(\Gamma), \quad (3)$$

where $\epsilon > 0$ and $H(\Gamma)$ is Shannon entropy.

Table 1. Alignment performance, using average FOSCTTM scores (lower is better), of SCOT, MMD-MA, and UnionCom for all datasets.

	Sim. 1	Sim. 2	Sim. 3	scGEM	SNAREseq
SCOT	**0.070**	**0.006**	**0.008**	0.206	0.198
MMD-MA	0.124	0.032	0.012	**0.201**	**0.149**
UnionCom	0.084	0.027	0.152	0.217	0.265

The addition of entropy diffuses the optimal coupling, meaning that more masses will be split. Equation 3 is a strictly convex optimization problem, and for some unknown vectors $u \in \mathbb{R}^{n_x}$ and $v \in \mathbb{R}^{n_y}$, the solution has the form $\Gamma^* = \text{diag}(u)K\text{diag}(v)$, with $K = \exp\left(-\frac{C}{\epsilon}\right)$, element-wise. This solution can be obtained efficiently via Sinkhorn's algorithm (Peyré et al., 2019).

2.2. Gromov-Wasserstein Optimal Transport

Classic optimal transport requires defining a cost function to move samples across domains, which can be difficult to implement for data in different dimensions. Gromov-Wasserstein distance allows for the comparison of distributions in different metric spaces by comparing pairwise distances between the samples across these domains. For this extension, we need to assume we have metric measure spaces $(\mathcal{X}, D^x, p)$ and $(\mathcal{Y}, D^y, q)$, where D^x and D^y are distance matrices on the two datasets with $D^x_{ij} = d_x(x_i, x_j)$ and $D^y_{ij} = d_y(y_i, y_j)$ for some distances d_x and d_y (Mémoli, 2011).

Given a cost function $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, defined over the distance matrices, the discrete Gromov-Wasserstein distance between p and q is defined by

$$GW(p, q) = \min_{\Gamma \in \Pi(p,q)} \sum_{i,j,k,l} \mathbf{L}_{ijkl}\Gamma_{ij}\Gamma_{kl}, \tag{4}$$

where $\mathbf{L} \in \mathbb{R}^{n_x \times n_x \times n_y \times n_y}$ is the fourth-order tensor defined by $\mathbf{L}_{ijkl} = L(D^x_{ik}, D^y_{jl})$. Intuitively, $L(D^x_{ik}, D^y_{jl})$ captures how transporting x_i onto y_j and x_k onto y_l would distort the original distances between x_i and x_k and between y_j and y_l. This change ensures that the optimal transport plan π will preserve some local geometry. For our problem, we use $L(x, y) = (x - y)^2$. As in the case of classic optimal transport, this problem can be solved efficiently through entropic regularization (Peyré et al., 2016):

$$\min_{\Gamma \in \Pi(p,q)} \sum_{i,j,k,l} \mathbf{L}_{ijkl}\Gamma_{ij}\Gamma_{kl} - \epsilon H(\Gamma). \tag{5}$$

Larger values of ϵ lead to an easier optimization problem but also a denser coupling matrix, meaning that more data points have correspondences with one another. Conversely, smaller values of ϵ lead to sparser solutions, meaning that the coupling matrix is more likely to find the correct one-to-one correspondences when they exist. However, it also means a harder (more non-convex) optimization problem (Alvarez-Melis & Jaakkola, 2018).

2.3. Single-Cell alignment using Optimal Transport (SCOT)

Our method, SCOT, works as follows. First, we compute the pairwise distances on our data by using graph distances as in (Nitzan et al., 2019). To do this, we construct a $k-$nearest neighbor graph based on the Euclidean distances within each data set. Then, we compute the shortest path distance on the graph between each pair of nodes and set the distance of any unconnected nodes to be the maximum (finite) distance in the graph.

Our approach is robust to the choice of k. Since we do not know the true distribution of the original data sets, we follow (Alvarez-Melis & Jaakkola, 2018) and set p and q to be the uniform distributions on the data points. With these graph distance matrices and marginal distributions, we solve for the optimal coupling Γ which minimizes Equation 5.

One of the major advantages of this approach is that we end up with a coupling matrix Γ with a probabilistic interpretation. The entries of the normalized row $n_x\Gamma_i$ are the probabilities that the fixed data point x_i corresponds to each y_j. However, to use the correspondence metrics previously used in the field to evaluate the alignment, we need to project the two datasets into the same space. We use a barycentric projection:

$$x_i \mapsto \frac{1}{p_i} \sum_{j=1}^{n_y} \Gamma_{ij} y_j. \tag{6}$$

This barycentric projection of point x_i is a weighted average of the y_j's, where the weight Γ_{ij} is the probability of correspondence between x_i and y_j. This projection averages over all the points, so it has a tendency to center the projected data onto the mean of the dataset it is being projected on. Figure 1 presents the schematic of the SCOT algorithm.

3. Experimental Setup

We benchmark our method on three different simulation schemes (from (Liu et al., 2019)): a bifurcated tree, a Swiss roll, and a circular frustum. We also align two single-cell co-assay datasets: (1) sc-GEM (Cheow et al., 2016), which simultaneously profiles gene expression and DNA methylation at multiple loci on human somatic cell samples undergoing conversion to induced pluripotent stem cells (iPSCs) and (2) SNARE-seq (Chen et al., 2019), which links chromatin accessibility with gene expression data on a mixture

Table 2. Running times (in seconds) of SCOT, MMD-MA, and UnionCom (CPU and GPU versions).

	Sim. 1	Sim. 2	Sim. 3	scGEM	SNAREseq
SCOT	**2.33**	**1.43**	**3.41**	**3.17**	**40.52**
MMD-MA	30.06	29.69	28.84	16.12	547.71
UnionCom	525.85	442.19	302.69	143.60	2169.74
UnionCom (GPU)	117.72	112.41	109.73	70.21	345.37

of BJ, H1, K562, GM12878 cells. All these datasets have 1–1 correspondence information, which we use only to evaluate the alignments quantitatively. For evaluation, we use the average "fraction of samples closer than the true match (FOSCTTM)" metric introduced by Liu *et al.* (Liu et al., 2019). For each sample in a domain, this metric calculates the fraction of samples that are positioned more closely to it than its true match after alignment. We average it across all the samples for each dataset and report the average FOSCTTM.

We benchmark the performance of SCOT against UnionCom and MMD-MA. Although none of these methods require any correspondence information for aligning data, they all require hyperparameter tuning for optimal alignment. For all methods, we define a grid of hyperparameters and select the set that yields the minimal average FOSCTTM. For SCOT, the grid is defined over the regularization weight (ϵ) and the number of neighbors (k) in the $k-$NN graph. For MMD-MA, it is defined over the width of the Gaussian for the initial kernel calculation (σ), the weights for the terms in the optimization problem (λ_1 and λ_2), and the dimensionality of the embedding space (p). For UnionCom, we tune the number of neighbors in the$k-$NN graph (k), the trade-off parameter (β), the regularization coefficient (ρ) and dimensionality of the embedding space (p). While not related to the algorithmic formulation, we also tune the learning rate to achieve a smoother convergence.

4. Results

4.1. SCOT successfully performs alignment

We report the alignment results from SCOT for both the simulated and real-world datasets, as compared with MMD-MA and UnionCom, in Table 1. For the simulated datasets, we observe that SCOT achieves the lowest average FOSCTTM metric (averaged over all samples in the datasets) and demonstrates its ability to recover the correct correspondences in simulations.

Next, we apply our method to real-world single-cell sequencing assays and observe that SCOT gives comparable performance to the baseline methods. For scGEM data, the best FOSCTTM values are 0.201, 0.217, and 0.2066 for MMD-MA, UnionCom, and ScOT, respectively. For the SNARE-seq dataset, MMD-MA yields the best FOSCTTM, followed by SCOT, with FOSCTTM values of 0.1499 and 0.1985. UnionCom achieves lower performance with a FOSCTTM value of 0.265. We note that this dataset consists of a larger number of cells (1047) compared to scGEM (177).

4.2. SCOT is faster and uses fewer hyperparameters

We compare the running times of SCOT with the baseline methods for the best performing hyperparameters. We ran the CPU versions of the algorithms on an Intel i5-8259U CPU (base frequency 2.30GHz) with 16GB memory. UnionCom also has a GPU version that we ran on a single NVIDIA GTX 1080ti with VRAM of 11GB. We observe that SCOT converges $\sim$15, $\sim$50, and $\sim$10 times faster than MMD-MA, UnionCom, and UnionCom-GPU, respectively, for the largest SNARE-Seq dataset (Table 2).

Unlike MMD-MA and UnionCom, which require the tuning of four parameters, SCOT requires the tuning of only two. Therefore, it drastically reduces the parameter search space making the algorithm an even faster tool for unsupervised single-cell alignment tasks.

5. Discussion

We have demonstrated that SCOT, which uses Gromov Wasserstein-based optimal transport to perform unsupervised integration of single-cell multi-omics data, performs well when compared to two state-of-the-art methods but in less time and with fewer hyperparameters.

To apply an evaluation metric and quantify the quality of alignment, we need to project the data into the same space. Here, we choose to use a barycentric projection to project one domain onto another, but there are various other ways to use the coupling matrix to infer alignment. For example, the coupling matrix could also be used with other dimension reduction methods such as t-SNE (as in UnionCom) to align the manifolds while embedding them both into new spaces. Additionally, depending on the application, a projection may not be required. For some downstream analyses, it may be sufficient to have probabilities relating the samples to one another. Our future work will focus on developing effective ways to utilize the coupling matrix, closely investigating the cases where SCOT and MMD-MA outperform each other, and extending our framework to handle more than two alignments at a time.

References

Alvarez-Melis, D. and Jaakkola, T. S. Gromov-wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018.

Amodio, M. and Krishnaswamy, S. MAGAN: Aligning biological manifolds. 2018.

Cao, K., Bai, X., Hong, Y., and Wan, L. Unsupervised topological alignment for single-cell multi-omics integration. *bioRxiv*, 2020.

Chen, S., Lake, B. B., and Zhang, K. High-throughput sequencing of transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology*, 37(12):1452–1457, 2019.

Cheow, L. F., Courtois, E. T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R. Z., Tan, D. S. Q., Robson, P., Yuin-Han, L., Quake, S. R., and Burkholder, W. F. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature Methods*, 13(10):833–836, 2016.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pp. 2292–2300, 2013.

Galichon, A. A survey of some recent applications of optimal transport methods to econometrics. *Econometrics Journal*, 20(2), 2017.

Liu, J., Huang, Y., Singh, R., Vert, J.-P., and Noble, W. S. Jointly embedding multiple single-cell omics measurements. *BioRxiv*, pp. 644310, 2019.

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

Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. Gene expression cartography. *Nature*, 576(7785): 132–137, 2019.

Peyré, G., Cuturi, M., and Solomon, J. Gromov-wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pp. 2664–2672, 2016.

Peyré, G., Cuturi, M., et al. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6): 355–607, 2019.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papelexi, E., III, W. M. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. Comprehensive integration of single-cell data. *Cell*, 77(7):1888–1902, 2019.

Welch, J. D., Hartemink, A. J., and Prins, J. F. Matcher: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. *Genome biology*, 18(1):138, 2017.

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E. Z. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*, 177(7):1873–1887, 2019.

Yang, K. D. and Uhler, C. Multi-domain translation by learning uncoupled autoencoders. *arXiv preprint arXiv:1902.03515*, 2019.

Yang, K. D., Damodaran, K., Venkatchalapathy, S., Soylemezoglu, A. C., Shivashankar, G., and Uhler, C. Autoencoder and optimal transport to infer single-cell trajectories of biological processes. *bioRxiv*, pp. 455469, 2018.