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 plans and potentials towards their original solution when both measures are weighted sums of Dirac masses.

Contents

*Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: `luca.nenna@universite-paris-saclay.fr`

†Institut Universitaire de France (IUF).

‡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 & ParMA, Inria Saclay, 91405, Orsay, France. Email: `louis.tocquec@universite-paris-saclay.fr`

Notations

The following notations we will be used throughout the article.

- $\mathscr{M}^+(\mathfrak{X})$ is the set of finite positive measures defined on $\mathfrak{X}$.

- $\langle\,\cdot\,|\,\cdot\,\rangle$ is the inner product associated with the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^d$.

- $\preceq$ stands for a component-wise inequality between vectors in $\mathbb{R}^{\mathfrak{X}}$.

- $\lesssim$ stands for $\leq$ up to a multiplicative universal positive constant.

- Δ_d is the $(d-1)$-dimensional simplex of $\mathbb{R}^d$ defined by

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

- $\sqcup$ stands for a disjoint union of sets.

- $\|\cdot\|_A$ is the norm given by $\|x\|_A = \sqrt{\langle Ax \,|\, x\rangle}$ where A is a positive definite self-adjoint operator.

- $\operatorname{dom} f$ denotes the domain $\{x : f(x) < +\infty\}$ of a convex function valued in $\mathbb{R} \cup \{+\infty\}$.

1 Introduction

Given two positive measures μ and ν on compact domains $\mathfrak{X} \subseteq \mathbb{R}^d$ with same total mass and $\mathfrak{Y} \subseteq \mathbb{R}^d$ and a cost function $c \colon \mathfrak{X} \times \mathfrak{Y} \to \mathbb{R}_+$, the Monge-Kantorovich optimal transport (OT) problem consists in

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

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

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

where Ent is the relative entropy.

    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 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 $\varepsilon \to 0$. 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; MS25; NP24], which provided asymptotic expansions 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,

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

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, depending on a parameter $\varepsilon > 0$, as follows

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

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 ε, let $\varepsilon \mapsto \gamma(\varepsilon)$ 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 under certain conditions on ϕ, the trajectory $\varepsilon \mapsto \gamma(\varepsilon)$ converges towards an optimizer $\bar\gamma$ of (1.3) at a rate of at least $\sqrt{\varepsilon}$, i.e.*

$$\|\bar\gamma - \gamma(\varepsilon)\| \lesssim \sqrt{\varepsilon}.$$

Theorem B (Simplified statement of Theorem 4.5))**.** *For positive ε, let $\varepsilon \mapsto \xi(\varepsilon)$ 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 under certain conditions on ϕ, the dual trajectory $\varepsilon \mapsto \xi(\varepsilon)$ converges towards the dual optimizer $\bar\xi$ at a rate of at least ε, i.e.*

$$\|\bar\xi - \xi(\varepsilon)\| \lesssim \varepsilon.$$

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 the announced asymptotic rates of the regularized problem when the regularization parameter tends to 0. 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 $\mu \in \mathcal{M}^+(\mathcal{X})$, $\nu \in \mathcal{M}^+(\mathcal{Y})$ and $\gamma \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$ 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 reformulate the marginal constraint 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,

$$\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}} \left\{ \langle c \,|\, \gamma \rangle \mid A\gamma = q \right\}, \tag{OT$_0$}$$

and for any non-negative ε,

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

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 $\operatorname{Ent} \colon \mathbb{R}^{\mathcal{X} \times \mathcal{Y}} \to \mathbb{R}$ is the discrete entropy, i.e. the entropy relative to the counting measure $\mathcal{H}^0$ defined by

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

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

Remark 2.1. Notice that, contrary to the problem stated in the general case in (1.2), we took the entropy with respect to $\mathcal{H}^0$ instead of $\mu \otimes \nu$. However in the case where $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $A\gamma = (\mu, \nu)$, they satisfy the following relation:

$$\operatorname{Ent}(\gamma \,|\, \mu \otimes \nu) = \operatorname{Ent}(\gamma) - \operatorname{Ent}(\mu) - \operatorname{Ent}(\nu) - 2.$$

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

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

Proposition 2.1. *For any $\varepsilon > 0$, (OT_ε) and its dual admit a unique solution γ_ε and ξ_ε (up to the kernel of A^*, which has dimension 1, for ξ_ε), respectively. Moreover, the following relation holds*

$$\gamma_\varepsilon = e^{\frac{1}{\varepsilon}(A^* \xi_\varepsilon - c)}.$$

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, Proposition 4.4]).

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 general function $F : \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \to \mathbb{R} \cup \{+\infty\}$ which is used as a penalization on the marginals of the transport coupling. The unbalanced optimal transport problem then takes the form:

$$\boxed{\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}} \{\langle c \,|\, \gamma \rangle + F(A\gamma)\}}. \tag{UOT_0}$$

In the same way as before, we consider a regularized version of this optimization problem which is given by the minimization of $\mathscr{C}_\varepsilon$ defined for every $\varepsilon > 0$ by

$$\mathscr{C}_\varepsilon(\gamma) \stackrel{\text{def}}{=} \langle c \,|\, \gamma \rangle + F(A\gamma) + \varepsilon \operatorname{Ent}(\gamma),$$

i.e. we consider

$$\boxed{\inf_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} \{\langle c \,|\, \gamma \rangle + F(A\gamma) + \varepsilon \operatorname{Ent}(\gamma)\}}. \tag{UOT_ε}$$

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

(H_0) F is a proper convex and lower semi-continuous function.

(H_{dom}) There exists $(\mu, \nu) \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$ such that $F(\mu, \nu) < +\infty$.

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 $q \in (\mathbb{R}_+)^N$ be a non-negative vector and $\phi \colon \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ an entropy function meaning that it is a convex and lower semi-continuous function such that $\operatorname{dom}\phi \subseteq \mathbb{R}_+$. The discrete Csiszàr ϕ-divergence is defined on $\mathbb{R}^N$ by

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

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

Remark 2.4. The choice of considering a general function F instead of a Csiszàr divergence is motivated by the fact that one may also want to consider variational problems given by the sum of an optimal transport term and a congestion on the marginals, which are not Csiszàr divergences. All assumptions that we make on F can easily be translated as assumptions on the entropy function ϕ when $F(\cdot) = D_\phi(\cdot|q)$ and $(\mu, \nu) = q \succ 0$. In that case (H_0) and (H_{dom}) are equivalent to the following:

(H_0') ϕ is a proper convex lower semi-continuous functions;

(H_{dom}') there exists $\alpha > 0$ such that $\phi(\alpha) < +\infty$.

2.2 Properties of the regularized unbalanced optimal transport

Let us start by introducing the dual formulations of the unbalanced OT problem and its regularization. We will use the adjoint operator of A defined for any $\xi = (\varphi, \psi) \in \mathbb{R}^{\mathfrak{X}} \oplus \mathbb{R}^{\mathfrak{Y}}$ by

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

The dual problems of (UOT_0) and (UOT_ε) are given by

$$\inf_{\xi \in \mathbb{R}^{\mathfrak{X}} \oplus \mathbb{R}^{\mathfrak{Y}}} \{F^*(-\xi) \mid A^*\xi \preceq c\}. \tag{DUOT$_0$}$$

and

$$\inf_{\xi \in \mathbb{R}^{\mathfrak{X}} \oplus \mathbb{R}^{\mathfrak{Y}}} F^*(-\xi) + \sum_{(x,y)\in\mathfrak{X}\times\mathfrak{Y}} \varepsilon \exp\left(\frac{1}{\varepsilon}\langle A^*\xi - c \mid e_{x,y}\rangle\right), \tag{DUOT$_\varepsilon$}$$

where $(e_{x,y})_{(x,y)\in\mathfrak{X}\times\mathfrak{Y}}$ denotes the canonical basis of $\mathbb{R}^{\mathfrak{X}\times\mathfrak{Y}}$. We denote by $\mathscr{K}_\varepsilon$ the functional of the dual problem, that is for $\xi = (\phi, \psi)$,

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

Remark 2.5. When $F(\cdot) = D_\phi(\cdot|q)$, F^* is given by

$$F^*(\xi) = \langle q \mid \phi^*(\xi)\rangle + I(\xi),$$

where $I(\xi) = 0$ if $\xi_z \leq \phi_\infty'$ for every $z \in \mathfrak{X} \sqcup \mathfrak{Y}$ such that $q_z = 0$, and $I(\xi) = +\infty$ otherwise.

We are now ready to state the basic results concerning existence, uniqueness and duality for these problems.

Proposition 2.2. *Let us assume* (H_0) *and* (H_{dom}). *Concerning the unbalanced OT problem, the following results hold:*

(i) *strong duality, i.e.*

$$\inf (UOT_0) + \inf (DUOT_0) = 0;$$

(ii) *existence for* (UOT_0) *if* F *is coercive, and uniqueness up to* $\ker A = \mathbb{R}(\mathbf{1}_{\mathbb{R}^{\mathfrak{X}}}, -\mathbf{1}_{\mathbb{R}^{\mathfrak{Y}}})$ *if* F *is strictly convex;*

(iii) *existence for* $(DUOT_0)$, *and uniqueness if* F^* *is strictly convex.*

Proof. We begin by introducing the convex 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$_0$) problem can be written as

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

By (H$_0$) and (H$_{\text{dom}}$), F is proper convex and lower semi-continuous, and there exists $(\mu, \nu) \in (\mathbb{R}_+^*)^{\mathcal{X}} \oplus (\mathbb{R}_+^*)^{\mathcal{Y}}$ such that $F(\mu, \nu) < +\infty$. By definition of A, there exists $\gamma \in (\mathbb{R}_+^*)^{\mathcal{X} \times \mathcal{Y}}$ such that $A\gamma = (\mu, \nu)$. Since both F and G are finite at γ and F is continuous at γ, we may apply FenchelRockafellar's Theorem [Bre11, Theorem 1.12] to assert that strong duality holds and that there exists a solution to the dual problem:

$$\inf_{\gamma \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}} G(\gamma) + F(A\gamma) = \max_{\xi \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} -G^*(A^*\xi) - F^*(-\xi),$$

where F^* and G^* are the LegendreFenchel transforms of F and G, respectively. Computing these transforms explicitly yields (i).

Now, assume that F is coercive. Notice that setting $\tilde{A}\gamma = (A\gamma, \langle c \,|\, \gamma \rangle)$ and $\tilde{\mathscr{C}}(p, t) = t + F(p)$ for $(p, t) \in (\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}) \oplus \mathbb{R}$, (UOT$_0$) can be equivalently written as

$$\inf_{\gamma \in (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}} \tilde{\mathscr{C}}(\tilde{A}\gamma).$$

By (H$_0$), it admits a competitor with finite cost. Let us notice that the cone $\Lambda = \tilde{A}\left((\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}}\right)$ is a closed convex cone, as a finitely generated convex cone (this can be shown by Caratheodory's theorem for cones). To get existence, we just have to show that $\tilde{\mathscr{C}}$ is coercive on Λ. Take a sequence $(A\gamma_n, \langle c \,|\, \gamma_n \rangle)_{n \in \mathbb{N}} \subseteq \Lambda$ which is unbounded. If $(A\gamma_n)$ is unbounded, $F(A\gamma_n) \to +\infty$ by coercivity of F, and since $\langle c \,|\, \gamma_n \rangle \geq 0$ for every n, $\tilde{\mathscr{C}}(\gamma_n) \to +\infty$. If on the contrary $(A\gamma_n)$ is bounded, then $(\langle c \,|\, \gamma_n \rangle)$ is unbounded, and necessarily it tends to $+\infty$. Since F is convex proper and lower semi-continuous, it is bounded from below by a linear map, and since $(A\gamma_n)$ is bounded, then $(F(\gamma_n))$ is lower bounded. This shows that $\tilde{\mathscr{C}}(\gamma_n) \to +\infty$ also in this case, hence $\tilde{\mathscr{C}}$ is coercive on the closed set Λ thus the primal problem problem admits a minimizer. Existence for the dual problem has been justified above, and the uniqueness statements are straightforward consequences of strict convexity, thus (ii) and (iii) hold true. $\qquad \square$

We shall need the definition of *slack variable* associated with this problem later on.

Definition 2.6 (Slack variable)**.** Assuming that (DUOT$_0$) has a unique solution $\bar{\xi} = (\bar{\varphi}, \bar{\psi})$ then we denote by κ the slack variable, i.e. the gap between $A^*\bar{\xi}$ and c, given for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ by

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

We denote by I_0 the set of saturated constraint where $\kappa = 0$, namely

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

Let us now deal with existence and duality for the regularized problem.

Proposition 2.3. *Under assumptions* (H_0) *and* (H_{dom})*, the following hold concerning the regularized unbalanced OT problem for every* $\varepsilon > 0$*:*

(i) strong duality, i.e.
$$\inf\left(\mathrm{UOT}_\varepsilon\right) + \inf\left(\mathrm{DUOT}_\varepsilon\right) = 0;$$

(ii) existence and uniqueness for $(\mathrm{UOT}_\varepsilon)$*;*

(iii) existence for $(\mathrm{DUOT}_\varepsilon)$*, and uniqueness if* F^* *is strictly convex.*

Proof. Let $\varepsilon > 0$. Rather than using Fenchel-Rockafellar as in the unregularized case, we are going to show existence for the primal problem first then use subdifferential calculus to deduce strong duality and existence for the dual problem. Since F is proper convex and lower semi-continuous, there exist $a \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}$ and $b \in \mathbb{R}$ 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

$$\mathscr{C}_\varepsilon(\gamma) = \langle c \,|\, \gamma \rangle + \mathrm{F}(A\gamma) + \varepsilon \operatorname{Ent}(\gamma) \geq \langle c + A^* a \,|\, \gamma \rangle + \varepsilon \operatorname{Ent}(\gamma) + b.$$

Using the superlinearity of the entropy functional Ent, we deduce that $\mathscr{C}_\varepsilon$ is coercive. Since it is also lower semi-continuous, the $(\mathrm{UOT}_\varepsilon)$ admits at least a solution. The strict convexity of Ent ensures that $\mathscr{C}_\varepsilon$ admits a unique solution, so that (ii) holds. Now, let γ_ε be the unique minimizer of $(\mathrm{UOT}_\varepsilon)$. Then, by first-order optimality conditions,

$$0 \in \partial \mathscr{C}_\varepsilon(\gamma_\varepsilon).$$

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

$$G_\varepsilon(\gamma) \stackrel{\mathrm{def}}{=} \langle c \,|\, \gamma \rangle + \varepsilon \operatorname{Ent}(\gamma) \quad \text{so that} \quad \mathscr{C}_\varepsilon = G_\varepsilon + \mathrm{F}.$$

Then the optimality condition becomes

$$0 \in A^* \,\partial\, \mathrm{F}(A\gamma_\varepsilon) + \partial G_\varepsilon(\gamma_\varepsilon),$$

where we used assumption (H_{dom}) (see [Roc70, Theorem 23.8]) to guarantee the existence of a point γ such that G_ε is continuous at γ and $A\gamma$ lies in the interior of $\operatorname{dom} \mathrm{F}$. Therefore, there exists $u \in \partial \mathrm{F}(A\gamma_\varepsilon)$ and $v \in \partial G_\varepsilon(\gamma_\varepsilon)$ such that

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

By the FenchelYoung equality, we have

$$\mathrm{F}(A\gamma_\varepsilon) + \mathrm{F}^*(u) = \langle u \,|\, A\gamma_\varepsilon \rangle, \quad \text{and} \quad G_\varepsilon(\gamma_\varepsilon) + G_\varepsilon^*(v) = \langle \gamma_\varepsilon \,|\, v \rangle = \langle \gamma_\varepsilon \,|\, \varepsilon - A^* u \rangle = -\langle u \,|\, A\gamma_\varepsilon \rangle.$$

Summing the two equalities yields:

$$\mathrm{F}(A\gamma_\varepsilon) + G_\varepsilon(\gamma_\varepsilon) + \mathrm{F}^*(u) + G_\varepsilon^*(-A^* u) = 0,$$

from which we get

$$\mathrm{F}(A\gamma_\varepsilon) + G_\varepsilon(\gamma_\varepsilon) = -\mathrm{F}^*(u) - G_\varepsilon^*(-A^* u).$$

This shows at the same time that u is a solution to the dual problem and strong duality holds. Since the Legendre transforms $\mathrm{F}^*, G_\varepsilon^*$ can be computed explicitly we have (i). Obviously, uniqueness holds if F^* is strictly convex, and (iii) holds. $\qquad \square$

Remark 2.7. Uniqueness of the solution for $(\mathrm{DUOT}_\varepsilon)$ holds modulo $\ker A^*$, even in the absence of strict convexity of F^*.

Remark 2.8. In the case where $\mathrm{F}(\cdot) = \mathrm{D}_\phi(\cdot|q)$ and $q \succ 0$, we have already seen that $(\mathrm{H}_{\mathrm{dom}})$ holds if and only if $\phi(\alpha) < +\infty$ for some $\alpha > 0$, in which case the previous proposition holds, and we have existence for $(\mathrm{DUOT}_\varepsilon)$. Let us show that existence for $(\mathrm{DUOT}_\varepsilon)$ may fail without the assumption $q \succ 0$. Consider a function ϕ such that $\phi'_\infty = +\infty$ and assume for example that $q_{x_0} = \mu_{x_0} = 0$ for some $x_0 \in \mathcal{X}$. By contradiction assume that there a solution $\xi(\varepsilon)$ of $(\mathrm{DUOT}_\varepsilon)$. We define a sequence $(\xi^n(\varepsilon))_{n \in \mathbb{N}}$ such that for any $n \in \mathbb{N}$,

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

Since $\phi'_\infty = +\infty$, by Remark 2.5 we have

$$\mathrm{F}^*(-\xi^n) = \langle q \,|\, \phi^*(-\xi^n(\varepsilon)) \rangle = \mathrm{F}^*(-\xi)$$

and

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

which is a contradiction.

Proposition 2.4. *Let $\varepsilon > 0$. Under (H_0) and $(\mathrm{H}_{\mathrm{dom}})$, the optimal regularized transport coupling for $(\mathrm{UOT}_\varepsilon)$ denoted $\gamma(\varepsilon)$ and any solution of $(\mathrm{DUOT}_\varepsilon)$ denoted ξ_ε are related through the following formula*

$$\gamma(\varepsilon) = \exp(\tfrac{1}{\varepsilon}(A^*\xi_\varepsilon - c)).$$

Proof. We observe that by strong duality, as stated in Proposition 2.3,

$$\mathrm{F}(\gamma(\varepsilon)) + \langle c \,|\, \gamma(\varepsilon) \rangle + \varepsilon \operatorname{Ent}(\gamma) + \mathrm{F}^*(-\xi_\varepsilon) - \varepsilon \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \exp\left(\frac{1}{\varepsilon} \langle A^*\xi_\varepsilon - c \,|\, e_{x,y} \rangle \right) = 0.$$

Denoting $(\gamma_{1,\varepsilon}, \gamma_{2,\varepsilon}) = A\gamma(\varepsilon)$ and $(\varphi_\varepsilon, \psi_\varepsilon) = \xi_\varepsilon$, by Fenchel-Young inequality $\mathrm{F}(\gamma(\varepsilon)) + \mathrm{F}^*(-\xi_\varepsilon) \geq \langle \gamma(\varepsilon) \,|\, \xi_\varepsilon \rangle = \langle \gamma_{1,\varepsilon} \,|\, \varphi_\varepsilon \rangle + \langle \gamma_{2,\varepsilon} \,|\, \psi_\varepsilon \rangle$ thus

$$\langle c \,|\, \gamma(\varepsilon) \rangle + \varepsilon \operatorname{Ent}(\gamma(\varepsilon)) \leq \langle \gamma_{1,\varepsilon} \,|\, \varphi_\varepsilon \rangle + \langle \gamma_{2,\varepsilon} \,|\, \psi_\varepsilon \rangle + \varepsilon \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \exp\left(\frac{1}{\varepsilon} \langle A^*\xi - c \,|\, e_{x,y} \rangle \right).$$

The left-hand side is the primal functional of $(\mathrm{OT}_\varepsilon)$ with fixed marginals $(\gamma_{1,\varepsilon}, \gamma_{2,\varepsilon})$ evaluated at $\gamma(\varepsilon)$, and the right-hand side is its dual functional evaluated at ξ_ε. Since the former is always greater than the latter, we have equality and $\gamma(\epsilon)$ and ξ_ε are respectively optimal for the primal and dual regularized OT problems with these fixed marginals. It classically follows that

$$\gamma(\varepsilon) = \exp\left(\tfrac{1}{\varepsilon}(A^*\xi_\varepsilon - c) \right).$$

$\square$

3 Convergence and regularity of the trajectories

This section is devoted to the study of the trajectory in ε of the solutions for the primal and dual problems, (UOT$_\varepsilon$) and (DUOT$_\varepsilon$) respectively. 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

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

Let us introduce a further series of hypotheses that we shall use several times throughout the paper.

(H$_1$) $W \cap (\mathbb{R}_+)^{\mathcal{X} \times \mathcal{Y}} \subseteq \operatorname{dom} F$, for some neighborhood W of the origin;

(H$_2$) F is super-linear at infinity;

(H$_3$) F* is strongly convex at every point of its domain;

(H$_4$) F* is $\mathcal{C}^2$ on $\operatorname{int}(\operatorname{dom} F^*)$.

Notice that (H$_1$) implies (H$_{\mathrm{dom}}$) and (H$_3$) implies strict convexity of F*.

Remark 3.3. In the case where $F(\cdot) = D_\phi(\cdot | q)$, with $q \succ 0$ fixed, all these assumptions on F are equivalent to the following assumptions on ϕ:

(H$_1'$) $\operatorname{dom} \phi$ contains some neighborhood of 0 in $\mathbb{R}_+$;

(H$_2'$) $\phi_\infty' = +\infty$;

(H$_3'$) ϕ^* is strongly convex at every point of its domain;

(H$_4'$) ϕ^* is $\mathcal{C}^2$ on $\operatorname{int}(\operatorname{dom} \phi^*)$.

This can be proven straighforwardly using the expression of F* given in Remark 2.5.

The following proposition shows in particular that under (H$_2$), F* has full domain. Its proof can be found for example in [San23], but we provide it for the sake of completeness.

Proposition 3.1. *Under assumptions* (H$_0$) *and* (H$_2$)*, the Legendre transform* F* *is bounded on every ball.*

Proof. Let $R > 0$. Recall that for every $\xi \in B(0, R)$ we have

$$F^*(\xi) \stackrel{\text{def}}{=} \sup_{v \in \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}} \langle v \,|\, \xi \rangle - F(v).$$

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

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

the infimum being a finite quantity since a proper lower semi-continuous convex function always has an affine minorant. We have shown that F^* is bounded from above on every ball, and since it is also proper convex and lower semi-continuous by (H$_0$), it is also bounded from below on every ball. $\qquad\square$

Lemma 3.2 (Uniform coercivity of $\mathscr{K}_\varepsilon$)**.** *Under assumptions* (H$_0$) *and* (H$_1$), *the functional* $\inf\limits_{\varepsilon > 0} \mathscr{K}_\varepsilon$ *is coercive.*

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

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

where $u_+ = \max\{0, u\}$ denotes the positive part of u. 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, $\xi_{z_0}^n \to -\infty$ then we have, using (H$_1$), that there exists $r \in \mathbb{R}_+^*$ such that $re_{z_0} \in \mathrm{dom}\,\mathrm{F}$. It leads to

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

On the other hand, if that is not the case, then for any $z \in \mathcal{X} \sqcup \mathcal{Y}$, ξ_z^n is bounded from below by some constant $M < 0$. Thus there exists $z_0 \in \mathcal{X} \sqcup \mathcal{Y}$ such that $\xi_{z_0}^n \to +\infty$. Without loss of generality we assume that $z_0 \in \mathcal{X}$. Then we have, using (H$_1$) again, that for any $y \in \mathcal{Y}$, there exists $r \in \mathbb{R}_+^*$ such that $re_y \in \mathrm{dom}\,\mathrm{F}$. For n large enough, it leads to

$$
\begin{aligned}
V(\xi^n) &\geq \mathrm{F}^*(-\xi^n) + (\varphi_{z_0}^n + \psi_y^n - c(z_0, y))_+ \\
&\geq -r\psi_y^n - \mathrm{F}(re_y) + \varphi_{z_0}^n + \psi_y^n - c(z_0, y) \\
&\geq |1 - r|M - F(re_y) + \varphi_{z_0}^n - c(z_0, y) \xrightarrow[n \to +\infty]{} +\infty.
\end{aligned}
$$

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. $\qquad\square$

3.1 Dual trajectory

In this section we show that the primal trajectory is regular, that it sastisfies an ODE (that will be used in Section 4), and that it converges as $\varepsilon \to 0$.

Proposition 3.3. *Assume that assumptions* (H_0)-(H_4) *hold. Then, there exists a unique solution $\xi(\varepsilon)$ to* $(\text{DUOT}_\varepsilon)$ *for every $\varepsilon > 0$ and the trajectory $\varepsilon \mapsto \xi(\varepsilon)$ is a $\mathcal{C}^1$ function on $\mathbb{R}_+^*$.*

Proof. The existence and uniqueness of the solution to $(\text{DUOT}_\varepsilon)$ comes from Proposition 2.3 and the strict convexity of the functional $\mathscr{K}_\varepsilon$, resulting from (H_3).

Let us first consider any $\varepsilon_0 \in \mathbb{R}_+^*$ and denote by $f: \mathbb{R}_+^* \times (\mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}}) \to \mathbb{R}$ the function defined by $f(\varepsilon, \xi) = D\mathscr{K}_\varepsilon(\xi)$. It is well-defined and of class $\mathcal{C}^2$ by Proposition 3.1 and (H_4). By optimality of $\xi(\varepsilon_0)$, $f(\varepsilon_0, \xi(\varepsilon_0)) = 0$ and by (H_3), we get $\det D_\xi f(\varepsilon_0, \xi(\varepsilon_0)) = \det D^2 \mathscr{K}_{\varepsilon_0}(\xi(\varepsilon_0)) > 0$. Thus the implicit function theorem allows to deduce that the trajectory is of class $\mathcal{C}^1$ on $\mathbb{R}_+^*$. $\qquad\square$

In the following $\xi(\varepsilon)$ will always denote the unique solution of $(\text{DUOT}_\varepsilon)$.

Proposition 3.4. *Assume that assumptions* (H_0)-(H_4) *hold. Then the trajectory $\varepsilon \mapsto \xi(\varepsilon)$ satisfies the following ordinary differential equation (ODE) on $\mathbb{R}_+^*$,*

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

which can be explicitly written, denoting $\gamma(\varepsilon)$ the unique solution to (UOT_ε)*, as*

$$A \operatorname{diag}\gamma(\varepsilon) A^* \dot\xi(\varepsilon) + \varepsilon D^2 \mathrm{F}^*(-\xi(\varepsilon))\dot\xi(\varepsilon) - A \operatorname{diag}\gamma(\varepsilon)\frac{A^*\xi(\varepsilon) - c}{\varepsilon} = 0. \qquad (\text{ODE})$$

Here $\operatorname{diag}\gamma$ denotes the linear operator on $\mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ given by $(\operatorname{diag}\gamma)\rho = (\gamma_{x,y}\rho_{x,y})_{x,y}$.

Proof. For every $\varepsilon > 0$, by using the optimality of $\xi(\varepsilon)$ for the functional $\mathscr{K}_\varepsilon$, one can obtain

$$D\mathscr{K}_\varepsilon(\xi(\varepsilon)) = 0.$$

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

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

Recalling the expression of $\mathscr{K}_\varepsilon$, i.e. $\mathscr{K}_\varepsilon(\gamma) = \mathrm{F}^*(-\xi) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \varepsilon \exp\left(\frac{1}{\varepsilon}\langle A^*\xi - c \,|\, e_{x,y}\rangle\right)$ and the relation between $\gamma(\varepsilon)$ and $\xi(\varepsilon)$ given in Proposition 2.4, we may compute $D\mathscr{K}_\varepsilon$, $D^2\mathscr{K}_\varepsilon$, $D\mathscr{K}_\varepsilon$ and $(\partial/\partial\varepsilon)(D\mathscr{K}_\varepsilon)$ explicitly, so as to obtain (ODE). $\qquad\square$

Proposition 3.5. *Under assumptions* (H_0)-(H_3)*, the trajectory $\varepsilon \mapsto \xi(\varepsilon)$ converges as $\varepsilon \to 0$ towards the unique solution $\bar\xi$ to* (DUOT_0)*.*

Proof. Recall that $\bar\xi$ is unique by strict convexity of F^*. Using the optimality of $\xi(\varepsilon)$ for $(\text{DUOT}_\varepsilon)$, we leverage the following inequality :

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

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

$$\mathscr{K}_\varepsilon(\xi(\varepsilon)) = \mathrm{F}^*(-\xi(\varepsilon)) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \varepsilon e^{\frac{1}{\varepsilon}\langle A^*\xi(\varepsilon) - c \,|\, e_{x,y}\rangle} \leq \mathrm{F}^*\left(-\bar\xi\right) + \varepsilon|\mathcal{X}||\mathcal{Y}|. \qquad (3.1)$$

12

From this, and the uniform coercivity of $\mathscr{K}_\varepsilon$ as stated in Lemma 3.2 (which applies thanks to (H_0) and (H_2)), we deduce that the trajectory $\varepsilon \mapsto \xi(\varepsilon)$ is bounded on $(0,1)$. By Bolzano-Weierstrass Theorem, we may consider a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\xi(\varepsilon_n) \to \tilde{\xi}$ as $n \to +\infty$, where $\tilde{\xi}$ is an accumulation point of the trajectory (at 0). Since $(\xi(\varepsilon_n))_{n\in\mathbb{N}}$ is bounded, by Proposition 3.1 we know that $|F^*(-\xi(\varepsilon_n))|$ is bounded by some constant $C > 0$ for any $n \in \mathbb{N}$. Thus by (3.1), for every n and every $(x,y) \in \mathfrak{X} \times \mathcal{Y}$,

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

If $\alpha \overset{\text{def}}{=} \langle A^*\tilde{\xi} - c\,|\,e_{x,y}\rangle > 0$ then for n large enough we would have $\varepsilon_n e^{\frac{1}{\varepsilon_n}\langle A^*\xi(\varepsilon_n)-c\,|\,e_{x,y}\rangle} \geq \varepsilon_n e^{\frac{\alpha}{2\varepsilon_n}} \to +\infty$: 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

$$F^*(-\xi(\varepsilon_n)) \leq F^*\left(-\bar{\xi}\right) + \varepsilon_n|\mathfrak{X}||\mathcal{Y}|,$$

and taking the limit as $n \to +\infty$, we obtain by lower semi-continuity of F^*:

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

This shows that any accumulation point of the trajectory is optimal for $(DUOT_0)$ thus equal to $\bar{\xi}$, hence $\xi(\varepsilon)$ converges to $\bar{\xi}$. □

3.2 Primal trajectory

The purpose of the following proposition is to show that the curve $\varepsilon \mapsto \gamma(\varepsilon)$ converges to $\bar{\gamma}$, the solution of (UOT_0) of minimal entropy (which is unique by strict convexity of Ent), under suitable assumptions.

Proposition 3.6. *Under assumptions (H_0)-(H_4), the trajectory $\varepsilon \mapsto \gamma(\varepsilon)$ converges towards $\bar{\gamma}$, the solution of (UOT_0) with minimal entropy Ent, as $\varepsilon \to 0$. Moreover, $\varepsilon \mapsto \gamma(\varepsilon)$ is a $\mathcal{C}^1$ function on $\mathbb{R}_+^*$.*

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

$$\langle c\,|\,\gamma(\varepsilon)\rangle + F(A\gamma(\varepsilon)) + \varepsilon \operatorname{Ent}(\gamma(\varepsilon)) \leq \langle c\,|\,\bar{\gamma}\rangle + F(A\bar{\gamma}) + \varepsilon \operatorname{Ent}(\bar{\gamma}). \tag{3.2}$$

Since $\bar{\gamma}$ is optimal for (UOT_0), we have

$$\langle c\,|\,\bar{\gamma}\rangle + F(A\bar{\gamma}) \leq \langle c\,|\,\gamma(\varepsilon)\rangle + F(A\gamma(\varepsilon))$$

and combining it with (3.2) we get

$$\operatorname{Ent}(\gamma(\varepsilon)) \leq \operatorname{Ent}(\bar{\gamma}). \tag{3.3}$$

By coercivity of Ent, we can conclude that the trajectory $(\gamma(\varepsilon))_{\varepsilon>0}$ is bounded. It implies that there exists $(\varepsilon_n)_{n\in\mathbb{N}}$ converging to 0 such that $\gamma(\varepsilon_n) \to \tilde{\gamma}$ as $n \to +\infty$, where $\tilde{\gamma}$ is an accumulation point at 0. Since Ent is lower bounded, $\operatorname{Ent}(\gamma(\varepsilon))$ is bounded, thus taking $\varepsilon = \varepsilon_n$ in (3.2) and passing to the limit shows that $\tilde{\gamma}$ is a solution of (UOT_0). Finally, using (3.3), taking $\varepsilon = \varepsilon_n$ and passing to the limit, by lower semicontinuity of Ent we obtain

$$\operatorname{Ent}(\tilde{\gamma}) \leq \operatorname{Ent}(\bar{\gamma}),$$

which implies that $\tilde{\gamma} = \bar{\gamma}$, and thus $\gamma(\varepsilon) \to \bar{\gamma}$ as $\varepsilon \to 0$. Finally, the regularity of the trajectory is a direct consequence of Proposition 2.4 and Proposition 3.3. □

4 Asymptotics for the regularized unbalanced problem

This section is devoted to an asymptotic analysis of the regularized unbalanced problem as $\varepsilon \to 0$. 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 (corresponding to entropy penalization in the primal). We adapt their strategy to the non-linearity induced 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 as set 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* (H$_0$) - (H$_4$), *the quantity* $D\mathrm{F}^*(-\bar{\xi})$ *belongs to* E_0.

Proof. Consider the curve $\varepsilon \mapsto \xi(\varepsilon)$ of solutions to (DUOT$_\varepsilon$). By (H$_0$), (H$_2$), (H$_4$) 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:

$$D\mathscr{K}_\varepsilon(\xi(\varepsilon)) = 0$$
$$\iff D\mathrm{F}^*(-\xi(\varepsilon)) + \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \gamma_{x,y}(\varepsilon) Ae_{x,y} = 0.$$

By Proposition 3.5 and Proposition 3.6, $\xi(\varepsilon) \to \bar{\xi}$ and $\gamma(\varepsilon) \to \bar{\gamma}$ as $\varepsilon \to 0$, thus taking the limit in the above expression leads to

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

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

$$\gamma_{x,y}(\varepsilon) = e^{\frac{1}{\varepsilon}\langle A^*\xi(\varepsilon)-c \,|\, e_{x,y}\rangle} \xrightarrow[\varepsilon\to 0]{} 0.$$

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

For any positive ε, we consider the quantity

$$d(\varepsilon) \overset{\text{def}}{=} \tfrac{1}{\varepsilon}(\xi(\varepsilon) - \bar{\xi})$$

and we look at $\mathscr{K}_\varepsilon$ as a function of d instead of ξ, namely we introduce the new functional $\tilde{\mathscr{K}}_\varepsilon \colon \mathbb{R}^{\mathcal{X}} \oplus \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}$ defined as

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

where κ is the gap between $A\bar{\xi}$ and c as defined in Definition 2.6. Notice that it is defined in such a way that $\tilde{\mathscr{K}}_\varepsilon(\frac{1}{\varepsilon}(\xi - \bar{\xi})) = \mathscr{K}_\varepsilon(\xi)$.

Proposition 4.2. *Under assumptions* (H$_0$)-(H$_4$), *the curve* $\varepsilon \mapsto d(\varepsilon)$ *is bounded for* ε *sufficiently small.*

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

$$\langle A^*d(\varepsilon)\,|\,e_{x,y}\rangle = \frac{1}{\varepsilon}\langle A^*\xi(\varepsilon) - A^*\bar{\xi}\,|\,e_{x,y}\rangle = \frac{1}{\varepsilon}\langle A^*\xi(\varepsilon) - c\,|\,e_{x,y}\rangle,$$

and thus by Proposition 3.6,

$$\langle A^*d(\varepsilon)\,|\,e_{x,y}\rangle = \log\gamma_{x,y}(\varepsilon) \xrightarrow[\varepsilon \to 0]{} \log\bar{\gamma}_{x,y}.$$

Since $\bar{\gamma}$ has finite entropy and $\bar{\gamma} \succ 0$, we deduce from this and from the regularity of $\xi(\varepsilon)$ (stated in Proposition 3.3), that $\langle A^*d(\varepsilon)\,|\,e_{x,y}\rangle$ is a bounded quantity on $(0,1)$ for any $(x,y) \in I_0$, which in turn implies that $d^{(0)}(\varepsilon)$, the projection of $d(\varepsilon)$ onto E_0, is bounded. Besides, since $\tilde{\mathscr{K}}_\varepsilon(\frac{1}{\varepsilon}(\xi - \bar{\xi})) = \mathscr{K}_\varepsilon(\xi)$ and $\xi(\varepsilon)$ is the unique minimizer of $\mathscr{K}_\varepsilon$, $d(\varepsilon)$ is the unique minimizer of $\tilde{\mathscr{K}}_\varepsilon$. Comparing it with $d^{(0)}(\varepsilon)$ yields

$$\tilde{\mathscr{K}}_\varepsilon(d(\varepsilon)) \le \tilde{\mathscr{K}}_\varepsilon(d^{(0)}(\varepsilon)),$$

which implies

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

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

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

Notice that $\varepsilon d(\varepsilon) = \xi(\varepsilon) - \bar{\xi} \to 0$ and $-\bar{\xi} \in \mathrm{dom}\,\mathrm{F}^*$, because F^* has full domain thanks to Proposition 3.1. Thus, by (H$_3$), for some $r \in (0,1)$, for all $\varepsilon \le r$ we have

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

where $\lambda_0(-\bar{\xi},r) > 0$, and by (H$_4$) and Taylor's expansion, we have

$$\mathrm{F}^*\left(-\varepsilon d^{(0)}(\varepsilon) - \bar{\xi}\right) \le \mathrm{F}^*\left(-\bar{\xi}\right) - \varepsilon\langle D\,\mathrm{F}^*\left(-\bar{\xi}\right)\,|\,d^{(0)}(\varepsilon)\rangle + C\varepsilon^2 \tag{4.3}$$

for some $C > 0$ and ε sufficiently small. By Lemma 4.1, we have

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

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

$$\lambda_0\|d(\varepsilon)\|^2 \le \sum_{(x,y)\notin I_0} \frac{2}{\varepsilon} e^{\langle A^*d^{(0)}(\varepsilon) - \frac{1}{\varepsilon}\kappa\,|\,e_{x,y}\rangle} + C.$$

We have seen that $d^{(0)}(\varepsilon)$ is bounded, and since $\kappa_{x,y} > 0$ for every $(x,y) \notin I_0$, the right-hand side is bounded, thus $d(\varepsilon)$ is bounded as well for $\varepsilon \le r$. $\qquad\square$

Proposition 4.3. *Under assumptions* (H_0)-(H_4), *the trajectory* $\varepsilon \mapsto d(\varepsilon)$ *converges towards a finite quantity denoted* $\bar{d}$.

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* $L : E \to E'$ *a linear operator. If the set of solutions* $\mathcal{U}$ *of the problem*

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

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

$$\mathcal{U} = x_0 + \ker L.$$

Proof. Let $x_0 \in \mathcal{U}$. If $x_1 = x_0 + y$, with $y \in \ker L$, then $Lx_0 = Lx_1$ thus $f(Lx_1) = f(Lx_0)$, and it follows that $x_1 \in \mathcal{U}$.

Conversely, let $x_1 \in \mathcal{U}$. We set $y = x_1 - x_0$ and $x_2 = \frac{1}{2}(x_0 + x_1) = x_0 + \frac{1}{2}y$. Assuming that $y \notin \ker L$, $Lx_1 \neq Lx_2$ 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 $x_2 \overset{\text{def}}{=} (x_0 + x_1)/2 = x_0 + \frac{1}{2}y$. This contradicts the optimality of x_0, therefore $y \in \ker L$ and $x_1 \in x_0 + \ker L$. $\qquad\square$

Proof of Proposition 4.3. Let $\bar{d}$ be an accumulation point at 0 of the curve $\varepsilon \mapsto d(\varepsilon)$, meaning that $d(\varepsilon_n) \to \bar{d}$ as $n \to +\infty$ for some sequence $\varepsilon_n \to 0$, which exists thanks to Proposition 4.2. Since $\xi(\varepsilon_n)$ is a minimizer of $(\text{DUOT}_\varepsilon)$ for $\varepsilon = \varepsilon_n$, we obtain the following equation:

$$DF^*\Big(-\varepsilon_n d(\varepsilon_n) - \bar{\xi}\Big) + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} e^{\langle A^* d(\varepsilon_n) - \frac{1}{\varepsilon_n} \kappa \,|\, e_{x,y}\rangle} A e_{x,y} = 0.$$

Taking the limit in this expression leads to

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

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

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

so that $\bar{d}$ is a minimizer of G. Notice that by Lemma 4.1, $\langle DF^*(-\bar{\xi}) \,|\, z\rangle$ 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

$$\mathcal{U} = \bar{d} + \bigcap_{(x,y) \in I_0} \ker A e_{x,y}.$$

Now, let us introduce $\bar{d}_1 \in \mathcal{U}$ such that $\bar{d}_1 \neq \bar{d}$. We define $d_1(\varepsilon)$ by

$$d_1(\varepsilon) \overset{\text{def}}{=} d(\varepsilon) + \bar{d}_1 - \bar{d}$$

which satisfies $d_1(\varepsilon_n) \to \bar{d}_1$ as $n \to +\infty$.

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

$$\tilde{\mathscr{K}}_\varepsilon(d(\varepsilon_n)) \leq \tilde{\mathscr{K}}_\varepsilon(d_1(\varepsilon_n)),$$

or equivalently

$$\mathrm{F}^* \left(-\varepsilon_n d(\varepsilon_n) - \bar{\xi} \right) + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \varepsilon_n e^{\langle A^* d(\varepsilon_n) \,|\, e_{x,y} \rangle}$$

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

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

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

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

where we have denoted $\|u\|_A \overset{\text{def}}{=} \langle Au \,|\, u \rangle$ and omitted the first-order terms, which cancelled thanks to Lemma 4.1. Besides, thanks to the fact that $d(\varepsilon_n) - d_1(\varepsilon_n) \in \bigcap_{(x,y) \in I_0} \ker A e_{x,y}$, 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:

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

Passing to the limit, we obtain

$$\|\bar{d}\|^2_{D^2\mathrm{F}^*(-\bar{\xi})} \leq \|\bar{d}_1\|^2_{D^2\mathrm{F}^*(-\bar{\xi})}.$$

Thus, we have shown that $\bar{d}$ is the unique solution of the strictly convex (by (H$_3$)) problem:

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

We have shown that $d(\varepsilon)$ has a unique accumulation point at 0, and since it is bounded on a neighborhood of 0, it converges to $\bar{d}$. $\qquad\square$

Now, we shall study the derivative of $d(\varepsilon)$ with respect to ε which is equal to $\dot{d}(\varepsilon) = -\frac{1}{\varepsilon}d(\varepsilon) + \frac{1}{\varepsilon}\dot{\xi}(\varepsilon)$.

Theorem 4.5. *Under assumptions* (H$_0$)-(H$_4$)*, the convergence rate of the curve $\varepsilon \mapsto \xi(\varepsilon)$ is at least of order $O(\varepsilon)$, i.e. for any ε sufficiently close to 0, one has*

$$\|\xi(\varepsilon) - \bar{\xi}\| \lesssim \varepsilon.$$

Proof. Let $\varepsilon > 0$. We define a functional Λ_ε by

$$\Lambda_\varepsilon(\sigma) \stackrel{\text{def}}{=} \left\| A^*\sigma + \frac{c - A^*\bar{\xi}}{\varepsilon^2} \right\|_{\text{diag}(\gamma(\varepsilon))}^2 + \varepsilon \left\| \sigma + \frac{d(\varepsilon)}{\varepsilon} \right\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2.$$

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

$$D\Lambda_\varepsilon(\sigma) = 2A\,\text{diag}(\gamma(\varepsilon))\left(A^*\sigma + \frac{c - A^*\bar{\xi}}{\varepsilon^2}\right) + 2\varepsilon D^2\mathrm{F}^*(-\xi(\varepsilon))\left(\sigma + \frac{d(\varepsilon)}{\varepsilon}\right)$$

and thus, using the ODE satisfied by $\varepsilon \mapsto \xi(\varepsilon)$ and established in Proposition 3.4,

$$D\Lambda_\varepsilon(\dot{d}(\varepsilon)) = \frac{2}{\varepsilon}\left(A\,\text{diag}(\gamma(\varepsilon))A^*\dot{\xi}(\varepsilon) + \varepsilon D^2\mathrm{F}^*(-\xi(\varepsilon))\dot{\xi}(\varepsilon) - A\,\text{diag}(\gamma(\varepsilon))\frac{A^*\xi(\varepsilon) - c}{\varepsilon}\right) = 0.$$

Using this property, we get the following,

$$\Lambda_\varepsilon(\dot{d}(\varepsilon)) \leq \Lambda_\varepsilon(0)$$

$$\Longleftrightarrow \quad \left\| A^*\frac{\dot{\xi}(\varepsilon)}{\varepsilon} \right\|_{\text{diag}(\gamma(\varepsilon))}^2 + \varepsilon \left\| \frac{\dot{\xi}(\varepsilon)}{\varepsilon} \right\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2 \leq \left\| \frac{c - A^*\bar{\xi}}{\varepsilon^2} \right\|_{\text{diag}(\gamma(\varepsilon))}^2 + \varepsilon \left\| \frac{d(\varepsilon)}{\varepsilon} \right\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2.$$

This in turn implies

$$\varepsilon \left\| \frac{\dot{\xi}(\varepsilon)}{\varepsilon} \right\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2 \leq \sum_{(x,y)\notin I_0} \frac{1}{\varepsilon^2}\gamma_{x,y}(\varepsilon)\langle c - A^*\bar{\xi}\,|\,e_{x,y}\rangle^2 + \varepsilon\left\| \frac{\bar{\xi} - \xi(\varepsilon)}{\varepsilon^2} \right\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2$$

$$\Longrightarrow \quad \|\dot{\xi}(\varepsilon)\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2 \leq \sum_{(x,y)\notin I_0} \frac{1}{\varepsilon}\gamma_{x,y}(\varepsilon)\langle c - A^*\bar{\xi}\,|\,e_{x,y}\rangle^2 + \frac{1}{\varepsilon^2}\|\bar{\xi} - \xi(\varepsilon)\|_{D^2\mathrm{F}^*(-\xi(\varepsilon))}^2$$

$$\Longrightarrow \quad \|\dot{\xi}(\varepsilon)\|^2 \leq \frac{C_1}{\varepsilon\lambda_{\min}(\varepsilon)} \max_{(x,y)\notin I_0} \gamma_{x,y}(\varepsilon) + \frac{\lambda_{\max}(\varepsilon)}{\lambda_{\min}(\varepsilon)}\|d(\varepsilon)\|^2,$$

where $C_1 \stackrel{\text{def}}{=} \sum_{(x,y)\notin I_0} \kappa_{x,y}^2$ and $\lambda_{\min}(\varepsilon) > 0$ and $\lambda_{\max}(\varepsilon) < +\infty$ are respectively the smallest and the largest eigenvalues of $D^2\mathrm{F}^*(-\xi(\varepsilon))$. We have already seen that the trajectory $\varepsilon \mapsto d(\varepsilon)$ is bounded for $\varepsilon \leq r$ for some small $r > 0$ thus there exists $C_2 > 0$ such that for $\varepsilon \leq r$, $\|d(\varepsilon)\| \leq C_2$, and for any $(x,y) \notin I_0$

$$\gamma_{x,y}(\varepsilon) = e^{\frac{1}{\varepsilon}\langle A^*\xi(\varepsilon)-c\,|\,e_{x,y}\rangle} = e^{\langle A^*d(\varepsilon)\,|\,e_{x,y}\rangle}e^{-\frac{1}{\varepsilon}\kappa_{x,y}} \leq C_2 e^{-\frac{1}{\varepsilon}\kappa^*},$$

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

$$\|\dot{\xi}(\varepsilon)\|^2 \leq \frac{C}{\varepsilon}e^{-\frac{1}{\varepsilon}\kappa^*} + K,$$

where

$$C \stackrel{\text{def}}{=} \sup_{0<\varepsilon\leq r} \frac{C_1 C_2}{\lambda_{\min}(\varepsilon)} < +\infty \quad \text{and} \quad K \stackrel{\text{def}}{=} \sup_{0<\varepsilon\leq r} C_2^2 \frac{\lambda_{\max}(\varepsilon)}{\lambda_{\min}(\varepsilon)} < +\infty,$$

these quantities being finite because F^* is of class $\mathcal{C}^2$ by (H$_4$). Moreover, $\varepsilon \mapsto \frac{1}{\varepsilon}e^{-\frac{1}{\varepsilon}\kappa^*}$ converges exponentially fast towards 0 and thus $\frac{1}{\varepsilon}e^{-\frac{1}{\varepsilon}\kappa^*} \lesssim 1$ when ε is sufficiently close to 0. Consequently, we have as ε goes to 0

$$\|\dot{\xi}(\varepsilon)\| \lesssim 1,$$

which leads to

$$\|\bar{\xi} - \xi(\varepsilon)\| \leq \int_0^\varepsilon \|\dot{\xi}(\omega)\|\,\mathrm{d}\omega \lesssim \int_0^\varepsilon \mathrm{d}\omega \lesssim \varepsilon.$$

$\square$

4.2 Asymptotic analysis of the primal trajectory

Theorem 4.6. *Under assumptions* (H$_0$)-(H$_4$), *the convergence rate of the curve* $\varepsilon \mapsto \gamma(\varepsilon)$ *is at least of order* $O(\sqrt{\varepsilon})$, *i.e. for any* ε *close to* 0, *one has*

$$\|\gamma(\varepsilon) - \bar{\gamma}\| \lesssim \sqrt{\varepsilon}.$$

Proof. Let $\varepsilon > 0$ sufficiently small in order to have $\|\xi(\varepsilon) - \bar{\xi}\| \lesssim \varepsilon$ as shown in Theorem 4.5. We have seen in Proposition 3.4 that $\dot{\xi}$ solves (ODE), which can be recognized as the first-order optimality condition of the following convex optimization problem

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

We compare $\dot{\xi}(t)$ with the competitor $h(\varepsilon) \overset{\text{def}}{=} \frac{1}{\varepsilon}(\xi(\varepsilon) - \bar{\xi})$ to obtain

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

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

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

Moreover, we compute

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

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

$$
\begin{aligned}
\gamma_{x,y}(\varepsilon) \left(\varepsilon \frac{\dot{\gamma}_{x,y}(\varepsilon)}{\gamma_{x,y}(\varepsilon)} \right)^2 &\lesssim \frac{1}{\varepsilon^2} \max_{(x,y) \notin I_0} \gamma_{x,y}(t) + \varepsilon \\
\Longleftrightarrow \qquad \varepsilon^2 \frac{\dot{\gamma}_{x,y}(\varepsilon)^2}{\gamma_{x,y}(t)} &\lesssim \frac{1}{\varepsilon^2} \max_{(x,y) \notin I_0} \gamma_{x,y}(\varepsilon) + \frac{1}{\varepsilon} \\
\Longrightarrow \qquad \dot{\gamma}_{x,y}(\varepsilon)^2 &\lesssim \frac{1}{\varepsilon^4} \max_{(x,y) \notin I_0} \gamma_{x,y}(\varepsilon) + \frac{1}{\varepsilon},
\end{aligned}
$$

19

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

$$\gamma(\varepsilon) = e^{\frac{1}{\varepsilon}(A^*\xi(\varepsilon)-c)} = e^{\frac{1}{\varepsilon}A^*(\xi(\varepsilon)-\bar{\xi})-\frac{1}{\varepsilon}\kappa} = e^{A^*d(\varepsilon)-\frac{1}{\varepsilon}\kappa},$$

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

$$\forall (x,y) \notin I_0, \quad \gamma_{x,y}(\varepsilon) \leq Ce^{-\frac{1}{\varepsilon}\kappa^*},$$

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

$$|\dot{\gamma}_{x,y}(\varepsilon)| \lesssim \frac{1}{\sqrt{\varepsilon}}.$$

Finally, we have

$$\|\gamma(\varepsilon) - \bar{\gamma}\| \leq \int_0^\varepsilon \|\dot{\gamma}(\omega)\|\, \mathrm{d}\omega \lesssim \int_0^\varepsilon \frac{1}{\sqrt{\omega}}\, \mathrm{d}\omega \lesssim \sqrt{\varepsilon}.$$

which allows to obtain

$$\|\gamma(\varepsilon) - \bar{\gamma}\| \lesssim \sqrt{\varepsilon}. \qquad \square$$

5 Numerical experiments

We consider two finite sets $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ 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 the dual problem. We consider two different datasets. The first dataset is composed of two discretized Gaussian measures on a regular 1-dimensional grid, also with unequal masses: $|\mu| = 11$ and $|\nu| = 10$. The second and third datasets consist of two point clouds $\mathcal{X}$ and $\mathcal{Y}$ with uniform weights, respectively of dimension 2 and 3. Notably, $\mathcal{Y}$ contains outliers points that are significantly distant from the others. The measures supported on these sets have different total masses, with $|\mu| = 11$ and $|\nu| = 10$.

The convergence rate plots are presented on a logarithmic scale and are compared to their theoretically expected rates.
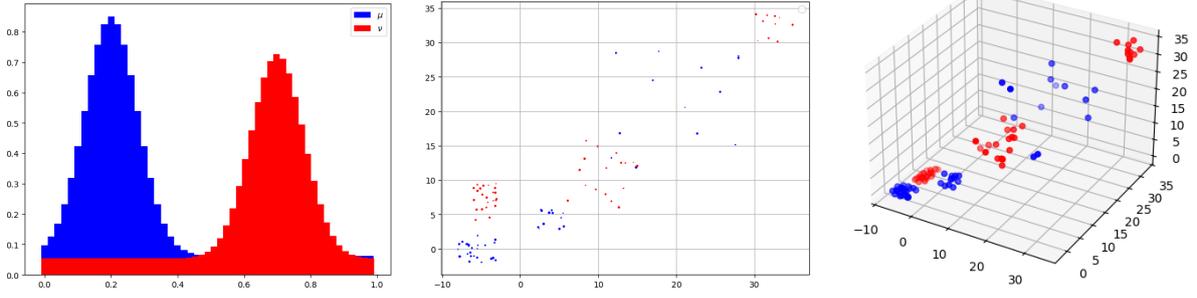
Figure 1: Three Datasets. The left panel shows two discretized Gaussian measures. The center panel shows two weighted point clouds in 2D. The right panel shows two weighted point clouds in 3D.

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

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

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

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

It follows that ϕ^* is a $\mathbb{C}^2$ function on $\mathbb{R}_+^*$ that is strongly convex at every point of its domain.
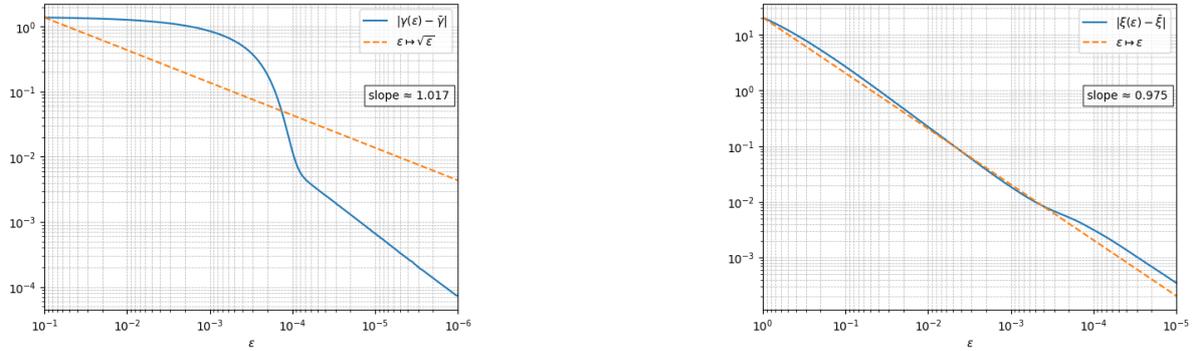
Figure 2: $F(\cdot) = D_\phi(\cdot|q)$ with $\phi(t) = t(\log t - 1)$, μ and ν are discretized Gaussian measures. The left plot (log scale) represents the behavior of $\varepsilon \mapsto \|\gamma(\varepsilon) - \bar{\gamma}\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $\varepsilon \mapsto \|\xi(\varepsilon) - \bar{\xi}\|$ (blue) compared to the expected rate (orange).
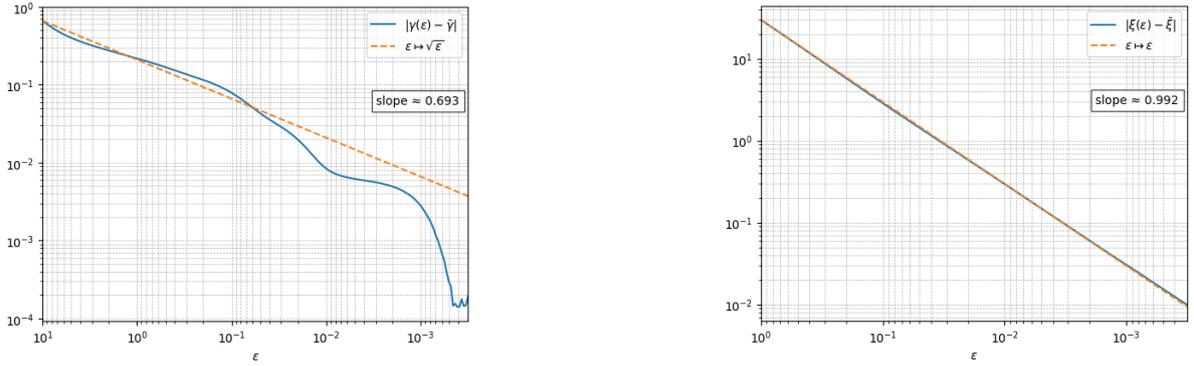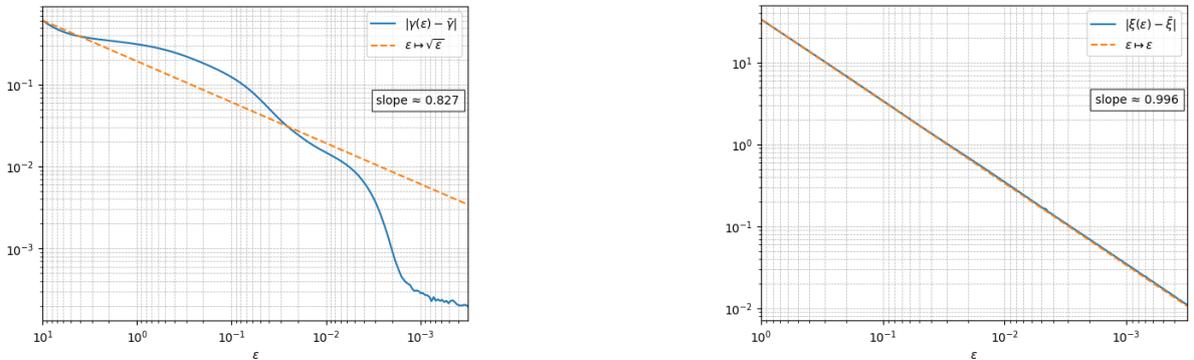
Figure 3: $\mathrm{F}(\cdot) = \tau \mathrm{D}_\phi(\cdot | q)$ with $\phi(t) = t(\log t - 1)$ and $\tau = 10$, μ and ν are measures supported on 2D point clouds. The left plot (log scale) represents the behavior of $\varepsilon \mapsto \|\gamma(\varepsilon) - \bar{\gamma}\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $\varepsilon \mapsto \|\xi(\varepsilon) - \bar{\xi}\|$ (blue) compared to the expected rate (orange).

Figure 4: $\mathrm{F}(\cdot) = \tau \mathrm{D}_\phi(\cdot | q)$ with $\phi(t) = t(\log t - 1)$ and and $\tau = 10$, μ and ν are measures supported on 3D point clouds. The left plot (log scale) represents the behavior of $\varepsilon \mapsto \|\gamma(\varepsilon) - \bar{\gamma}\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $\varepsilon \mapsto \|\xi(\varepsilon) - \bar{\xi}\|$ (blue) compared to the expected rate (orange).

We observe that the theoretical rate appears to be sharp for the dual problem, as indicated by the corresponding slopes when ε is sufficiently small. In contrast, for the primal problem, the observed convergence seems to exceed the theoretical prediction, suggesting a faster rate.

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 $\phi(x) \overset{\text{def}}{=} \frac{1}{2}|x - 1|^2$ 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 given for any $y \in \mathbb{R}$ as:

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

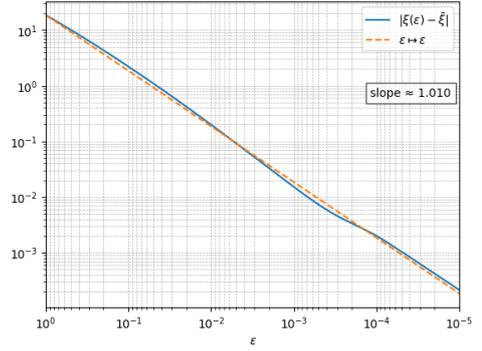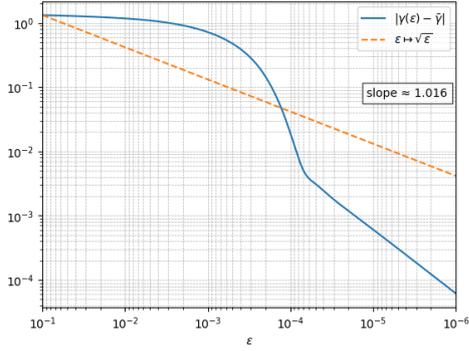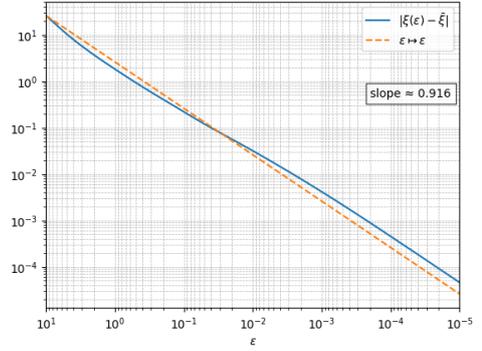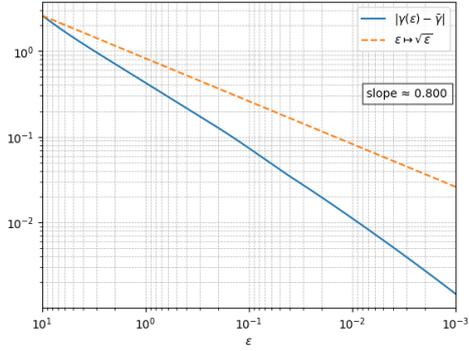which is $\mathcal{C}^2$ and strongly-convex on $\mathbb{R}$.

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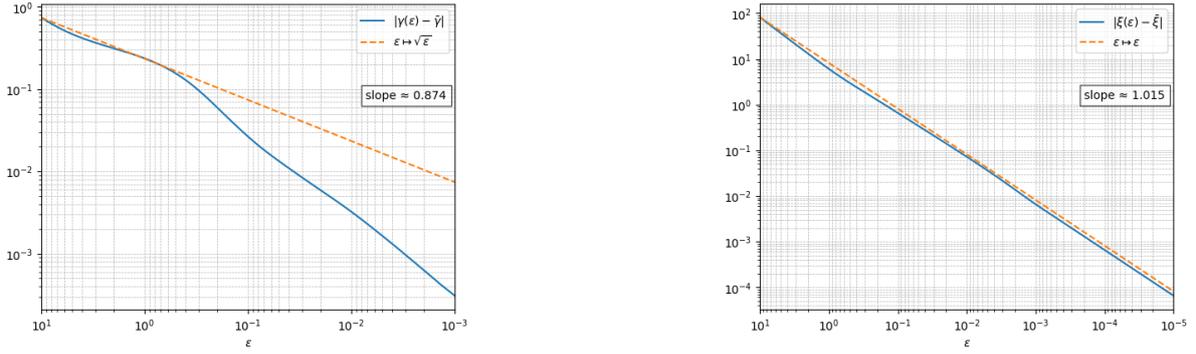

Figure 6: $F(\cdot) = D_\phi(\cdot|q)$ with $\phi(t) = \frac{1}{2}|t-1|^2$, μ and ν are measures supported on 2D point clouds. The left plot (log scale) represents the behavior of $\varepsilon \mapsto \|\gamma(\varepsilon) - \bar{\gamma}\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $\varepsilon \mapsto \|\xi(\varepsilon) - \bar{\xi}\|$ (blue) compared to the expected rate (orange).

Figure 7: $F(\cdot) = D_\phi(\cdot|q)$ with $\phi(t) = \frac{1}{2}|t - 1|^2$, μ and ν are measures supported on 2D point clouds. The left plot (log scale) represents the behavior of $\varepsilon \mapsto \|\gamma(\varepsilon) - \bar{\gamma}\|$ (blue) compared to the expected rate (orange). The right plot (log scale) represents $\varepsilon \mapsto \|\xi(\varepsilon) - \bar{\xi}\|$ (blue) compared to the expected rate (orange).

In this example, we see that the theoretical rate remains sharp for the dual problem, but once again does not accurately describe the behavior of the primal problem.

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.

[Bre11]   H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. New York, NY: Springer, 2011. DOI: 10.1007/978-0-387-70914-7.

[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.

[MS25]   H. Malamut and M. Sylvestre. "Convergence rates of the regularized optimal transport: Disentangling suboptimality and entropy". In: *SIAM Journal on Mathematical Analysis* 57.3 (2025), pp. 2533–2558.

[NP24]   L. Nenna and P. Pegon. "Convergence rate of entropy-regularized multi-marginal 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 Trendső 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.