Low-rank Optimal Transport through Factor Relaxation with Latent Coupling

Peter Halmos*, Xinhao Liu*, Julian Gold*, Benjamin J Raphael

PRINCETON UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Background

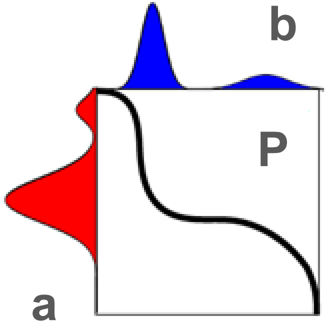
- *Optimal transport* finds a least-cost coupling **P** between two probability distributions **a** and **b**

- A coupling between **a** and **b** is a member of the set $\Pi_{\boldsymbol{a},\boldsymbol{b}}$ which is given as $\Pi_{\boldsymbol{a},\boldsymbol{b}} := \Pi_{\boldsymbol{a},\cdot} \cap \Pi_{\cdot,\boldsymbol{b}}.$

  for $\Pi_{\boldsymbol{a},\cdot} := \{\boldsymbol{P} \in \mathbb{R}_+^{n\times m} : \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}\},\ \Pi_{\cdot,\boldsymbol{b}} := \{\boldsymbol{P} \in \mathbb{R}_+^{n\times m} : \boldsymbol{P}^{\mathrm{T}}\mathbf{1}_n = \boldsymbol{b}\},$

  P has marginal distribution **a**

  P has marginal distribution **b**



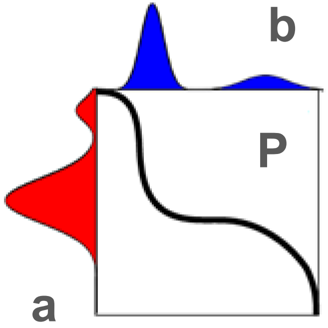
(source: Peyre & Cuturi)

Background

- The primal problem in optimal transport finds the least-cost coupling **P** for a distance **C**

$$\min_{\boldsymbol{P} \in \Pi_{\boldsymbol{a},\boldsymbol{b}}} \langle \boldsymbol{C}, \boldsymbol{P} \rangle_F$$

- The value of this cost is the *Wasserstein distance* between **a** and **b**

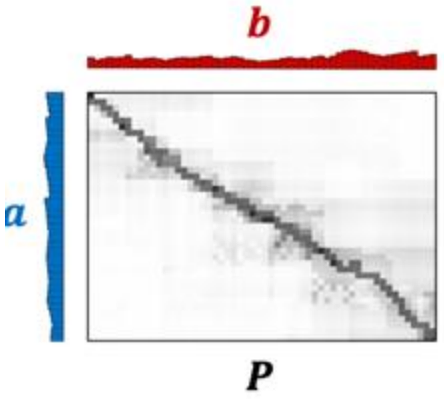


(source: Peyre & Cuturi)

Scaling OT

- Sinkhorn algorithm (Cuturi ‘13) scaled OT in terms of time-complexity (to $\mathrm{poly}(1/\eta\epsilon)$) from classical methods

- Existing methods have quadratic space complexity, limiting application to massive datasets



Full-rank P

Low rank OT

- OT with linear space complexity (Scetbon '21, Forrow '19, Lin '21)

- Scetbon '21 computes a coupling **P** subject to a rank-r constraint

$$\mathrm{W}_r(\mu, \nu) := \min_{\boldsymbol{P} \in \Pi_{\boldsymbol{a},\boldsymbol{b}}(r)} \langle \boldsymbol{C}, \boldsymbol{P} \rangle_F$$

where one defines: $\Pi_{\mathbf{a},\mathbf{b}}(r) = \left\{ \mathbf{P} \in \Pi_{\mathbf{a},\mathbf{b}} : \mathrm{rank}_+[\mathbf{P}] \leq r \right\}$

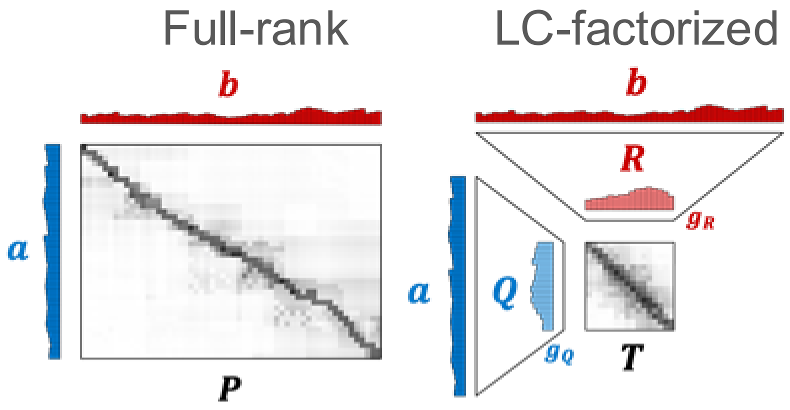
Low-rank factorization of the Coupling

- (Scetbon '21) factor the coupling **P** with sub-coupling factors and an inner diagonal matrix

$$\mathbf{P} = \mathbf{Q}\mathrm{diag}(1/\mathbf{g})\mathbf{R}^T$$

- Following Lin et al '21, we parametrize the coupling **P** with a latent coupling (LC) factorization with latent coupling **T**

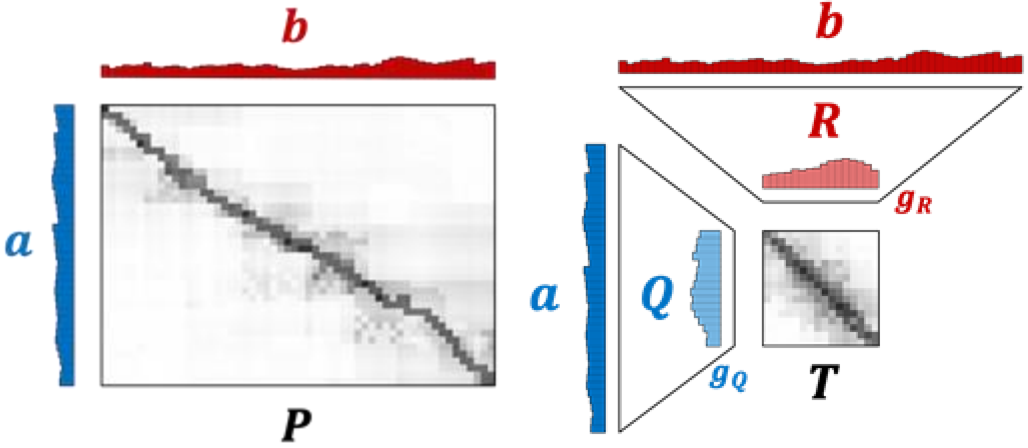
$$\mathbf{P} = \mathbf{Q}\mathrm{diag}(1/\mathbf{g}_Q)\mathbf{T}\mathrm{diag}(1/\mathbf{g}_R)\mathbf{R}^T$$



The LC-factorization

This coarsens the coupling **P** and the distributions **a**, **b** to latent representations of each

$$(\boldsymbol{P}, \boldsymbol{a}, \boldsymbol{b}) \rightarrow (\boldsymbol{T}, \boldsymbol{g}_Q, \boldsymbol{g}_R)$$



Full-rank LC-factorized

Factor Relaxation with Latent Coupling (FRLC)

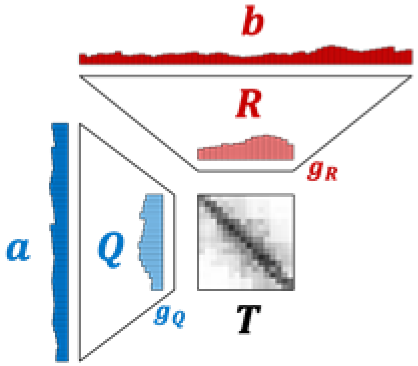
- *Coordinate* mirror-descent algorithm with three OT sub-problems
- Algorithm sketch for balanced OT:

$$\boldsymbol{Q}_k \leftarrow \tilde{\Pi}_{\boldsymbol{a},\boldsymbol{g}_Q}(\boldsymbol{Q}_k \odot \exp(-\gamma_k \nabla_{\boldsymbol{Q}})) \quad \text{(Semi-Relaxed OT)}$$

$$\boldsymbol{R}_k \leftarrow \tilde{\Pi}_{\boldsymbol{b},\boldsymbol{g}_R}(\boldsymbol{R}_k \odot \exp(-\gamma_k \nabla_{\boldsymbol{R}})) \quad \text{(Semi-Relaxed OT)}$$

$$\boldsymbol{g}_Q, \boldsymbol{g}_R = \boldsymbol{Q}_k^{\mathrm{T}} \mathbf{1}_n, \boldsymbol{R}_k^{\mathrm{T}} \mathbf{1}_m$$

$$\boldsymbol{T}_k \leftarrow \Pi_{\boldsymbol{g}_R,\boldsymbol{g}_Q}(\boldsymbol{T}_k \odot \exp(-\gamma_{\boldsymbol{T}} \nabla_{\boldsymbol{T}})) \quad \text{(Balanced OT)}$$



Convergence Guarantees

- Extend non-asymptotic stationary convergence of mirror descent for smooth objectives (Ghadimi et al ‘14) to coordinate mirror descent

- Use to show convergence of FRLC in (**Q**, **R**, **T**)

$$\min_k \Delta(\mathbf{x}_k, \mathbf{x}_{k-1}) \leq \frac{D^2 L}{N(\alpha^2/2L)} = \frac{2D^2L^2}{N\alpha^2},$$

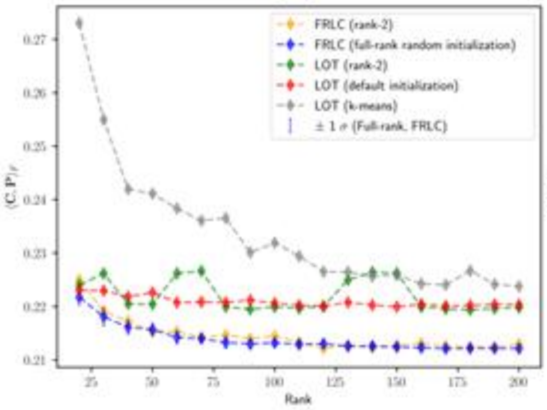
for criterion:

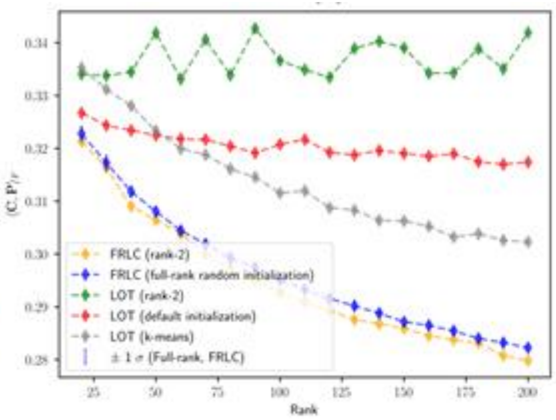
$$\Delta_k(\boldsymbol{x}_k, \boldsymbol{x}_{k+1}) := \frac{1}{\gamma_k^2}\left[\|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k\|_F^2 + \|\boldsymbol{R}_{k+1} - \boldsymbol{R}_k\|_F^2 + \|\boldsymbol{T}_{k+1} - \boldsymbol{T}_k\|_F^2\right]$$

Experimental Validation

FRLC has lower OT cost and higher performance on downstream metrics across synthetic and real datasets compared to LOT (Scetbon et al ‘21)

Benchmark on standard synthetic datasets





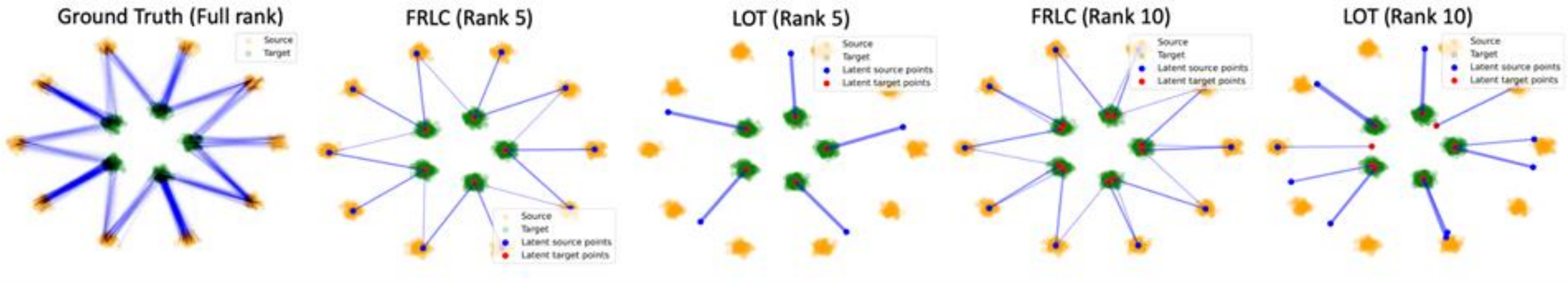
Benchmark on large-scale spatial transcriptomics alignment (Chen ‘22)

Objective	Algorithm	Spearman ρ	ARI	AMI
W	LOT-U	0.394	0.332	0.397
	FRLC-U	**0.465**	**0.466**	**0.484**
	LOT-SR	0.391	0.319	0.396
	FRLC-SR	**0.467**	**0.475**	**0.492**
GW	LOT-U	0.005	0.0	0.0
	FRLC-U	**0.266**	**0.299**	**0.364**
	LOT-SR	0.004	0.0	0.0
	FRLC-SR	**0.275**	**0.318**	**0.381**
FGW	LOT-U	**0.391**	0.330	0.400
	FRLC-U	0.368	**0.391**	**0.420**
	LOT-SR	0.396	0.337	0.401
	FRLC-SR	**0.465**	**0.469**	**0.499**

U=unbalanced, SR=semi-relaxed;
W=Wasserstein, GW=Gromov Wasserstein, FGW=Fused Gromow-Wasserstein

Visualization of LC-projections

LC-projection = barycentric projection of both datasets; can visualize these aligned by the latent coupling

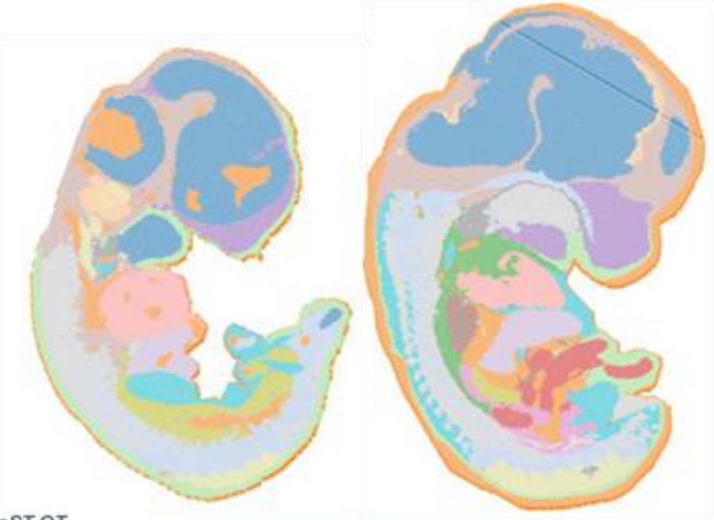


PRINCETON UNIVERSITY

Other Applications

Applications include co-clustering and the inference of growth-rates on large-scale spatial transcriptomics (Chen ‘22)

Large-scale computation of growth rates (*DeST-OT, Halmos & Liu et al 24*) using semi-relaxed FRLC on mouse embryogenesis dataset (Chen '22)



Co-clustering of massive-scale (100k+ points) mouse embryogenesis dataset (Chen '22)

Thank you!

Code: https://github.com/raphael-group/FRLC