Structured Matching via Cost-Regularized Unbalanced Optimal Transport

Emanuele Pardini
University of Pisa

Katerina Papagiannouli
University of Pisa

Abstract

Unbalanced optimal transport (UOT) provides a flexible way to match or compare non-negative finite Radon measures. However, UOT requires a predefined ground transport cost, which may misrepresent the data's underlying geometry. Choosing such a cost is particularly challenging when datasets live in heterogeneous spaces, often motivating practitioners to adopt Gromov–Wasserstein formulations. To address this challenge, we introduce cost-regularized unbalanced optimal transport (CR-UOT), a framework that allows the ground cost to vary while allowing mass creation and removal. We show that CR-UOT incorporates unbalanced Gromov–Wasserstein–type problems through families of inner-product costs parameterized by linear transformations, enabling the matching of measures (or point clouds) across Euclidean spaces. We develop algorithms for such CR-UOT problems using entropic regularization and demonstrate that this approach improves the alignment of heterogeneous single-cell omics profiles, especially when many cells lack direct matches.

1 INTRODUCTION

Optimal Transport (OT) has become a central tool in machine learning and related fields, providing a principled way to compare probability measures while respecting geometric structure. OT has been successfully applied to generative modeling (Bréchet et al., 2023; Genevay et al., 2018; Montavon et al., 2016; Tong, 2023), adversarial training (Sinha et al., 2018; Wong et al., 2019), domain adaptation (Courty et al., 2017; Fatras et al., 2021a), neuroscience (Janati et al., 2020), and single-cell biology (Bunne, 2023; Demetci et al., 2022a,b; Schiebinger et al., 2019a). In recent years, OT has become a powerful framework for addressing the graph matching problem. It is often viewed as a continuous relaxation of the Quadratic Assignment Problem, formulated through the Gromov-Wasserstein (GW) distance (Montavon et al., 2016; Xu et al., 2019), which extends the classical Wasserstein distance to compare distributions defined on different metric spaces. Several variants have been developed to handle labeled graphs, such as the Fused Gromov-Wasserstein (FGW) distance (Vayer et al., 2019).

Despite its successes, classical OT rests on two restrictive assumptions: (i) perfect mass preservation, and (ii) a fixed ground cost function. These assumptions often break down in applications. First, the *mass preservation* assumption requires that the marginals of the transport plan exactly match the input measures. This is unrealistic when data are noisy, incomplete, or inherently heterogeneous. To address this limitation, the framework of unbalanced optimal transport (UOT) has been developed, which relaxes the strict mass conservation constraint and allows for comparisons between measures of different total mass (Chizat et al., 2018; Demetci et al., 2022a; Kondratyev et al., 2016; Liero et al., 2016). A notable benefit of UOT is its robustness to outliers, since unmatched mass can be discarded rather than transported. This property has made UOT valuable in diverse applications, including deep learning theory (Chizat et al., 2018; Mazelet et al., 2025; Rotskoff and Vanden-Eijnden, 2019), single-cell biology (Demetci et al., 2022a,b; Schiebinger et al., 2019b), and domain adaptation (Fatras et al., 2021b). Unbalanced GW (Séjourné et al., 2019) and Fused Unbalanced GW (FUGW) (Thual et al., 2022) were proposed to generalize the GW and FGW distances to unbalanced settings with application in positive unlabeled learning and brain alignment.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

Second, OT requires specifying a *ground cost* that quantifies discrepancies between source and target points. In many applications—especially when measures lie in different ambient spaces or dimensions—this cost is unknown or may misrepresent the true geometry of the problem. Gromov–Wasserstein (GW) distances (Mémoli, 2011) circumvent this issue by aligning distributions through the comparison of relational structures within each space. While elegant, GW poses two difficulties: it results in a nonconvex quadratic optimization problem with high computational cost, and it forfeits some of the interpretability and guarantees available in linear OT settings (Dumont et al., 2022; Vayer, 2020).

Despite its flexibility, computing UOT remains computationally demanding: it requires solving a linear program whose complexity scales cubically with the number of samples (Pele and Werman, 2009; Peyré and Cuturi, 2019). Moreover, empirical estimation of UOT distances is challenging due to the curse of dimensionality (Dudley, 1969). To mitigate these issues, several tractable variants have been proposed with reduced complexity and improved statistical properties, such as entropic OT (Cuturi, 2013; Pham et al., 2020), minibatch OT (Fatras et al., 2020, 2021b), sliced UOT (Bonet et al., 2024), and cost regularized OT (Sebbouh et al., 2024).

In this work, we focus on extending the ideas of cost regularized OT to develop UOT methods. *This work addresses both challenges–unbalancedness and unknown ground cost–simultaneously and at the same time time computational efficiency.* We introduce *cost-regularized unbalanced OT (CR-UOT)* inspired by (Sebbouh et al., 2024), a framework that allows the ground cost to vary while relaxing marginal constraints. Our approach unifies and extends UOT and certain GW problems by introducing convex regularizers over costs, yielding families of parameterized linear inner-product costs across spaces. This formulation admits efficient algorithms via entropic regularization, while retaining connections to Monge maps through theoretical grounded approximation and convergent results. The proofs of the results are given in the Appendix.

Contributions. Our main contributions are:

- **Formulation:** We introduce CR-UOT, a framework combining convex cost regularization with unbalanced OT, unifying and extending existing UOT and GW formulations. In particular
- **Theory:** We prove existence of minimizers and establish convergence results of values and minimizers of the entropic regularized problem. Focusing on inner-product costs parameterized by linear transformations across spaces, we introduce a simple block coordinate descent algorithm to solve the associated CR-UOT problem. We show that, under mild conditions, optimal couplings are induced by deterministic Monge maps. We introduce entropic unbalnced Monge maps across spaces and show that they converge to the ground truth Monge maps under suitable assumptions.
- **Applications:** We demonstrate that the use of such entropic maps improves alignment of heterogeneous single-cell multiomics datasets, particularly when modalities lack direct correspondence or differ in proportions across cell types similar to (Demetci et al., 2022a,b).

2 BACKGROUND

Notations. In what follows, we consider $\mathcal{X}$ and $\mathcal{Y}$ to be compact metric spaces, and $\alpha \in \mathcal{M}^+(\mathcal{X})$ $\beta \in \mathcal{M}(\mathcal{Y})^+$ to be finite positive Radon measures satisfying $m(\alpha)m(\beta) \neq 0$, where $m(\mu)$ is the total mass of the measure μ. $\mathcal{C}(\mathcal{X})$, $\mathcal{C}_b(\mathcal{X})$ are continuous functions and bounded continuous functions on $\mathcal{X}$ respectively. Given $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$, we define its marginals $\pi_i = p_{i_\#}\pi$ for $i = 1, 2$.

Optimal Transport Problem (OT). Given the family of all possible *couplings* between α and β

$$\Pi(\alpha, \beta) = \left\{\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) : \ \pi_1 = \alpha, \ \pi_2 = \beta\right\},$$

we denote the linear OT problem between α and β with cost $c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})$ as

$$\mathsf{OT}(\alpha, \beta) \triangleq \min_{\pi \in \Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, \mathrm{d}\pi(x, y),$$

which is a *linear* problem in π, see (Santambrogio, 2015). When $c = d_{\mathcal{X}}^p$ and $\mathcal{X} = \mathcal{Y}$, the OT defines a distance between probability measures for all $p \geq 1$, see (Villani, 2008).

Unbalanced Optimal Transport (UOT) We recall the static formulation of UOT proposed by (Liero et al., 2018), which uses φ-divergence as penalty terms.

Definition 2.1 (φ-divergences)**.** *Let* $(\alpha, \beta) \in \mathcal{M}^+(\mathcal{X}) \times \mathcal{M}^+(\mathcal{X})$. *Let* $\varphi : [0, +\infty) \to [0, \infty]$ *be an entropy function, i.e.* φ *is convex and lower semicontinuous (lsc),* $\mathrm{dom}(\varphi) \triangleq \{x \in \mathbb{R}, \varphi(x) < +\infty\} \subset [0, +\infty)$ *and* $\varphi(1) = 0$. *Denote*

$$\varphi'_\infty \triangleq \lim_{x\to+\infty} \frac{\varphi(x)}{x}.$$

The φ-divergence between α and β is

$$\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int_{\mathbb{R}^d} \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right) \mathrm{d}\beta(x) + \varphi'_\infty \int_{\mathbb{R}^d} \mathrm{d}\alpha^\perp(x),$$

where $\alpha^\perp$ is defined as $\alpha = \frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\beta + \alpha^\perp$ in Lebesgue decomposition form. We call φ superlinear when $\varphi'_\infty = +\infty$.

A special case of φ-divergence, which we use later in the experiments and the formulation of entropic regularizers, is the Kullback-Leibler (KL) divergence:

$$\mathrm{D}_{\mathrm{KL}}(\alpha|\beta) = \begin{cases} \int_{\mathcal{X}} \varphi_{\mathrm{KL}}\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\right) \mathrm{d}\beta & \text{if } \alpha \ll \beta \\ +\infty & \text{otherwise,} \end{cases}$$

where $\varphi_{\mathrm{KL}}(x) = x\log(x) - x + 1$.

Problem 2.2 (UOT)**.** *For a lower semicontinuous cost $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and entropy functions φ_1, φ_2, we denote the unbalanced OT problem between α and β as*

$$\mathsf{UOT}(\alpha,\beta) \triangleq \inf_{\pi \in \mathcal{M}^+(\mathcal{X},\mathcal{Y})} \int c(x,y)\, d\pi(x,y) + \mathrm{D}_{\varphi_1}(\pi_1 \mid \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \mid \beta).$$

Observe that when $\varphi_1 = \varphi_2 = \iota_{\{1\}}$, with $\iota_{\{1\}}(1) = 0$ and $\iota_{\{1\}}(s) = +\infty$ for every $s \neq 1$, we find again the balanced $\mathsf{OT}(\alpha,\beta)$ problem.

In the results that follow we will usually assume one of the following compatibility conditions:

$$(m(\alpha)\mathrm{dom}(\varphi_1)) \cap (m(\beta)\mathrm{dom}(\varphi_2)) \neq \emptyset \tag{1}$$

and the stronger one

$$\begin{aligned}&[\mathrm{Int}(m(\alpha)\mathrm{dom}(\varphi_1)) \cap (m(\beta)\mathrm{dom}(\varphi_2))] \\ &\quad \cup [(m(\alpha)\mathrm{dom}(\varphi_1)) \cap \mathrm{Int}(m(\beta)\mathrm{dom}(\varphi_2))] \neq \emptyset.\end{aligned} \tag{2}$$

3 COST-REGULARIZED UOT

We now allow the cost itself to vary under a convex regularizer $\mathcal{R}$.

Problem 3.1 ($\mathsf{C\mathcal{R}UOT}$)**.** *Suppose we are given two entropy functions $\varphi_1, \varphi_2 : [0,+\infty) \to [0,+\infty]$ and a convex function $\mathcal{R} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$. We define the $\mathcal{R}$-regularized $\boldsymbol{\varphi}$-unbalanced optimal transport problem as*

$$\mathsf{C\mathcal{R}UOT}(\alpha,\beta) \triangleq \inf_{\pi,c} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, \mathrm{d}\pi(x,y) + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \mathcal{R}(c).$$

Remarkably, we can make connections between $\mathsf{C\mathcal{R}UOT}$ and another important family of problems involving concave functions of measures.

Problem 3.2 ($\mathsf{UO\mathcal{Q}T}$)**.** *Suppose we are given two entropy functions $\varphi_1, \varphi_2 : [0,+\infty) \to [0,+\infty]$ and a concave function $\mathcal{Q} : \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$. We define the $\boldsymbol{\varphi}$-unbalanced optimal $\mathcal{Q}$-transport problem as*

$$\mathsf{UO\mathcal{Q}T}(\alpha,\beta) \triangleq \inf_{\pi \in \mathcal{M}^+(\mathcal{X},\mathcal{Y})} \mathcal{Q}(\pi) \quad + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta).$$

In the appendix we show how solving a $\mathsf{C\mathcal{R}UOT}$ problem is the same as solving an $\mathsf{UO\mathcal{Q}T}$ problem from a certain concave functional $\mathcal{Q}$ built from $\mathcal{R}$. We say that two minimization problems are equivalent, or that one is an instance of the other, when they have the same minimizers in $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$.

Entropic regularization. In practice, a preferred way to solve UOT problems is using entropic regularization (Chizat et al., 2018). We consider adding such regularization also to $\mathsf{C\mathcal{R}UOT}$ problems.

Problem 3.3. *Suppose we are given $\varepsilon > 0$, two entropy functions $\varphi_1, \varphi_2 : [0,+\infty) \to [0,+\infty]$ and a convex function $\mathcal{R} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to [0,+\infty]$. We define the ε-entropic $\mathcal{R}$-regularized $\boldsymbol{\varphi}$-unbalanced optimal transport problem as*

$$\begin{aligned}\mathsf{C\mathcal{R}UOT}_\varepsilon(\alpha,\beta) \triangleq \inf_{\pi,c} &\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, \mathrm{d}\pi(x,y) \\ &+ \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) \\ &+ \mathcal{R}(c) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha \otimes \beta).\end{aligned}$$

3.1 Existence of Minimizers

Our first result consists in establishing the existence of optimal solutions for the CRUOT problems. We will need the following definition.

Definition 3.4 (Cost-Parametrized Regularizers)**.** *A convex function $\mathcal{R} : \mathcal{C}(\mathcal{X}\times\mathcal{Y}) \to [0,+\infty]$ is called* cost-parametrized regularizer *if there exists $\mathcal{F}$ a compact subset of $\mathbb{R}^d$ and a family of costs $\{c_\theta\}_{\theta\in\mathcal{F}} \subset \mathcal{C}(\mathcal{X}\times\mathcal{Y})$ s.t.*

$$\mathcal{R}(c) = \begin{cases} \tilde{\mathcal{R}}(\theta) & \text{if } c = c_\theta \text{ for some } \theta \in \mathcal{F} \\ +\infty & \text{otherwise,} \end{cases}$$

with $\tilde{\mathcal{R}} : \mathcal{F} \to [0,+\infty)$ a suitable lower semicontinuous function.

For this family of cost-parametrized regularizers we show that we can find an optimal solution for $\mathsf{C\mathcal{R}UOT}(\alpha,\beta)$ problems.

Theorem 3.5 (Existence)**.** *Let (φ_1, φ_2) be a pair of superlinear entropy functions satisfying (5) and $\varepsilon \geq 0$. Assume a cost-parametrized regularizer $\mathcal{R}$ as defined in Definition A.7 with $\{c_\theta\}_{\theta\in\mathcal{F}}$ a uniformly bounded from below family of continuous costs s.t. $c_{\theta_k} \to c_\theta$ uniformly whenever $\theta_k \to \theta$. Then the problem $\mathsf{C\mathcal{R}UOT}_\varepsilon(\alpha, \beta)$ admits at least one minimizer in $\mathcal{F} \times \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$.*

3.2 Convergence of entropic minimizers

Remarkably, the next result guarantees that, when the cost-regularization involves a sufficiently regular cost-parametrized regularizer (Definition A.7), the entropy-regularized $\mathsf{C\mathcal{R}UOT}_\varepsilon$ problem actually converges to the original $\mathsf{C\mathcal{R}UOT}$ when $\varepsilon \to 0$.

Theorem 3.6. *Let $\varepsilon_n \to 0$ and suppose that the assumptions of Theorem 3.5 to hold with φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5). Then the following hold.*

1. $\mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha, \beta) \overset{n\to+\infty}{\longrightarrow} \mathsf{C\mathcal{R}UOT}(\alpha, \beta)$.
2. *Consider a sequence $(\theta_*^{\varepsilon_n}, \pi_*^{\varepsilon_n})_{n\in\mathbb{N}} \subset \mathcal{F}\times\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$ s.t. $(\theta_*^{\varepsilon_n}, \pi_*^{\varepsilon_n})$ minimizes $\mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha, \beta)$ for every $n \in \mathbb{N}$. There exists a subsequence $(\theta_*^{\varepsilon_{n_k}}, \pi_*^{\varepsilon_{n_k}})_{k\in\mathbb{N}}$ s.t.*
$$\theta_*^{\varepsilon_{n_k}} \to \theta_*, \qquad \pi_*^{\varepsilon_{n_k}} \rightharpoonup \pi_*,$$
where (θ_*, π_*) is optimal for $\mathsf{C\mathcal{R}UOT}(\alpha, \beta)$.

3.3 IP-Cost-Regularized UOT

Fix $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{Y} \subset \mathbb{R}^q$ two compact domains of the respective euclidean space.

The Gromov-Wasserstein problem. Let us begin by recalling the definition of GW problem.

Problem 3.7 (GW)**.** *Fix $p \in [1, \infty)$. Consider two continuous cost functions $c_\mathcal{X} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $c_\mathcal{Y} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$. The GW_p problem is defined as*
$$\inf_{\pi\in\Pi(\alpha,\beta)} \left[\int_{(\mathcal{X}\times\mathcal{Y})^2} |c_\mathcal{X}(x, x') - c_\mathcal{Y}(y, y')|^p \,\mathrm{d}(\pi \otimes \pi)\right]^{\frac{1}{p}}.$$

We are particularly interested in the following particular case.

Problem 3.8 (GW-IP)**.** *Let $c_\mathcal{X}(x, x') = -\langle x, x'\rangle$ and $c_\mathcal{Y}(y, y') = -\langle y, y'\rangle$, we denote by $\mathsf{GW\text{-}IP}(\alpha, \beta)$ the problem*
$$\inf_{\pi\in\Pi(\alpha,\beta)} \int_{(\mathcal{X}\times\mathcal{Y})^2} |\langle x, x'\rangle - \langle y, y'\rangle|^2 \,\mathrm{d}(\pi \otimes \pi).$$

The main reason for our interest is the following result (Vayer, 2020) connecting GW-IP and cost-regularized optimal transport problems.

Proposition 3.9. *Let $r > 0$. Denote $\mathcal{F}_r := \{M \in \mathbb{R}^{q\times p} \mid \|M\|_F \leq r\}$. Then, GW-IP and the problem*
$$\min_{\pi\in\Pi(\alpha,\beta)} \min_{M\in\mathcal{F}_r} -\int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle \,\mathrm{d}\pi(x, y)$$
are equivalent, i.e. they have the same minimizers in π.

Inspired by the previous proposition, we define the following class of cost-regularized unbalanced optimal transport problems using inner product costs parametrized by linear transformations.

Problem 3.10 ($\mathsf{C\mathcal{R}_rUOT}$)**.** *Suppose we are given $r > 0$ and $\varepsilon \geq 0$. Consider the family of matrices $\mathcal{F}_r = \{M \in \mathbb{R}^{q\times p} \mid \|M\|_F \leq r\}$. We define the following problem $\mathsf{C\mathcal{R}_rUOT}_\varepsilon(\alpha, \beta)$*
$$\inf_{\substack{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})\\ M\in\mathcal{F}_r}} -\int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle \,\mathrm{d}\pi + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha \otimes \beta).$$
When $\varepsilon = 0$ we will use the alternative notation $\mathsf{C\mathcal{R}_rUOT}(\alpha, \beta)$.

Remark 3.11. *Clearly the previous problem is an instance of the general cost-regularized optimal transport Problem 3.1 with cost-parametrized regularizer $\mathcal{R}_r : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to [0, +\infty]$ defined by*
$$\mathcal{R}_r(c) = \begin{cases} 0 & \text{if } c(x, y) = -\langle Mx, y\rangle \text{ for } M \in \mathcal{F}_r \\ +\infty & \text{otherwise.} \end{cases}$$

Interestingly, it still has a connection with GW-IP.

Theorem 3.12. *Suppose $p \geq q$, φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5), fix $r > 0$ and $\varepsilon \geq 0$. For every $\pi \in \mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$ denote*
$$M(\pi) = \begin{cases} \frac{r}{\|C(\pi)\|_F} C(\pi) & \text{if } C(\pi) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$
where $C(\pi) = \int_{\mathcal{X}\times\mathcal{Y}} yx^T \,\mathrm{d}\pi(x, y)$. The problem $\mathsf{C\mathcal{R}_rUOT}_\varepsilon(\alpha, \beta)$ admits minimizers $(M_\varepsilon^*, \pi_\varepsilon^*)$ with $M_\varepsilon^* = M(\pi_\varepsilon^*)$. Moreover, if (M^*, π^*) minimizes $\mathsf{C\mathcal{R}_rUOT}(\alpha, \beta)$, then π^* minimizes $\mathsf{GW\text{-}IP}(\pi_1^*, \pi_2^*)$.

4 ENTROPIC MAPS

We show that under mild regularity, optimal couplings are induced by deterministic maps.

Definition 4.1. *A map $T : \mathcal{X} \to \mathcal{Y}$ is a* Monge map *for the problem $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha,\beta)$ if there exists $(M^*,\pi^*) \in \mathcal{F}_r \times \mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$ s.t. (M^*,π^*) is optimal for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha,\beta)$ and $\pi^* = (\mathrm{id}, T)_\# \pi_1^*$.)*

Denote $B_r := \{y \in \mathbb{R}^q \,|\, \|y\| \le r \max_{x\in\mathcal{X}} \|x\|\}$. Observe that $M_\#\mu \in \mathcal{M}^+(B_r)$ for every $M \in \mathcal{F}_r$ and $\mu \in \mathcal{M}^+(\mathcal{X})$.

Theorem 4.2. *Assume $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{Y} \subset \mathbb{R}^q$ compact domains, $p \ge q$, φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5) and α absolutely continuous w.r.t. the Lebesgue measure on $\mathcal{X}$. Then, there exists a Monge map for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha,\beta)$. In particular, for every optimal couple (M^*,π^*) for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha,\beta)$ there exists a map T_* s.t. $\pi^* = (\mathrm{id}, T_*)_\#\pi_1^*$. Moreover, if M^* is surjective then we can take*

$$T_* = -\nabla f_* \circ M^*$$

with $f_* \in \mathcal{C}(\mathbb{R}^q)$ an optimal Kantorovich potential for the problem $\mathsf{OT}^{c_{\mathrm{ip}}}(M^*_\#\pi_1^*, \pi_2^*)$ differentiable $M^*_\#\pi_1^*$-a.e., where $c_{\mathrm{ip}} = -\langle y', y\rangle$ for every $y, y' \in \mathbb{R}^q$.

Clarified the existence of a Monge map for the problem $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha,\beta)$ we turn to the task of its estimation using entropic regularization to leverage the computational advantages that it carries. We denote $(M^\varepsilon, \pi^\varepsilon)$ solutions to $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi},\varepsilon}(\alpha,\beta)$ and $\alpha^\varepsilon, \beta^\varepsilon$ the marginals of π^ε. We fix two sequences $(\varepsilon_n)_{n\in\mathbb{N}}, (\varepsilon'_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ s.t. $\varepsilon_n, \varepsilon'_j \to 0$.

Definition 4.3. *For every $j, n \in \mathbb{N}$ we define the* entropic map *$T_{j,n} : \mathbb{R}^p \to \mathbb{R}^q$ as follows*

$$T_{j,n}(x) = \frac{\int_{\mathcal{Y}} y \exp\left[\frac{1}{\varepsilon_n}(g_{j,n}(y) + \langle M^{\varepsilon'_j} x, y\rangle)\right] \mathrm{d}\beta^{\varepsilon'_j}(y)}{\int_{\mathcal{Y}} \exp\left[\frac{1}{\varepsilon_n}(g_{j,n}(y) + \langle M^{\varepsilon'_j} x, y\rangle)\right] \mathrm{d}\beta^{\varepsilon'_j}(y)}, \tag{3}$$

where $(f_{j,n}, g_{j,n}) \in \mathcal{C}(B_r) \times \mathcal{C}(\mathcal{Y})$ are optimal for

$$\sup_{f,g\in\mathcal{C}(B_r)\times\mathcal{C}(\mathcal{Y})} \int_{B_r} f \, dM^{\varepsilon'_j}_\# \alpha^{\varepsilon'_j} + \int_{\mathcal{Y}} g \, d\beta^{\varepsilon'_j} - \varepsilon_n \int_{B_r\times\mathcal{Y}} \left[\exp\left(\frac{f \oplus g - c_{\mathrm{ip}}}{\varepsilon_n}\right) - 1\right] d\big(M^{\varepsilon'_j}_\# \alpha^{\varepsilon'_j} \otimes \beta^{\varepsilon'_j}\big).$$

Theorem 4.4. *Assume $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{Y} \subset \mathbb{R}^q$ compact domains, $p \ge q$, φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5) and α absolutely continuous w.r.t. the Lebesgue measure on $\mathcal{X}$. Assume also $B_r \subset \mathcal{Y}$. Then there exists a subsequence $(\varepsilon'_{j_h})_{h\in\mathbb{N}}$ s.t. $M^{\varepsilon'_{j_h}} \to M^*$ optimal for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha,\beta)$. Moreover, suppose M^* surjective and that $M^*_\#\alpha^*$ and β^* satisfy the Assumptions A1-A3 in (Pooladian and Niles-Weed, 2021) for $\alpha \ge 2$. Then*

$$\lim_{n\to+\infty} \lim_{h\to+\infty} T_{j_h,n} = T_* \quad \text{in } L^2(\alpha^*),$$

where T_* is a Monge map for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha,\beta)$ which pushes α^* to β^*.

5 BLOCK COORDINATE DESCENT ALGORITHM FOR ENTROPIC IP-COST-REGULARIZED UOT

Consider the setup of the previous section. To approximate a solution for Problem 3.10 we use the following block coordinate descent algorithm (Bertsekas, 1997; Tseng, 2001). Initializing at $M_0 \in \mathcal{F}_r$, let

$$\begin{aligned}\pi^{k+1} &= \operatorname*{arg\,min}_{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})} -\int_{\mathcal{X}\times\mathcal{Y}} \langle M_k x, y\rangle \, d\pi(x,y) \\ &\quad + \mathrm{D}_{\varphi_1}(\pi_1 \mid \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \mid \beta) \\ &\quad + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi \mid \alpha\otimes\beta), \\ M_{k+1} &= \operatorname*{arg\,min}_{M\in\mathcal{F}_r} -\int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle \, d\pi^{k+1}(x,y).\end{aligned}$$

It is a well-studied problem in the optimization and machine learning literature. Adapting Lemma 4.2.2 in (Vayer, 2020) we get

Lemma 5.1. *Fix $r \in (0,+\infty)$ and $\pi \in \Pi(\alpha,\beta)$. Then*

$$r\left\|\int_{\mathcal{X}\times\mathcal{Y}} yx^T \, \mathrm{d}\pi(x,y)\right\|_F = \sup_{M\in\mathcal{F}_r} \int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle \, \mathrm{d}\pi(x,y)$$

and supremum is achieved by

$$M(\pi) = \begin{cases} \frac{r}{\|C(\pi)\|_F} C(\pi) & \text{if } C(\pi) \ne 0 \\ 0 & \text{otherwise} \end{cases}$$

where $C(\pi) = \int_{\mathcal{X}\times\mathcal{Y}} yx^T \, \mathrm{d}\pi(x,y)$.

By Lemma 5.1 the block coordinate descent algorithm becomes

$$\begin{aligned}\pi^{k+1} &= \operatorname*{arg\,min}_{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})} -\int_{\mathcal{X}\times\mathcal{Y}} \langle M_k x, y\rangle \, d\pi(x,y) \\ &\quad + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha\otimes\beta) \\ M_{k+1} &= M(\pi^{k+1})\end{aligned} \tag{4}$$

We now give a simple convergence result for the above algorithm in the practical case of discrete measures.

Theorem 5.2. *Let $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^p$, $\mathcal{Y} = \{y_j\}_{j=1}^m \subset \mathbb{R}^q$, $\alpha = \sum_{i=1}^n a_i \delta_{x_i}$, $\beta = \sum_{j=1}^m b_j \delta_{y_j}$ and $\varepsilon, r > 0$. Suppose $\{a_i\}_{i=1}^n, \{b_j\}_{j=1}^m \subset (0,+\infty)$ and the entropy functions φ_1 and φ_2 to be superlinear. Then, any limit point of the sequence $((M_k, \pi^k))_{k\in\mathbb{N}}$ defined by the block coordinate descent scheme (8) is a stationary point of the objective function of $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_\varepsilon(\alpha,\beta)$.*

Computational complexity. Let us analyze the computational complexity of the algorithm in (8). Consider two pointclouds $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^p$, $\mathcal{Y} = \{y_j\}_{j=1}^n \subset \mathbb{R}^q$. The π-step of alternate minimization algorithm above can be solved using the unbalanced Sinkhorn algorithm (Chizat et al., 2018; Pham et al., 2020; Séjourné et al., 2019), whose iterations are made of parallelizable operations and have a quadratic computational cost in n. In particular, as detailed in (Pham et al., 2020), the computational complexity of computing an ε-approximate solution via the unbalanced Sinkhorn algorithm is $O(n^2/\varepsilon)$, this represents a clear advantage of the unbalanced formulation when compared to the $O(n^2/\varepsilon^2)$ of the balanced Sinkhorn algorithm.

The M-step has instead the above closed form and its computation has a cost of $O(n^2pq)$.

6 APPLICATIONS TO SINGLE-CELL MULTI-OMICS ALIGNMENTS

We evaluate the effectiveness of our Algorithm (8) on two **single-cell multi-omics datasets** (Argelaguet, 2021). Each dataset consists of two tables that record different cellular characteristics (modalities), measured on cells of distinct types. The two modalities live in Euclidean spaces of different dimensions, and our goal is to align the cells across modalities with respect to their type.

Formally, we assign uniform probability measures α and β to the source and target datasets, respectively. For each choice of $\varepsilon, \varepsilon', \lambda > 0$, we compute an entropic map $T_{\varepsilon,\varepsilon'}$ (see (3)) that approximates a Monge map for the cost-regularized problem $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha, \beta)$ (see Theorem 4.4). Since alignments are always computed from the higher- to the lower-dimensional modality, we denote by *source modality data* the table containing the higher-dimensional measurements, and by *target modality data* the table containing the lower-dimensional ones. Importantly, in these datasets, the two modalities admit a one-to-one correspondence: each source measurement has a unique paired target measurement from the same cell, and every target cell appears in the source data.

To evaluate performance in more challenging conditions, we additionally simulate **unbalancedness** by subsampling the source and target data with cell-type-dependent proportions, thereby breaking the one-to-one correspondence.

All experiments are carried out in a **supervised setting**, where the cell type (label) is available for both modalities. Performance is quantified using **Label Transfer Accuracy (LTA)** (Demetci et al., 2022a,b), defined as the accuracy of predictions on aligned source data (in the target space) obtained by a k-nearest neighbors classifier trained on the target modality.

In all experiments, we set the entropy functions to $\varphi_1 = \varphi_2 = \lambda\varphi_{\mathrm{KL}}$, where $\varphi_{\mathrm{KL}}(x) = x\log(x) - x + 1$. We fix $k = 5$ for the k-NN classifier used in computing LTA, and set $\varepsilon = 5 \times 10^{-3}$ and $r = 1$ for the constraint set $\mathcal{F}_r$. The remaining hyperparameters ε' and λ are tuned by grid search.

6.1 scGEM dataset

The first dataset we use is the scGEM dataset (Cheow et al., 2016), (Demetci et al., 2022a,b) containing the gene expression and DNA methylation modalities of 177 human somatic cells. The source modality is the gene expression one, which have dimension $p = 34$, while the DNA methylation is the target modality, of dimension $q = 27$. The task is to match source and target datasets using an entropic map from the source to the target. In the left column of Table 1 are described the results of $\mathsf{C}\mathcal{R}_r\mathsf{UOT}$ on the full scGEM dataset when varying the parameter λ and the same kind of results is reported in the right column of Table 1 for the randomly subsampled scGEM dataset. In particular, for the latter experiment, we randomly pick two cell types and subsampled at 30% the cells of the first type in the gene expression domain (source) and at 30% the cells of the second type in the DNA methylation domain (target).

Note that the case $\lambda = +\infty$ corresponds to the problem $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_\varepsilon(\alpha, \beta)$ with $\varphi_1 = \varphi_2 = \iota_{\{1\}}$, which can be seen as an entropic-regularized version of the Gromov-Wasserstein problem $\mathsf{GW\text{-}IP}(\alpha, \beta)$ (see Proposition 3.9).

We observe that the unbalanced alignments ($\lambda < +\infty$), consistently outperform those obtained with the balanced formulation ($\lambda = +\infty$, (Sebbouh et al., 2024)). In particular, on the subsampled scGEM dataset, introducing unbalancedness yields a substantial improvement: the method effectively compensates for differences in cell-type proportions and the lack of one-to-one correspondence caused by subsampling. This highlights the importance of relaxing the mass conservation constraint in scenarios where the datasets exhibit sampling biases or partial overlap.

We also observe that our approach outperforms all the methods reported in (Demetci et al., 2022a).

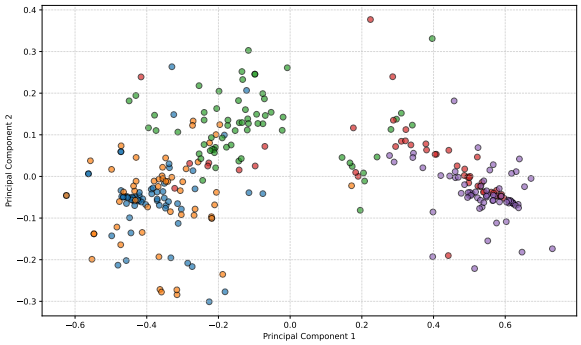


Figure 1: Visualization of the entropic map alignment of the subsampled scGM dataset with $\lambda = 1.0$ using two-dimensional PCA. Different colours refer to different cell types.

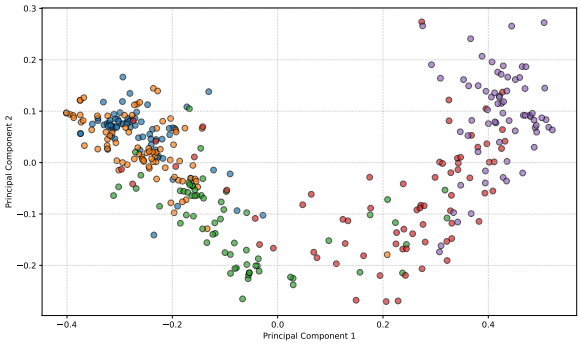


Figure 2: Visualization of the entropic map alignment of the full scGM dataset with $\lambda = 1.3$ using two-dimensional PCA. Different colours refer to different cell types.

Subsampled scGEM		Full scGEM	
λ	LTA	λ	LTA
$+\infty$	0.477	$+\infty$	0.661
1.5	0.638	2.0	0.667
1.3	**0.665**	1.5	0.672
1.0	0.664	1.3	**0.689**
0.7	0.604	1.0	0.661

Table 1: Left Column: LTA accuracies for subsampled scGEM; Right Column: LTA accuracies for the full scGEM

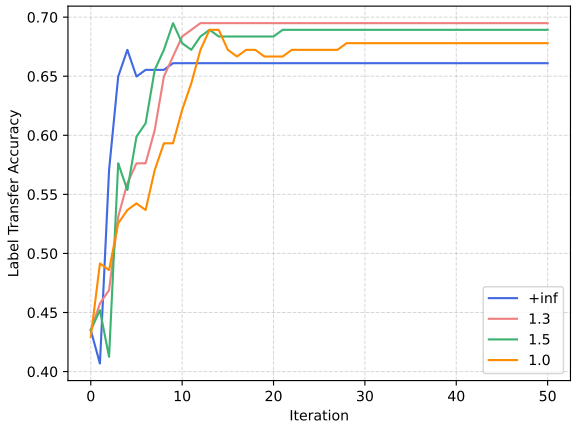


Figure 3: Plots of the LTA of the alignments for the full scGEM dataset obtained using the entropic map associated to the couple (M, P) at each iteration of the Algorithm 8. We report the balanced case and the three different choices of the unbalancedness hyperparameter λ.

6.2 SNAREseq dataset

The second dataset we use is the SNAREseq dataset (Chen et al., 2019), (Demetci et al., 2022a,b), containing the chromatine accessibility (ATAC-seq) and gene expression (RNA-seq) of 1047 single cells of 4 different types. The source ATAC-seq modality has dimension $p = 19$, while the target RNA-seq modality has dimension $q = 10$. As in the previous experiment, we align the source and target modality datasets using an entropic map from the source to the target. Results on both the full dataset and an unbalanced subsampling show that our method remains robust when cell-type proportions differ, with complete results and implementation details provided in the Appendix.

7 CONCLUSION AND FUTURE WORK

We introduced a unified framework that combines unbalanced optimal transport (UOT) with cost learning, supported by theoretical guarantees on the existence of minimizers, convergence, and Monge maps. On the computational side, we highlighted in the appendix the potential of low-rank parametrization of transport plans and regularizers which forcing sparcity in line with recent advances on scalable OT methods (Sebbouh et al., 2024). A systematic treatment of these approximations within our framework could substantially reduce memory and runtime complexity while preserving theoretical guarantees.

Acknowledgments

KP acknowledges the partial support of the project PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme. KP would like to thank Dario Trevisan and Andrea Agazzi for helpful discussions. KP was supported by Max Planck Institute for Mathematics in the Sciences in Leipzig.

References

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

R. Argelaguet. *Statistical methods for the integrative analysis of single-cell multi-omics data*. PhD thesis, 2021.

J. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge–kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 2000.

Y. Benyamini and J. Lindenstrauss. *Geometric nonlinear functional analysis*, volume 48. American Mathematical Soc., 1998.

D. Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

C. Bonet, K. Nadjahi, T. Séjourné, K. Fatras, and N. Courty. Slicing unbalanced optimal transport. *Transactions on Machine Learning Research*, 2024.

P. Bréchet, K. Papagiannouli, J. An, and G. Montúfar. Critical points and convergence analysis of generative deep linear networks trained with bures-wasserstein loss. In *International Conference on Machine Learning*, pages 3106–3147. PMLR, 2023.

C. Bunne. Learning single-cell multimodal data with optimal transport. *Nature Biotechnology*, 2023.

G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.

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

L. Cheow, E. Courtois, Y. Tan, R. Viswanathan, Q. Xing, Rui Z. Tan, D. SW Tan, P. Robson, Y. Loh, and S. Quake. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature methods*, 13(10):833–836, 2016.

S. Chewi, J. Niles-Weed, and P. Rigollet. Statistical optimal transport. *arXiv preprint arXiv:2407.18163*, 2024.

L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *NeurIPS*, pages 3040–3050, 2018.

L. Chizat, G. Peyré, B. Schmitzer, and F. Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314): 2563–2609, 2018.

S. Chowdhury and F. Mémoli. The gromov–wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.

N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems*, 2017.

M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NeurIPS*, pages 2292–2300, 2013.

P. Demetci, R. Santorella, M. Chakravarthy, B. Sandstede, and R. Singh. Scotv2: Single-cell multiomic alignment with disproportionate cell-type representation. *Journal of Computational Biology*, 29(11): 1213–1228, 2022a.

P. Demetci, R. Santorella, B. Sandstede, W. Noble, and R. Singh. Scot: single-cell multi-omics alignment with optimal transport. *Journal of computational biology*, 29(1):3–18, 2022b.

R. Dudley. The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40 (1):40–50, 1969.

L. Dumont, N. Papadakis, and G. Peyré. Existence and uniqueness for gromov–wasserstein optimal transport. *SIAM Journal on Mathematical Analysis*, 54(2):2022–2059, 2022.

T. Dumont, T. Lacombe, and F. Vialard. On the existence of monge maps for the gromov–wasserstein problem. *Foundations of Computational Mathematics*, pages 1–48, 2024.

Luca Eyring, Dominik Klein, Théo Uscidda, Giovanni Palla, Niki Kilbertus, Zeynep Akata, and Fabian Theis. Unbalancedness in neural monge maps improves unpaired domain translation. *arXiv preprint arXiv:2311.15100*, 2023.

K. Fatras, N. Courty, R. Flamary, and M. Cuturi. Minibatch optimal transport; computational and statistical aspects. In *NeurIPS*, pages 1–12, 2020.

K. Fatras, T. Séjourné, R. Flamary, and N. Courty. Unbalanced minibatch optimal transport; applications to domain adaptation. In *International conference on machine learning*, pages 3186–3197. PMLR, 2021a.

K. Fatras, Y. Zine, R; Flamary, R. Gribonval, and N; Courty. Unbalanced minibatch optimal transport; applications to domain adaptation. In *ICML*, pages 3186–3197, 2021b.

A. Genevay, G. Peyré, and M. Cuturi. Learning generative models with sinkhorn divergences. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1608–1617, 2018.

H. Janati, M. Cuturi, and A. Gramfort. Multi-subject dictionary learning and wasserstein means for alignment of medical data. *Advances in Neural Information Processing Systems*, 2020.

A. Klenke. *Probability theory: a comprehensive course*. Springer, 2008.

S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite radon measures. *Advances in Differential Equations*, 21(11/12):1117–1164, 2016.

M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869–2911, 2016.

M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

S. Mazelet, R. Flamary, and Thirion B. Unsupervised learning for optimal transport plan prediction between unbalanced graphs. 2025.

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

G. Montavon, K. Müller, and M. Cuturi. Wasserstein training of restricted boltzmann machines. In *Advances in Neural Information Processing Systems*, 2016.

M. Nutz. Introduction to entropic optimal transport. *Lecture notes, Columbia University*, 2021.

M. Nutz and Johannes Wiesel. Entropic optimal transport: Convergence of potentials. *Probability Theory and Related F.*, 184(1):401–424, 2022.

O. Pele and M. Werman. Fast and robust earth mover's distances. *2009 IEEE 12th International Conference on Computer Vision*, pages 460–467, 2009.

G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6): 355–607, 2019.

Khiem Pham, Khang Le, Nhat Ho, Tung Pham, and Hung Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, pages 7673–7682. PMLR, 2020.

M. Pham and N. Gozlan. Unbalanced optimal transport with entropic regularization. *arXiv preprint arXiv:2009.04269*, 2020.

A. Pooladian and J. Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

G. Rotskoff and E. Vanden-Eijnden. Trainability and accuracy of neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 72(6):1262–1298, 2019.

W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991. ISBN 9780070542365.

F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87. Springer, 2015.

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

G. Schiebinger, Jenny Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Andrew Solomon, Jesse Gould, Stanley Liu, Shirley Lin, Pierre Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019b.

O. Sebbouh, M. Cuturi, and G. Peyré. Structured transforms across spaces with cost-regularized optimal transport. In *International Conference on Artificial Intelligence and Statistics*, pages 586–594. PMLR, 2024.

T. Séjourné, J. Feydy, F. Vialard, A. Trouvé, and G. Peyré. Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958*, 2019.

A. Sinha, H. Namkoong, and J. Duchi. Certifying some distributional robustness with principled adversarial

training. In *International Conference on Learning Representations (ICLR)*, 2018.

K. Sturm. *The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces*, volume 290. American Mathematical Society, 2023.

A. Thual, Q. Tran, T. Zemskova, N. Courty, R. Flamary, S. Dehaene, and B. Thirion. Aligning individual brains with fused unbalanced gromov wasserstein. *Advances in neural information processing systems*, 35:21792–21804, 2022.

A. Tong. Improving and generalizing flow-based generative models with minibatch optimal transport. In *International Conference on Machine Learning (Workshop Track)*, 2023.

P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.

T. Vayer. A contribution to optimal transport on incomparable spaces. *arXiv preprint arXiv:2011.04447*, 2020.

T. Vayer, N. Courty, R. Tavenard, and R. Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

E. Wong, L. Schmidt, and J Z. Kolter. Wasserstein adversarial examples via projected sinkhorn iterations. In *International Conference on Machine Learning (ICML)*, 2019.

H. Xu, D. Luo, H. Zha, and L. Duke. Gromov-wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pages 6932–6941. PMLR, 2019.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes. We analyze the algorithm in terms of accuracy.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries.
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes. Included in the supplementary material.
   (c) Clear explanations of any assumptions. Yes
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). Yes
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Yes
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not Applicable
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. Yes
   (b) The license information of the assets, if applicable. Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable
   (d) Information about consent from data providers/curators. Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. Not Applicable
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

A APPENDIX FOR SECTION 3

Notation. We write $\mathcal{M}^+(Z)$ for finite nonnegative Radon measures on a compact metric space Z, and $\mathcal{C}(Z)$ for continuous real-valued functions on Z. For $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$, denote marginals by $\pi_i = p_{i\#}\pi$, $i = 1, 2$. For entropy functions φ_1, φ_2, we use the φ-divergences $\mathrm{D}_{\varphi_i}(\cdot \,|\, \cdot)$, and $\mathrm{D}_{\mathrm{KL}}(\cdot \,|\, \cdot)$ is the Kullback-Leibler divergence. We keep $\rho = \alpha \otimes \beta$ in entropic terms, where α, β is the source and target measure respectively.

A.1 Cost Regularized Unbalanced Optimal Transport

Problem A.1 ($\mathsf{C\mathcal{R}UOT}$ problems)**.** *Let $\mathcal{R} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be a convex function. We define the following cost-regularised problem*

$$\mathsf{C\mathcal{R}UOT}(\alpha, \beta) \triangleq \inf_{\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \Big\{ \int_{\mathcal{X} \times \mathcal{Y}} c \, \mathrm{d}\pi + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \mathcal{R}(c) \Big\}.$$

For $\varepsilon > 0$, we define the entropic cost-regularised unbalanced optimal transport

$$\mathsf{C\mathcal{R}UOT}_\varepsilon(\alpha, \beta) \triangleq \inf_{\pi, c} \Big\{ \int c \, \mathrm{d}\pi + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \mathcal{R}(c) + \varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho) \Big\}.$$

In the balanced case $\boldsymbol{\varphi} = (\iota_{(=)}, \iota_{(=)})$ we write $\mathsf{C\mathcal{R}OT}_\varepsilon(\alpha, \beta)$ and $\mathsf{C\mathcal{R}OT}(\alpha, \beta)$.

For completeness we recall the compatibility conditions

$$(m(\alpha)\mathrm{dom}(\varphi_1)) \cap (m(\beta)\mathrm{dom}(\varphi_2)) \neq \emptyset \tag{5}$$

and the stronger one

$$\begin{aligned} &[\mathrm{Int}(m(\alpha)\mathrm{dom}(\varphi_1)) \cap (m(\beta)\mathrm{dom}(\varphi_2))] \\ &\quad \cup [(m(\alpha)\mathrm{dom}(\varphi_1)) \cap \mathrm{Int}(m(\beta)\mathrm{dom}(\varphi_2))] \neq \emptyset. \end{aligned} \tag{6}$$

Remark A.2 (Feasibility conditions)**.**
Balanced case. *If $\varphi = (\iota_=, \iota_=)$, feasibility requires the compatibility condition* (1)*. Indeed, if the supports of α and β do not overlap under the marginal maps $m(\cdot)$, then there exists no admissible coupling with the prescribed marginals, and consequently $\Pi(\alpha, \beta) = \emptyset$, making $\mathrm{CROT}(\alpha, \beta) = +\infty$. Condition* (1) *thus ensures that at least some mass from α can be transported to β without violating the marginal constraints.*

Unbalanced case. *When general entropy functions φ_1, φ_2 are used, the feasibility of the relaxed formulation requires either* (1) *or the stronger condition* (2)*. The latter guarantees that the effective domains of the divergences are compatible, so that partial mass transfer is possible even when the supports of α and β do not perfectly coincide. In practice,* (2) *prevents degenerate situations where both divergences assign infinite cost to any nontrivial measure, ensuring that the unbalanced OT functional admits at least one finite value.*

A.2 From convex to concave functionals on plans

Problem A.3 ($\mathsf{UO\mathcal{Q}T}$ problems)**.** *Given entropy functions φ_1, φ_2 and a concave $\mathcal{Q} : \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$, define*

$$\mathsf{UO\mathcal{Q}T}(\alpha, \beta) \triangleq \inf_{\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \mathcal{Q}(\pi) + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta),$$

and $\mathsf{UO\mathcal{Q}T}_\varepsilon$ by adding $\varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho)$.

For fixed $\varepsilon \geq 0$, set

$$\mathcal{J}_{\boldsymbol{\varphi}, \mathcal{R}}(\pi) := \inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \int c \, \mathrm{d}\pi + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \mathcal{R}(c) + \varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho),$$

$$\mathcal{I}_{\boldsymbol{\varphi}, \mathcal{Q}}(\pi) := \mathcal{Q}(\pi) + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho).$$

Proposition A.4. ***(From $\mathcal{R}$ to $\mathcal{Q}$).***

Fix $\varepsilon \geq 0$. Let $\varphi_1, \varphi_2 : [0, +\infty) \to [0, +\infty]$ be entropy functions and $\mathcal{R} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to [0, +\infty]$ be convex such that, for every $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$,

$$\inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} c \, \mathrm{d}\pi + \mathcal{R}(c) \;\in\; \mathbb{R}.$$

Define the concave functional $\mathcal{Q}_{\mathcal{R}} : \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ by

$$\mathcal{Q}_{\mathcal{R}}(\pi) \;\triangleq\; \inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} c \, \mathrm{d}\pi + \mathcal{R}(c).$$

Then, for all $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$,

$$\mathcal{J}_{\boldsymbol{\varphi}, \mathcal{R}}(\pi) \;=\; \mathcal{I}_{\boldsymbol{\varphi}, \mathcal{Q}_{\mathcal{R}}}(\pi),$$

and in particular the minimizers coincide and

$$\mathsf{C}\mathcal{R}\mathit{UOT}_{\varepsilon}(\alpha, \beta) \;=\; \mathsf{UO}\mathcal{Q}_{\mathcal{R}}\,\mathit{T}_{\boldsymbol{\varphi}, \varepsilon}(\alpha, \beta).$$

Proof. $\mathcal{Q}_{\mathcal{R}}$ is concave as an infimum of affine maps of π. The identity $\mathcal{J}_{\boldsymbol{\varphi}, \mathcal{R}} = \mathcal{I}_{\boldsymbol{\varphi}, \mathcal{Q}_{\mathcal{R}}}$ follows immediately from the definitions, hence the equality of values and minimizers. □

Proposition A.5. ***(From $\mathcal{Q}$ to $\mathcal{R}$).*** *Fix $\varepsilon \geq 0$. Let $\mathcal{Q} : \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ be concave and weakly upper semicontinuous. Define $\bar{\mathcal{Q}} : \mathcal{M}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \cup \{-\infty\}$ by*

$$\bar{\mathcal{Q}}(\gamma) \;=\; \begin{cases} \mathcal{Q}(\gamma), & \gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}), \\ -\infty, & \textit{otherwise}, \end{cases}$$

and the convex functional $\mathcal{R}_{\mathcal{Q}} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathcal{R}_{\mathcal{Q}}(c) \;\triangleq\; (-\bar{\mathcal{Q}})^*(-c).$$

Then, for all $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$,

$$\mathcal{J}_{\boldsymbol{\varphi}, \mathcal{R}_{\mathcal{Q}}}(\pi) \;=\; \mathcal{I}_{\boldsymbol{\varphi}, \mathcal{Q}}(\pi),$$

and, in particular,

$$\mathsf{UO}\mathcal{Q}\mathsf{T}_{\varepsilon}(\alpha, \beta) \;=\; \mathsf{C}\mathcal{R}_{\mathcal{Q}}\mathsf{UOT}_{\varepsilon}(\alpha, \beta),$$

with the same set of minimizers.

Proof. Since $-\bar{\mathcal{Q}}$ is proper, convex and weakly lower semicontinuous on $\mathcal{M}(\mathcal{X} \times \mathcal{Y})$, Fenchel–Moreau yields

$$\bar{\mathcal{Q}}(\pi) = -(-\bar{\mathcal{Q}})^{**}(\pi) = -\sup_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \Big\{ \textstyle\int c \, \mathrm{d}\pi - (-\bar{\mathcal{Q}})^*(c) \Big\} = \inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \Big\{ \textstyle\int c \, \mathrm{d}\pi + (-\bar{\mathcal{Q}})^*(-c) \Big\}.$$

Adding $\mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho)$ on both sides gives

$$\mathcal{I}_{\boldsymbol{\varphi}, \mathcal{Q}}(\pi) = \inf_{c \in \mathcal{C}(\mathcal{X} \times \mathcal{Y})} \Big\{ \textstyle\int c \, \mathrm{d}\pi + \mathcal{R}_{\mathcal{Q}}(c) \Big\} + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho) = \mathcal{J}_{\boldsymbol{\varphi}, \mathcal{R}_{\mathcal{Q}}}(\pi).$$

□

Remark A.6. *The conjugate is taken with respect to the duality $\langle \gamma, c \rangle = \int c \, \mathrm{d}\gamma$ between $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ and $\mathcal{C}(\mathcal{X} \times \mathcal{Y})$. Weak topologies are the ones induced by this pairing.*

A.3 Proof of Theorem 3.5: Existence of minimizers for $\mathsf{C}\mathcal{R}\mathsf{UOT}$

For completeness we recall the definition of the class of cost-parametrized regularizers and the statement of Theorem 3.5.

Definition A.7 (Cost-Parametrized Regularizers)**.** *A convex function $\mathcal{R} : \mathcal{C}(\mathcal{X} \times \mathcal{Y}) \to [0, +\infty]$ is called* cost-parametrized regularizer *if there exist $\mathcal{F}$ a compact subset of $\mathbb{R}^d$ and a family of costs $\{c_\theta\}_{\theta\in\mathcal{F}} \subset \mathcal{C}(\mathcal{X} \times \mathcal{Y})$ s.t.*

$$\mathcal{R}(c) = \begin{cases} \tilde{\mathcal{R}}(\theta) & \text{if } c = c_\theta \text{ for some } \theta \in \mathcal{F} \\ +\infty & \text{otherwise,} \end{cases}$$

with $\tilde{\mathcal{R}} : \mathcal{F} \to [0, +\infty)$ a lower semicontinuous function.

Theorem A.8 (Existence)**.** *Let (φ_1, φ_2) be a pair of superlinear entropy functions satisfying (5) and $\varepsilon \geq 0$. Assume a cost-parametrized regularizer $\mathcal{R}$ as defined in Definition A.7 with $\{c_\theta\}_{\theta\in\mathcal{F}}$ a uniformly bounded from below family of continuous costs s.t. $c_{\theta_k} \to c_\theta$ uniformly whenever $\theta_k \to \theta$. Then the problem $\mathsf{C}\mathcal{R}\mathsf{UOT}_\varepsilon(\alpha, \beta)$ admit at least one minimizer in $\mathcal{F} \times \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$.*

Proof. Consider

$$J(\theta, \pi) \;=\; \int_{\mathcal{X}\times\mathcal{Y}} c_\theta \, d\pi \;+\; \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) \;+\; \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \, \mathrm{KL}(\pi \,|\, \rho) \;+\; \tilde{\mathcal{R}}(\theta),$$

where ρ is the reference measure in $\mathcal{X} \times \mathcal{Y}$ and $\rho = \alpha \times \beta$. Using the facts that $c_\theta \geq L$ and the convexity of φ_i, one gets the standard mass–coercivity bound

$$J(\theta, \pi) \geq m(\pi) \left(L + \frac{m(\alpha)}{m(\pi)} \varphi_1\Big(\frac{m(\pi)}{m(\alpha)}\Big) + \frac{m(\beta)}{m(\pi)} \varphi_2\Big(\frac{m(\pi)}{m(\beta)}\Big) \right),$$

which tends to $+\infty$ uniformly in θ when $m(\pi) \to \infty$, since φ_i are superlinear. Hence, the minimizers lie in

$$\mathcal{A} \;\triangleq\; \mathcal{F} \times \mathcal{B}_R^+, \qquad \mathcal{B}_R^+ \triangleq \{\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}) : m(\pi) \leq R\},$$

for some $R > 0$. Note that $\mathcal{B}_R^+$ is weakly compact by weak closedness and Banach-Alaoglou theorem.

So $\mathcal{A}$ is compact for the product topology $\tau \triangleq \tau_{\text{eucl}} \times \tau_{\text{weak}}$, where τ_{eucl} is the Euclidean topology on $\mathcal{F}$ and τ_{weak} is the weak topology in $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$.

It remains to show J is τ–l.s.c. Take a sequence $(\theta_k, \pi^k)_k \subset \mathcal{F} \times \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ such that $(\theta_k, \pi^k) \to (\theta, \pi)$ in τ. Then $m(\pi^k)$ is bounded, π^k weakly converges to π, and π_i^k weakly converges to π_i. By the uniform convergence $c_{\theta_k} \to c_\theta$ and the boundedness of the masses $m(\pi^k)$,

$$\liminf_{k\to\infty} \int c_{\theta_k} \, d\pi^k \geq \liminf_{k\to\infty} \left(- \|c_{\theta_k} - c_\theta\|_\infty \, m(\pi^k) + \int c_\theta \, d\pi^k \right) = \int c_\theta \, d\pi.$$

The mappings $\pi \mapsto \mathrm{D}_{\varphi_i}(\pi_i \,|\, \cdot)$ are weakly l.s.c. in the marginals. Also, $\mathrm{KL}(\cdot \,|\, \rho)$ is weakly l.s.c. on compact metric spaces. Therefore J is l.s.c. on $\mathcal{A}$, and by Weierstrass theorem there exists a minimizer of J on $\mathcal{A}$, see (Santambrogio, 2015, Box 1.1). Hence, $\mathrm{CRUOT}_\varepsilon(\alpha, \beta)$ admits at least one minimizer in $\mathcal{F} \times \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$.

□

A.4 Proof of Theorem 3.6: Convergence of Entropic Minimizers

For completeness, we restate Theorem 3.6.

Theorem A.9. *Let $\varepsilon_n \to 0$ and suppose that the assumptions of Theorem 3.5 hold with φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5). Then the following hold.*

1. $\mathsf{C}\mathcal{R}\mathsf{UOT}_{\varepsilon_n}(\alpha, \beta) \overset{n\to+\infty}{\longrightarrow} \mathsf{C}\mathcal{R}\mathsf{UOT}(\alpha, \beta)$.

2. *Consider a sequence* $(\theta_*^{\varepsilon_n}, \pi_*^{\varepsilon_n})_{n\in\mathbb{N}} \subset \mathcal{F} \times \mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$ *s.t.* $(\theta_*^{\varepsilon_n}, \pi_*^{\varepsilon_n})$ *minimizes* $\mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha,\beta)$ *for every* $n \in \mathbb{N}$. *There exists a subsequence* $(\theta_*^{\varepsilon_{n_k}}, \pi_*^{\varepsilon_{n_k}})_{k\in\mathbb{N}}$ *s.t.*

$$\theta_*^{\varepsilon_{n_k}} \to \theta_*, \qquad \pi_*^{\varepsilon_{n_k}} \rightharpoonup \pi_*,$$

where (θ_*, π_*) *is optimal for* $\mathsf{C\mathcal{R}UOT}(\alpha,\beta)$.

In order to prove Theorem 3.6 we need the following results. First Lemma gives us convergence of values. The last two Lemmas are well-known results in the literature of unbalanced optimal transport problems about optimal marginals (Liero et al., 2018).

Lemma A.10. *Suppose the same assumptions as in Theorem 3.6 hold. Let* $(\varepsilon_n)_{n\in\mathbb{N}} \subset (0,+\infty)$ *with* $\varepsilon_n \to 0$. *Assume that there exists a sequence* $(\eta_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ *with* $\eta_j \to 0$ *such that for every* $j \in \mathbb{N}$ *there exists* $\pi^j \in \mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$, $\theta_j \in \mathcal{F}$ *with* $\mathrm{D_{KL}}(\pi^j \,|\, \alpha\otimes\beta) < +\infty$ *and*

$$\int_{\mathcal{X}\times\mathcal{Y}} c_{\theta_j}\,\mathrm{d}\pi^j + \mathrm{D}_{\varphi_1}((\pi^j)_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}((\pi^j)_2 \,|\, \beta) + \tilde{\mathcal{R}}(\theta_j) \le \mathrm{C\mathcal{R}UOT}(\alpha,\beta) + \eta_j.$$

Then $\mathrm{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha,\beta) \to \mathrm{C\mathcal{R}UOT}(\alpha,\beta)$ *as* $n \to \infty$.

Proof. For each n we have

$$\begin{aligned}\mathsf{C\mathcal{R}UOT}(\alpha,\beta) \le \mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha,\beta) \le \int_{\mathcal{X}\times\mathcal{Y}} c_{\theta_j}\,\mathrm{d}\pi^j + \mathrm{D}_{\varphi_1}((\pi^j)_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}((\pi^j)_2 \,|\, \beta)\\ + \tilde{\mathcal{R}}(\theta_j) + \varepsilon_n\,\mathrm{D_{KL}}(\pi^j \,|\, \alpha\otimes\beta).\end{aligned}$$

Hence

$$\mathsf{C\mathcal{R}UOT}(\alpha,\beta) \le \liminf_{n\to\infty} \mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha,\beta) \le \limsup_{n\to\infty} \mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha,\beta) \le \mathsf{C\mathcal{R}UOT}(\alpha,\beta) + \eta_j.$$

Letting $j \to \infty$ gives the claim. □

Lemma A.11 (Fenchel–Kantorovich duality and optimal marginals, (Eyring et al., 2023; Liero et al., 2018))**.** *Let* $\varphi_1, \varphi_2 : [0,+\infty) \to [0,+\infty]$ *be proper l.s.c. strictly convex entropy functions, and* $c \in \mathcal{C}(\mathcal{X}\times\mathcal{Y})$. *Consider the unbalanced optimal transport problem*

$$\mathsf{UOT}^c(\alpha,\beta) = \inf_{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})} \int c\,\mathrm{d}\pi + \mathrm{D}_{\varphi_1}(\pi_1|\alpha) + \mathrm{D}_{\varphi_2}(\pi_2|\beta).$$

Then its Fenchel–Kantorovich dual reads

$$\mathsf{D}^c(\alpha,\beta) = \sup\left\{ -\int \varphi_1^*(-f)\,\mathrm{d}\alpha - \int \varphi_2^*(-g)\,\mathrm{d}\beta \,\middle|\, (f,g) \in \mathcal{C}(\mathcal{X})\times\mathcal{C}(\mathcal{Y}),\ f(x)+g(y) \le c(x,y) \right\}.$$

If (π_*, f_*, g_*) *are optimal for the primal and dual problems, then the optimal marginals satisfy*

$$\frac{\mathrm{d}\pi_{*,1}}{\mathrm{d}\alpha} = (\varphi_1^*)'(-f_*), \qquad \frac{\mathrm{d}\pi_{*,2}}{\mathrm{d}\beta} = (\varphi_2^*)'(-g_*).$$

Equivalently,

$$\alpha_* = (\varphi_1^*)'(-f_*)\,\alpha, \qquad \beta_* = (\varphi_2^*)'(-g_*)\,\beta.$$

This lemma guarantees we can approximate π_* by discrete (simple) plans with the same marginals α_*, β_*, while preserving continuity of the cost term.

Lemma A.12 (Block approximation, (Nutz, 2021))**.** *Suppose* $\mathcal{X}$ *and* $\mathcal{Y}$ *are compact metric spaces, and let* $\mu \in \mathcal{M}^+(\mathcal{X})$, $\nu \in \mathcal{M}^+(\mathcal{Y})$. *Fix a plan* $\pi \in \Pi(\mu,\nu)$. *Then, for every* $\delta > 0$, *there exists a plan* $\pi^\delta \in \Pi(\mu,\nu)$ *such that*

$$\pi^\delta \ll \mu\otimes\nu, \qquad \frac{\mathrm{d}\pi^\delta}{\mathrm{d}(\mu\otimes\nu)} \text{ is bounded}, \qquad \pi^\delta \rightharpoonup \pi \text{ as } \delta \to 0.$$

In particular, for any continuous cost $c \in \mathcal{C}(\mathcal{X}\times\mathcal{Y})$,

$$\int_{\mathcal{X}\times\mathcal{Y}} c\,\mathrm{d}\pi^\delta \longrightarrow \int_{\mathcal{X}\times\mathcal{Y}} c\,\mathrm{d}\pi \qquad \text{as } \delta \to 0.$$

Proof of Theorem 3.6 Let $(\theta_*, \pi_*) \in \mathcal{F} \times \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ be an optimal couple for the unregularized problem $\mathsf{C\mathcal{R}UOT}(\alpha, \beta)$ and denote by $\alpha_* := \pi_{*,1}$ and $\beta_* := \pi_{*,2}$ the first and second marginals of π_*. If we are in the balanced case, then $\alpha_* = \alpha$ and $\beta_* = \beta$. Otherwise, in the unbalanced setting, the optimal marginals α_*, β_* are reweighted versions of α, β determined by the optimal dual potentials (f_*, g_*) of the problem. Indeed, by the Fenchel–Kantorovich duality for the unbalanced problem (cf. Lemma A.11), the optimal marginals of the unregularized plan π_* satisfy

$$\frac{\mathrm{d}\pi_{*,1}}{\mathrm{d}\alpha} = (\varphi_1^*)'(-f_*), \qquad \frac{\mathrm{d}\pi_{*,2}}{\mathrm{d}\beta} = (\varphi_2^*)'(-g_*),$$

where (f_*, g_*) are the optimal dual potentials. Thus, we can write $\alpha_* = (\varphi_1^*)'(-f_*)\,\alpha$ and $\beta_* = (\varphi_2^*)'(-g_*)\,\beta$.

Let us set $\sigma_1 := (\varphi_1^*)'(-f_*)$ and $\sigma_2 := (\varphi_2^*)'(-g_*)$, so that $\alpha_* = \sigma_1\alpha$ and $\beta_* = \sigma_2\beta$. These σ_i are bounded positive densities (since f_*, g_* are bounded).

By Lemma A.12, there exists a sequence of couplings $(\pi_*^\delta)_{\delta>0} \subset \Pi(\alpha_*, \beta_*)$ such that $\pi_*^\delta \rightharpoonup \pi_*$.

Hence, for every $\eta > 0$, we can find a plan $\pi^\eta \in \Pi(\alpha_*, \beta_*)$ such that

$$\int c_{\theta_*}\,\mathrm{d}\pi^\eta + \mathrm{D}_{\varphi_1}(\alpha_* \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\beta_* \,|\, \beta) + \tilde{\mathcal{R}}(\theta_*) \le \mathsf{C\mathcal{R}UOT}(\alpha, \beta) + \eta.$$

This π^η is an η–optimal coupling for the unregularized problem.

Next we verify that π^η has finite Kullback–Leibler divergence with respect to $\alpha \otimes \beta$. Using the change–of–measure formula,

$$\frac{\mathrm{d}\pi^\eta}{\mathrm{d}(\alpha \otimes \beta)} = \frac{\mathrm{d}\pi^\eta}{\mathrm{d}(\alpha_* \otimes \beta_*)} \frac{\mathrm{d}(\alpha_* \otimes \beta_*)}{\mathrm{d}(\alpha \otimes \beta)} = \frac{\mathrm{d}\pi^\eta}{\mathrm{d}(\alpha_* \otimes \beta_*)}\,\sigma_1\sigma_2.$$

This decomposition uses the Radon–Nikodym derivative.The first factor is the density of π^η w.r.t. its own marginals. The second factor comes from the change of measures $\alpha_* \otimes \beta_* = \sigma_1\sigma_2(\alpha \otimes \beta)$.

Since σ_1, σ_2 are bounded, the product density above is bounded, so $\mathrm{D}_{\mathrm{KL}}(\pi^\eta \,|\, \alpha \otimes \beta) < +\infty$.This shows the approximating sequence satisfies the finite–KL condition required by Lemma A.10.

Applying Lemma A.10 with this family $(\pi^\eta)_\eta$ and θ_* yields the convergence of values (1).

Finally, for the convergence of minimizers (point (2)), the sequence $(\pi_*^{\varepsilon_n})_n$ is tight, since the coercivity estimate in Theorem A.8 implies a uniform bound $m(\pi_*^{\varepsilon_n}) \le R$. Thus, up to a subsequence, $\pi_*^{\varepsilon_n} \rightharpoonup \bar{\pi}$ weakly in $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$. The coercivity of the functional gives uniform mass bounds. Moreover, since $\mathcal{F}$ is compact we can also suppose $\theta_*^{\varepsilon_n} \to \bar{\theta} \in \mathcal{F}$.

By the uniform convergence of the costs c_θ and weak lower semicontinuity of the divergences, we have

$$\int c_{\bar{\theta}}\,\mathrm{d}\bar{\pi} + \mathrm{D}_{\varphi_1}(\bar{\pi}_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\bar{\pi}_2 \,|\, \beta) + \tilde{\mathcal{R}}(\bar{\theta}) \le \liminf_{n\to\infty} \mathsf{C\mathcal{R}UOT}_{\varepsilon_n}(\alpha, \beta).$$

Using the convergence of the values from (1), we conclude that $(\bar{\theta}, \bar{\pi})$ is optimal for the limit problem $\mathsf{C\mathcal{R}UOT}(\alpha, \beta)$.

A.5 Proof of Theorem 3.12

We restate here Theorem 3.12

Theorem A.13. *Suppose $p \ge q$. Fix $r > 0$ and $\varepsilon \ge 0$. For every $\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ denote*

$$C(\pi) := \int_{\mathcal{X}\times\mathcal{Y}} yx^\top\, d\pi(x, y) \in \mathbb{R}^{q\times p}, \qquad M(\pi) \;:=\; \frac{r}{\|C(\pi)\|_{\mathrm{F}}}\, C(\pi) \quad (\textit{with } M(\pi) := 0 \textit{ if } C(\pi) = 0).$$

Then, the problem

$$\mathsf{C\mathcal{R}_r UOT}_\varepsilon(\alpha, \beta) = \inf_{\substack{\pi \in \mathcal{M}^+(\mathcal{X}\times\mathcal{Y}) \\ \|M\|_{\mathrm{F}} \le r}} \Big\{ -\int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle\, d\pi \;+\; \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) \;+\; \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) \;+\; \varepsilon\, \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha \otimes \beta) \Big\}$$

admits minimizers $(M_\varepsilon^*, \pi_\varepsilon^*)$ *with* $M_\varepsilon^* = M(\pi_\varepsilon^*)$. *Moreover, if* (M^*, π^*) *minimizes* $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_\varepsilon(\alpha, \beta)$, *then* π^* *minimizes the reduced functional*

$$\mathcal{G}_\varepsilon(\pi) \;:=\; -r\,\|C(\pi)\|_{\mathrm{F}} + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon\,\mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha \otimes \beta) \quad \textit{over } \pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}).$$

Proof. We prove the Theorem in several steps. The existence part follows from Theorem A.8. Let us prove the second part.

Step 1: Optimal M for fixed π.
Fix π and set $C := C(\pi) = \int yx^\top \, d\pi$. Using $\langle Mx, y\rangle = \mathrm{tr}(yx^\top M^\top)$,

$$\int \langle Mx, y\rangle \, d\pi = \mathrm{tr}\Big(\Big(\int yx^\top \, d\pi\Big) M^\top\Big) = \langle C, M\rangle_{\mathrm{F}}.$$

Hence, for fixed π, the inner minimization in M is

$$\inf_{\|M\|_{\mathrm{F}} \le r} \{-\langle C, M\rangle_{\mathrm{F}}\} \;=\; -\sup_{\|M\|_{\mathrm{F}} \le r} \langle C, M\rangle_{\mathrm{F}}.$$

By Cauchy–Schwarz in the Frobenius inner-product, $\sup_{\|M\|_{\mathrm{F}} \le r} \langle C, M\rangle_{\mathrm{F}} = r\,\|C\|_{\mathrm{F}}$, attained at $M = M(\pi)$. Thus, for every fixed π,

$$\inf_{\|M\|_{\mathrm{F}} \le r} \Big\{ -\int \langle Mx, y\rangle \, d\pi \Big\} \;=\; -r\,\|C(\pi)\|_{\mathrm{F}}\,, \quad \text{with minimizer } M(\pi).$$

Step 2: Reduction to a problem on π.
Plugging the optimal $M(\pi)$ back gives the reduced functional

$$\mathcal{G}_\varepsilon(\pi) := -r\,\|C(\pi)\|_{\mathrm{F}} + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon\,\mathrm{D}_{\mathrm{KL}}(\pi \,|\, \alpha \otimes \beta).$$

Therefore

$$\mathsf{C}\mathcal{R}_r\mathsf{UOT}_\varepsilon(\alpha, \beta) \;=\; \inf_{\pi \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})} \mathcal{G}_\varepsilon(\pi),$$

Step 3: Conclusion.
Suppose (M^*, π^*) optimal for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_\varepsilon(\alpha, \beta)$, we need to prove that π^* minimizes $\mathcal{G}_\varepsilon$. If we could find $\tilde{\pi} \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ s.t. $\mathcal{G}_\varepsilon(\tilde{\pi}) < \mathcal{G}_\varepsilon(\pi^*)$, then $(M(\tilde{\pi}), \tilde{\pi})$ would give a strictly smaller joint value (by Step 2), contradicting optimality. Hence π^* minimizes the reduced functional. □

B APPENDIX FOR SECTION 4

B.1 Proof of Theorem 4.2

Theorem 4.2 will give us the existence of a Monge map for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}$ problems. We state the theorem here for completeness.

Theorem B.1. *Assume $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{Y} \subset \mathbb{R}^q$ are compact domains, $p \ge q$, and either*

1. *φ_1, φ_2 are superlinear, strictly convex and the strong compatibility (2) holds; or*
2. *(balanced case) $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ and the compatibility (1) holds.*

Assume moreover that α is absolutely continuous w.r.t. the Lebesgue measure on $\mathcal{X}$. Then every optimal couple (M^*, π^*) for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha, \beta)$ there exists a map T_* such that

$$\pi^* = (\mathrm{id}, T_*)_\# \pi_1^*.$$

If in addition M^* is surjective, there exists a convex Kantorovich potential $f_* \in \mathcal{C}(\mathbb{R}^q)$ for the linear OT problem on $\mathbb{R}^q$ with cost $c_{\mathrm{ip}}(y', y) := -\langle y', y\rangle$ between $M_\#^*\pi_1^*$ and π_2^*, differentiable $M_\#^*\pi_1^*$-a.e., such that

$$T_* \;=\; -\nabla f_* \circ M^* \qquad \pi_1^*\text{-a.e.}$$

In order to prove Theorem 4.2, we need the following results.

Proposition B.2 (Fenchel–Kantorovich duality for $-\langle\cdot,\cdot\rangle$)**.** *Let $\mu,\nu\in\mathcal{M}^+(\mathbb{R}^q)$ be finite measures with compact support (or with finite first moments). Consider the linear OT problem*

$$\mathrm{OT}^{c_{\mathrm{ip}}}(\mu,\nu) \;:=\; \inf_{\gamma\in\Pi(\mu,\nu)}\int_{\mathbb{R}^q\times\mathbb{R}^q} -\langle y',y\rangle\,\mathrm{d}\gamma(y',y).$$

Then the Kantorovich dual is

$$\mathrm{OT}^{c_{\mathrm{ip}}}(\mu,\nu) \;=\; \sup_{f\in\Gamma(\mathbb{R}^q)}\Big\{-\int_{\mathbb{R}^q} f(y')\,\mathrm{d}\mu(y') \;-\int_{\mathbb{R}^q} f^*(y)\,\mathrm{d}\nu(y)\Big\},$$

where $\Gamma(\mathbb{R}^q)$ denotes proper l.s.c. convex functions and f^* is the convex conjugate of f. Moreover, dual optimizers exist and there is no duality gap.

Proof. By Fenchel–Young, inequality for every f and every (y',y), $f(y')+f^*(y) \;\geq\; \langle y',y\rangle \iff -\langle y',y\rangle \;\geq\; -f(y')-f^*(y)$. Integrating now against any $\gamma\in\Pi(\mu,\nu)$ we get

$$\int -\langle y',y\rangle\,\mathrm{d}\gamma \;\geq\; -\int f\,\mathrm{d}\mu - \int f^*\,\mathrm{d}\nu.$$

Taking the infimum in γ and the supremum in f yields weak duality. Under the stated compactness assumption, the standard Kantorovich duality theorem applies to the l.s.c. cost $-\langle\cdot,\cdot\rangle$. This follows from Theorem Fenchel–Moreau on $\mathcal{M}(\mathbb{R}^q)\times\mathcal{C}(\mathbb{R}^q)$. □

Corollary B.3 (Optimality/KKT conditions)**.** *Let $\gamma^*\in\Pi(\mu,\nu)$ and $f_*\in\Gamma(\mathbb{R}^q)$ be primal/dual optimizers for Proposition B.2. Then:*

1. Support condition
$$\operatorname{spt}\gamma^* \;\subset\; \{(y',y)\in\mathbb{R}^q\times\mathbb{R}^q:\; y\in-\partial f_*(y')\}.$$
Equivalently, $y'\in\partial f_*^*(-y)$ on $\operatorname{spt}\gamma^*$.
2. Measurable selection: *There exists a measurable map $T^\sim:\mathbb{R}^q\to\mathbb{R}^q$ with $\gamma^*=(\mathrm{id},T^\sim)_\#\mu$ and $T^\sim(y')\in -\partial f_*(y')$ μ-a.e.*
3. Gradient form (a.e. differentiability): *if μ is absolute continuous w.r.t Lebesgue measure, then f_* is differentiable μ-a.e. and*
$$\gamma^*=(\mathrm{id},-\nabla f_*)_\#\mu,\qquad T^\sim(y')=-\nabla f_*(y')\quad \mu\text{-a.e.}$$

Proof. Optimality forces equality in Fenchel–Young γ^*-a.e., i.e., $f_*(y')+f_*^*(y)=\langle y',y\rangle$, which is equivalent to $y\in-\partial f_*(y')$. This gives (1). Disintegrating γ^* w.r.t. μ and choosing a measurable selector from the monotone set $-\partial f_*$ yields (2). If μ absolute continues with respect to Lebesgue measure, Alexandrov/Rademacher imply f_* is a.e. differentiable and the subgradient is single-valued a.e., giving (3). □

Proof of Theorem 4.2 We prove the theorem in several steps. *Step 1: Reduce the coupling to $\mathbb{R}^q\times\mathbb{R}^q$.* By the existence theorem for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}$ (Theorem A.8), there exists an optimal pair (M^*,π^*). We set the measures on $\mathbb{R}^q$

$$\mu \;\triangleq\; M^*_\#\pi^*_1,\qquad \nu \;\triangleq\; \pi^*_2.$$

Consider the pushforward plan on $\mathbb{R}^q\times\mathbb{R}^q$ defined by

$$\gamma^* \;\triangleq\; (M^*,\mathrm{id})_\#\pi^*.$$

Then $\gamma^*\in\Pi(\mu,\nu)$ and, by change of variables,

$$\int_{\mathcal{X}\times\mathcal{Y}} -\langle M^*x,y\rangle\,d\pi^*(x,y) \;=\; \int_{\mathbb{R}^q\times\mathbb{R}^q} -\langle y',y\rangle\,d\gamma^*(y',y). \tag{7}$$

Optimality of (M^*,π^*) implies that, for fixed M^*, π^* minimizes the UOT problems with cost $c_{M^*}(x,y)=-\langle M^*x,y\rangle$. Hence, by (7), γ^* is optimal for the *linear* OT problem on $\mathbb{R}^q$ between μ and ν with cost $c_{\mathrm{ip}}(y',y)=-\langle y',y\rangle$.

Step 2: Duality on $\mathbb{R}^q$ and graph structure. With $\mu = M^*_\# \pi^*_1$ and $\nu = \pi^*_2$ from B.3, the pushed-forward optimal plan $\gamma^* = (M^*, \mathrm{id})_\# \pi^*$ solves $\mathrm{OT}_{\mathrm{ip}}(\mu, \nu)$. By Corollary B.3, γ^* is a graph $(\mathrm{id}, T^\sim)_\# \mu$ with $T^\sim \in -\partial f_*$. Lifting back to $\mathcal{X}$ via $y' = M^* x$ gives $\pi^* = (\mathrm{id}, T_*)_\# \pi^*_1$ with $T_*(x) = T^\sim(M^* x)$, and if μ absolute continuous with respect to Lebesgue measure then $T_*(x) = -\nabla f_*(M^* x)$ π^*_1-a.e.

The Kantorovich dual for $c_{\mathrm{ip}}(y', y) = -\langle y', y \rangle$ is

$$\sup_{f \in \mathcal{C}(\mathbb{R}^q)} \Big\{ -\int f(y')\, d\mu(y') - \int f^*(y)\, d\nu(y) \Big\}.$$

Let f_* be an optimal potential. By Fenchel optimality, γ^* is concentrated on the set

$$\mathcal{G}_* \triangleq \{(y', y) \in \mathbb{R}^q \times \mathbb{R}^q \ : \ y \in -\partial f_*(y')\},$$

i.e. $y \in -\partial f_*(y')$ μ-a.e. (equivalently, $y' \in \partial f^*_*(-y)$). In particular, there exists $T^\sim : \mathbb{R}^q \to \mathbb{R}^q$ with

$$\gamma^* = (\mathrm{id}, T^\sim)_\# \mu \qquad \text{and} \qquad T^\sim(y') \in -\partial f_*(y') \quad \mu\text{-a.e.}$$

If f_* is differentiable μ-a.e. (this will be the case when μ is a.c. on $\mathbb{R}^q$), then $T^\sim(y') = -\nabla f_*(y')$ μ-a.e.

Step 3: Lift the graph back to $\mathcal{X}$. Define $T_* : \mathcal{X} \to \mathcal{Y}$ by

$$T_*(x) \ := \ T^\sim\big(M^* x\big).$$

Then

$$(M^*, \mathrm{id})_\# \big((\mathrm{id}, T_*)_\# \pi^*_1\big) = (\mathrm{id}, T^\sim)_\# \big(M^*_\# \pi^*_1\big) = (\mathrm{id}, T^\sim)_\# \mu = \gamma^*.$$

But $(M^*, \mathrm{id})_\# \pi^* = \gamma^*$. Since disintegration of measures with respect to the map $x \mapsto M^* x$ is unique up to π^*_1-null sets, and y under an optimal plan on $\mathbb{R}^q \times \mathbb{R}^q$ depends only on $y' = M^* x$, it follows that $\pi^* = (\mathrm{id}, T_*)_\# \pi^*_1$.

Step 4: Surjective case and differentiability. If M^* is surjective and α is absolute continuous w.r.t the Lebesgue measure, then $\pi^*_1 \ll \alpha$ in both the balanced case ($\pi^*_1 = \alpha$) and in the unbalanced case (first-order optimality gives π^*_1 absolute continuous with respect to α with continuous density). Hence $\mu = M^*_\# \pi^*_1$ is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^q$. By Alexandrov theorem, the optimal potential f_* is differentiable μ-a.e., and the optimal γ^* is induced by the map $y' \mapsto -\nabla f_*(y')$. Therefore, $T^\sim = -\nabla f_*$ μ-a.e., and the representation from Step 3 yields

$$T_*(x) = T^\sim(M^* x) = -\nabla f_*\big(M^* x\big) \qquad \pi^*_1\text{-a.e.},$$

as claimed.

B.2 Proof of Theorem 4.4

Clarified the existence of a Monge map for the problem $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha, \beta)$, we turn to the task of its approximation using entropic regularization to leverage the computational advantages.

In the following, for every $\varepsilon > 0$, we will denote $(\pi^\varepsilon, M^\varepsilon)$ an optimal couple for $\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi},\varepsilon}(\alpha, \beta)$ s.t. $M^\varepsilon = M(\pi^\varepsilon)$ and we name $\alpha^\varepsilon := \pi^\varepsilon_1$ and $\beta^\varepsilon := \pi^\varepsilon_2$. Observe that α^ε has support in $\mathcal{X}$, indeed α has support in $\mathcal{X}$ and $\pi^\varepsilon \ll \alpha \otimes \beta$ imples $\alpha^\varepsilon \ll \alpha$. Moreover, it will be useful to note that, since $\|Mx\| \le r \max_{x \in \mathcal{X}} \|x\|$ for every $M \in \mathcal{F}_r$ and $x \in \mathcal{X}$, the measure $M^\varepsilon_\# \alpha^\varepsilon$ has support contained in the compact ball $B_r := \{y \in \mathbb{R}^q \mid \|y\| \le r \max_{x \in \mathcal{X}} \|x\|\}$ for every $\varepsilon > 0$.

We fix two sequences $(\varepsilon_n)_{n \in \mathbb{N}}, (\varepsilon'_j)_{j \in \mathbb{N}} \subset (0, +\infty)$ s.t. $\varepsilon_n, \varepsilon'_j \to 0$.

Definition B.4. *For every $j, n \in \mathbb{N}$ we define the* entropic map $T_{n,j} : \mathbb{R}^p \to \mathbb{R}^q$ *as follows*

$$T_{n,j}(x) = \frac{\int_{\mathcal{Y}} y \exp\left[\frac{1}{\varepsilon_n}(g_{j,n}(y) + \langle M^{\varepsilon'_j} x, y \rangle)\right] \mathrm{d}\beta^{\varepsilon'_j}(y)}{\int_{\mathcal{Y}} \exp\left[\frac{1}{\varepsilon_n}(g_{j,n}(y) + \langle M^{\varepsilon'_j} x, y \rangle)\right] \mathrm{d}\beta^{\varepsilon'_j}(y)},$$

where $(f_{j,n}, g_{j,n}) \in \mathcal{C}(B_r) \times \mathcal{C}(\mathcal{Y})$ are optimal for $\mathsf{D}^{c_{\mathrm{ip}}}_{\varepsilon_n}(M^{\varepsilon'_j}_\# \alpha^{\varepsilon'_j}, \beta^{\varepsilon'_j})$, where

$$\mathsf{D}^{c_{\mathrm{ip}}}_{\varepsilon_n}(M^{\varepsilon'_j}_{\#}\alpha^{\varepsilon'_j},\beta^{\varepsilon'_j}) = \sup_{f,g\in\mathcal{C}(B_r)\times\mathcal{C}(\mathcal{Y})} \int_{B_r} f\,dM^{\varepsilon'_j}_{\#}\alpha^{\varepsilon'_j} + \int_{\mathcal{Y}} g\,d\beta^{\varepsilon'_j} - \varepsilon_n \int_{B_r\times\mathcal{Y}} \left[\exp\left(\frac{f\oplus g - c_{\mathrm{ip}}}{\varepsilon_n}\right) - 1\right] d\big(M^{\varepsilon'_j}_{\#}\alpha^{\varepsilon'_j}\otimes\beta^{\varepsilon'_j}\big).$$

Note that, in our setting, the hypothesis of Theorem 3.6 are satisfied, hence we can find a subsequence $(\varepsilon'_{j_h})_{h\in\mathbb{N}}$, independend of n, s.t. $\pi^{\varepsilon'_{j_h}} \rightharpoonup \pi^*$ and $M^{\varepsilon'_{j_h}} \to M^*$ with (π^*, M^*) optimal for $\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha,\beta)$. In particular, denoting for every $A\in\mathbb{R}^{q\times p}$ the cost $c_A(x,y) = -\langle Ax, y\rangle$, we have $c_{M^{\varepsilon'_{j_h}}} \to c_{M^*}$ uniformly. We name α^* and β^* the marginals of π^*. We have $M^{\varepsilon'_{j_h}}_{\#}\alpha^{\varepsilon'_{j_h}} \rightharpoonup M^*_{\#}\alpha^*$, indeed by weak convergence the family of measures $(\alpha^{\varepsilon'_{j_h}})_{h\in\mathbb{N}}$, and consequently also $(M^{\varepsilon'_{j_h}}_{\#}\alpha^{\varepsilon'_{j_h}})_{h\in\mathbb{N}}$, is bounded, therefore it suffices to prove

$$\int_{\mathbb{R}^q} \phi\,\mathrm{d}M^{\varepsilon'_{j_h}}_{\#}\alpha^{\varepsilon'_{j_h}} \to \int_{\mathbb{R}^q} \phi\,\mathrm{d}M^*_{\#}\alpha^*$$

for every $\phi\in\mathcal{C}_b(\mathbb{R}^q)$ Lipschitz continuous (Klenke, 2008, Theorem 13.16). Fix $\phi\in\mathcal{C}_b(\mathbb{R}^q)$ Lipschitz continuous and note that actually

$$\begin{aligned}\left|\int_{\mathbb{R}^q} \phi\,\mathrm{d}M^{\varepsilon'_{j_h}}_{\#}\alpha^{\varepsilon'_{j_h}} - \int_{\mathbb{R}^q} \phi\,\mathrm{d}M^*_{\#}\alpha^*\right| &\le \int_{\mathcal{X}} |\phi(M^{\varepsilon'_{j_h}}x) - \phi(M^*x)|\,\mathrm{d}\alpha^{\varepsilon'_{j_h}}(x) \\ &\quad + \left|\int_X \phi(M^*x)\,\mathrm{d}\left(\alpha^{\varepsilon'_{j_h}} - \alpha^*\right)(x)\right| \\ &\le L_\phi m(\alpha^{\varepsilon'_{j_h}})\|M^{\varepsilon'_{j_h}} - M^*\|_F \max_{x\in\mathcal{X}}\|x\| \\ &\quad + \left|\int_{\mathcal{X}} \phi(M^*x)\,\mathrm{d}\left(\alpha^{\varepsilon'_{j_h}} - \alpha^*\right)(x)\right| \\ &\to 0\end{aligned}$$

where L_ϕ is the Lipschitz constant of ϕ.

Proposition B.5. *For every $n\in\mathbb{N}$ define $T_n:\mathcal{X}\to\mathcal{Y}$ as*

$$T_n(x) = \frac{\int_{\mathcal{Y}} y\exp\left[\frac{1}{\varepsilon_n}(g_n(y) + \langle M^*x, y\rangle)\right]\,\mathrm{d}\beta^*(y)}{\int_{\mathcal{Y}} \exp\left[\frac{1}{\varepsilon_n}(g_n(y) + \langle M^*x, y\rangle)\right]\,\mathrm{d}\beta^*(y)},$$

for some $(f_n, g_n)\in\mathcal{C}(B_r)\times\mathcal{C}(\mathcal{Y})$ optimal for $\mathsf{D}^{c_{\mathrm{ip}}}_{\varepsilon_n}(M^*_{\#}\alpha^*,\beta^*)$. Then $T_{j_h,n}\to T_n$ in $L^2(\alpha^*)$ for every $n\in\mathbb{N}$.

Proof. From the previous discussion we know that $M^{\varepsilon'_{j_h}}_{\#}\alpha^{\varepsilon'_{j_h}} \rightharpoonup M^*_{\#}\alpha^*$ and $\beta^{\varepsilon'_{j_h}} \rightharpoonup \beta^*$ as $h\to\infty$. Fix $x_0\in\mathcal{X}$ and note that, up to replacing $(f_{j,n}, g_{j,n})$ by $(f_{j,n} - f_{j,n}(x_0),\, g_{j,n} + f_{j,n}(x_0))$, we may assume $f_{j,n}(x_0) = 0$ for every $j, n\in\mathbb{N}$.

In particular, by the compactness argument in (Genevay et al., 2018), we can find $(f_n, g_n)\in\mathcal{C}(B_r)\times\mathcal{C}(\mathcal{Y})$ optimal for $\mathsf{D}^{c_{\mathrm{ip}}}_{\varepsilon_n}(M^*_{\#}\alpha^*,\beta^*)$ and a subsequence $(\tilde{j}_h)_{h\in\mathbb{N}}$ of $(j_h)_{h\in\mathbb{N}}$ such that $f_{\tilde{j}_h,n}\to f_n$ and $g_{\tilde{j}_h,n}\to g_n$ uniformly on their domains as $h\to\infty$, for every fixed $n\in\mathbb{N}$.

To ease notation, for every $x\in\mathcal{X}$, $n\in\mathbb{N}$, $g\in\mathcal{C}(\mathcal{Y})$ and $A\in\mathbb{R}^{q\times p}$ define

$$F^x_n(g, A)(y) = \exp\left(\frac{1}{\varepsilon_n}\big(g(y) - c_A(x,y)\big)\right), \qquad y\in\mathcal{Y},$$

where $c_A(x,y) = -\langle Ax, y\rangle$ is the inner-product cost associated to A. Observe that

$$\begin{aligned}\left\|F^x_n(g_{\tilde{j}_h,n}, M^{\varepsilon'_{\tilde{j}_h}}) - F^x_n(g_n, M^*)\right\|_\infty &\le \omega_n\left(\left\|g_{\tilde{j}_h,n} - g_n\right\|_\infty + \left\|c_{M^{\varepsilon'_{\tilde{j}_h}}} - c_{M^*}\right\|_\infty\right) \\ &\longrightarrow 0 \quad \text{as } h\to\infty,\end{aligned}$$

where ω_n is the modulus of continuity of $t \mapsto \exp(t/\varepsilon_n)$ on the compact interval where the uniformly bounded functions $g_{\tilde{j}_h,n} - c_{M^{\varepsilon'_{\tilde{j}_h}}}$ and $g_n - c_{M^*}$ take their values.

Hence, for every $\phi \in \mathcal{C}(\mathcal{Y})$ and $n \in \mathbb{N}$,

$$\begin{aligned}
&\left| \int_{\mathcal{Y}} \phi\, F_n^x(g_{\tilde{j}_h,n}, M^{\varepsilon'_{\tilde{j}_h}})\, \mathrm{d}\beta^{\varepsilon'_{\tilde{j}_h}} - \int_{\mathcal{Y}} \phi\, F_n^x(g_n, M^*)\, \mathrm{d}\beta^* \right| \\
&\quad \leq \|\phi\|_\infty \left\| F_n^x(g_{\tilde{j}_h,n}, M^{\varepsilon'_{\tilde{j}_h}}) - F_n^x(g_n, M^*) \right\|_\infty \beta^{\varepsilon'_{\tilde{j}_h}}(\mathcal{Y}) \\
&\qquad + \left| \int_{\mathcal{Y}} \phi\, F_n^x(g_n, M^*)\, \mathrm{d}(\beta^{\varepsilon'_{\tilde{j}_h}} - \beta^*) \right| \longrightarrow 0
\end{aligned}$$

as $h \to \infty$, since $\phi F_n^x(g_n, M^*) \in \mathcal{C}(\mathcal{Y})$ and $\beta^{\varepsilon'_{\tilde{j}_h}} \rightharpoonup \beta^*$.

Consequently, we deduce that $T_{\tilde{j}_h,n}(x) \to T_n(x)$ pointwise for all $x \in \mathcal{X}$. Moreover, by Jensen's inequality,

$$\left\| T_{\tilde{j}_h,n}(x) \right\|^2, \ \|T_n(x)\|^2 \leq \max_{y \in \mathcal{Y}} \|y\|^2 < +\infty \quad \text{for all } x \in \mathcal{X},$$

so by the dominated convergence theorem we obtain $T_{\tilde{j}_h,n} \to T_n$ in $L^2(\alpha^*)$.

Finally, note that the above argument can be applied to every subsequence of $(T_{j_h,n})_{h \in \mathbb{N}}$, and that T_n is independent of the particular choice of optimal potential g_n in its expression (all such g_n differ only by an additive constant, see e.g. (Genevay et al., 2018). Therefore the whole sequence satisfies

$$T_{j_h,n} \to T_n \quad \text{in } L^2(\alpha^*)$$

for every fixed $n \in \mathbb{N}$. □

We restate here Theorem 4.4.

Theorem B.6. *Assume $\mathcal{X} \subset \mathbb{R}^p$, $\mathcal{Y} \subset \mathbb{R}^q$ compact domains, $p \geq q$, φ_1, φ_2 superlinear strictly convex satisfying (6) or $\varphi_1 = \varphi_2 = \iota_{\{1\}}$ satisfying (5) and α absolutely continuous w.r.t. the Lebesgue measure on $\mathcal{X}$. Assume also $B_r \subset \mathcal{Y}$. Then there exists a subsequence $(\varepsilon'_{j_h})_{h \in \mathbb{N}}$ s.t. $M^{\varepsilon'_{j_h}} \to M^*$ optimal for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha, \beta)$. Moreover, suppose M^* surjective and that $M^*_\# \alpha^*$ and β^* satisfy the Assumptions A1-A3 in (Pooladian and Niles-Weed, 2021) for $\alpha \geq 2$. Then*

$$\lim_{n \to +\infty} \lim_{h \to +\infty} T_{j_h,n} = T_* \quad \text{in } L^2(\alpha^*),$$

where T_* is a Monge map for $\mathsf{C}\mathcal{R}_r\mathsf{UOT}(\alpha, \beta)$ which pushes α^* to β^*.

Proof. From Proposition B.5 we have $T_{j_h,n} \to T_n$ in $L^2(\alpha^*)$ for every $n \in \mathbb{N}$. From Theorem B.1 we know that the Monge map for $\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi}}(\alpha, \beta)$ associated to the minimisers (M^*, π^*) is $T_* = -\nabla f_* \circ M^*$ where $f_* \in \mathcal{C}(B_r)$ is an optimal Kantorovich potential for $\mathsf{OT}^{c_{\mathrm{ip}}}(M^*_\# \alpha^*, \beta^*)$. In particular, $-\nabla f_*$ is the Monge map for the problem $\mathsf{OT}^{c_{\mathrm{ip}}}(M^*_\# \alpha^*, \beta^*)$, therefore $-\nabla f_* = \nabla \phi$ $M^*_\# \alpha^*$-a.e. by the hypothesis. Moreover, it is easy to see that $T_n = -\nabla f_n \circ M^*$ with (f_n, g_n) optimal for $\mathsf{D}^{c_{\mathrm{ip}}}_{\varepsilon_n}(M^*_\# \alpha^*, \beta^*)$ (see Theorem 2.7 and 3.16 in (Carlier et al., 2017) for the convergence of entropic OT potentials to Kantorovich potentials in the balanced case with cost c_{ip}). The claim follows by applying Corollary 1 in (Pooladian and Niles-Weed, 2021), indeed

$$\begin{aligned}
\int_{\mathcal{X}} \|T_n - T_*\|^2 \, \mathrm{d}\alpha^* &= \int_{\mathcal{X}} \| -\nabla f_n \circ M^* + \nabla f_* \circ M^* \|^2 \, \mathrm{d}\alpha^* \\
&= \int_{\mathcal{X}} \| -\nabla f_n - \nabla \phi \|^2 \, \mathrm{d}M^*_\# \alpha^* \to 0 \quad \text{as } n \to +\infty.
\end{aligned}$$

□

C APPENDIX FOR SECTION 5

To approximate a solution for Problem $\mathsf{C}\mathcal{R}_{\mathsf{r}}\mathsf{UOT}_{\varepsilon}$ we use the following block coordinate descent algorithm (Bertsekas, 1997):

$$\begin{aligned}\pi^{k+1} &= \underset{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})}{\arg\min} -\int_{\mathcal{X}\times\mathcal{Y}} \langle M_k x, y\rangle \,\mathrm{d}\pi(x,y) + \mathrm{D}_{\varphi_1}(\pi_1\,|\,\alpha) + \mathrm{D}_{\varphi_2}(\pi_2\,|\,\beta) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi\,|\,\alpha\otimes\beta)\\ M_{k+1} &= \underset{M\in\mathcal{F}_r}{\arg\min} -\int_{\mathcal{X}\times\mathcal{Y}} \langle Mx, y\rangle \,\mathrm{d}\pi^{k+1}(x,y),\end{aligned}$$

that adapting Lemma 4.2.2 in (Vayer, 2020) the algorithm becomes:

$$\begin{aligned}\pi^{k+1} &= \underset{\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})}{\arg\min} -\int_{\mathcal{X}\times\mathcal{Y}} \langle M_k x, y\rangle \,\mathrm{d}\pi(x,y) + \mathrm{D}_{\varphi_1}(\pi_1\,|\,\alpha) + \mathrm{D}_{\varphi_2}(\pi_2\,|\,\beta) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi\,|\,\alpha\otimes\beta)\\ M_{k+1} &= \frac{r}{\left\|\int_{\mathcal{X}\times\mathcal{Y}} yx^T \,\mathrm{d}\pi^{k+1}(x,y)\right\|_F} \int_{\mathcal{X}\times\mathcal{Y}} yx^T \,\mathrm{d}\pi^{k+1}(x,y).\end{aligned} \tag{8}$$

C.1 Proof of Theorem 5.2

Theorem C.1 (Theorem 5.2)**.** *Let $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^p$, $\mathcal{Y} = \{y_j\}_{j=1}^m \subset \mathbb{R}^q$, $\alpha = \sum_{i=1}^n a_i\delta_{x_i}$, $\beta = \sum_{j=1}^m b_j\delta_{y_j}$ and $\varepsilon, r > 0$. Suppose $\{a_i\}_{i=1}^n, \{b_j\}_{j=1}^m \subset (0,+\infty)$ and the entropy functions φ_1 and φ_2 to be superlinear. Then, any limit point of the sequence $((M_k,\pi^k))_{k\in\mathbb{N}}$ defined by the block coordinate descent scheme (8) is a stationary point of the objective function of $\mathsf{C}\mathcal{R}_{\mathsf{r}}\mathsf{UOT}_{\varepsilon}(\alpha,\beta)$.*

Proof. The objective function of $\mathsf{C}\mathcal{R}_{\mathsf{r}}\mathsf{UOT}_{\varepsilon}(\alpha,\beta)$ can seen as the function $J : \mathbb{R}^{n\times m}\times\mathbb{R}^{q\times p}\to\mathbb{R}$ s.t.

$$\begin{aligned}J(P,M) = -\sum_{i=1}^n\sum_{j=1}^m \langle Mx_i, y_j\rangle P_{i,j} &+ \sum_{i=1}^n \varphi_1\left(\frac{P_i}{a_i}\right)a_i + \sum_{j=1}^m \varphi_2\left(\frac{P^j}{b_j}\right)b_j\\ &+ \varepsilon\sum_{i=1}^n\sum_{j=1}^m\left[\frac{P_{i,j}}{a_ib_j}\log\frac{P_{i,j}}{a_ib_j} - \frac{P_{i,j}}{a_ib_j} + 1\right]a_ib_j\\ &+ \delta_{[0,+\infty)^{n\times m}}(P) + \delta_{\mathcal{F}_r}(M)\end{aligned}$$

where we identify any plan $\pi\in\mathcal{M}^+(\mathcal{X}\times\mathcal{Y})$ with the matrix $P = (\pi(\{(x_i,y_j)\}))_{i,j}\in[0,+\infty)^{n\times m}$ and we denote $P_i = \sum_{j=1}^m P_{i,j}$, $P^j = \sum_{i=1}^n P_{i,j}$ for every i and j.

Let us now rewrite the the function J as

$$J(P,M) = \langle C(M), P\rangle + \sum_i a_i\phi_1(P_i/a_i) + \sum_j b_j\phi_2(P^j/b_j) + \varepsilon\,\mathrm{KL}(P, ab^\top) + \delta_{[0,+\infty)^{n\times m}}(P) + \delta_{F_r}(M),$$

where $C(M) = \left(y_jx_i^\top M\right)_{i,j}$. So, we are able to decompose J as $J(P,M) = J_0(P,M) + g(P) + h(M)$, where

$$J_0(P,M) = \langle C(M), P\rangle + \sum_i a_i\phi_1(P_i/a_i) + \sum_j b_j\phi_2(P^j/b_j) + \varepsilon\,\mathrm{KL}(P, ab^\top),$$

$$g(P) = \delta_{[0,+\infty)^{n\times m}}(P), \qquad h(M) = \delta_{F_r}(M).$$

The smooth part J_0 is continuously differentiable, while g and h are proper, convex and lower semicontinuous.

Because ϕ_1,ϕ_2 are superlinear and the KL term controls the total mass of P, the quantity $\sum_{i,j}P_{ij}$ is uniformly bounded on every sublevel set of J. Since F_r is compact and $[0,+\infty)^{n\times m}$ is closed, all sublevel sets of J are compact. Now for fixed M, the subproblem

$$\min_{P\geq 0}\ J_0(P,M)$$

is strictly convex because of the KL regularization and the superlinear functions ϕ_1, ϕ_2. As a result, it has a unique minimizer $P^\star(M)$. Moreover, for $\varepsilon > 0$ the minimizer satisfies $P^\star_{ij}(M) > 0$ for all i, j, so $g(P)$ does not play a role in the opimization and J_0 is differentiable at $P^\star(M)$.

Now we need to consider the exact minimization in the M–block. For fixed P, the subproblem

$$\min_{M \in F_r} \langle C(M), P \rangle$$

is linear over the Frobenius ball $F_r = \{M : \|M\|_F \leq r\}$. Let $C(P) = \sum_{i,j} P_{ij}\, x_i y_j^\top$. If $C(P) \neq 0$, the unique minimizer is $M(P) = r\,\frac{C(P)}{\|C(P)\|_F}$. If $C(P) = 0$, every element of F_r is optimal, in order to keep the block-coordinate map single-valued we set $M(P) = 0$ in this case. Thus the M–update is always uniquely defined. Now, we get the convergence by applying Lemma 3.1 and Theorem 4.1 in (Tseng, 2001). The function J has the form $J = f + g + h$ with

$$f = J_0 \text{ (smooth)}, \qquad g(P) = \delta_{[0,+\infty)^{n\times m}}(P), \qquad h(M) = \delta_{F_r}(M).$$

Each block subproblem is solved exactly, and all sublevel sets of J are compact. Therefore, every limit point of the alternating minimization sequence (P_k, M_k) is a stationary point of J. This concludes the proof.

□

The practical pseudocode implementation of the alternate minimization scheme (8) in the discrete case is the following:

Algorithm 1: BCD for $\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi},\varepsilon}(\alpha, \beta)$

Input: Entropy functions φ_1, φ_2, numbers $\varepsilon, r > 0$, source $\alpha = \sum_{i=1}^n a_i \delta_{x_i}$ and target $\beta = \sum_{j=1}^m b_j \delta_{y_j}$ with $\mathcal{X} = \{x_i\}_{i=1}^n \subset \mathbb{R}^p$, $\mathcal{Y} = \{y_j\}_{j=1}^m \subset \mathbb{R}^q$, $a = (a_i)_{i=1}^n \in (0, +\infty)^n$, $b = (b_j)_{j=1}^m \in (0, +\infty)^m$
Output: M^ε, P^ε optimal for $\mathcal{R}_r\mathsf{UOT}_{\boldsymbol{\varphi},\varepsilon}(\alpha, \beta)$
$M \leftarrow M_0$;
$k \leftarrow 1$;
while $k \leq N$ *and* err $\leq$ tol **do**
  $C \leftarrow (-\langle M x_i, y_j \rangle)_{i,j}$;
  $P \leftarrow \text{Sinkhorn}(a, b, C, \varphi_1, \varphi_2, \varepsilon)$;
  $M \leftarrow \frac{k}{\|\sum_{i,j} y_j x_i^T P_{i,j}\|_F} \sum_{i,j} y_j x_i^T P_{i,j}$;
end
return M, P

D ADDITIONAL EXPERIMENTS

In this section, we provide further experiments to evaluate the effectiveness of our Algorithm 1. First, Figure 4 provides further insights to better grasp the effect of unbalancedness on the entropic map. We gradually increase the admitted unbalancedness by decreasing the parameter λ.

D.1 SNAREseq dataset

The second dataset we use is the SNAREseq dataset, containing the chromatine accessibility (ATAC-seq) and gene expression (RNA-seq) of 1047 single cells of 4 different types. The source ATAC-seq modality has dimension $p = 19$, while the target RNA-seq modality has dimension $q = 10$. Again, the task is to match source and target modality datasets using an entropic map from the source to the target. In Table 2 we report the results of $\mathsf{C}\mathcal{R}_r\mathsf{UOT}$ on the full SNAREseq dataset when varying the parameter λ and the same type of results are reported in Table 3 for the randomly subsampled SNAREseq dataset. For the experiment with the subsampled SNAREseq dataset, we randomly pick two cell types: in the source dataset we subsample them at 50% and the other two types at 75%; in the target dataset we subsample them at 75% and the other two types at 50%.

λ	LTA
$+\infty$	**0.944**
5.0	**0.944**
2.5	0.941
1.0	0.941
0.5	0.938

Table 2: Full SNAREseq dataset results.

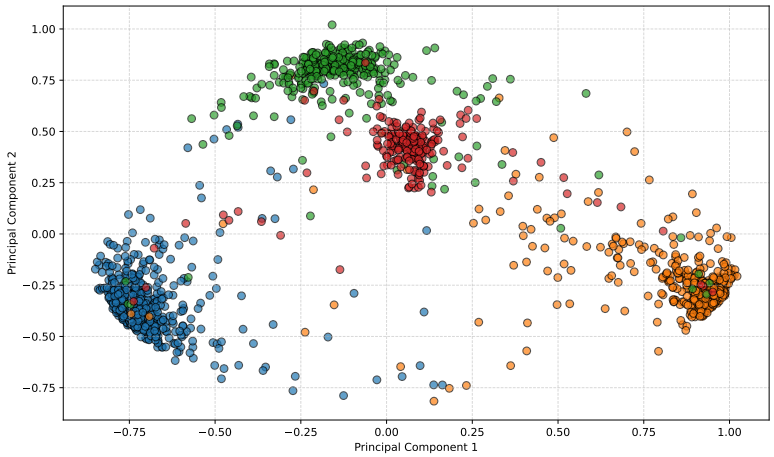


Figure 5: Visualization of the entropic map alignment of the full SNAREseq dataset with $\lambda = 5.0$ using two-dimensional PCA. Different colours refer to different cell types.

λ	LTA
$+\infty$	0.582
1.0	0.656
0.5	0.695
0.1	0.752
0.07	**0.761**

Table 3: Subsampled SNAREseq dataset results.

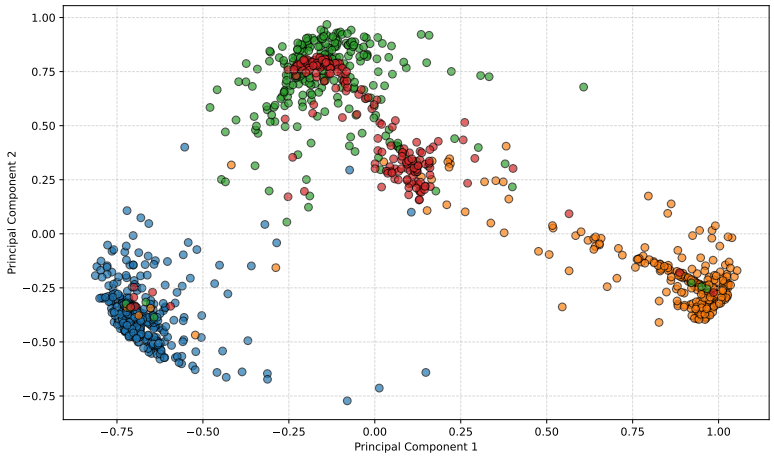
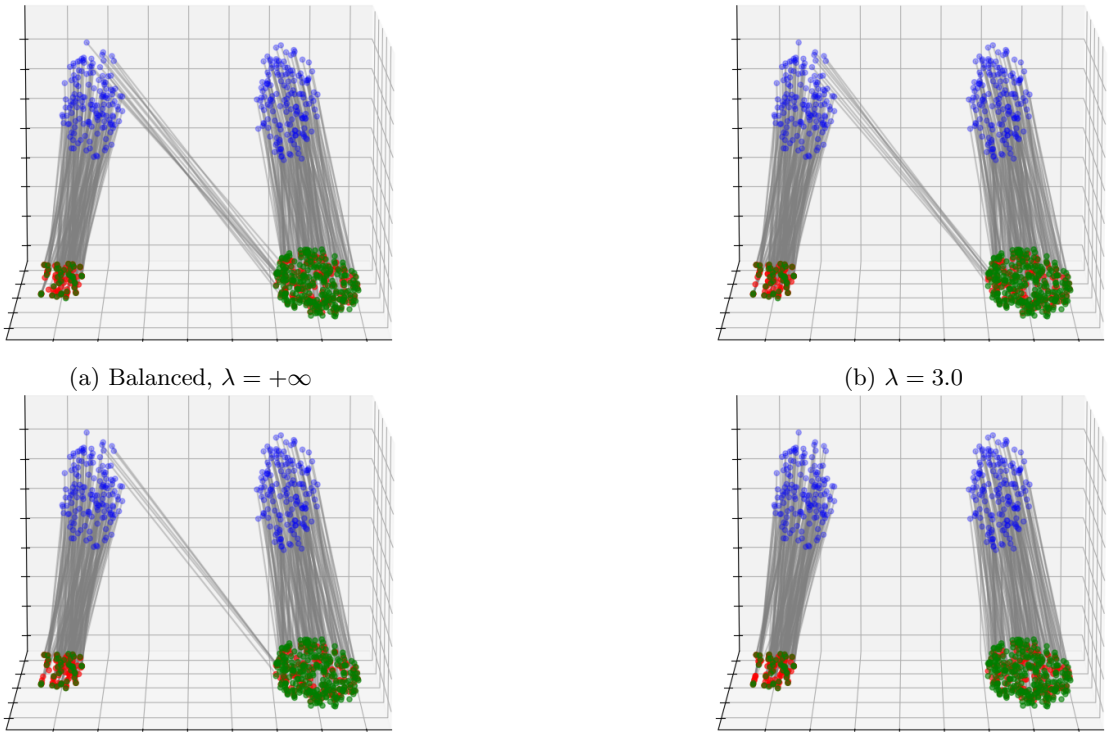


Figure 6: Visualization of the entropic map alignment of the subsampled SNAREseq dataset with $\lambda = 0.07$ using two-dimensional PCA. Different colours refer to different cell types.



(a) Balanced, $\lambda = +\infty$

(b) $\lambda = 3.0$

(c) $\lambda = 1.5$

(d) $\lambda = 0.5$

Figure 4: The source data (blue) is generated sampling from a balanced mixture of uniform distributions on two ellipsoids in $\mathbb{R}^3$, while the target data (green) is obtained by sampling from an unbalanced mixture of a square $\mathcal{S}$ and an ellipse $\mathcal{E}$ in $\mathbb{R}^2$. For visualization, $\mathbb{R}^2$ is embedded into $\mathbb{R}^3$ by setting the third coordinate to zero. The aligned source points are shown in red.

D.2 Details on the entropic map in the case where M^* is not surjective

The low-rank (or sparse) regularizations extend naturally to the cost-regularized unbalanced optimal transport problem with inner-product cost. Specifically, for costs of the form

$$c_M(x, y) = -\langle Mx,\, y\rangle \qquad \text{and} \qquad \mathcal{R}(M) = \tfrac{1}{2}\,\|M\|_F^2 + \lambda\, g(M),$$

where g encodes a structure constraint (e.g., nuclear norm, ℓ_1, or $\ell_{1,2}$-penalty, or an explicit factorization $M = M_2^\top M_1$), the corresponding $\mathcal{R}\mathsf{UOT}$ problem

$$\inf_{M,\pi\geq 0} \int_{\mathcal{X}\times\mathcal{Y}} -\langle Mx, y\rangle \,\mathrm{d}\pi + \mathcal{R}(M) + \mathrm{D}_{\varphi_1}(\pi_1 \,|\, \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 \,|\, \beta) + \varepsilon \mathrm{D}_{\mathrm{KL}}(\pi \,|\, \rho)$$

admits minimizers under the same hypotheses as in Theorem A.8. For fixed M, the minimization over π is precisely the entropic UOT with cost c_M, solvable via the unbalanced Sinkhorn algorithm. For fixed π, the update in M is a proximal step on $\int yx^\top \,\mathrm{d}\pi$ and takes the same closed form as in the balanced case.

On Monge maps. When α is absolute constinuous with respect to the and the learned linear operator M^* is *surjective* (i.e. of full column rank q), the assumptions of Theorem B.1 apply and the optimal coupling π^* is induced by a Monge map

$$\pi^* = (\mathrm{id}, T_*)_\# \pi_1^*, \qquad T_* = -\nabla f_* \circ M^*,$$

where f_* is the Kantorovich potential associated with the inner-product cost between $M^*_\# \pi_1^*$ and π_2^*. If the regularizer $g(M)$ enforces a low-rank structure ($\mathrm{rank}(M^*) = r < q$), then M^* is not surjective and Monge maps are no longer guaranteed to exist globally. In this case, one may still interpret the optimal plan as acting on the lower-dimensional image measure $\mu^* = M^*_\# \pi_1^*$, through a map $\tilde{T} : \mathrm{Im}(M^*) \to \mathcal{Y}$ optimal for the cost

$c_{\mathrm{ip}}(y', y) = -\langle y', y \rangle$, and write formally

$$\pi^* = (\mathrm{id}, \tilde{T} \circ M^*)_{\#} \pi_1^*.$$

The theoretical guarantees of Monge map require the full-rank assumption on M^*, while the low-rank and sparse parametrizations remain fully valid from the optimization and numerical perspectives. In Figure 7 we observe how low-rank affects the optimal transport plan across different levels of unbalancedness. Each subplot shows the learned transport map $M_\varepsilon \alpha$ (green), the ground-truth map $M_* \alpha$ (teal), and the target samples β (red). Orange lines represent barycentric displacements induced by the optimal plan.When $\lambda \to \infty$ (upper-left plot), the problem reduces to the balanced and the model transports the entire source mass. As λ decreases ($3.0 \to 1.5 \to 0.5$), the marginal penalty weakens, allowing partial mass creation or removal: the transport plan concentrates on geometrically consistent regions while ignoring unmatched components. The rightmost plot reports the total transported mass, which decreases monotonically with λ, confirming the progressive relaxation of the mass constraint. Although from Figure 7, we can see that the learned map aligns closely with the ground-truth low-rank map we need to investigate further statistical guarantees of the learned transport map in case M^* is not full rank.

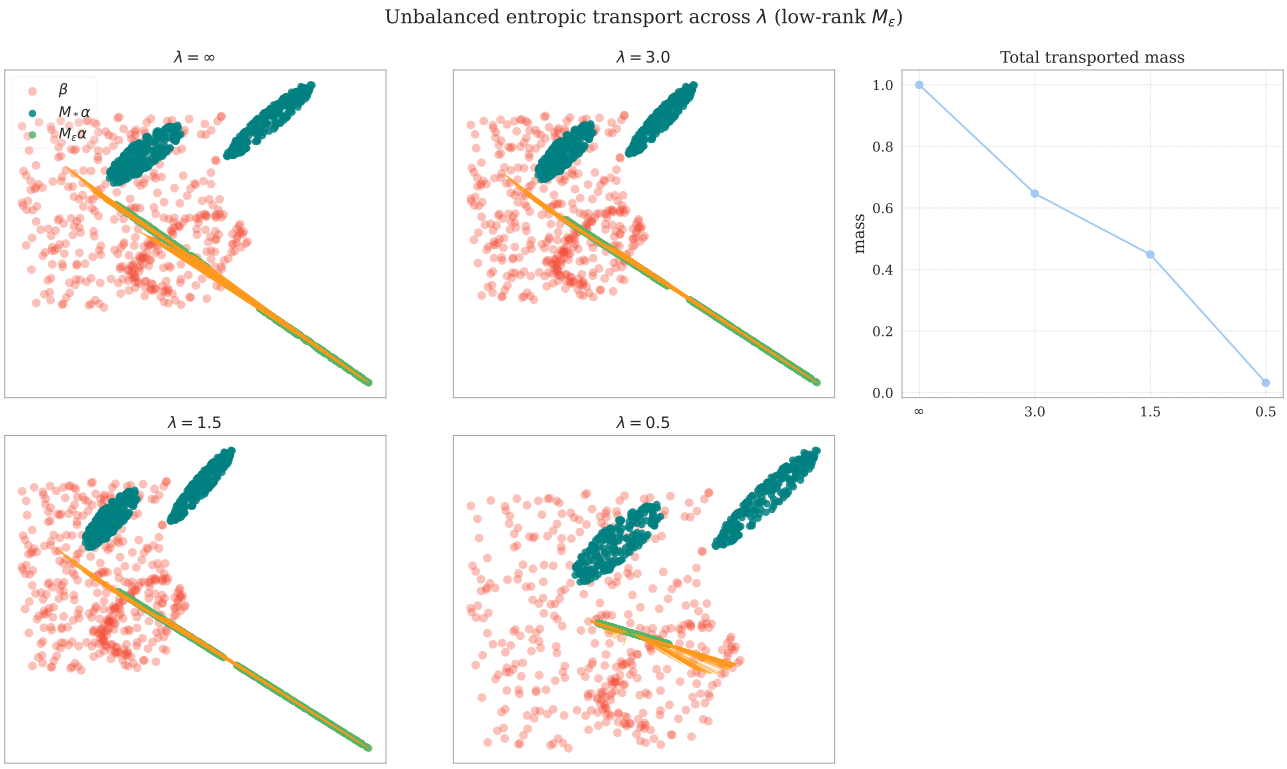


Figure 7: **Low-rank unbalanced optimal transport across unbalancedness levels λ.** The learned map $M_\varepsilon \alpha$ (green) approaches the ground-truth $M_\star \alpha$ (teal) while ignoring unmatched mass in the target β (red)as λ decreases. The total transported mass (right figure) decreases monotonically, reflecting the transition from balanced to strongly unbalanced transport.