Low-Rank Optimal Transport through Factor Relaxation with Latent Coupling

Peter Halmos, Xinhao Liu, Julian Gold, Benjamin Raphael

Yessin Moakher, Augustin Kheng, Nathan Boughalem-Salier

Outline

1. Problem formulation and state of the art
2. Low-Rank Optimal Transport with Latent Coupling
3. Experimental Results

Outline

1. Problem forumulation and state of the art
2. Low-Rank Optimal Transport with Latent Coupling
3. Experimental Results

Background : Discrete Optimal Transport

Primal problem formulation:

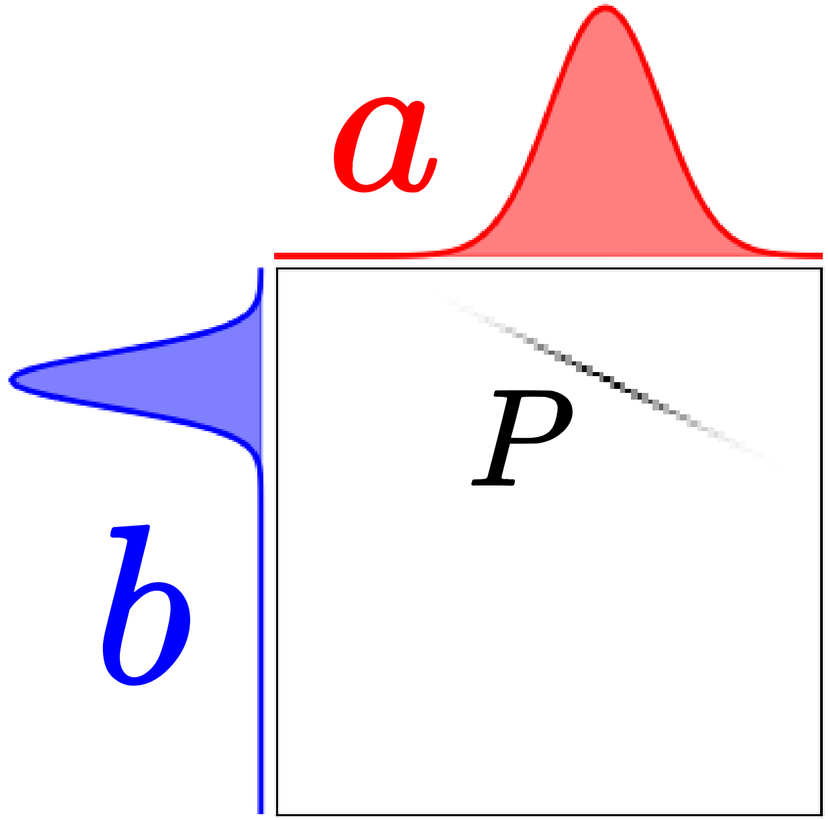
$$\min_{P \geq 0} \langle P, C \rangle$$

$$\text{s.t.} \quad P \in \Pi_{a,.}$$

$$P \in \Pi_{.,b}$$

Where :

- P is the transport plan
- C is the cost matrix
- a and b are the marginal distributions

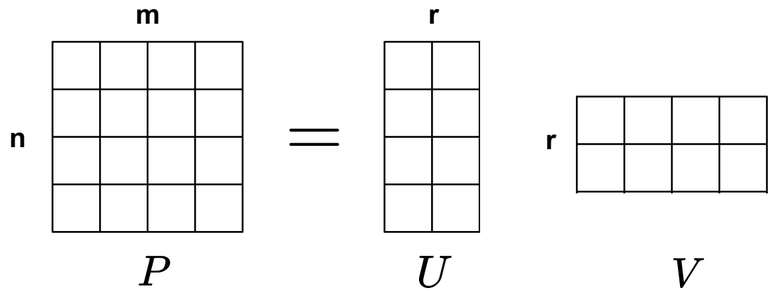


Scaling Optimal Transport

Low rank factorisation:

Idea : Working in the space of matrices of rank <= r

Low rank factorisation



$\Longrightarrow$ **Problem**: Can’t simply transfer the contraints to P on U and V.

Scaling Optimal Transport

Low rank factorisation:

-[Altschuler'18] Propose : factorize the Kernel.

-[Scetborn ICML'21] Propose : $P = Q diag(1/g) R^T$

$$\text{Where} : Q \in \Pi_{a,g} \text{ and } R \in \Pi_{b,g}$$

-**[Halmos Neurips'24]** Propose : $P = Q \operatorname{diag}\left(\frac{1}{g_Q}\right) T \operatorname{diag}\left(\frac{1}{g_R}\right) R^T$

where g_Q and g_R are the inner marginals of Q and $R, Q \in \Pi_{a,.}, R \in \Pi_{b,.}, T \in \Pi_{g_Q, g_R}$.

Low rank Optimal Transport

Difference between the two methods

(1) [Scetbon ICML'21]

$$P = Q diag(1/g) R^T$$

(2) [Halmos Neurips'24]

$$P = Q \,\mathrm{diag}\left(\frac{1}{g_Q}\right) T \,\mathrm{diag}\left(\frac{1}{g_R}\right) R^T$$

+Both optimize over the same space, but (2) has more parameters.

+They use different optimization algorithms:

- (1) Mirror descent followed by Dykstra's algorithm
- (2) Coordinate mirror descent

The optimization literature includes results on the equivalence of Dykstra's algorithm and coordinate descent (e.g., [Tibshirani NIPS'17] for regularized regression).

+Non convex problems, but we have stationary convergence (not necessarily to the minimum) thanks to [Ghadimi'13].

Outline

1. Problem forumulation and state of the art
2. Low-Rank Optimal Transport with Latent Coupling
3. Experimental Results

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate descent:

Recall our problem is :

$$\min_{(\mathbf{Q},\mathbf{R},\mathbf{T})} \mathcal{L}_{\mathrm{LC}} = \langle \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \mathbf{M}\rangle_F$$

$$\text{s.t.} \quad g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m,$$

$$\mathbf{Q} \in \Pi_{a,\cdot}, \quad \mathbf{R} \in \Pi_{b,\cdot}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \mathbf{Q} \in \mathbb{R}^+_{n,r}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r}, \quad \mathbf{T} \in \mathbb{R}^+_{r,r},$$

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate descent:

Recall our problem is :

$$\min_{(\mathbf{Q},\mathbf{R},\mathbf{T})} \mathcal{L}_{\mathrm{LC}} = \langle \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \mathbf{M}\rangle_F$$

$$\text{s.t.} \quad g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m,$$

$$\mathbf{Q} \in \Pi_{a,.}, \quad \mathbf{R} \in \Pi_{b,.}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \mathbf{Q} \in \mathbb{R}^+_{n,r}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r}, \quad \mathbf{T} \in \mathbb{R}^+_{r,r},$$

Coordinate descent

- $(Q_{k+1}, R_{k+1}) \leftarrow \underset{\mathbf{Q}\in\Pi_{a,.},\mathbf{R}\in\Pi_{b,.},\mathbf{Q}\geq 0,\mathbf{R}\geq 0}{\arg\min} \mathcal{L}_{\mathrm{LC}}(Q, R, T_k)$

- $T_{k+1} \leftarrow \underset{\mathbf{T}\in\Pi_{g_{Q_{k+1}},g_{R_{k+1}}},\mathbf{T}\geq 0}{\arg\min} \mathcal{L}_{\mathrm{LC}}(Q_{k+1}, R_{k+1}, T)$

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate descent:

Recall our problem is :

$$\min_{(\mathbf{Q},\mathbf{R},\mathbf{T})} \mathcal{L}_{\mathrm{LC}} = \langle \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \mathbf{M}\rangle_F$$

$$\text{s.t.} \quad g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m,$$

$$\mathbf{Q} \in \Pi_{a,.}, \quad \mathbf{R} \in \Pi_{b,.}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \mathbf{Q} \in \mathbb{R}^+_{n,r}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r}, \quad \mathbf{T} \in \mathbb{R}^+_{r,r},$$

Coordinate descent

- $(Q_{k+1}, R_{k+1}) \leftarrow \underset{\mathbf{Q}\in\Pi_{a,.},\mathbf{R}\in\Pi_{b,.},\mathbf{Q}\geq 0,\mathbf{R}\geq 0}{\arg\min} \mathcal{L}_{\mathrm{LC}}(Q, R, T_k)$
- $T_{k+1} \leftarrow \underset{\mathbf{T}\in\Pi_{g_{Q_{k+1}},g_{R_{k+1}}},\mathbf{T}\geq 0}{\arg\min} \mathcal{L}_{\mathrm{LC}}(Q_{k+1}, R_{k+1}, T)$

Optimize with one step of mirror descent

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate Mirror descent:

$$(Q_{k+1}, R_{k+1}) \leftarrow \underset{\mathbf{Q}\in\Pi_{a,\cdot},\mathbf{R}\in\Pi_{b,\cdot},\mathbf{Q}\geq 0,\mathbf{R}\geq 0}{\arg\min} \mathcal{L}_{\mathrm{LC}}(Q, R, T_k)$$

- $(Q_{k+1}, R_{k+1}) \leftarrow \arg\min_{\mathbf{Q}\in\Pi_{a,\cdot},\mathbf{R}\in\Pi_{b,\cdot}} \langle (Q, R), \nabla_{Q,R}\mathcal{L}_{\mathrm{LC}} \rangle + \frac{1}{\gamma_k}\mathrm{KL}((Q, R) \parallel (Q_k, R_k)).$

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate Mirror descent:

- $(Q_{k+1}, R_{k+1}) \leftarrow \arg\min_{\mathbf{Q}\in\Pi_{a,\cdot},\mathbf{R}\in\Pi_{b,\cdot}} \langle (Q,R), \nabla_{Q,R}\mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}((Q,R) \parallel (Q_k,R_k))$.

$$\equiv \begin{cases} Q_{k+1} \leftarrow \arg\min_{\mathbf{Q}\in\Pi_{a,\cdot}} \langle Q, \nabla_Q \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}(Q \parallel Q_k). \\ R_{k+1} \leftarrow \arg\min_{\mathbf{R}\in\Pi_{b,\cdot}} \langle R, \nabla_R \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}((R \parallel R_k). \end{cases}$$

- $T_{k+1} \leftarrow \arg\min_{T\in\Pi_{g_{Q_{k+1}},g_{R_{k+1}}}} \langle T, \nabla_T \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}(T \parallel T_k)$.

Low-Rank Optimal Transport with Latent Coupling [Halmos Neurips'24]

Coordinate Mirror descent:

- $(Q_{k+1}, R_{k+1}) \leftarrow \arg\min_{\mathbf{Q}\in\Pi_{a,\cdot},\mathbf{R}\in\Pi_{b,\cdot}} \langle (Q,R), \nabla_{Q,R}\mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}((Q,R) \parallel (Q_k,R_k))$.

$$\equiv \begin{cases} Q_{k+1} \leftarrow \arg\min_{\mathbf{Q}\in\Pi_{a,\cdot}} \langle Q, \nabla_Q \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}(Q \parallel Q_k). \\ R_{k+1} \leftarrow \arg\min_{\mathbf{R}\in\Pi_{b,\cdot}} \langle R, \nabla_R \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}((R \parallel R_k). \end{cases}$$

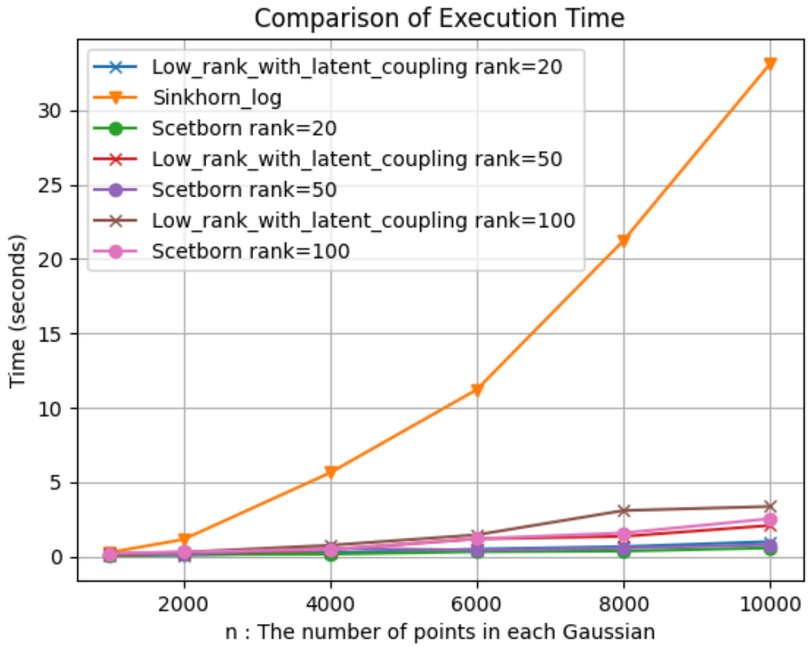
- $T_{k+1} \leftarrow \arg\min_{T\in\Pi_{g_{Q_{k+1}},g_{R_{k+1}}}} \langle T, \nabla_T \mathcal{L}_{\mathrm{LC}}\rangle + \frac{1}{\gamma_k}\mathrm{KL}(T \parallel T_k)$.

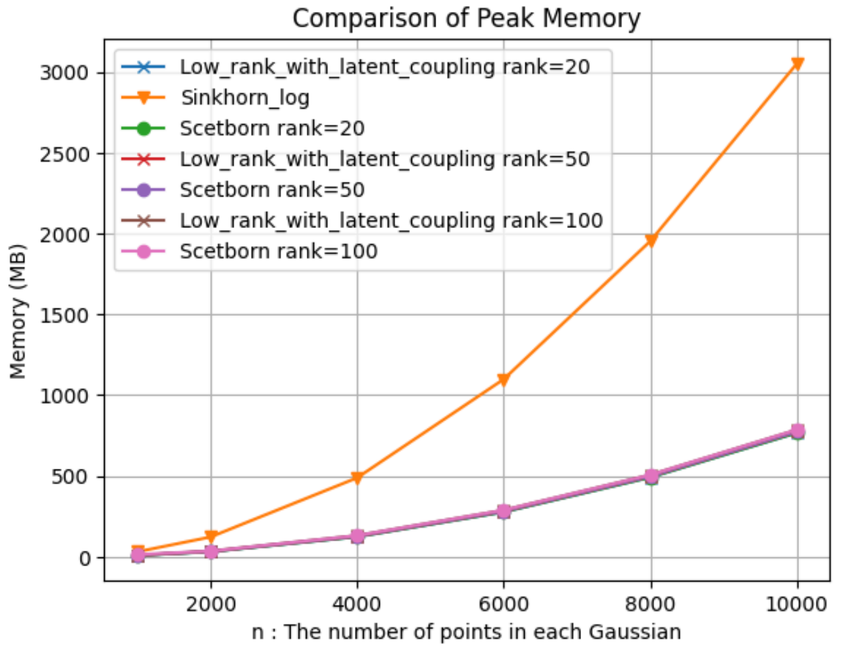
$\Longrightarrow$ Optimal Transport problems that we can solve using Sinkhorn algorithm.

Outline

1. Problem forumulation and state of the art
2. Low-Rank Optimal Transport with Latent Coupling
3. Experimental Results

Benchmarking the Scaling





*We verified that the two methods have the same linear loss and respect the marginals

*Low-rank with latent coupling : tau=10,gamma=10,epsilon=1e-9,max_iter=3000

*Sinkhorn : reg=10

*Low-Rank Sinkhorn Factorization : gamma_0=10

Visualizing the projection

(1) [Scetbon ICML'21]

$$P = Q diag(1/g) R^T$$

(2) [Halmos Neurips'24]

$$P = Q \,\mathrm{diag}\left(\frac{1}{g_Q}\right) T \,\mathrm{diag}\left(\frac{1}{g_R}\right) R^T$$

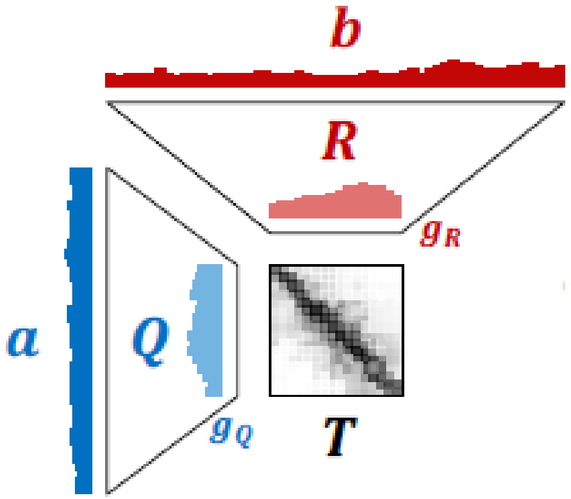
Visualizing the projection

(1) [Scetbon ICML'21]

$$P = Q diag(1/g) R^T$$

(2) [Halmos Neurips'24]

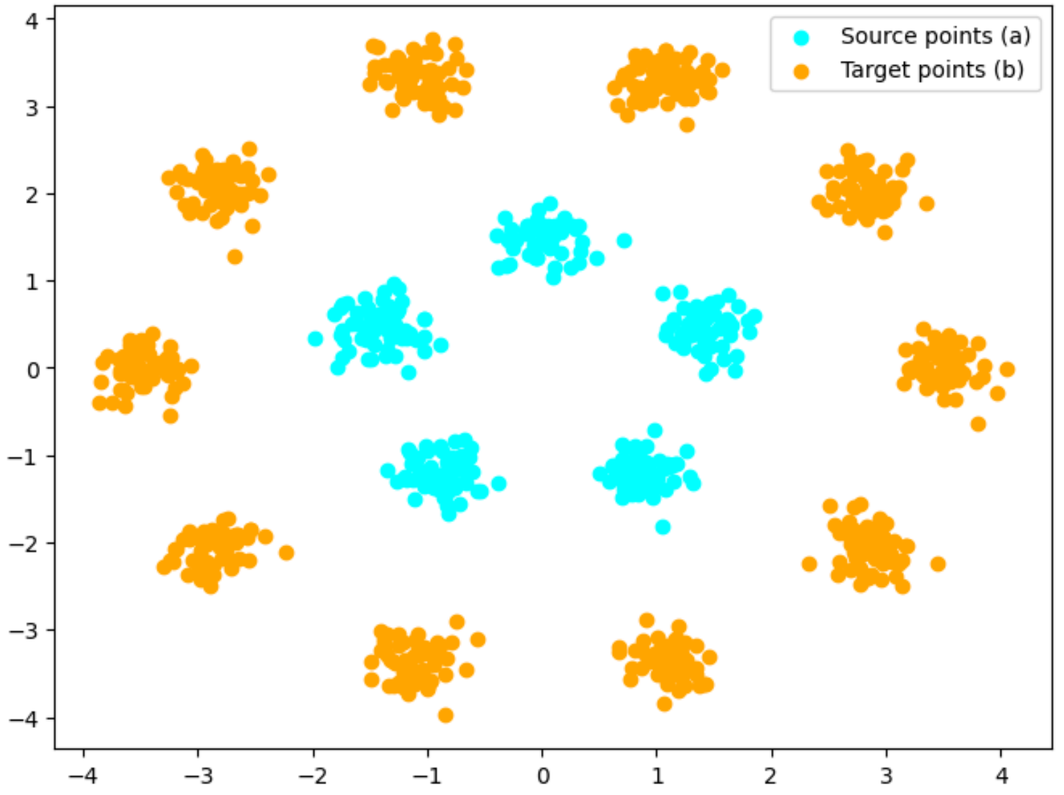
$$P = Q \,\mathrm{diag}\left(\frac{1}{g_Q}\right) T \,\mathrm{diag}\left(\frac{1}{g_R}\right) R^T$$



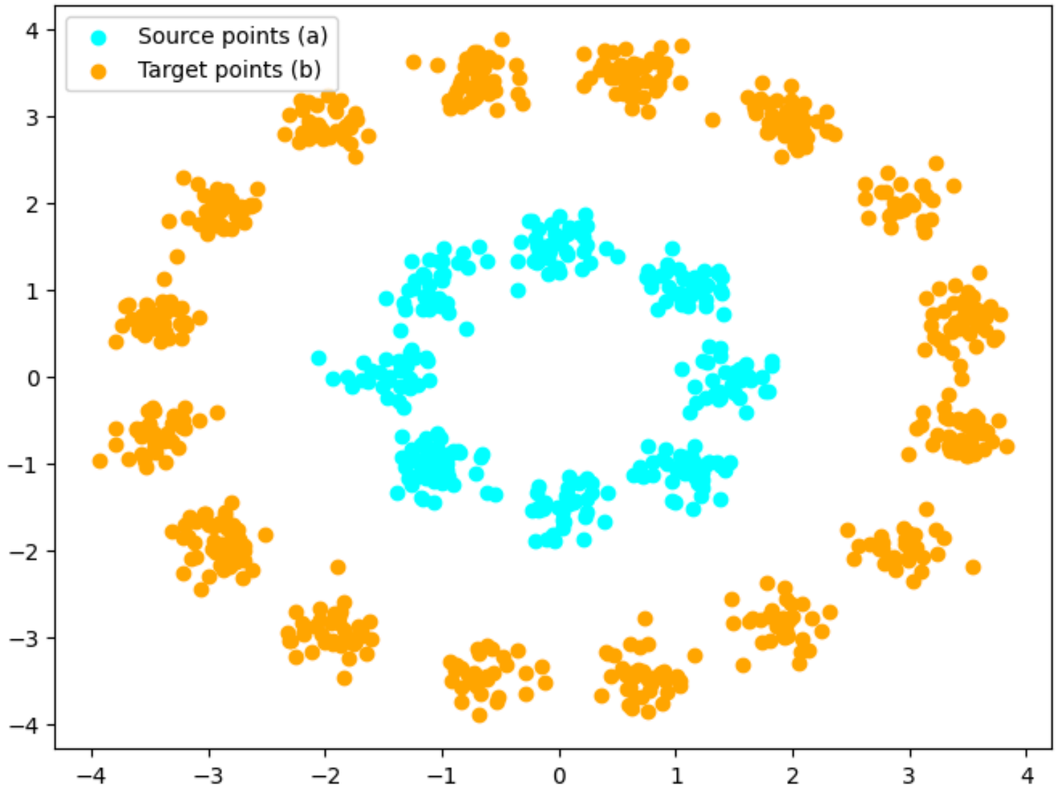
$$Y^a = \mathrm{diag}(1/g_Q) Q^T Z^a$$

$$Y^b = \mathrm{diag}(1/g_R) R^T Z^b$$

Visualizing the projection

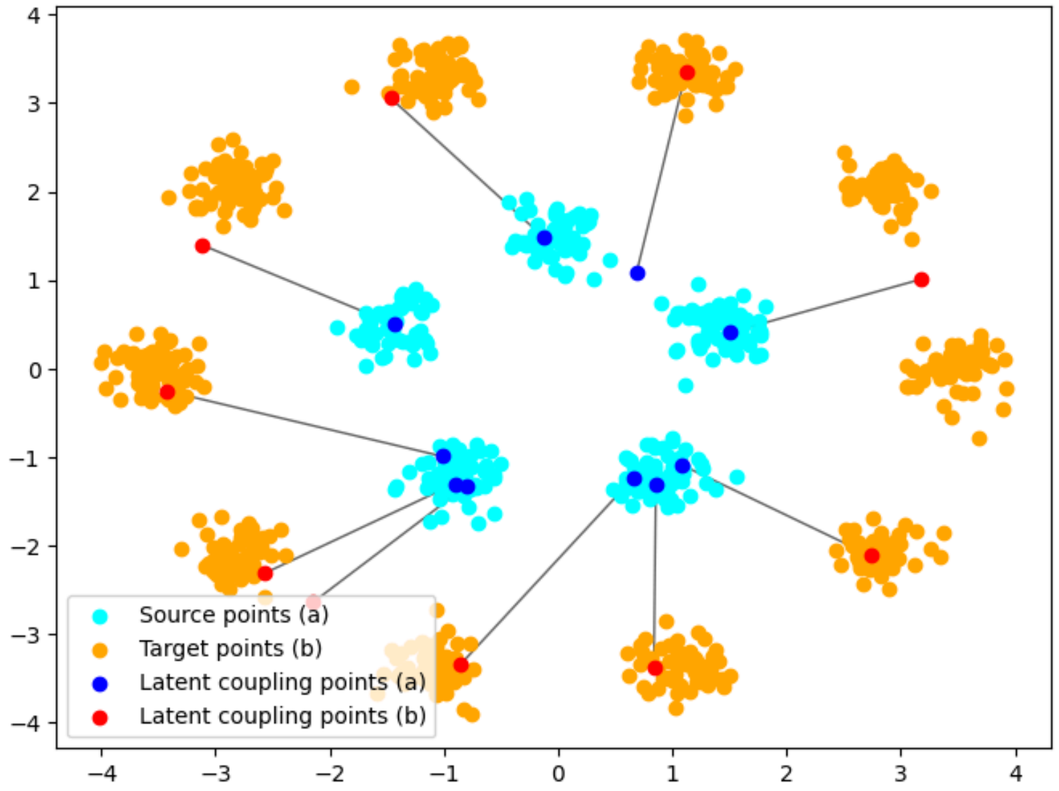


5 source points, 10 target points



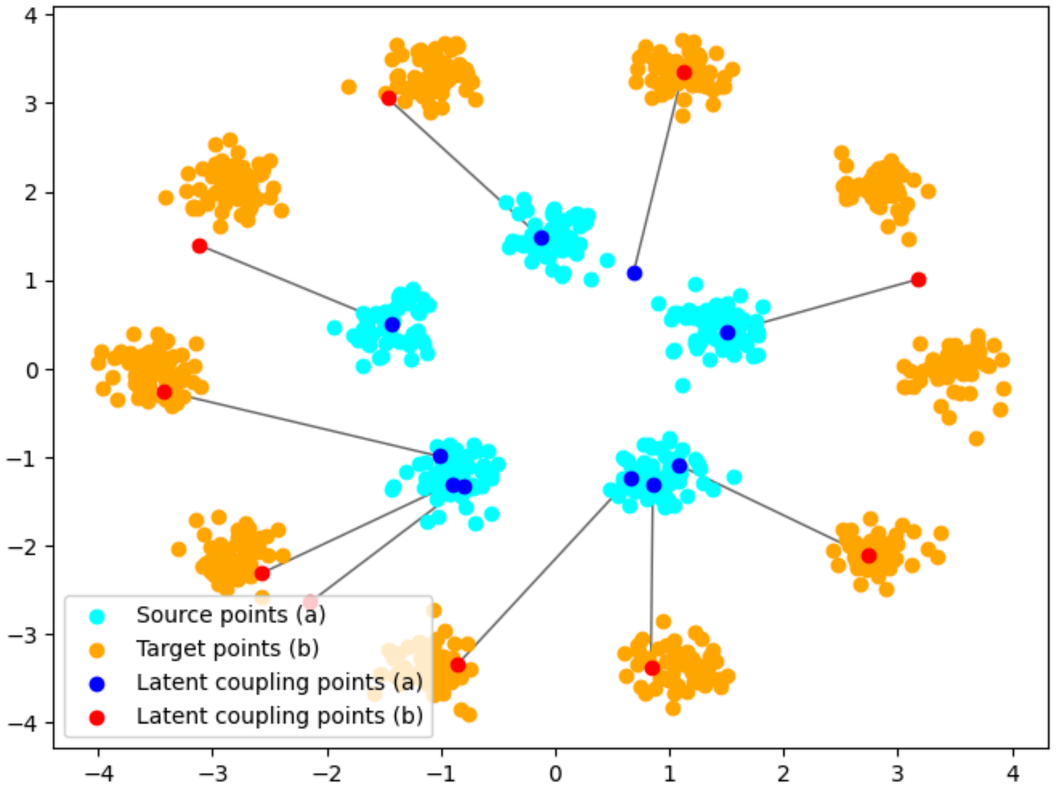
8 source points, 16 target points

Visualizing the projection: 5-10 points

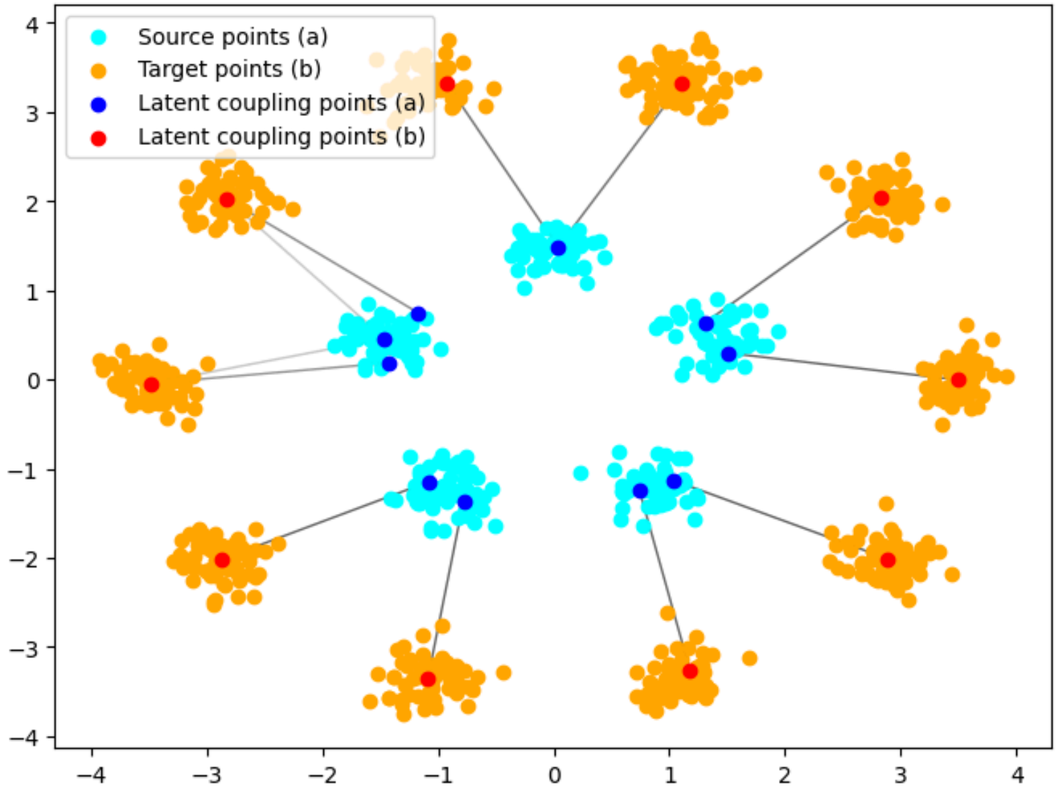


LOT (Scetbon 2021)

Visualizing the projection: 5-10 points

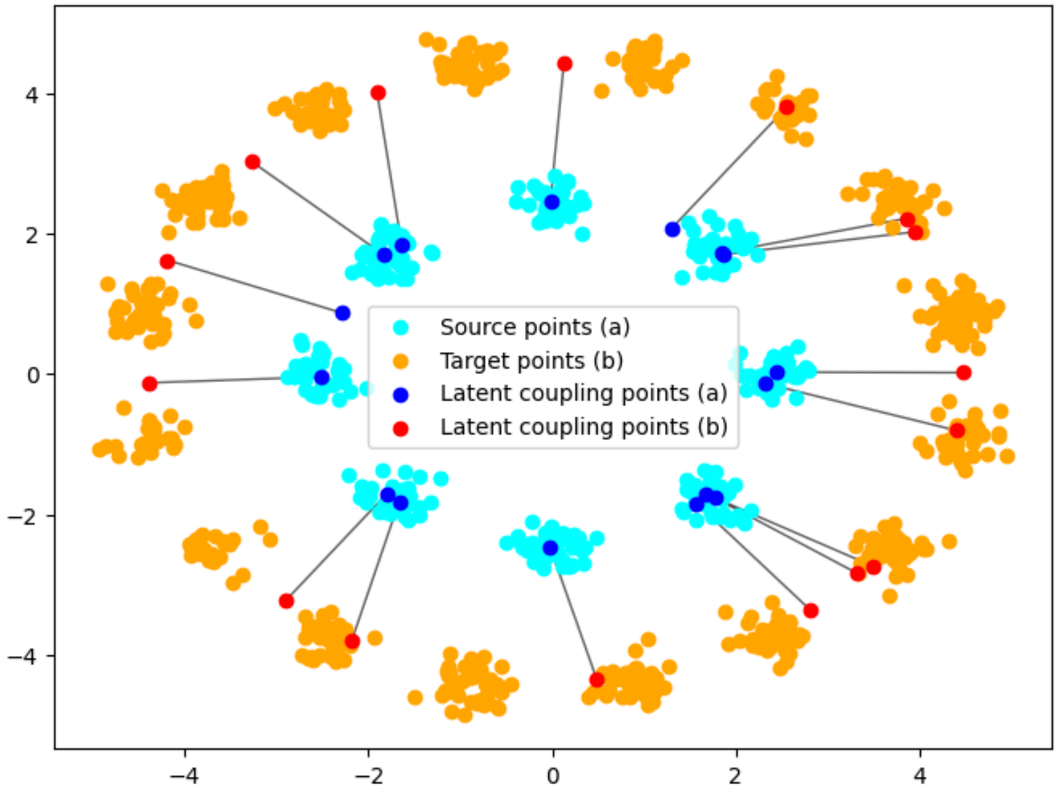


LOT (Scetbon 2021)



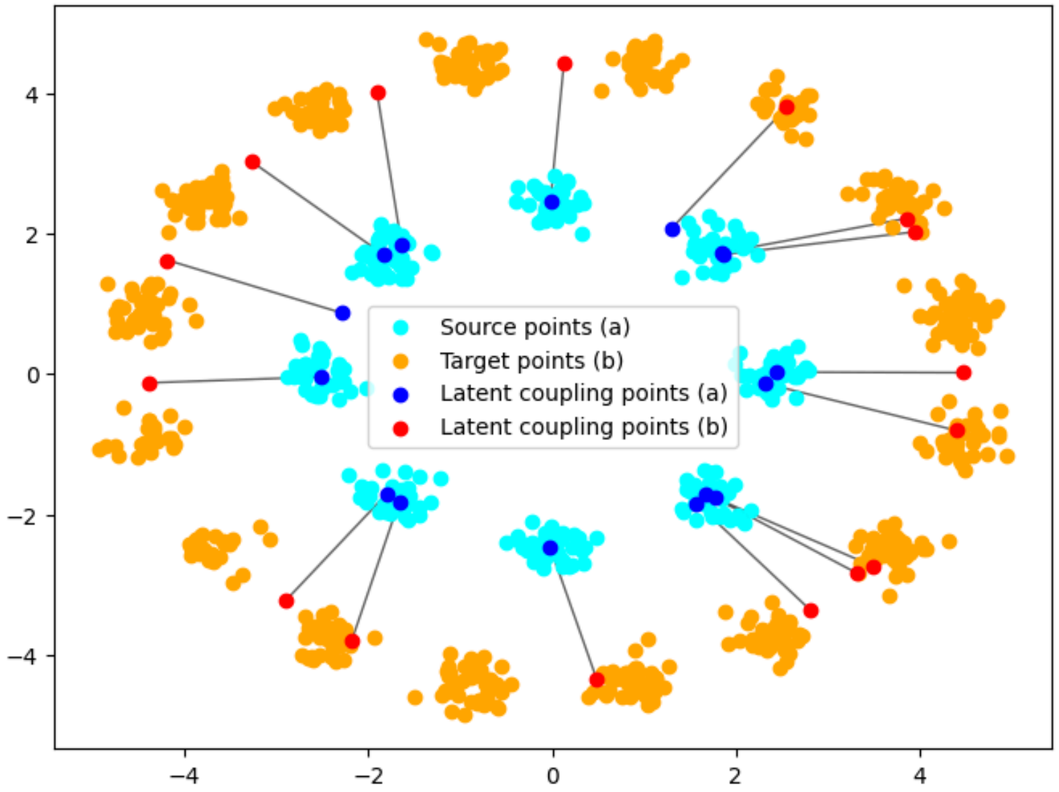
FRLC (Halmos 2024)

Visualizing the projection: 8-16 points

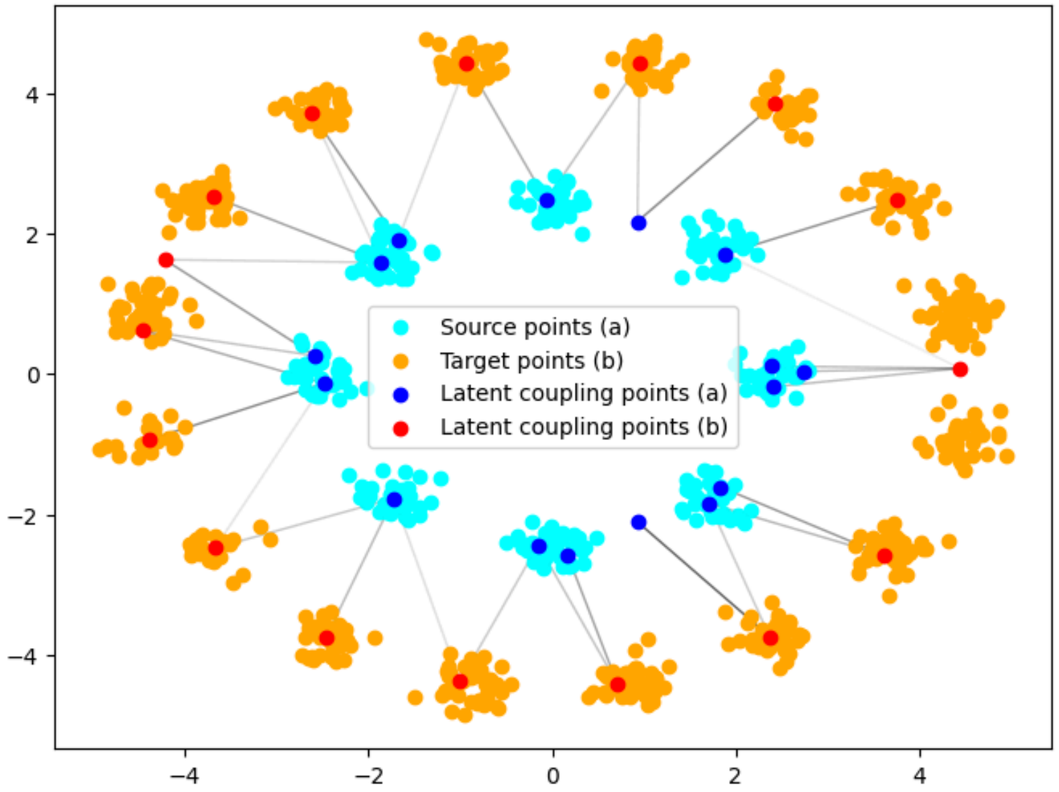


LOT (Scetbon 2021)

Visualizing the projection: 8-16 points



LOT (Scetbon 2021)



FRLC (Halmos 2024)

Visualizing the projection: non square T

T is not necessarely square !

$$\min_{(\mathbf{Q},\mathbf{R},\mathbf{T})} \mathcal{L}_{\mathrm{LC}} = \langle \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \mathbf{M}\rangle_F$$

$$\text{s.t.} \quad g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m,$$

$$\mathbf{Q} \in \Pi_{a,\cdot}, \quad \mathbf{R} \in \Pi_{b,\cdot}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \mathbf{Q} \in \mathbb{R}^+_{n,r_1}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r_2}, \quad \mathbf{T} \in \mathbb{R}^+_{r_1,r_2},$$

Visualizing the projection: non square T

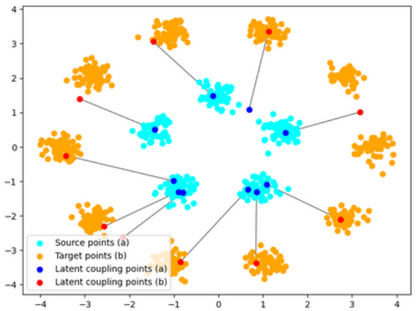
T is not necessarely square !

$$\min_{(\mathbf{Q},\mathbf{R},\mathbf{T})} \mathcal{L}_{\mathrm{LC}} = \langle \mathbf{Q}\,\mathrm{diag}\left(\frac{1}{g_Q}\right)\mathbf{T}\,\mathrm{diag}\left(\frac{1}{g_R}\right)\mathbf{R}^T, \mathbf{M}\rangle_F$$

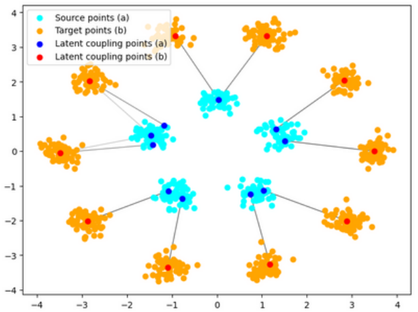
$$\text{s.t.} \quad g_Q := \mathbf{Q}^T 1_n, \quad g_R := \mathbf{R}^T 1_m,$$

$$\mathbf{Q} \in \Pi_{a,\cdot}, \quad \mathbf{R} \in \Pi_{b,\cdot}, \quad \mathbf{T} \in \Pi_{g_Q,g_R}, \mathbf{Q} \in \mathbb{R}^+_{n,r_1}, \quad \mathbf{R} \in \mathbb{R}^+_{m,r_2}, \quad \mathbf{T} \in \mathbb{R}^+_{r_1,r_2},$$

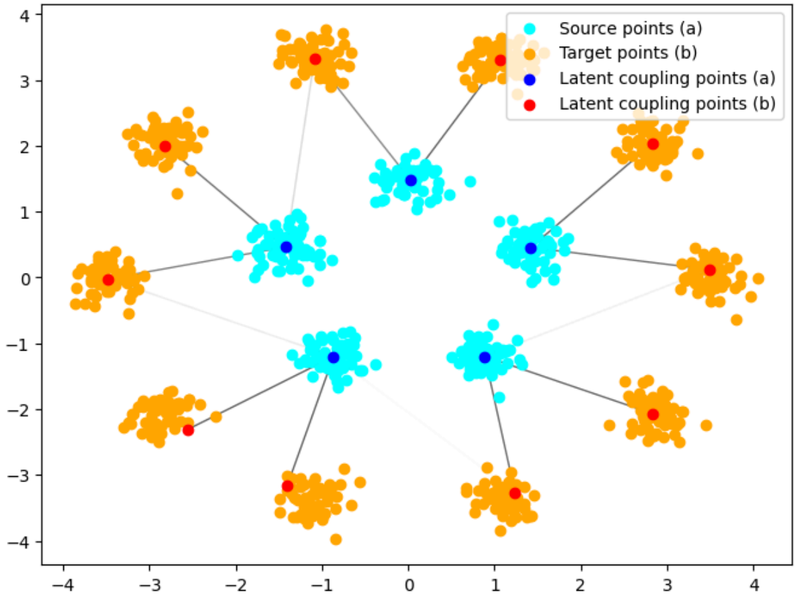
Recall, with T square :



LOT (Scetbon 2021)



FRLC (Halmos 2024)



FRLC (Halmos 2024)

Benchmark : real datasets

-Single-cell RNA sequencing captures cell encoded as vectors at different time points, but due to its destructive nature, the progression of individual cells over time cannot be tracked.

-For a pair of time points (ti, tj), the problem is determining which descendants of cell x at time ti give rise to at time tj.

$\Longrightarrow$ [Schiebinger'19] proposes unbalanced optimal transport problem.

Benchmark : real datasets

-Single-cell RNA sequencing captures cell encoded as vectors at different time points, but due to its destructive nature, the progression of individual cells over time cannot be tracked.

-For a pair of time points (ti, tj), the problem is determining which descendants of cell x at time ti give rise to at time tj.

$\Longrightarrow$ [Schiebinger'19] proposes unbalanced optimal transport problem.

Method	$\langle P, C \rangle$	$\|P1_m - a\|_2$	$\|P^T 1_n - b\|_2$	Time
Sinkhorn(POT)	0.406	0.0001	6.37×10^{-15}	0.8s
Sinkhorn(Ott)	0.24	0.1626	9.13×10^{-8}	2.1s
LOT(Ott)[r=5]	0.406	0.0003	6.9×10^{-9}	40s
LOT(Ott)[r=10]	0.406	0.0003	5.97×10^{-9}	1m
FRLC(original)[r=5]	0.3834	0.001	0.0009	4.5s
FRLC(original)[r=10]	0.3731	0.0012	0.0012	4.5s

Table 1: Comparison of methods on single-cell trajectory inference problem.

*Here we have n=4556 m=3449, we did PCA(30) as preprocessing as done in the original paper.

*epsilon = 5, reg_a = 1 (equivalent to tau_a = 1/(1+epsilon) in ott)

Benchmark : real datasets

If the problem where balanced:

Method	$\langle P, C \rangle$	$\|P1_m - a\|_2$	$\|P^T 1_n - b\|_2$	Time
EMD2	0.3309	1.1×10^{-16}	9×10^{-18}	2s
Sinkhorn(POT)	0.4067	5.132×10^{-15}	5.874×10^{-15}	0.9s
Sinkhorn(Ott)	0.3491	3.92×10^{-8}	1.39×10^{-5}	2.5s
LOT(Ott)[r=5]	0.4068	6.96×10^{-9}	6.52×10^{-9}	0.7s
LOT(Ott)[r=50]	0.4068	7.08×10^{-9}	6.37×10^{-9}	22.3s
FRLC(original)[r=5]	0.3822	0.0009	0.0008	4.5s
FRLC(original)[r=50]	0.3581	0.001	0.001	5.8s
FRLC(our code)[r=5]	0.4068	8.96×10^{-7}	1×10^{-6}	**0.2s**
FRLC(our code)[r=50]	0.4067	2×10^{-6}	2.24×10^{-6}	**0.4s**

Table 2: Comparison of methods on single-cell trajectory inference balanced problem.

*epsilon = 10

Limits

High sensitivity to hyperparmeters :

- **number of iterations** required for a satisfactory solution highly dependent on the dataset,
- may fall outside the feasibility domain quickly if **step size** and **inner marginals regularization** not tuned.

Conclusion

- **Effective complexity** compared to other methods,
- **Validation of the method** over artificial and real datasets,
- **Capturing data structure** by identifing latent coupling points,
- **Hyperparameter Sensitivity** is the major drawback.

References

- Halmos, P., Liu, X., Gold, J., & Raphael, B. J. (2024). Low-Rank Optimal Transport through Factor Relaxation with Latent Coupling. In Advances in Neural Information Processing Systems 38 (NeurIPS 2024).
- Scetbon, M., Cuturi, M., & Peyré, G. (2021). Low-Rank Sinkhorn Factorization. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021).
- Scetbon, M., Klein, M., Palla, G., & Cuturi, M. (2023). Unbalanced Low-Rank Optimal Transport Solvers. In Advances in Neural Information Processing Systems 37 (NeurIPS 2023).
- Altschuler, J., Bach, F., Rudi, A., and Niles-Weed, J. Massively scalable sinkhorn distances via the nyström method, 2018.
- Tibshirani, R. J. (2017). Dykstra's Algorithm, ADMM, and Coordinate Descent: Connections, Insights, and Extensions. In Advances in Neural Information Processing Systems 31 (NeurIPS 2017).
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., & Lander, E. S. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming.
- Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T. H., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., & Vayer, T. (2021). POT: Python Optimal Transport.
- Cuturi, M., Meng-Papaxanthos, L., Tian, Y., Bunne, C., Davis, G., & Teboul, O. (2022). Optimal Transport Tools (OTT): A JAX Toolbox for all things Wasserstein.
- Forrow, A., Hütter, J.-C., Nitzan, M., Rigollet, P., Schiebinger, G., & Weed, J. (Year). Statistical Optimal Transport via Factored Couplings. MIT, Harvard University, Broad Institute.