Convergence rates for regularized unbalanced optimal transport: the discrete case

Luca Nenna*, Paul Pegon† and Louis Tocquec‡

2025

Unbalanced optimal transport (UOT) is a natural extension of optimal transport (OT) allowing comparison between measures of different masses. It arises naturally in machine learning by offering a robustness against outliers. The aim of this work is to provide convergence rates of the regularized transport cost and plans towards their original solution when both measures are weighted sums of Dirac masses.

Keywords. optimal transport, unbalanced optimal transport, entropic regularization, convex analysis, Csiszàr divergence, asymptotic analysis.
2020 Mathematics Subject Classification. Primary: 49Q22 ; Secondary: 49N15, 94A17, 49K40.

Contents

*Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: `luca.nenna@universite-paris-saclay.fr`
†CEREMADE, Université Paris–Dauphine, Université PSL, CNRS & MOKAPLAN, Inria Paris, 75016 Paris, France. Email: `pegon@ceremade.dauphine.fr`
‡Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay & MOKAPLAN, Inria Paris, 75016 Paris, France. Email: `louis.tocquec@universite-paris-saclay.fr`

Notations

In this section we define some notations that we will be used throughout the article.

- $\mathscr{M}^+(\mathcal{X})$ is the set of finite positive measures defined on $\mathcal{X}$.
- $\langle\cdot\,|\,\cdot\rangle$ is the inner product associated with the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^N$.
- $\preceq$ stands for a component-wise inequality.
- $\lesssim$ stands for $\leq$ up to a multiplicative universal positive constant.
- Δ_N is the $(N-1)$-dimensional simplex of $\mathbb{R}^N$ defined by

$$\Delta_N \overset{\text{def}}{=} \left\{ z \in (\mathbb{R}_+)^N \,\middle|\, \sum_{i=1}^N z_i = 1 \right\}.$$

- $\sqcup$ stands for the disjoint union.

1 Introduction

Given two probability measures μ and ν on compact domains $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}^d$ and a cost function $c\colon \mathcal{X}\times\mathcal{Y} \to \mathbb{R}_+$, the Monge-Kantorovich optimal transport problem consists in solving

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\gamma(x,y) \tag{1.1}$$

where $\Gamma(\mu,\nu)$ denotes the set of all probability measures on the product space $\mathcal{X}\times\mathcal{Y}$ having μ and ν as marginals. A popular way to solve numerically this problem consists in adding an entropy regularizing term, tuned via a nonnegative parameter t, which leads to the following *entropic optimal transport* problem

$$\inf_{\gamma\in\Gamma(\mu,\nu)} t\left(\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\gamma(x,y)\right) + \mathrm{Ent}(\gamma|\mu\otimes\nu), \tag{1.2}$$

where Ent is the relative entropy. In the literature, one often finds a parameter $\varepsilon \geq 0$ placed in front of the entropy rather than in front of the cost function. In this case, the focus is on the limit as ε vanishes. These two formulations are equivalent as long as the parameters remain positive, corresponding in our case to $t^{-1} = \varepsilon$ and letting $t \to +\infty$.

Over the last decade this kind of approach has witnessed an impressive increasing interest since it has proved to be an efficient way to approximate OT problems. From the computational point of view, (1.2) can be solved by an iterative projections method, which turns out to correspond to the celebrated Sinkhorn's algorithm [Sin67; Sin64], successfully developed in the pioneering works [Cut13; GS09; Ben+15] The simplicity of the implementation and the good convergence guarantees [FL89; MG20] determined the success of the entropic optimal transport for a wide range of applications, see for instance [PC+19] and the references therein. Notice now if we look at (1.2) as a perturbation of (1.1), then it is quite natural to study the behavior of both optimal values and minimizers, when the regularization parameter varies, and in particular when $t \to +\infty$. According to this let us mention some related works [Mik90; Léo12; Léo14; Car+17], which established Γ-convergence results of (1.2) to (1.1), [Pal19; CT21; ANS22; EN22; CPT23; MS23; NP24], which provided asymptotic expansion for the cost. Here we want to focus on the convergence rate of the primal and dual variable in the spirit of [CM94; Wee18] in the discrete setting for a generalized OT problem: unbalanced OT.

By definition, classical optimal transport requires the two given measures to have the same total mass, which can be a limitation for certain practical applications, particularly in machine learning, or imaging. An extension of the classical framework, known as unbalanced optimal transport, addresses this limitation by relaxing the marginal constraints and incorporating a divergence term into the objective function to penalise the marginals. It was simultaneously introduced by [Chi+18b; LMS18; KMV16] and may be expressed in the following form,

$$\boxed{\inf_{\gamma \in \mathscr{M}^+(\mathcal{X}\times\mathcal{Y})} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\gamma(x,y) + \mathrm{D}_\phi(\gamma_1|\mu) + \mathrm{D}_\phi(\gamma_2|\nu)}, \tag{1.3}$$

where γ_i is the i-th marginal of the coupling γ and D_ϕ is a ϕ-divergence, also known as Csiszàr divergence (see Definition 2.3).

As in the usual optimal transport problem, one can introduce an entropy regularized version of the unbalanced optimal transport problem, as follows

$$\boxed{\inf_{\gamma \in \mathscr{M}^+(\mathcal{X}\times\mathcal{Y})} t\left(\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\gamma(x,y) + \mathrm{D}_\phi(\gamma_1|\mu) + \mathrm{D}_\phi(\gamma_2|\nu)\right) + \mathrm{Ent}(\gamma|\mu\otimes\nu)}, \tag{1.4}$$

This paper focuses on this problem when both measures have finite support. The aim is to establish convergence rates of the regularized optimal variables, for the primal and dual problems, towards the original ones. The main contributions can be summarized in the following two theorems.

Theorem A. *(Simplified statement of Theorem 4.6)*
For positive t, let $t \mapsto \gamma(t)$ be the curve of optimal solutions to the regularized unbalanced optimal transport problem (1.4). Under the assumption that μ and ν are atomic

measures, and given certain conditions, the trajectory $t \mapsto \gamma(t)$ converges towards an optimizer γ^* of (1.3) at a rate of at least $1/\sqrt{t}$, i.e.

$$\|\gamma^* - \gamma(t)\| \lesssim \frac{1}{\sqrt{t}}.$$

Theorem B. *(Simplified statement of Theorem 4.5)*
For positive t, let $t \mapsto \xi(t)$ be the curve of optimal solutions to the regularized dual problem to (1.4). Under the assumption that μ and ν are atomic measures, and given certain conditions, the dual trajectory $t \mapsto \xi(t)$ converges towards the dual optimizer ξ^* at a rate of at least $1/t$, i.e.

$$\|\xi^* - \xi(t)\| \lesssim \frac{1}{t}.$$

Outline of the paper. The paper is organized as follows: in section 2.2 we introduce the discrete balanced and unbalanced optimal transport problems, their dual problems, and their regularized counterparts. In particular we focus on the dual unbalanced problem and give some preliminary results. Section 3 is devoted to the convergence of both the primal and the dual trajectory of solutions to the regularized unbalanced OT. In section 4 we provide an asymptotic analysis of the regularized problem when the regularization parameter tends to $+\infty$. Section 5 is devoted to numerical experiments and comparisons with the established asymptotics.

2 Discrete unbalanced optimal transport

2.1 The setting

Since we study the case of measures μ, ν with finite support, we assume from now on that $\mathcal{X}, \mathcal{Y}$ are finite sets. Notice that measures $\mu \in \mathscr{M}^+(\mathcal{X})$, $\nu \in \mathscr{M}^+(\mathcal{Y})$ and $\gamma \in \mathscr{M}^+(\mathcal{X} \times \mathcal{Y})$ can now be written as

$$\mu = \sum_{x \in \mathcal{X}} \mu_x \delta_x, \quad \nu = \sum_{y \in \mathcal{Y}} \nu_y \delta_y, \quad \text{and} \quad \gamma = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \gamma_{x,y} \delta_{(x,y)},$$

and can thus be identified with their respective weight vectors $(\mu_x)_x \in \mathbb{R}^{\mathcal{X}}$, $(\nu_y)_y \in \mathbb{R}^{\mathcal{Y}}$ and $(\gamma_{x,y})_{(x,y)} \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$. In the same way, the cost c can be identified with the vector $(c(x,y))_{(x,y)} \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$. By introducing the linear map $A \colon \mathbb{R}^{\mathcal{X} \times \mathcal{Y}} \to \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$ defined by

$$A\gamma = \left(\left(\sum_{y \in \mathcal{Y}} \gamma_{x,y} \right)_{x \in \mathcal{X}}, \left(\sum_{x \in \mathcal{X}} \gamma_{x,y} \right)_{y \in \mathcal{Y}} \right),$$

one can re-formulated the constraint $\gamma \in \Gamma(\mu, \nu)$ as $A\gamma = (\mu, \nu) \overset{\text{def}}{=} q$ and the optimal transport problem and its regularized formulation can then be expressed as finite-dimensional problems as follows,

$$\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}} \left\{ \langle c | \gamma \rangle \mid A\gamma = q \right\}, \tag{OT}$$

and for any non-negative t,

$$\inf_{\gamma\in(\mathbb{R}_+)^{\mathcal{X}}} \{t\langle c|\gamma\rangle + \mathrm{Ent}(\gamma) \mid A\gamma = q\}\,, \tag{EOT$_t$}$$

where $\langle c|\gamma\rangle \overset{\text{def}}{=} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} c(x,y)\gamma_{x,y}$ denotes the canonical inner product on $\mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$ and $\mathrm{Ent} : \mathbb{R}^{\mathcal{X}\times\mathcal{Y}} \to \mathbb{R}$ is the discrete entropy, i.e. the entropy relative with respect to the counting measure $\mathcal{H}^0$ defined by

$$\mathrm{Ent}(\gamma) \overset{\text{def}}{=} \mathrm{Ent}(\gamma|\mathcal{H}^0) = \begin{cases} \displaystyle\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \gamma_{x,y}(\log\gamma_{x,y} - 1) & \text{if } \gamma_{x,y} \geq 0 \text{ for every } (x,y) \\ +\infty & \text{otherwise,} \end{cases}$$

with a slight abuse of notation.

Remark 2.1. We take the convention that $0\log 0 = 0$ providing a continuous extension of $t \mapsto t\log t$ on $\mathbb{R}_+$.

Notice that, in the case where $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, the entropy relative with respect to the product measure $\mu\otimes\nu$ and the discrete entropy satisfy the following relation:

$$\inf\,(\text{EOT}_t) = \inf\,(1.2) - \mathrm{Ent}(\mu) - \mathrm{Ent}(\nu) - 2.$$

One can also express problems (OT) and (EOT$_t$) respectively as follows,

$$\inf_{\gamma\in(\mathbb{R}_+)^{\mathcal{X}\times\mathcal{Y}}} \{\langle c|\gamma\rangle \mid A\gamma = q\} \quad \text{and} \quad \inf_{\gamma\in\mathbb{R}^{\mathcal{X}\times\mathcal{Y}}} \{t\langle c|\gamma\rangle + \mathrm{Ent}(\gamma) \mid A\gamma = q\}\,.$$

Remark 2.2. Assuming that $\mathcal{X}$ and $\mathcal{Y}$ are finite subsets of $\mathbb{R}^d$, say $\mathcal{X} = \{x_1, \cdots, x_N\}$ and $\mathcal{Y} = \{y_1, \cdots, y_M\}$, the linear map A can be associated with its matrix A_{mat} in the bases $\mathcal{B} = (e_{x_1,y_1}, e_{x_1,y_2}, \cdots, e_{x_n,y_M})$ and $\mathcal{B}' = (e_{x_1}, \cdots, e_{x_n}, e_{y_1}, \cdots e_{y_M})$ defined as

$$A_{mat} = \begin{pmatrix} \mathbb{1}_M & 0_M & \cdots & 0_M & \mathrm{I}_M \\ 0_M & \mathbb{1}_M & \cdots & 0_M & \mathrm{I}_M \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_M & 0_M & \cdots & \mathbb{1}_M & \mathrm{I}_M \end{pmatrix}^T.$$

Proposition 2.1. *For any $t > 0$, (EOT$_t$) and its dual admit a unique solution γ_t and ξ_t (up to the kernel of A^*, which has dimension 1, for ξ_t), respectively. Moreover, the following relation holds*

$$\gamma_t = e^{t(A^*\xi_t - c)}.$$

We will not detail the proof since it is well known in the literature (for instance it is a by-product of the result in [PC+19]).

The unbalanced optimal transport problem is a natural extension of classical optimal transport problem where one can consider measures of different masses by relaxing the mass conservation constraint, for more details we refer the reader to the seminal works [Chi+18b; Chi+18a; LMS18; KMV16]. This approach, in particular, allows the creation or destruction of mass. To formalize this problem, we introduce a function $\mathrm{F}\colon \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \to \mathbb{R} \cup \{+\infty\}$ which is used as a penalization on the marginals of the transport coupling. The unbalanced optimal transport problem then reads as:

$$\boxed{\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}} \{\langle c|\gamma\rangle + \mathrm{F}(A\gamma)\}}. \tag{UOT}$$

In the same way as before, we consider a regularized version of this optimization problem which is given, for any non-negative t, by

$$\inf_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} \{t\langle c|\gamma\rangle + t\,\mathrm{F}(A\gamma) + \mathrm{Ent}(\gamma)\}.$$

For practical reasons, it will often be more convenient to work with its re-scaled version, defined for any positive t as

$$\boxed{\inf_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} \left\{\langle c|\gamma\rangle + \mathrm{F}(A\gamma) + \frac{1}{t}\,\mathrm{Ent}(\gamma)\right\}}. \tag{UOT$_t$}$$

For any positive t, we denote by $\mathscr{C}_t$ the functional of the regularized unbalanced optimal transport problem, i.e.,

$$\mathscr{C}_t \stackrel{\text{def}}{=} \langle c|\gamma\rangle + \mathrm{F}(A\gamma) + \frac{1}{t}\,\mathrm{Ent}(\gamma).$$

Throughout all this paper we will use the following assumptions for F:

(H1) F is a proper convex and lower semi-continuous function;

(H2) There exists $(\gamma_1, \gamma_2) \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$ such that $\mathrm{F}((\gamma_1, \gamma_2)) < +\infty$.

In the literature, for instance [SVP22], F is often taken to be a Csiszàr ϕ-divergence defined as follows.

Definition 2.3. (Discrete Csiszàr ϕ-divergence)
Let p, $q \in (\mathbb{R}_+)^N$ be two non-negative vectors and $\phi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ an entropy function meaning that it is a convex and lower semi-continuous function such that $\phi(1) = 0$. The discrete Csiszàr ϕ-divergence is defined as

$$\mathrm{D}_\phi(p \mid q) \stackrel{\text{def}}{=} \sum_{i\,:\,q_i>0} q_i \phi\left(\frac{p_i}{q_i}\right) + \phi'_\infty \sum_{i\,:\,q_i=0} p_i,$$

where $\phi'_\infty \stackrel{\text{def}}{=} \lim_{t\to\infty} \phi(t)/t$ is the recession constant.

Remark 2.4. The choice of considering a general function F instead of the Csiszàr divergence is motivated by the fact that one can also consider variational problems given by the sum of an optimal transport term and a congestion on the marginal. For this reason, in the following, we are going to make assumptions only on the function F that one can easily translate as assumptions on the entropy function ϕ when $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$.

Remark 2.5. In the case where $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ with fixed q, assuming that ϕ is a proper, convex, and lower semi-continuous function is sufficient to ensure (H1). The existence of some $\alpha > 0$ such that $\phi(\alpha) < +\infty$ guarantees (H2). Indeed, for any $x \in \mathcal{X}$, define

$$\gamma_1^x = \begin{cases} \mu_x \alpha & \text{if } \mu_x > 0, \\ \alpha & \text{if } \mu_x = 0, \end{cases}$$

and build γ_2 similarly. This ensures that $(\gamma_1, \gamma_2) \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$ and that $\mathrm{F}((\gamma_1, \gamma_2)) < +\infty$. Moreover, if ϕ is also superlinear (see definition 3.1), which implies that $\phi'_\infty = +\infty$, then it is necessary that $q \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$.

2.2 Properties of the regularized unbalanced optimal transport

Before studying in more details the properties of the solutions let us give some results concerning the primal and dual unbalanced problems. We introduce $A^* : \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ as the adjoint operator of A defined for any $\xi = (\varphi, \psi) \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$ as

$$A^* \xi = \varphi \oplus \psi = (\varphi_x + \psi_y)_{(x,y) \in \mathcal{X} \times \mathcal{Y}}.$$

We start our analysis by introducing the dual formulations of (UOT) and (UOT$_t$).

Proposition 2.2. *Under assumptions (H1) and (H2), the dual of (UOT) is given by*

$$\min_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \left\{ \mathrm{F}^*(-\xi) \mid A^* \xi \preceq c \right\}. \tag{Dual}$$

Moreover, if F *is assumed to be coercive then strong duality holds, that is*

$$\min_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \left\{ \mathrm{F}^*(-\xi) \mid A^* \xi \preceq c \right\} = - \min_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} \langle c | \gamma \rangle + F(A\gamma).$$

Proof. We begin by introducing the function $G : \mathbb{R}^{\mathcal{X} \times \mathcal{Y}} \to \mathbb{R}$ defined by

$$G(\gamma) = \langle c, \gamma \rangle + \iota_{(\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}}(\gamma),$$

where $\iota_{(\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}}$ denotes the indicator function of the non-negative orthant in $\mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$. With this definition, the unbalanced optimal transport (UOT) problem can be written as

$$\inf_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} G(\gamma) + \mathrm{F}(A\gamma).$$

By the Fenchel–Rockafellar duality theorem, strong duality holds, and there exists a solution to the dual problem, which is given by

$$\sup_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} -G^*(A^* \xi) - \mathrm{F}^*(-\xi) = \max_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} -G^*(A^* \xi) - \mathrm{F}^*(-\xi),$$

where G^* and F^* are the Legendre–Fenchel transforms of G and F, respectively. Now, assume that F is coercive. Consider a sequence $(\gamma^n)_{n \in \mathbb{N}}$ in $\mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ such that $\|\gamma^n\| \to +\infty$ as $n \to +\infty$. Since $c \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}$, we have

$$\langle c, \gamma^n \rangle + \mathrm{F}(A\gamma^n) \geq \mathrm{F}(A\gamma^n).$$

Moreover, because A is injective (i.e., $\ker A = \{0\}$), it follows that $\|A\gamma^n\| \to +\infty$, and hence the coercivity of F implies

$$\mathrm{F}(A\gamma^n) \to +\infty.$$

Therefore, the objective functional is coercive, and it follows that the primal problem admits a solution. $\square$

Definition 2.6. Take φ and ψ solution to (Dual) then we denote by I_0 the set of saturated constraint,

$$I_0 \overset{\text{def}}{=} \{(x,y)\in\mathcal{X}\times\mathcal{Y} \mid \varphi^*_x + \psi^*_y = c(x,y)\}.$$

The slack variable κ is the gap between $A^*\xi^*$ and c. For any $(x,y)\in\mathcal{X}\times\mathcal{Y}$ we define it as,

$$\kappa_{x,y} \overset{\text{def}}{=} c(x,y) - \varphi^*_x - \psi^*_y.$$

Proposition 2.3. *For any positive t, and under assumptions (H1) and (H2), the problem (UOT$_t$) admits a unique solution. Moreover, strong duality holds, and the corresponding dual problem is given by*

$$\min_{\xi\in\mathbb{R}^{\mathcal{X}}\oplus\mathbb{R}^{\mathcal{Y}}} \mathrm{F}^*(-\xi) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t}\exp\left(t\langle A^*\xi - c|e_{x,y}\rangle\right), \tag{Dual$_t$}$$

where $(e_{x,y})_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ denotes the canonical basis of $\mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$. The dual problem admits a solution, and if F^* is strictly convex, then it is unique.

Proof. Let $t > 0$. Since F is convex, there exist $a, b \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$ such that for any $\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$, we have

$$\mathrm{F}(\xi) \geq \langle a, \xi\rangle + b.$$

This implies that

$$\mathscr{C}_t(\gamma) = \langle c, \gamma\rangle + \mathrm{F}(A\gamma) + \frac{1}{t}\,\mathrm{Ent}(\gamma) \geq \langle c + A^*a, \gamma\rangle + \frac{1}{t}\,\mathrm{Ent}(\gamma) + b.$$

Using the superlinearity of the entropy functional Ent, we deduce that $\mathscr{C}_t$ is coercive. Therefore, the rescaled regularized unbalanced optimal transport problem (UOT$_t$) admits at least one solution. Moreover, since Ent is strictly convex, it follows that $\mathscr{C}_t$ is strictly convex as well, implying that the solution is unique. Now, let γ_t be the unique minimizer of (UOT$_t$). Then, by first-order optimality conditions,

$$0 \in \partial\mathscr{C}_t(\gamma_t).$$

Define the functional $G_t : \mathbb{R}^{\mathcal{X}\times\mathcal{Y}} \to \mathbb{R}$ by

$$G_t(\gamma) \overset{\text{def}}{=} \langle c, \gamma\rangle + \frac{1}{t}\,\mathrm{Ent}(\gamma).$$

Then the optimality condition becomes

$$0 \in A^* \partial \mathrm{F}(A\gamma_t) + \partial G_t(\gamma_t),$$

where we used assumption (H2) (see [Roc70][Theorem 23.8]). Therefore, there exists $u \in \partial \mathrm{F}(A\gamma_t)$ and $v \in \partial G_t(\gamma_t)$ such that

$$0 = A^* u + v.$$

By the Fenchel–Young equality, we have

$$\mathrm{F}(A\gamma_t) + \mathrm{F}^*(u) = \langle u, A\gamma_t\rangle, \quad \text{and} \quad G_t(\gamma_t) + G_t^*(v) = \langle \gamma_t, v\rangle = \langle \gamma_t, -A^*u\rangle = -\langle u, A\gamma_t\rangle.$$

Summing the two equalities yields:

$$\mathrm{F}(A\gamma_t) + G_t(\gamma_t) + \mathrm{F}^*(u) + G_t^*(-A^*u) = 0,$$

from which we get

$$\mathrm{F}(A\gamma_t) + G_t(\gamma_t) = -\,\mathrm{F}^*(u) - G_t^*(-A^*u),$$

and this concludes the proof. □

Remark 2.7. Although the Fenchel–Rockafellar theorem could also be used to prove proposition 2.3, our proof provides a more direct approach without requiring additional assumptions.

Remark 2.8. Uniqueness of the solution for (Dual$_t$) holds modulo $\ker A^*$, even in the absence of strict convexity of F^*.

We denote by $\mathscr{K}_t$ the functional of the dual problem, that is

$$\mathscr{K}_t(\xi) = \mathrm{F}^*(-\xi) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t} e^{t(\varphi_x + \psi_y - c(x,y))}.$$

Remark 2.9. In the case where $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$, the assumption $q \succ 0$ is necessary to have the existence of a solution for (Dual$_t$). Indeed, let $\xi(t)$ a solution of (Dual$_t$) with $t > 0$ and assume without loss of generality, that there exists $x_0 \in \mathcal{X}$ such that $q_{x_0} = \mu_{x_0} = 0$. We define a sequence $(\xi^n(t))_{n\in\mathbb{N}}$ such that for any $n \in \mathbb{N}$,

$$\xi_z^n(t) = \begin{cases} -n & \text{if } z = x_0 \\ \xi_z(t) & \text{otherwise.} \end{cases}$$

Then, one has

$$\begin{aligned}
\mathscr{K}_t(\xi^n(t)) &= \mathrm{F}^*(-\xi^n(t)) + \frac{1}{t} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} e^{t\langle A^*\xi^n(t) - c | e_{x,y}\rangle} \\
&= \mathscr{K}_t(\xi(t)) + \frac{1}{t}\sum_{y\in\mathcal{Y}} e^{t(\psi_y(t) - c(x_0,y)) - nt} - e^{t(\varphi_{x_0}(t) + \psi_y(t) - c(x_0,y))} \\
&\xrightarrow[n\to+\infty]{} \mathscr{K}_t(\xi(t)) - \frac{1}{t}\sum_{y\in\mathcal{Y}} e^{t(\varphi_{x_0}(t) + \psi_y(t) - c(x_0,y))} < \mathscr{K}_t(\xi(t)),
\end{aligned}$$

which is a contradiction. It implies that it cannot exist a solution to (Dual$_t$) as soon as q has a null element.

Proposition 2.4. *For any positive t and assuming (H1) and (H2), the optimal regularized transport coupling for (UOT_t) denoted $\gamma(t)$ and any solution of (Dual_t) denoted ξ_t are linked by the following formula*

$$\gamma(t) = e^{t(A^*\xi_t - c)}.$$

Proof. For any positive t, we observe that

$$\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X}\times\mathcal{Y}}} \mathscr{C}_t(\gamma) = \inf_{(p_1,p_2)\in\mathbb{R}^{\mathcal{X}}\oplus\mathbb{R}^{\mathcal{Y}}} \left(\inf_{\gamma\in\Gamma(p_1,p_2)} \langle c|\gamma\rangle + \frac{1}{t}\,\mathrm{Ent}(\gamma) \right) + \mathrm{F}((p_1,p_2)),$$

Fixing $p \stackrel{\text{def}}{=} (p_1, p_2)$ in the above formulation leads to an entropy-regularized optimal transport problem where the marginals are given by p, and the solution is denoted by γ_t^p. Using proposition 2.1, there also exists a solution ξ_t^p for the dual problem and it satisfies

$$\gamma_t^p = \exp\left(t(A^*\xi_t^p - c)\right).$$

As stated in proposition 2.3, there exists an unique $\gamma(t)$ solution of (UOT_t). Thus, taking $p^* = A\gamma(t)$ in the above formula allows to conclude. □

3 Convergence and regularity of the trajectories

This section is devoted to the study of the trajectory in t of the solutions for the primal and the dual problem. Before making more assumptions on the function F, let us recall some definitions for real valued functions defined on a normed vector space E.

Definition 3.1. A function $f\colon E \to \mathbb{R}\cup\{+\infty\}$ is *super-linear at infinity* if

$$\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Definition 3.2. A proper convex and lower semi-continuous function $f\colon E \to \mathbb{R}\cup\{+\infty\}$ is *strongly convex* at $x_0 \in \operatorname{dom} f$, if there exists $r > 0$, $x_0^* \in \partial f(x_0)$ and $\lambda(x_0, r) > 0$ such that

$$\forall x \in B(x_0, r)\cap \operatorname{dom} f, \quad f(x) \geq f(x_0) + \langle x_0^*|x - x_0\rangle + \frac{\lambda(x_0,r)}{2}\|x - x_0\|^2.$$

(H3) F is super-linear at infinity;

(H4) $W \cap (\mathbb{R}_+)^{\mathcal{X}\times\mathcal{Y}} \subset \operatorname{dom}\mathrm{F}$, with W a neighborhood of the origin;

(H5) F^* is strongly convex at every point of its domain;

(H6) F^* is $\mathcal{C}^2$ on $\operatorname{int}(\operatorname{dom}\mathrm{F}^*)$.

Remark 3.3. Assumption (H4) implies assumption (H2).

The following proposition shows in particular that under (H3), F^* has full domain. Its proof can be found for example in [San23], but we provide it for self-containedness.

Proposition 3.1. *Under assumptions (H1) and (H3), the Legendre transform* F^* *is bounded on every ball.*

Proof. Let $R > 0$. For any $\xi \in B(0, R)$ we have

$$\mathrm{F}^*(\xi) \overset{\text{def}}{=} \sup_{v \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \langle v | \xi \rangle - \mathrm{F}(v) \leq \max \left\{ \sup_{\substack{v \in \operatorname{dom} \mathrm{F} \\ v \neq 0}} \|v\| \left(\|\xi\| - \frac{\mathrm{F}(v)}{\|v\|} \right), -F(0) \right\}.$$

By super-linearity at infinity of F, there exists $M > 0$ such that $\mathrm{F}(v) \leq R\|v\|$ whenever $\|v\| \geq M$, in which case $\|v\|\|\xi\| - \mathrm{F}(v) \leq 0$, while for $v \leq M$ we have

$$\langle v \,|\, \xi \rangle - F(v) \leq M\|\xi\| - \inf_{\overline{B}(0,M)} F,$$

the infimum being a finite quantity since a lower semicontinuous convex function always has an affine minorant. We have shown that F^* is bounded from above on every ball, and since it is convex lsc, it is also globally bounded from below by a continuous affine map, and the result follows. □

Remark 3.4. In the case where $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$, with q fixed, ϕ being superlinear guarantees (H3) and $\phi'_\infty = +\infty$. Moreover, assume that there exists W a neighborhood of 0 such that $W \cap \mathbb{R}_+ \subset \operatorname{dom} \phi$ is sufficient to ensure (H4).

Proposition 3.2. *Assume* F *is of the form* $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$*, with* $q \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$ *fixed and* ϕ *being super-linear. Suppose* ϕ^* *is locally convex and of class* $\mathcal{C}^2$ *on* $\operatorname{int}(\operatorname{dom} \phi)$ *then* F *satisfies (H5) and (H6).*

Proof. ϕ^* being of class $\mathcal{C}^2$ on $\operatorname{int} \operatorname{dom} \phi$ is clearly sufficient to ensure (H6). Let now $\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$. Since ϕ^* is locally strongly convex, for any $u \in \operatorname{dom} \phi^*$, there exists $r > 0$ and $\lambda(r, u) > 0$ such that

$$\forall v \in B(u, r) \cap \operatorname{dom} \phi^*, \quad \phi^*(v) \geq \phi^*(u) + \eta(u - v) + \frac{\lambda(r, u)}{2}(u - v)^2,$$

with $\eta \in \partial\phi(u)$. Thus, given $\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$, one has

$$\begin{aligned}
\forall \tilde{\xi} \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}, \quad \mathrm{F}^*(\tilde{\xi}) &= \sum_{z \in \mathcal{X} \sqcup \mathcal{Y}} q_z \phi^*(\tilde{\xi}_z) \\
&\geq \sum_{z \in \mathcal{X} \sqcup \mathcal{Y}} q_z \left(\phi^*(\xi_z) + \eta_z(\xi_z - \tilde{\xi}_z) + \frac{\lambda(r_z, \xi_z)}{2}(\xi_z - \tilde{\xi}_z)^2 \right) \\
&\geq \mathrm{F}^*(\xi) + \langle \eta | \xi - \tilde{\xi} \rangle + \frac{\lambda(r, \xi)}{2} \|\xi - \tilde{\xi}\|^2,
\end{aligned}$$

where $\eta \in \partial \mathrm{F}^*(\xi)$, $r \overset{\text{def}}{=} \min_z r_z$ and $\lambda(r, \xi) \overset{\text{def}}{=} \min_z \lambda(r, \xi_z)$. It means that F^* is also locally strongly-convex. □

Lemma 3.3 (Uniform coercivity of $\mathscr{K}_t$)**.** *Under assumptions (H1) and (H4), the functional* $\inf_{t>0} \mathscr{K}_t$ *is coercive.*

Proof. For any $\xi = (\varphi, \psi) \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$, one has

$$\begin{aligned}
\inf_{t>0} \mathscr{K}_t(\xi) &= \mathrm{F}^*(-\xi) + \inf_{t>0} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t} e^{t(\varphi_x+\psi_y-c(x,y))} \\
&\geq \mathrm{F}^*(-\xi) + \inf_{t>0} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t} e^{t(\varphi_x+\psi_y-c(x,y))_+} \\
&\geq \mathrm{F}^*(-\xi) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (\varphi_x+\psi_y-c(x,y))_+ \stackrel{\text{def}}{=} V(\xi).
\end{aligned}$$

Let us show that V is coercive. Let $(\xi^n)_{n\in\mathbb{N}} = ((\varphi^n, \psi^n))_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$ such that $\|\xi^n\| \to +\infty$ as $n \to +\infty$. On one hand, if there exists $z_0 \in \mathcal{X} \sqcup \mathcal{Y}$ such that, after taking a subsequence, $\xi^n_{z_0} \to -\infty$ then we have, using (H4), that there exists $r \in \mathbb{R}^*_+$ such that $re_{z_0} \in \operatorname{dom} \mathrm{F}$. It leads to

$$V(\xi^n) \geq \mathrm{F}^*(-\xi^n) \geq -\xi^n_{z_0} - \mathrm{F}(re_{z_0}) \to +\infty.$$

On the other hand, if that is not the case, then for any $z \in \mathcal{X} \sqcup \mathcal{Y}$, ξ^n_z is bounded from below by some constant $M < 0$. Thus there exists $z_0 \in \mathcal{X} \sqcup \mathcal{Y}$ such that $\xi^n_{z_0} \to +\infty$. Without loss of generality we assume that $z_0 \in \mathcal{X}$. Then we have, using (H4) again, that for any $y \in \mathcal{Y}$, there exists $r \in \mathbb{R}^*_+$ such that $re_y \in \operatorname{dom} \mathrm{F}$. It leads to

$$\begin{aligned}
V(\xi^n) &\geq \mathrm{F}^*(-\xi^n) + \varphi^n_{z_0} + \psi^n_y - c(z_0, y) \\
&\geq -r\psi^n_y - F(re_y) + \varphi^n_{z_0} + \psi^n_y - c(z_0, y) \\
&\geq |1-r|M - F(re_y) + \varphi^n_{z_0} - c(z_0, y) \underset{n\to+\infty}{\longrightarrow} +\infty.
\end{aligned}$$

We have shown that $V(\xi^n) \to +\infty$ along a subsequence, but since the same reasoning applies to any arbitrary subsequence taken in the beginning, we get the result for the whole sequence, which establishes the coercivity of V. □

3.1 Trajectory of the regularized dual of UOT problem

Proposition 3.4. *Assume that assumptions (H1) – (H6) hold. Then, there exists a unique solution* $\xi(t)$ *to* (Dual_t) *for every* $t > 0$ *and the trajectory* $t \mapsto \xi(t)$ *is a* $\mathcal{C}^1$ *function on* $\mathbb{R}^*_+$.

Proof. The existence and uniqueness of the solution to (Dual_t) comes from Proposition 2.3 and the strict convexity of the functional $\mathscr{K}_t$, resulting from (H5).

Let us first consider any point $t_0 \in \mathbb{R}^*_+$ and denote by $f : \mathbb{R}^*_+ \times (\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}) \to \mathbb{R}$ the $\mathcal{C}^1$ function defined by $f(t, \xi) = D\mathscr{K}_t(\xi)$. By optimality of $\xi(t_0)$, $f(t_0, \xi(t_0)) = 0$ and by (H5), we get $\det D_\xi f(t_0, \xi(t_0)) = \det D^2\mathscr{K}_{t_0}(\xi(t_0)) > 0$. Thus the implicit function theorem allows to deduce that the trajectory is $\mathcal{C}^1$ on $\mathbb{R}^*_+$. □

Proposition 3.5. *Assume that assumptions (H1) – (H6) hold. Then the trajectory $t \mapsto \xi(t)$ satisfies the following ordinary differential equation (ODE) on $\mathbb{R}_+^*$:*

$$D^2\mathscr{K}_t(\xi(t))\dot{\xi}(t) + \left(\frac{\partial}{\partial t}D\mathscr{K}_t\right)(\xi(t)) = 0$$

which can be explicitly written as

$$A\,\mathrm{diag}\,\gamma(t)A^*\dot{\xi}(t) + \frac{1}{t}D^2\mathrm{F}^*(-\xi(t))\dot{\xi}(t) + \frac{1}{t^2}A\,\mathrm{diag}\,\gamma(t)\log\gamma(t) = 0. \tag{ODE}$$

Proof. For any positive t, by using the optimality of $\xi(t)$ for the functional $\mathscr{K}_t$, one can obtain

$$D\mathscr{K}_t(\xi(t)) = 0.$$

The function $t \mapsto D\mathscr{K}_t(\xi(t))$ is $\mathcal{C}^1$ on $\mathbb{R}_+^*$. Differentiating the previous equality with respect to t and using the chain rule leads to

$$D^2\mathscr{K}_t(\xi(t))\dot{\xi}(t) + \frac{\partial}{\partial t}D\mathscr{K}_t(\xi(t)) = 0.$$

By computing $D^2\mathscr{K}_t$ and $D\mathscr{K}_t$ one then obtains (ODE). □

Proposition 3.6. *Under assumptions (H1) – (H5), the trajectory $t \mapsto \xi(t)$ d converges towards the solution ξ^* to (Dual), as $t \to +\infty$.*

Proof. Using the optimality of $\xi(t)$ for (Dual$_t$), we leverage the following inequality

$$\mathscr{K}_t(\xi_t) = \mathrm{F}^*(-\xi(t)) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t}e^{t\langle A^*\xi(t)-c|e_{x,y}\rangle} \leq \mathrm{F}^*(-\xi^*) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t}e^{t\langle A^*\xi^*-c|e_{x,y}\rangle}.$$

The constraint $A^*\xi^* \preceq c$ allows to deduce that

$$\mathscr{K}_t(\xi_t) = \mathrm{F}^*(-\xi(t)) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t}e^{t\langle A^*\xi(t)-c|e_{x,y}\rangle} \leq \mathrm{F}^*(-\xi^*) + \frac{|\mathcal{X}||\mathcal{Y}|}{t}. \tag{3.1}$$

From this, and the uniform coercivity of $\mathscr{K}_t$ as stated in Lemma 3.3, we deduce that the trajectory is bounded on $\mathbb{R}_+^* \setminus (0,1)$. By Bolzano-Weierstrass Theorem, we may consider a sequence $(t_n)_{n\in\mathbb{N}}$ such that $\xi(t_n) \to \tilde{\xi}$ as $n \to +\infty$, where $\tilde{\xi}$ is an accumulation point of the trajectory (at infinity). Since $(\xi(t_n))$ is bounded, by Proposition 3.1 we know that $|F^*(\xi(t_n))|$ is bounded by some constant $C > 0$. Thus by (3.1), for every n and every $(x,y) \in \mathcal{X}\times\mathcal{Y}$,

$$\frac{1}{t_n}e^{t_n\langle A^*\xi(t_n)-c|e_{x,y}\rangle} \leq C + \mathrm{F}^*(-\xi^*) + \frac{|\mathcal{X}||\mathcal{Y}|}{t_n}.$$

If $\alpha \overset{\text{def}}{=} \langle A^*\tilde{\xi} - c\,|\,e_{x,y}\rangle > 0$ then for n large enough we would have $\frac{1}{t_n}e^{t_n\langle A^*\xi(t_n)-c|e_{x,y}\rangle} \geq \frac{e^{\alpha t_n/2}}{t_n} \to +\infty$: a contradiction with the boundedness of the right-hand side of the previous equation. We conclude that $A^*\tilde{\xi} \preceq c$, meaning that any accumulation point is feasible for the non-regularized dual problem. Moreover by (3.1) we also have

$$\mathrm{F}^*(-\xi(t_n)) \leq \mathrm{F}^*(-\xi^*) + \frac{|\mathcal{X}||\mathcal{Y}|}{t_n},$$

and taking the limit as $n \to +\infty$, we obtain

$$\mathrm{F}^*(-\tilde{\xi}) \leq \mathrm{F}^*(-\xi^*).$$

This shows that any accumulation point of the trajectory is optimal for the non-regularized dual problem. But since by (H5), F^* is strictly convex, (Dual) admits a unique solution ξ^*, thus $\xi(t_n)$ converges to ξ^*. □

3.2 Trajectory of the regularized UOT problem

The purpose of the following proposition is to show that the curve $t \mapsto \gamma(t)$ converges to γ^*, the solution of (UOT) of minimal entropy (which is unique by strict convexity of Ent), under suitable assumptions.

Proposition 3.7. *Under assumptions (H1) – (H6), the trajectory $t \mapsto \gamma(t)$ converges towards γ^*, the solution of (UOT) with minimal entropy* Ent*, as $t \to +\infty$. Moreover, $t \mapsto \gamma(t)$ is a $\mathcal{C}^1$ function on $\mathbb{R}_+^*$.*

Proof. We start by showing that the trajectory is bounded. For any positive t, by optimality of $\gamma(t)$ for (UOT_t) we have that

$$\langle c|\gamma(t)\rangle + \mathrm{F}(A\gamma(t)) + \frac{1}{t}\,\mathrm{Ent}(\gamma(t)) \leq \langle c|\gamma^*\rangle + \mathrm{F}(A\gamma^*) + \frac{1}{t}\,\mathrm{Ent}(\gamma^*). \tag{3.2}$$

Since γ^* is optimal for (UOT), we have

$$\langle c|\gamma^*\rangle + \mathrm{F}(A\gamma^*) \leq \langle c|\gamma(t)\rangle + \mathrm{F}(A\gamma(t))$$

and combining it with (3.2) we get

$$\mathrm{Ent}(\gamma(t)) \leq \mathrm{Ent}(\gamma^*). \tag{3.3}$$

By coercivity of Ent, we can conclude that the trajectory $(\gamma(t))_{t>0}$ is bounded. It implies that there exists $(t_n)_{n\in\mathbb{N}}$ such that $\gamma(t_n) \to \tilde{\gamma}$ as $n \to +\infty$, where $\tilde{\gamma}$ is an accumulation point at infinity. Taking $t = t_n$ in (3.2) and passing to the limit, we obtain that $\tilde{\gamma}$ is a solution of (UOT). Finally, using (3.3), taking $t = t_n$ and passing to the limit, we obtain

$$\mathrm{Ent}(\tilde{\gamma}) \leq \mathrm{Ent}(\gamma^*),$$

which implies that $\tilde{\gamma} = \gamma^*$, and thus $\gamma(t) \to \gamma^*$ as $t \to +\infty$. Finally, the regularity of the trajectory is a direct consequence of Proposition 2.4 and Proposition 3.4. □

4 Asymptotics for the regularized unbalanced problem

This section is devoted to an asymptotic analysis of the regularized unbalanced problem as $t \to +\infty$. We divide our study into two parts: we start with the dual problem and then we tackle the primal problem. The proofs are strongly inspired from [CM94], which dealt with linear problems with exponential penalization (entropy penalization in the primal). We adapt their strategy to the non-linearity introduced by the function F.

4.1 Asymptotic analysis of the dual trajectory

We start by introducing the linear space $E_0 \overset{\text{def}}{=} \operatorname{span}\{Ae_{x,y} \mid (x,y) \in I_0\}$, where I_0 is the set of saturated constraint defined in Definition 2.6. It corresponds to the space of marginals of (signed) measures concentrated on the set of saturated constraints.

Lemma 4.1. *Under assumptions (H1) – (H6), the quantity $D\mathrm{F}^*(-\xi^*)$ belongs to E_0.*

Proof. Consider the curve $t \mapsto \xi(t)$ of solutions to (Dual$_t$). By (H1), (H3), (H6) and Proposition 3.1, F^* is differentiable on $\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$, thus the first order optimality condition reads as:

$$
\begin{aligned}
& D\mathscr{K}_t(\xi(t)) = 0 \\
\iff D\mathrm{F}^*(-\xi(t)) + & \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \gamma_{x,y}(t) Ae_{x,y} = 0.
\end{aligned}
$$

By Proposition 3.6 and Proposition 3.7, $\xi(t) \to \xi^*$ and $\gamma(t) \to \gamma^*$ as $t \to +\infty$, thus taking the limit in the above expression leads to

$$
D\mathrm{F}^*(-\xi^*) + \sum_{(x,y)\in I_0} \gamma^*_{x,y} Ae_{x,y} = 0,
$$

where we have used the fact that for any $(x,y) \notin I_0$, one has

$$
\gamma_{x,y}(t) = e^{t(A^*\xi(t)-c)} \underset{t\to+\infty}{\longrightarrow} 0.
$$

From this, we can conclude that $D\mathrm{F}^*(-\xi^*) \in E_0$. □

For any positive t, we consider the quantity $d(t) \overset{\text{def}}{=} t(\xi(t) - \xi^*)$ and we look at $\mathscr{K}_t$ as a function of d instead of ξ, namely we introduce the new functional $\tilde{\mathscr{K}_t} \colon \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}$ defined as

$$
\tilde{\mathscr{K}_t}(d) \overset{\text{def}}{=} \mathrm{F}^*\left(-\frac{1}{t}d - \xi^*\right) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t} e^{\langle A^*d - t\kappa | e_{x,y}\rangle},
$$

where κ is the gap between $A\xi^*$ and c as defined in Definition 2.6. Notice that it is defined in such a way that $\tilde{\mathscr{K}_t}(t(\xi - \xi^*)) = \mathscr{K}_t(\xi)$.

Proposition 4.2. *Under assumptions (H1) – (H6), for any $\varepsilon > 0$ the trajectory $t \mapsto d(t)$ is uniformly bounded on $\mathbb{R}_+ \backslash (0, \varepsilon)$.*

Proof. Let $\varepsilon > 0$ and $t > \varepsilon$. Notice now that for any $(x,y) \in I_0$, one has

$$\langle A^* d(t)|e_{x,y}\rangle = t\langle A^*\xi(t) - A^*\xi^*|e_{x,y}\rangle = t\langle A^*\xi(t) - c|e_{x,y}\rangle,$$

and thus

$$\langle A^* d(t)|e_{x,y}\rangle = \log \gamma_{x,y}(t) \xrightarrow{t\to+\infty} \log \gamma^*_{x,y}.$$

We deduce from this and from the regularity of $\xi(t)$, see Proposition 3.4, that $\langle A^* d(t)|e_{x,y}\rangle$ is a bounded quantity on $\mathbb{R}_+\backslash(0,\varepsilon)$ for any $(x,y) \in I_0$, which in turn implies that $d^{(0)}(t)$, the projection of $d(t)$ onto E_0, is bounded. Besides, since $\tilde{\mathscr{K}}_t(t(\xi - \xi^*)) = \mathscr{K}_t(\xi)$ and $\xi(t)$ is the unique minimizer of $\mathscr{K}_t$, $d(t)$ is the unique minimizer $\tilde{\mathscr{K}}_t$. Comparing it with $d^{(0)}(t)$ yields

$$\tilde{\mathscr{K}}_t(d(t)) \leq \tilde{\mathscr{K}}_t(d^{(0)}(t)),$$

which implies

$$\begin{aligned}\mathrm{F}^*\left(-\frac{1}{t}d(t) - \xi^*\right) + \sum_{(x,y)\in I_0} \frac{1}{t} e^{\langle A^* d(t) - t\kappa|e_{x,y}\rangle}& \\ \leq \mathrm{F}^*\left(-\frac{1}{t}d^{(0)}(t) - \xi^*\right) &+ \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t} e^{\langle A^* d^{(0)}(t) - t\kappa|e_{x,y}\rangle}.\end{aligned}$$

Using that $\langle A^* d(t)|e_{x,y}\rangle = \langle d(t)|Ae_{x,y}\rangle = \langle A^* d^{(0)}(t)|e_{x,y}\rangle$ for any $(x,y) \in I_0$, we obtain

$$\mathrm{F}^*\left(-\frac{1}{t}d(t) - \xi^*\right) \leq \mathrm{F}^*\left(-\frac{1}{t}d^{(0)}(t) - \xi^*\right) + \sum_{(x,y)\notin I_0} \frac{1}{t} e^{\langle A^* d^{(0)}(t) - t\kappa|e_{x,y}\rangle}. \tag{4.1}$$

Notice that $d(t)/t = \xi(t) - \xi^* \to 0$ and $-\xi^* \in \operatorname{dom} \mathrm{F}$. Thus, by (H5), for some $r > \varepsilon$ and all $t \geq r$, we have

$$\mathrm{F}^*\left(-\frac{1}{t}d(t) - \xi^*\right) \geq \mathrm{F}^*(-\xi^*) - \frac{1}{t}\langle D\mathrm{F}^*(-\xi^*)\,|d(t)\rangle + \frac{\lambda_0(-\xi^*, r)}{2t^2}\|d(t)\|^2, \tag{4.2}$$

where $\lambda_0 \overset{\text{def}}{=} \lambda_0(-\xi^*, r) > 0$, and by (H6) and Taylor's expansion, we have

$$\mathrm{F}^*\left(-\frac{1}{t}d^{(0)}(t) - \xi^*\right) \leq \mathrm{F}^*(-\xi^*) - \frac{1}{t}\langle D\,\mathrm{F}^*(-\xi^*)\,|d^{(0)}(t)\rangle + \frac{C}{t^2} \tag{4.3}$$

for some $C > 0$ and every $t \geq r$. By Lemma 4.1, we have

$$\langle D\mathrm{F}^*(-\xi^*)\,|d(t)\rangle = \langle D\mathrm{F}^*(-\xi^*)\,|d^{(0)}(t)\rangle,$$

and combining this with (4.1), (4.2) and (4.3), we obtain

$$\lambda_0\|d(t)\|^2 \leq \sum_{(x,y)\notin I_0} 2te^{\langle A^* d^{(0)}(t) - t\kappa|e_{x,y}\rangle} + C.$$

We have seen that $d^{(0)}(t)$ is bounded, and since $\kappa_{x,y} > 0$ for every $(x,y) \notin I_0$, the right-hand side is bounded, thus $d(t)$ is bounded as well for $t \geq r \geq \varepsilon$. We obtain the result since d is bounded on any compact interval included in $\mathbb{R}^*_+$, by regularity of $t \mapsto \xi(t)$. $\square$

Proposition 4.3. *Under assumptions (H1) – (H6), the trajectory $t \mapsto d(t)$ converges towards a finite quantity denoted d^*.*

Before going into the proof, we introduce the following useful lemma.

Lemma 4.4. *Let E' be a normed vector space, $f : E' \to \mathbb{R}$ a strictly convex function and $L \colon E \to E'$ a linear operator and F an affine subspace of direction $\vec{F}$. If the set of solutions $\mathcal{U}$ of the problem*

$$\inf_{x \in F} f(Lx)$$

is nonempty, then for any $x_0 \in \mathcal{U}$, one has

$$\mathcal{U} = x_0 + \left(\vec{F} \cap \ker L\right).$$

Proof. Let $x_0 \in \mathcal{U}$. The affine subspace F can be expressed as $F = x_0 + \vec{F}$. Then, any $x_1 \in \mathcal{U}$ can be expressed as $x_1 = x_0 + y$, with $y \in \vec{F}$. Assuming that $y \notin \ker L$, $Lx_1 \leq Lx_2$ and by strict convexity of f we have

$$f(Lx_0) = \frac{1}{2} f(Lx_0) + \frac{1}{2} f(Lx_1) > f(Lx_2),$$

where $x_2 \stackrel{\text{def}}{=} (x_0 + x_1)/2 = x_0 + \frac{1}{2}y$. Since $x_2 \in F$, this contradicts the optimality of x_0 and x_1. Therefore $y \in \ker L$.

Conversely, let $x_1 \in F$ defined as $x_1 = x_0 + y$, with $y \in \vec{F} \cap \ker L$. Then, since $f(Lx_1) = f(Lx_0)$, it follows that $x_1 \in \mathcal{U}$. $\square$

Proof of Proposition 4.3. Let d^* be an accumulation point of the curve $t \mapsto d(t)$ at $+\infty$, meaning that $d(t_n) \to d^*$ as $n \to +\infty$ for some sequence $t_n \to +\infty$. We know that there exists such an accumulation point thanks to Proposition 4.2. Since $\xi(t_n)$ is a minimizer of (Dual$_t$) for $t = t_n$, we obtain the following equation:

$$D\mathrm{F}^*\left(-\frac{1}{t_n} d(t_n) - \xi^*\right) + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} e^{\langle A^* d(t_n) - t_n \kappa | e_{x,y} \rangle} A e_{x,y} = 0.$$

Taking the limit in this expression leads to

$$D\mathrm{F}^*(-\xi^*) + \sum_{(x,y) \in I_0} e^{\langle A^* d^* | e_{x,y} \rangle} A e_{x,y} = 0.$$

We recognize that is the first order optimality condition of the following differentiable convex function

$$\mathrm{G} : z \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \mapsto \langle D\mathrm{F}^*(-\xi^*) | z \rangle + \sum_{(x,y) \in I_0} e^{\langle A^* z | e_{x,y} \rangle},$$

so that d^* is a minimizer of G. Notice that by Lemma 4.1, $\langle D\mathrm{F}^*(-\xi^*) | z \rangle$ is a linear function of the quantities $\langle A^* z \,|\, e_{x,y} \rangle$ for $(x, y) \in I_0$, and since the exponential is strictly convex, we may apply Lemma 4.4 to assert that the set of optimal solutions $\mathcal{U}$ of G is given by

$$\mathcal{U} = d^* + \bigcap_{(x,y)\in I_0} \ker Ae_{x,y}.$$

Now, let us introduce $d_1^* \in \mathcal{U}$ such that $d_1^* \neq d^*$. We define $d_1(t)$ by

$$d_1(t) \stackrel{\text{def}}{=} d(t) + d_1^* - d^*$$

which satisfies $d_1(t_n) \to d_1^*$ as $n \to +\infty$.

We know that $d(t)$ is a minimizer solution of the functional $\tilde{\mathscr{K}_t}$, thus

$$\tilde{\mathscr{K}}(d(t_n)) \leq \tilde{\mathscr{K}}(d_1(t_n)),$$

or equivalently

$$\mathrm{F}^*\left(-\frac{1}{t_n}d(t_n) - \xi^*\right) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t_n} e^{\langle A^* d(t_n)|e_{x,y}\rangle} \\ \leq \mathrm{F}^*\left(-\frac{1}{t_n}d_1(t_n) - \xi^*\right) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t_n} e^{\langle A^* d_1(t_n)|e_{x,y}\rangle}.$$

Since $d(t_n)$ and $d_1(t_n)$ are bounded, we can write the second order Taylor expansion of F^* at $-\xi^*$ to get

$$\frac{\|d(t_n)\|^2_{D^2\,\mathrm{F}^*(-\xi^*)}}{2} + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} t_n e^{\langle A^* d(t_n) - t_n\kappa|e_{x,y}\rangle} \\ \leq \frac{\|d_1(t_n)\|^2_{D^2\,\mathrm{F}^*(-\xi^*)}}{2} + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} t_n e^{\langle A^* d_1(t_n) - t_n\kappa|e_{x,y}\rangle} + o(1),$$

noticing that the first-order term cancelled thanks to Lemma 4.1. Besides, thanks to the same lemma and the fact that $d(t_n) - d_1(t_n) \in \bigcap_{(x,y)\in I_0} \ker Ae_{x,y}$, the terms corresponding to $(x,y) \in I_0$ in the sum are equal on both sides. By non-negativity of the exponential, we forget the terms corresponding to $(x,y) \notin I_0$ in the left-hand side, and we get:

$$\|d(t_n)\|^2_{D^2\mathrm{F}^*(-\xi^*)} \leq \|d_1(t_n)\|^2_{D^2\mathrm{F}^*(-\xi^*)} + \sum_{(x,y)\notin I_0} \frac{2}{t} e^{\langle Ad_1(t_n) - t\kappa|e_{x,y}\rangle} + o(1).$$

Passing to the limit, we obtain

$$\|d^*\|^2_{D^2\mathrm{F}^*(-\xi^*)} \leq \|d_1^*\|^2_{D^2\mathrm{F}^*(-\xi^*)}.$$

Thus, we have shown that d^* is the unique solution of the strictly convex (by (H5)) problem:

$$\inf_{z\in\mathcal{U}} \|z\|^2_{D^2\mathrm{F}^*(-\xi^*)}.$$

Since every accumulation point of the trajectory is also a solution, it follows that there is an unique accumulation point d^*. □

Now, one shall compute the derivative of $d(t)$ with respect to t. A simple computation gives us that $\dot{d}(t) = \frac{1}{t}d(t) + t\dot{\xi}(t)$. Notice that this quantity consists of two parts: $d(t)$, which converges by the previous proposition, and $\dot{\xi}(t)$.

Theorem 4.5. *Under assumptions (H1) – (H6), the convergence rate of the curve $t \mapsto \xi(t)$ is at least of order $O(t^{-1})$, i.e. for any t sufficiently large, one has*

$$\|\xi(t) - \xi^*\| \lesssim \frac{1}{t}.$$

Proof. Let $\varepsilon > 0$ and $t \geq \varepsilon$. We define the functional Λ_t by

$$\Lambda_t(\xi) = \|A^*\xi + A^*\xi^* - c\|^2_{\operatorname{diag}(\gamma(t))} + \frac{1}{t}\|\xi - \xi(t) + \xi^*\|^2_{D^2\mathrm{F}^*(-\xi(t))}.$$

Note that $\dot{d}(t)$ is optimal for Λ_t, since the latter is convex and

$$D\Lambda_t(\xi) = 2A\operatorname{diag}(\gamma(t))(A^*\xi + A^*\xi^* - c) + \frac{2}{t}D^2\mathrm{F}^*(-\xi(t))(\xi - \xi(t) + \xi^*)$$

and thus

$$\begin{aligned} D\Lambda_t(\dot{d}(t)) &= 2t\left(A\operatorname{diag}(\gamma(t))A^*\dot{\xi}(t) + \frac{1}{t}D^2\mathrm{F}^*(-\xi(t))\dot{\xi}(t) + \frac{1}{t^2}A\operatorname{diag}(\gamma(t))\log\gamma(t)\right) \\ &= 0. \end{aligned}$$

Using this property, we get the following inequality

$$\begin{aligned} & \Lambda_t(\dot{d}(t)) \leq \Lambda_t(0) \\ \implies\ & \frac{1}{t}\|t\dot{\xi}(t)\|^2_{D^2\mathrm{F}^*(-\xi(t))} \leq \sum_{(x,y)\notin I_0} \gamma_{x,y}(t)\langle A^*\xi^* - c|e_{x,y}\rangle^2 + \frac{1}{t}\|\xi(t) - \xi^*\|^2_{D^2\mathrm{F}^*(-\xi(t))} \\ \implies\ & \|\dot{\xi}(t)\|^2_{D^2\mathrm{F}^*(-\xi(t))} \leq \frac{1}{t}\sum_{(x,y)\notin I_0} \gamma_{x,y}(t)\langle A^*\xi^* - c|e_{x,y}\rangle^2 + \frac{1}{t^2}\|\xi(t) - \xi^*\|^2_{D^2\mathrm{F}^*(-\xi(t))} \\ \implies\ & \|\dot{\xi}(t)\|^2 \leq \frac{1}{t}\frac{C_1}{\lambda_{\min}(t)}\max_{(x,y)\notin I_0}\gamma_{x,y}(t) + \frac{1}{t^4}\frac{\lambda_{\max}(t)}{\lambda_{\min}(t)}\|d(t)\|^2, \end{aligned}$$

where $C_1 \overset{\text{def}}{=} \sum_{(x,y)\notin I_0}\kappa^2_{x,y}$ and $\lambda_{\min}(t) > 0$ and $\lambda_{\max}(t) < +\infty$ are respectively the smallest and the largest eigenvalues of $D^2\mathrm{F}^*(-\xi(t))$. We have already seen that the trajectory $t \mapsto d(t)$ is bounded for $t \geq \varepsilon$ which implies that there exists $C_2 > 0$ such that, for any $(x, y) \notin I_0$, one has

$$\gamma_{x,y}(t) = e^{t\langle A^*\xi(t) - c|e_{x,y}\rangle} = e^{\langle A^*d(t)|e_{x,y}\rangle}e^{-t\kappa_{x,y}} \leq C_2 e^{-t\kappa^*},$$

where $\kappa^* \overset{\text{def}}{=} \min_{(x,y)\notin I_0}\kappa_{x,y}$. One obtains

$$\|\dot{\xi}(t)\|^2 \leq \frac{1}{t}Ce^{-t\kappa^*} + \frac{1}{t^4}K,$$

where

$$C \stackrel{\text{def}}{=} \sup_{t>\varepsilon} \frac{C_1 C_2}{\lambda_{\min}(t)} < +\infty \quad \text{and} \quad K \stackrel{\text{def}}{=} \sup_{t>\varepsilon} \frac{\lambda_{\max}(t)}{\lambda_{\min}(t)} \|d(t)\|^2 < +\infty.$$

Moreover, $t \mapsto t^{-1}e^{-t\kappa^*}$ converges exponentially fast towards 0 and thus $t^{-1}e^{-t\kappa^*} \lesssim t^{-4}$ when t is sufficiently large. Consequently, we have as t goes to $+\infty$

$$\|\dot{\xi}(t)\| \lesssim t^{-2},$$

which leads to

$$\|\xi^* - \xi(t)\| \le \int_t^{+\infty} \|\dot{\xi}(\omega)\| \, \mathrm{d}\omega \lesssim \int_t^{+\infty} \frac{1}{\omega^2} \, \mathrm{d}\omega \lesssim \frac{1}{t}.$$

□

4.2 Asymptotic analysis of the regularized transport plan

Theorem 4.6. *Under assumptions (H1) – (H6), the convergence rate of the curve $t \mapsto \gamma(t)$ is at least of order $O(t^{-\frac{1}{2}})$, i.e. for any t sufficiently large, one has*

$$\|\gamma(t) - \gamma^*\| \lesssim \frac{1}{\sqrt{t}}.$$

Proof. Let $t > 0$ sufficiently large in order to have $\|\xi(t) - \xi^*\| \lesssim t^{-1}$ as shown in Theorem 4.5. We have seen in Proposition 3.5 that $\dot{\xi}$ solves (ODE), which can be recognized as the first-order optimality condition of the following convex optimization problem

$$\inf_{\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \Psi_t(\xi) \stackrel{\text{def}}{=} \inf_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \|A^*\xi + \frac{1}{t^2} \log \gamma(t)\|^2_{\operatorname{diag}(\gamma(t))} + \frac{1}{t} \|\xi\|^2_{D^2\mathrm{F}^*(-\xi(t))}.$$

We compare $\dot{\xi}(t)$ with the competitor $h(t) \stackrel{\text{def}}{=} \frac{\xi^* - \xi(t)}{t}$ to obtain

$$\begin{aligned}
\Psi_t(\dot{\xi}(t)) &\le \Psi_t(h(t)) \\
&\le \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \gamma_{x,y}(t) \left(\langle A^* h(t) \,|\, e_{x,y} \rangle + \frac{\log \gamma_{x,y}(t)}{t^2} \right)^2 + \frac{1}{t} \|h(t)\|^2_{D^2\mathrm{F}^*(-\xi(t))} \\
&\le \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{t^2} \gamma_{x,y}(t) \langle A^*\xi^* - A^*\xi(t) + A^*\xi(t) - c | e_{x,y} \rangle^2 + \frac{1}{t^3} \|\xi^* - \xi(t)\|^2_{D^2\mathrm{F}^*(-\xi(t))} \\
&\le \sum_{(x,y)\notin I_0} \frac{1}{t^2} \gamma_{x,y}(t) \langle A^*\xi^* - c | e_{x,y} \rangle^2 + \frac{1}{t^3} \lambda_{\max}(t) \|\xi^* - \xi(t)\|^2 \\
&\lesssim \frac{1}{t^2} \max_{(x,y)\notin I_0} \gamma_{x,y}(t) + \frac{1}{t^5},
\end{aligned}$$

where the last inequality is obtained using that $\sum_{(x,y)\notin I_0}\langle A^*\xi^*-c|e_{x,y}\rangle^2$ and $\lambda_{\max}(t)$, the largest eigenvalue of $D^2\mathrm{F}^*(-\xi(t))$, are bounded quantities. From this, for any $(x,y)\in\mathcal{X}\times\mathcal{Y}$, one can obtain

$$\gamma_{x,y}(t)\langle A^*\dot{\xi}(t)+\frac{\log\gamma(t)}{t^2}|e_{x,y}\rangle^2 \lesssim \frac{1}{t^2}\max_{(x,y)\notin I_0}\gamma_{x,y}(t)+\frac{1}{t^5}.$$

Moreover, we compute

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{t}\log\gamma(t)\right) &= -\frac{1}{t^2}\log\gamma(t)+\frac{1}{t}\frac{\dot{\gamma}(t)}{\gamma(t)}\\ &= -\frac{1}{t}(A^*\xi(t)-c)+\frac{1}{t}\frac{(tA\dot{\xi}(t)+A\xi(t)-c)\gamma(t)}{\gamma(t)}\\ &= A^*\dot{\xi}(t),\end{aligned}$$

and in particular $A^*\dot{\xi}(t)+\frac{\log\gamma(t)}{t^2}=\frac{1}{t}\frac{\dot{\gamma}(t)}{\gamma(t)}$. Thus, we have

$$\begin{aligned} & \gamma_{x,y}(t)\left(\frac{1}{t}\frac{\dot{\gamma}_{x,y}(t)}{\gamma_{x,y}(t)}\right)^2 \lesssim \frac{1}{t^2}\max_{(x,y)\notin I_0}\gamma_{x,y}(t)+\frac{1}{t^5}\\ \iff \quad & \frac{\dot{\gamma}_{x,y}(t)^2}{\gamma_{x,y}(t)} \lesssim \max_{(x,y)\notin I_0}\gamma_{x,y}(t)+\frac{1}{t^3}\\ \implies \quad & \dot{\gamma}_{x,y}(t)^2 \lesssim \max_{(x,y)\notin I_0}\gamma_{x,y}(t)+\frac{1}{t^3},\end{aligned}$$

where the last inequality uses that $t\mapsto\gamma(t)$ is bounded when t is sufficiently large. Since $d(t)$ is bounded for t large enough and recalling that the transport coupling can be expressed in terms of the dual variable as

$$\gamma(t)=e^{t(A^*\xi(t)-c)}=e^{tA^*(\xi(t)-\xi^*)-t\kappa}=e^{A^*d(t)-t\kappa},$$

it follows that there exists $C>0$ such that

$$\forall (x,y)\notin I_0,\quad \gamma_{x,y}(t)\le Ce^{-t\kappa^*},$$

with $\kappa^*=\min_{(x,y)\notin I_0}\kappa_{x,y}$. It allows to conclude that for any $(x,y)\notin I_0$, $\gamma_{x,y}(t)$ converges exponentially fast towards 0. Consequently, we obtain the asymptotic estimate

$$|\dot{\gamma}_{x,y}(t)|\lesssim t^{-\frac{3}{2}}.$$

Finally, we have

$$\|\gamma^*-\gamma(t)\|\le\int_t^{+\infty}\|\dot{\gamma}(\omega)\|\,\mathrm{d}\omega\lesssim\int_t^{+\infty}\frac{1}{\omega^{\frac{3}{2}}}\,\mathrm{d}\omega\lesssim\frac{1}{\sqrt{t}}.$$

which allows to obtain

$$\|\gamma(t)-\gamma^*\|\lesssim\frac{1}{\sqrt{t}}.$$

□

5 Numerical experiments

We consider two finite sets $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$ with cardinality N and M, respectively and divergences which take the form $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ where q represents the marginals and ϕ is an entropy function.

For each divergence F considered in this section, we provide numerical illustrations of the expected upper bounds on the convergence rates for both the primal and dual problem. We consider two different datasets. The first dataset consists of two point clouds $\mathcal{X}$ and $\mathcal{Y}$ with uniform weights. Notably, $\mathcal{Y}$ contains outliers—points that are significantly distant from the others. The measures supported on these sets have different total masses, with $|\mu| = 13$ and $|\nu| = 15$.

The second dataset is composed of two discretized Gaussian measures on a regular 1-dimensional grid, also with unequal masses: $|\mu| = 11$ and $|\nu| = 10$.

The convergence rate plots are presented on a logarithmic scale and are compared to their theoretically expected rates.

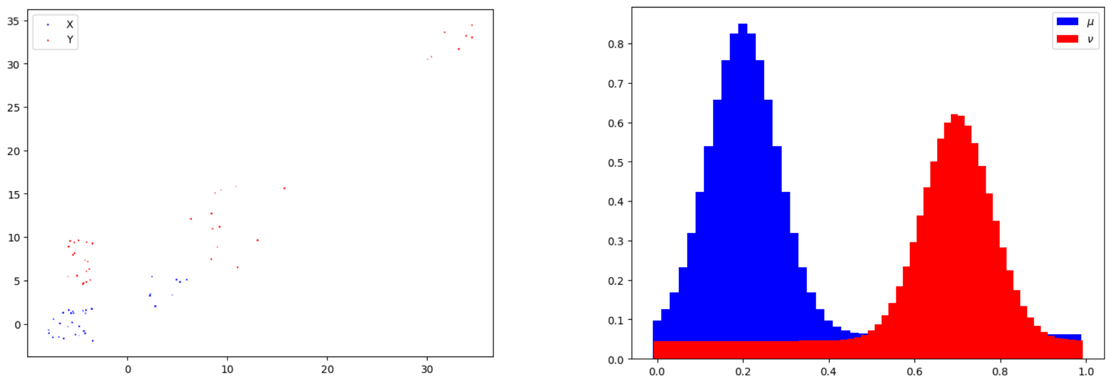

Figure 1: Two Datasets. The left plot represents two weighted point clouds. The right plot represents two discretized Gaussian measures.

5.1 Kullback-Leibler divergence

The most widely used case corresponds to the function ϕ defined as $\phi(x) \stackrel{\text{def}}{=} x(\log x - 1)$ on $\mathbb{R}_+$ and $+\infty$ otherwise. It gives the so called Kullback-Leibler divergence, defined for any $\psi \in \mathbb{R}^{N+M}$ by

$$\mathrm{F}(\psi) = \begin{cases} \displaystyle\sum_{i=1}^{N+M} \psi_i \left(\log \frac{\psi_i}{q_i} - 1\right) & \text{if } \psi_i \geq 0 \ (\forall i) \\ +\infty & \text{otherwise.} \end{cases}$$

The Legendre transform of ϕ is defined for any $y \in \mathbb{R}$ by

$$\phi^*(y) = e^y.$$

It follows that ϕ^* is a $\mathcal{C}^2$ function on $\mathbb{R}_+^*$ that is strongly convex at every point of its domain.

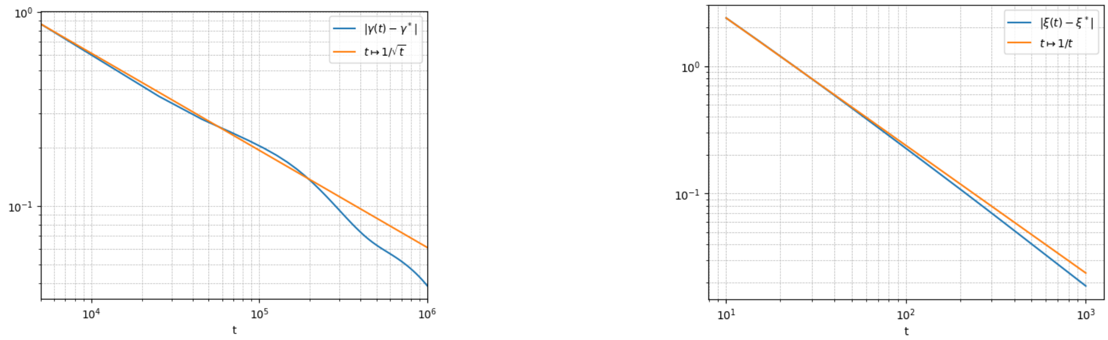

Figure 2: $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ with $\phi(t) = t(\log t - 1)$, μ and ν are measures supported on point clouds. The left plot (log scale) represents the behavior of $t \mapsto \|\gamma(t) - \gamma^*\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $t \mapsto \|\xi(t) - \xi^*\|$ (blue) compared to the expected rate (orange).

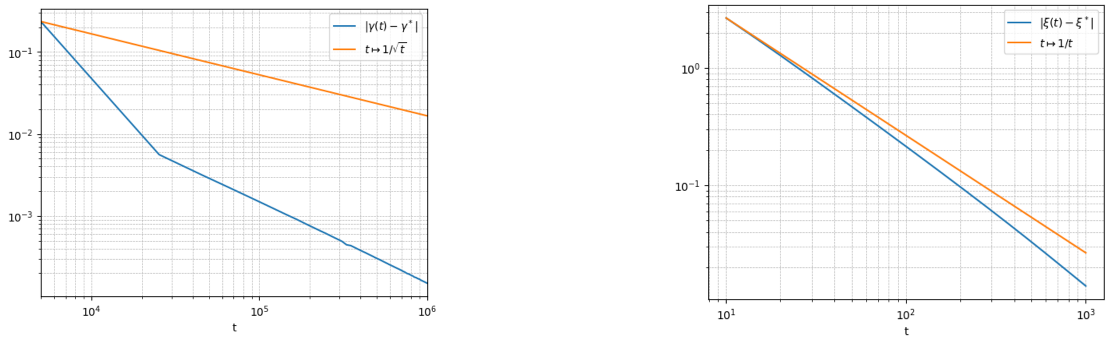

Figure 3: $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ with $\phi(t) = t(\log t - 1)$, μ and ν are discretized Gaussian measures. The left plot (log scale) represents the behavior of $t \mapsto \|\gamma(t) - \gamma^*\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $t \mapsto \|\xi(t) - \xi^*\|$ (blue) compared to the expected rate (orange).

We notice that our theoretical rate seems to be sharp for the dual problem but not for the primal problem.

5.2 Distance to the marginals

In the same framework as previously defined, another divergence function found in the literature is based on the Euclidean norm distance to q. Taking the entropy function defined by $\phi(x) \stackrel{\text{def}}{=} \frac{1}{2}|x-1|^2$ on $\mathbb{R}^{N+M}$ leads for any $\psi \in \mathbb{R}^{N+M}$ to the following divergence:

$$\mathrm{F}(\psi) = \frac{1}{2}\|\psi - q\|_2^2.$$

The Legendre transform of ϕ is defined for any $y \in \mathbb{R}$ as:

$$\phi^*(y) = \frac{1}{2}y^2 + y,$$

which is $\mathcal{C}^2$ and strongly-convex on $\mathbb{R}$.

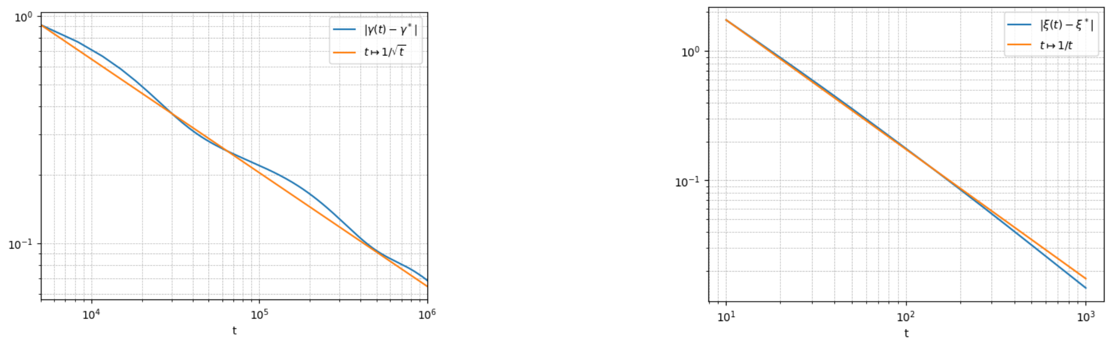

Figure 4: $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ with $\phi(x) = \frac{1}{2}|x-1|^2$, μ and ν are measures supported on point clouds. The left plot (log scale) represents the behavior of $t \mapsto \|\gamma(t) - \gamma^*\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $t \mapsto \|\xi(t) - \xi^*\|$ (blue) compared to the expected rate (orange).

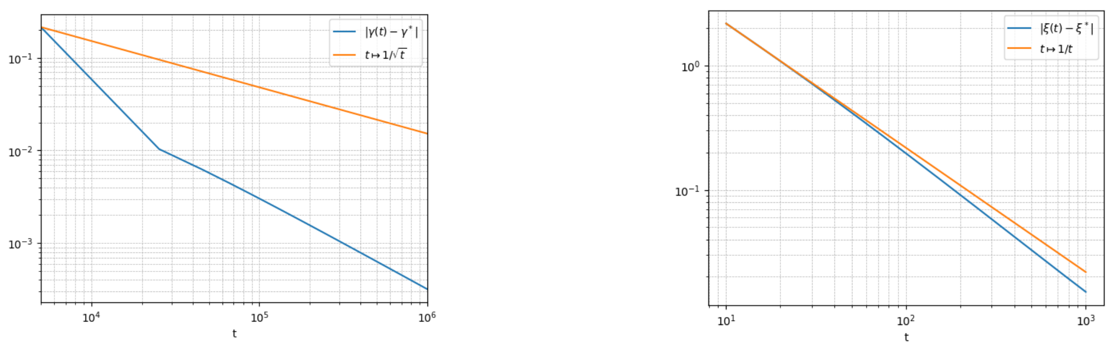

Figure 5: $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ with $\phi(x) = \frac{1}{2}|x-1|^2$, μ and ν are discretized Gaussian measures. The left plot (log scale) represents the behavior of $t \mapsto \|\gamma(t) - \gamma^*\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $t \mapsto \|\xi(t) - \xi^*\|$ (blue) compared to the expected rate (orange).

In this example again, we notice that our theoretical rate seems to be sharp for the dual problem but not for the primal problem. Either our analysis for the primal problem can be improved, or it is sharp for general entropies but not for the ones considered in these examples.

Acknowledgments. The authors acknowledge the support of the FMJH Program PGMO and the ANR project GOTA (ANR-23-CE46-0001).

References

[ANS22] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme. "Asymptotics for semidiscrete entropic optimal transport". In: *SIAM Journal on Mathematical Analysis* 54.2 (2022), pp. 1718–1741.

[Ben+15] J.-D. Benamou et al. "Iterative Bregman Projections for Regularized Transportation Problems". In: *SIAM Journal on Scientific Computing* 37.2 (Jan. 2015), A1111–A1138. ISSN: 1064-8275. DOI: 10.1137/141000439.

[Car+17] G. Carlier et al. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (Jan. 1, 2017), pp. 1385–1418. ISSN: 0036-1410. DOI: 10.1137/15M1050264.

[Chi+18a] L. Chizat et al. "An interpolating distance between optimal transport and Fisher–Rao metrics". In: *Foundations of Computational Mathematics* 18 (2018), pp. 1–44.

[Chi+18b] L. Chizat et al. "Unbalanced optimal transport: Dynamic and Kantorovich formulations". In: *Journal of Functional Analysis* 274.11 (2018), pp. 3090–3123.

[CM94] R. Cominetti and J. S. Martín. "Asymptotic analysis of the exponential penalty trajectory in linear programming". In: *Mathematical Programming* 67 (1994), pp. 169–187.

[CPT23] G. Carlier, P. Pegon, and L. Tamanini. "Convergence rate of general entropic optimal transport costs". In: *Calculus of Variations and Partial Differential Equations* 62.4 (2023), p. 116.

[CT21] G. Conforti and L. Tamanini. "A formula for the time derivative of the entropic cost and applications". In: *Journal of Functional Analysis* 280.11 (2021), p. 108964.

[Cut13] M. Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in neural information processing systems* 26 (2013).

[EN22] S. Eckstein and M. Nutz. *Convergence Rates for Regularized Optimal Transport via Quantization*. Comment: Fixed a typo in Theorem 3.8. Aug. 31, 2022. DOI: 10.48550/arXiv.2208.14391. URL: http://arxiv.org/abs/2208.14391. preprint.

[FL89] J. Franklin and J. Lorenz. "On the Scaling of Multidimensional Matrices". In: *Linear Algebra and its Applications*. Special Issue Dedicated to Alan J. Hoffman 114–115 (Mar. 1, 1989), pp. 717–735. ISSN: 0024-3795. DOI: 10.1016/0024-3795(89)90490-4.

[GS09] A. Galichon and B. Salanié. *Matching with Trade-Offs: Revealed Preferences over Competing Characteristics*. Preprint SSRN-1487307, 2009.

[KMV16] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. "A new optimal transport distance on the space of finite Radon measures". In: *Advances in Differential Equations* 21.11/12 (2016), pp. 1117–1164. DOI: 10.57262/ade/1476369298.

[Léo12] C. Léonard. "From the Schrödinger problem to the Monge–Kantorovich problem". In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.

[Léo14] C. Léonard. "A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport". In: *Discrete and Continuous Dynamical Systems* 34.4 (2014), p. 1533. DOI: 10.3934/dcds.2014.34.1533.

[LMS18] M. Liero, A. Mielke, and G. Savaré. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". In: *Inventiones mathematicae* 211.3 (2018), pp. 969–1117.

[MG20] S. D. Marino and A. Gerolin. "An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm". In: *Journal of Scientific Computing* 85.2 (Oct. 19, 2020), p. 27. ISSN: 1573-7691. DOI: 10.1007/s10915-020-01325-7.

[Mik90] T. Mikami. "Variational Processes from the Weak Forward Equation". In: *Communications in mathematical physics* 135.1 (1990), pp. 19–40.

[MS23] H. Malamut and M. Sylvestre. "Convergence rates of the regularized optimal transport: Disentangling suboptimality and entropy". In: *arXiv preprint arXiv:2306.06940* (2023).

[NP24] L. Nenna and P. Pegon. "Convergence rate of entropy-regularized multimarginal optimal transport costs". In: *Canadian Journal of Mathematics* (2024), pp. 1–21.

[Pal19] S. Pal. "On the difference between entropic cost and the optimal transport cost". In: *arXiv preprint arXiv:1905.12206* (2019).

[PC+19] G. Peyré, M. Cuturi, et al. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[Roc70] R. Rockafellar. "Convex Analysis". In: *Princeton Mathematical Series* 28 (1970).

[San23] F. Santambrogio. *A Course in the Calculus of Variations: Optimization, Regularity, and Modeling*. Springer Nature, 2023.

[Sin64] R. Sinkhorn. "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices". In: *The Annals of Mathematical Statistics* 35.2 (June 1964), pp. 876–879. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177703591.

[Sin67] R. Sinkhorn. "Diagonal Equivalence to Matrices with Prescribed Row and Column Sums". In: *The American Mathematical Monthly* 74.4 (1967), pp. 402–405.

[SVP22] T. Séjourné, F.-X. Vialard, and G. Peyré. “Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 4995–5021.

[Wee18] J. Weed. “An explicit analysis of the entropic penalty in linear programming”. In: *Conference On Learning Theory*. PMLR. 2018, pp. 1841–1855.