Entropic approximation of ∞-optimal transport problems

Camilla Brizzi, Guillaume Carlier, Luigi de Pascale

Entropic approximation of ∞-optimal transport problems

Camilla Brizzi * Guillaume Carlier† Luigi De Pascale ‡

February 24, 2023

Abstract

We propose an entropic approximation approach for optimal transportation problems with a supremal cost. We establish Γ-convergence for suitably chosen parameters for the entropic penalization and that this procedure selects ∞-cyclically monotone plans at the limit. We also present some numerical illustrations performed with Sinkhorn's algorithm.

Keywords: ∞-optimal transport, ∞-cyclical monotonicity, entropic approximation.

MS Classification: 49Q22, 65K10.

1 Introduction

The usual Monge-Kantorovich optimal transport problem consists, given a transportation cost and distribution of sources and targets, in finding a transport plan making the average transport cost minimal. It has attracted a considerable amount of attention in the last three decades, as can be seen from the textbooks of Villani [17], [18] and Santambrogio [15]. In optimal transport probems with a supremal cost (also called L^∞ optimal transport), one

*Dipartimento di Matematica e Informatica, Università di Firenze, Viale Morgagni 67/a, 50134 Firenze, Italy, camilla.brizzi@unifi.it

†CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, France and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

‡Dipartimento di Matematica e Informatica, Università di Firenze, Viale Morgagni 67/a, 50134 Firenze, Italy, luigi.depacale@unifi.it

rather looks for transport plans which minimize the essential supremum of the cost. Whereas the usual Monge-Kantorovich problem is linear programming, L^∞ optimal transport leads to non-convex optimization (eventhough the supremal cost has convex sublevel sets), which to a large extent, explains why there are less theoretical results and numerical methods (with the notable exception of the recent combinatorial approach of Bansil and Kitagawa [1]) to address them. As in the Calculus of Variations with a supremal functional, L^∞ optimal transport may admit many minimizers and selecting special ones which satisfy tractable optimality conditions is an important issue, which was studied first by Champion, De Pascale and Juutinen in [6]. In contrast with the classical Monge-Kantorovich problem, where restrictions of optimal plans remain optimal between their marginals, this might be false for L^∞ optimal transport. This is why the authors of [6] have introduced the notion of restrictable optimal and shown that such restrictable solutions are characterized by a remarkable property of ∞-cyclical monotonicity of their support. This was the starting point for the existence of optimal maps for L^∞ optimal transport under various conditions on the cost and the marginals, see [6], [10], [3].

Among numerical methods for optimal transport (see Cuturi and Peyré [14], Benamou [2], Mérigot and Thibert [12]), the entropic penalization approach and the Sinkhorn algorithm have gained a lot of popularity since Cuturi's paper [7]. Entropic optimal transport, which has connections with large deviations and the so-called Schrödinger bridge problem, see Léonard [11] has also stimulated an intensive stream of recent theoretical research, see the lecture notes of Nutz [13] and the references therein. A recent breakthrough in the field is the work of Bernton, Ghosal and Nutz [8] where a large deviations principle, related to cyclical monotonicity is established for entropic optimal plans.

The goal of the present paper is to investigate, theoretically and numerically, whether the entropic approximation strategy can be used for L^∞ optimal transport as well. We will in particular see how the results of [8] can be used to show that this approximation selects at the limit the distinguished restrictable ∞-cyclically monotone minimizers introduced in [6].

The article is organized as follows. The setting is introduced in Section 2. Section 3 is devoted to Γ-convergence towards the supremal cost functional. In Section 4, we study how the entropic penalization selects ∞-cyclically monotone plans in the limit. In Section 5, we give some quantitative convergence estimates and a large deviations upper bound in the spirit of [8]. Finally, we present some numerical illustrations in Section 6.

2 Assumptions and notations

In the sequel, we will always assume that the transportation cost is continuous and nonnegative, $c \in C(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$, and that the fixed marginals of the problem, μ, ν are two Borel probability measures on $\mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, with compact support. Let $\Pi(\mu, \nu)$ be the set of transport plans between μ and ν i.e. the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as marginals. More precisely, a Borel probability measure γ on $\mathbb{R}^d \times \mathbb{R}^d$ belongs to $\Pi(\mu, \nu)$ when

$$\gamma(A \times \mathbb{R}^d) = \mu(A) \text{ and } \gamma(\mathbb{R}^d \times A) = \nu(A),$$

for every Borel subset A of $\mathbb{R}^d$. Note that every γ in $\Pi(\mu, \nu)$ has its support in $\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)$ and that c is uniformly continuous on $\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)$. We are interested in the following ∞-optimal transport problem (see [6]):

$$\inf_{\gamma \in \Pi(\mu,\nu)} \gamma - \operatorname{ess\,sup} c = \|c\|_{L^\infty(\gamma)}. \tag{∞-OT}$$

In contrast with classical optimal transport where one minimizes an integral cost,

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma, \tag{OT}$$

(∞-OT) is a non-convex and presumably harder problem.
Due to the success of entropic approximation of optimal transport with regularization parameter $\varepsilon > 0$

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) d\gamma + \varepsilon H(\gamma | \mu \otimes \nu), \tag{ε-EOT}$$

recalled in the introduction, it seems natural to introduce, for $\varepsilon > 0$ and exponent $p \geq 1$ the following functional $J_{p,\varepsilon} : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$

$$J_{p,\varepsilon}(\gamma) := \begin{cases} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)^p d\gamma(x, y) + \varepsilon H(\gamma | \mu \otimes \nu) \right)^{\frac{1}{p}} & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise,} \end{cases}$$

where H stands for relative entropy:

$$H(\gamma | \mu \otimes \nu) = \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} \log \left(\frac{\mathrm{d}\gamma}{\mathrm{d}\mu \otimes \nu} \right) \mathrm{d}\gamma & \text{if } \gamma \ll \mu \otimes \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

Note that due to strict convexity of the entropy, for every $\varepsilon > 0$ and $p \geq 1$, $J_{p,\varepsilon}$ admits a unique minimizer. We now denote by $J_\infty : \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$, the supremal functional

$$J_\infty(\gamma) := \begin{cases} \gamma - \operatorname{ess\,sup} c & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise.} \end{cases}$$

Finally, let us set

$$J_p := J_{p,1}.$$

Since $H(\gamma|\mu \otimes \nu) \geq 0$ with an equality exactly when $\gamma = \mu \otimes \nu$, $J_{p,\varepsilon}(\gamma) \geq \|c\|_{L^p(\gamma)}$ but also $\|c\|_{L^p(\gamma)} \leq J_\infty(\gamma)$. So, roughly speaking both approximations play in opposite directions: adding the entropic term is an approximation *from above* but approximating $\|c\|_{L^\infty(\gamma)}$ by $\|c\|_{L^p(\gamma)}$ is an approximation *from below*.
We also observe that letting $p \to \infty$ and $\varepsilon \to 0$ is not enough to ensure that minimizers of $J_{p,\varepsilon}$ converge to a minimizer of J_∞ (i.e. a solution of ∞-OT). Indeed, if $\|c\|_\infty \leq \frac{1}{2}$ and $\varepsilon = \frac{1}{p}$ the minimizer γ_p of $J_{p,\frac{1}{p}}$ satisfies

$$H(\gamma_p|\mu \otimes \nu) \leq p 2^{-p}$$

hence γ_p converges (actually strongly by Pinsker's inequality, see e.g. Lemma 2.5 in [16]) to $\mu \otimes \nu$ which in general is not a minimizer of J_∞. On the one hand, this suggests that Γ-convergence of the regularizations above to J_∞ require conditions relating ε to p. On the other hand, in the previous example, we see that the range of c^p compared to the size of the entropic penalization ε is crucial. But the solutions of the ∞-optimal transport problem are invariant when one replaces c by an increasing function of c, in particular one can replace c by $c + 2$ (say) so that c^p will typically dominate the entropic term and one can expect Γ-convergence as $p \to \infty$ for a fixed (or even large) value of ε (see the next section for more details).

3 Γ-convergence

Our first result concerns the Γ-convergence of $J_{p,\varepsilon}$ to J_∞:

Theorem 3.1. *Under the general assumptions of Section 2 we have:*

1. *J_{p,ε_p} Γ-converges (for the weak star topology of $\mathcal{P}(\operatorname{spt}(\mu) \times \operatorname{spt}(\nu))$) to J_∞ as $p \to \infty$ provided $\varepsilon_p^{\frac{1}{p}} \to 0$ as $p \to \infty$,*

2. *if, in addition,* $c \geq 1 + \lambda$ *with* $\lambda \geq 0$*, then* J_{p,ε_p} Γ*-converges to* J_∞ *as* $p \to \infty$ *provided*

$$\lim_{p\to\infty} \frac{1}{p} \log\Big(1 + \varepsilon_p \frac{\log(p)}{(1+\lambda)^p}\Big) = 0. \tag{3.1}$$

In particular, in this case, $J_{p,1}$ *and* $J_{p,p}$ Γ*-converge to* J_∞ *as* $p \to \infty$*.*

Proof. 1. Let $\gamma_p \in \Pi(\mu,\nu)$ converge weakly star to γ. By nonnegativity of $H(\gamma_p|\mu\otimes\nu)$, we have

$$\liminf_p J_{p,\varepsilon_p}(\gamma_p) \geq \liminf_p \|c\|_{L^p(\gamma_p)}.$$

Hence, for fixed q, since $\|c\|_{L^p(\gamma_p)} \geq \|c\|_{L^q(\gamma_p)}$ for $p \geq q$, we have

$$\liminf_p J_{p,\varepsilon_p}(\gamma_p) \geq \liminf_p \|c\|_{L^q(\gamma_p)} = \|c\|_{L^q(\gamma)}$$

taking the supremum with respect to q thus yields the desired Γ-liminf inequality

$$\liminf_p J_{p,\varepsilon_p}(\gamma_p) \geq \|c\|_{L^\infty(\gamma)} = J_\infty(\gamma).$$

Let us now prove the Γ-limsup inequality. For any $\gamma \in \Pi(\mu,\nu)$ we consider γ^δ, the block approximation of γ at scale $\delta \in (0,1)$ defined by (3.3) below, whose convergence to γ is guaranteed by the first inequality in (3.4). By concavity, we first have for $p \geq 1$,

$$\begin{aligned} J_{p,\varepsilon_p}(\gamma^\delta) &\leq \|c\|_{L^p(\gamma^\delta)} + \varepsilon_p^{\frac{1}{p}} H(\gamma^\delta|\mu\otimes\nu)^{\frac{1}{p}} \\ &\leq \|c\|_{L^\infty(\gamma^\delta)} + \varepsilon_p^{\frac{1}{p}} H(\gamma^\delta|\mu\otimes\nu)^{\frac{1}{p}}. \end{aligned}$$

Denoting by ω a modulus of continuity of c on $\mathrm{spt}(\mu)\times\mathrm{spt}(\nu)$, thanks to the first inequality in (3.4), we have

$$\|c\|_{L^\infty(\gamma^\delta)} \leq \|c\|_{L^\infty(\gamma)} + \omega(\sqrt{2d}\delta),$$

being $\sqrt{2d}\delta$ the diameter of the cubes of the approximation. Moreover, by the second inequality in (3.4), we have

$$H(\gamma^\delta|\mu\otimes\nu)^{\frac{1}{p}} \leq d^{\frac{1}{p}} \log(L/\delta)^{\frac{1}{p}}$$

so if we define γ_p as the block approximation of γ at scale $\delta = \frac{1}{p}$ (say), we obtain

$$\limsup_p J_{p,\varepsilon_p}(\gamma_p) \leq J_\infty(\gamma) + \limsup_p \Big(\omega\Big(\frac{\sqrt{2d}}{p}\Big) + d^{\frac{1}{p}} \varepsilon_p^{\frac{1}{p}} \log(Lp)^{\frac{1}{p}}\Big) = J_\infty(\gamma),$$

since we have assumed that $\varepsilon_p^{\frac{1}{p}} \to 0$ as $p \to +\infty$.

2. Let us now assume that $c \geq 1 + \lambda$, the proof of the Γ-liminf inequality for J_{p,ε_p} is exactly as above. For $\gamma \in \Pi(\mu,\nu)$ and γ_p the block approximation of γ at scale $\frac{1}{p}$, we have

$$\begin{aligned} J_{p,\varepsilon_p}(\gamma_p) &\leq \|c\|_{L^\infty(\gamma_p)} \Big(1 + \frac{d\varepsilon_p \log(Lp)}{(1+\lambda)^p}\Big)^{\frac{1}{p}} \\ &\leq \Big(J_\infty(\gamma) + \omega\Big(\frac{\sqrt{2d}}{p}\Big)\Big)\Big(1 + \frac{d\varepsilon_p \log(Lp)}{(1+\lambda)^p}\Big)^{\frac{1}{p}} \end{aligned} \tag{3.2}$$

so that, as soon as (3.1) holds, one has

$$\limsup_p J_{p,\varepsilon_p}(\gamma_p) \leq J_\infty(\gamma).$$

□

Remark 3.2. Notice that in case $c \geq 1 + \lambda$ for some $\lambda > 0$, Γ-convergence of J_{p,ε_p} to J_∞ is guaranteed even for fastly increasing ε_p like $\varepsilon_p = p^m(1+\lambda)^p$ with $m \geq 0$. On the contrary, in the general case, the condition $\varepsilon_p^{\frac{1}{p}} \to 0$ requires to choose values of ε way too small to be used in practice for numerical computations. This suggests in practice to rescale the cost so that it is bounded from below by 1.

Remark 3.3. We observe that in (3.2) it is sufficient that $||c||_{L^\infty(\gamma_p)} \geq 1 + \lambda$, therefore the conclusion of case 2. in Theorem 3.1 remains valid under the weaker assumption that $v_\infty = \min_{\Pi(\mu,\nu)} J_\infty \geq 1 + \lambda$.

For the Γ-limsup inequality, we have used the block approximation introduced in [4], which is defined as follows:

Definition 3.4. *Let $\gamma \in \Pi(\mu,\nu)$. For $\delta > 0$ and $k \in \mathbb{Z}^d$, we denote by Q_k^δ the cube $\delta(k + [0,1)^d)$. The block approximation of γ at scale $\delta \in (0,1)$ is then defined by*

$$\gamma^\delta := \sum_{k,l \in \mathbb{Z}^d \,:\, \mu(Q_k^\delta)>0,\, \nu(Q_l^\delta)>0} \gamma(Q_k^\delta \times Q_l^\delta)\mu_k^\delta \otimes \nu_l^\delta \tag{3.3}$$

where μ_k^δ and ν_l^δ are defined by

$$\mu_k^\delta(A) = \frac{\mu(Q_k^\delta \cap A)}{\mu(Q_k^\delta)},\ \nu_l^\delta(A) = \frac{\nu(Q_l^\delta \cap A)}{\nu(Q_l^\delta)}$$

for every Borel subset A of $\mathbb{R}^d$.

For the sake of completeness, we give a short proof of the properties of the block approximation that we have used in the proof of Theorem 3.1 (see [4] and [5] for related results):

Lemma 3.5. *Let $\gamma \in \Pi(\mu,\nu)$ and γ^δ be the block approximation of γ at scale $\delta \in (0,1)$, then $\gamma^\delta \in \Pi(\mu,\nu)$ and*

$$W_\infty(\gamma^\delta,\gamma) \le \sqrt{2d}\delta,\ H(\gamma^\delta|\mu\otimes\nu) \le d\log\Big(\frac{L}{\delta}\Big), \tag{3.4}$$

where L is a constant depending only on $\mathrm{spt}(\mu)$ *(actually on its diameter).*

Proof. The fact that $\gamma^\delta \in \Pi(\mu,\nu)$ is easy to check by construction (see [4]). Now observe that by (3.3) the density of γ^δ with respect to $\mu\otimes\nu$ is

$$\frac{d\gamma^\delta}{d\mu\otimes\nu}(x,y) = \begin{cases} \frac{\gamma(Q_k^\delta\times Q_l^\delta)}{\mu(Q_k^\delta)\nu(Q_l^\delta)} & \text{if } (x,y)\in Q_k^\delta\times Q_l^\delta, \text{ and } \mu(Q_k^\delta),\,\nu(Q_j^\delta)>0,\\ 0 & \text{otherwise.}\end{cases}$$

Therefore

$$\begin{aligned} H(\gamma^\delta|\mu\otimes\nu) &= \sum_{k,l\in\mathbb{Z}^d\,:\,\mu(Q_k^\delta)>0,\,\nu(Q_l^\delta)>0} \int_{Q_k^\delta\times Q_l^\delta} \log\left(\frac{\gamma(Q_k^\delta\times Q_l^\delta)}{\mu(Q_k^\delta)\nu(Q_l^\delta)}\right) d\gamma^\delta\\ &\le \sum_{k,l\in\mathbb{Z}^d\,:\,\mu(Q_k^\delta)>0,\,\nu(Q_l^\delta)>0} \int_{Q_k^\delta\times Q_l^\delta} \log\left(\frac{1}{\mu(Q_k^\delta)}\right) d\gamma^\delta\\ &= \sum_{k\in\mathbb{Z}^d\,:\,\mu(Q_k^\delta)>0} \mu(Q_k^\delta)\log\left(\frac{1}{\mu(Q_k^\delta)}\right), \end{aligned}$$

where the inequality is due to the fact that $\frac{\gamma(Q_k^\delta\times Q_l^\delta)}{\nu(Q_l^\delta)} \le 1$, while the last equality is obtained summing over l. If $L\ge 1$ is such that $\mathrm{spt}\,\mu$ is contained in a cube of side $L-1$, the number of cubes Q_k^δ with positive μ-measure is not greater than $N_\delta := \big(\frac{L}{\delta}\big)^d$. Therefore, applying Jensen's inequality to the concave function $f(z) = z\log(\frac{1}{z})$, we have

$$\begin{aligned} H(\gamma^\delta|\mu\otimes\nu) &\le \sum_{k=1}^{N_\delta} \mu(Q_k^\delta)\log\left(\frac{1}{\mu(Q_k^\delta)}\right)\\ &\le N_\delta\left(\frac{1}{N_\delta}\sum_{k=1}^{N_\delta}\mu(Q_k^\delta)\log\left(\frac{1}{\sum_{k=1}^{N_\delta}\frac{1}{N_\delta}\mu(Q_k^\delta)}\right)\right)\\ &= \log(N_\delta) = d\log(L) - d\log(\delta), \end{aligned}$$

which proves the second inequality in (3.4).

By construction $\gamma(Q_k^\delta \times Q_l^\delta) = \gamma^\delta(Q_k^\delta \times Q_l^\delta)$, for any k, l. Let J be the set of pairs of indices (k,l) such that $\gamma^\delta(Q_k^\delta \times Q_l^\delta) > 0$ and set $\overline{Q}_j = Q_k^\delta \times Q_l^\delta$, for any $j = (k,l) \in J$. We define

$$\eta^\delta := \sum_{j\,:\,\gamma(\overline{Q}_j)>0} \gamma(\overline{Q}_j)\gamma_j \otimes \gamma_j^\delta,$$

where $\gamma_j(A) := \frac{\gamma(A\cap\overline{Q}_j)}{\gamma(\overline{Q}_j)}$ and $\gamma_j^\delta(A) := \frac{\gamma^\delta(A\cap\overline{Q}_j)}{\gamma^\delta(\overline{Q}_j)}$. By construction $\eta^\delta \in \Pi(\gamma,\gamma^\delta)$, thus

$$W_\infty(\gamma,\gamma^\delta) \leq ||x-y||_{L^\infty(\eta^\delta)} \leq \operatorname{diam}(\overline{Q}_j) = \sqrt{2d}\delta.$$

□

4 Selection of plans with ∞-cyclically monotone support

As shown in [6] and [10], restrictable minimizers of J_∞ are supported on ∞-cyclically monotone sets, such sets are defined as follows:

Definition 4.1. *A set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be ∞-cyclically monotone if we have that*

$$\max_{i=1,\dots,k} \{c(x_i,y_i)\} \leq \max_{i=1,\dots,k} \{c(x_i,y_{i+1})\},$$

for all $k \in \mathbb{N}^*$ and $\{(x_i,y_i)\}_{i=1}^k \subset \Gamma$, where $y_{k+1} = y_1$. A transport plan γ is said to be ∞-cyclically monotone if $\operatorname{spt}\gamma$ is an ∞-cyclically monotone set.

Since every permutation can be obtained as composition of cycles on disjoint sets and trivial cycles on fixed points, one can see that ∞-cyclical monotonicity of a set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is equivalent to the fact that for every $k \in \mathbb{N}^*$, every $\{(x_i,y_i)\}_{i=1}^k \subset \Gamma$ and every $\sigma \in \Sigma(k)$ (where $\Sigma(k)$ is the permutation group of $\{1,\dots,k\}$), one has

$$\max_{i=1,\dots,k} \{c(x_i,y_i)\} \leq \max_{i=1,\dots,k} \left\{c(x_i,y_{\sigma(i)})\right\}.$$

Usually, in the literature, the previous definition is called ∞-c-cylical monotonicity, to keep notations simple, we have omitted the dependence on the cost c; let us remark that ∞-cyclical monotonicity is invariant by replacing c by a strictly increasing transformations of c (like c^p with $p > 0$), contrarily

to the usual notion of c-cyclical monotonicity. We recall that a nonempty subset Γ of $\mathbb{R}^d \times \mathbb{R}^d$ is called c-cyclically monotone when for every $k \in \mathbb{N}^*$, every $(x_i, y_i)_{i=1}^k \subset \Gamma$ and every permutation $\sigma \in \Sigma(k)$, one has

$$\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_i, y_{\sigma(i)}). \tag{4.1}$$

Our goal in this section is to investigate the convergence of the entropic approximation to ∞-cyclically monotone plans. We shall make use of the analysis of the landmark recent article [8]. Let us first recall the notion of (c, ε)-cyclically invariance introduced in [8]:

Definition 4.2. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$ be a measurable function. A coupling $\gamma \in \Pi(\mu, \nu)$ is called (c, ε)-cyclically invariant if $\gamma \ll \mu \otimes \nu$ and its density admits a representative $\frac{d\gamma}{d\mu \otimes \nu} : \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$ such that*

$$\prod_{i=1}^{k} \frac{d\gamma}{d\mu \otimes \nu}(x_i, y_i) = \exp\left(-\frac{1}{\varepsilon}\left[\sum_{i=k}^{k}(c(x_i, y_i) - c(x_i, y_{i+1}))\right]\right) \prod_{i=1}^{k} \frac{d\gamma}{d\mu \otimes \nu}(x_i, y_{i+1}),$$

for all $k \in \mathbb{N}^*$ and $\{(x_i, y_i)\}_{i=1}^k \subset \mathbb{R}^d \times \mathbb{R}^d$, where $y_{k+1} = y_1$.

In [8] (Proposition 2.2), it is shown that whenever (ε-EOT) is finite, the (unique) solution γ_ε of (ε-EOT) is characterized by being (c, ε)-cyclically invariant. The next lemma, which is a part of Lemma 3.1 in [8], provides an estimate for (c, ε)-cyclically invariant couplings, which will be useful for our purpose. For the reader's convenience we provide also here the proof.

Lemma 4.3. *Let $\varepsilon > 0$ and $\gamma_\varepsilon \in \Pi(\mu, \nu)$ be (c, ε)-cyclical invariant. For every fixed $k \geq 2$, $k \in \mathbb{N}$, and $\delta \geq 0$, let $A_{k,c}(\delta)$ be the set defined by*

$$A_{k,c}(\delta) := \left\{ (x_i, y_i)_{i=1}^k \in \left(\mathbb{R}^d \times \mathbb{R}^d\right)^k \,:\, \sum_{i=1}^{k} c(x_i, y_i) - \sum_{i=1}^{k} c(x_i, y_{i+1}) \geq \delta \right\} \tag{4.2}$$

where $y_{k+1} = y_1$. Let $A \subset A_{k,c}(\delta)$ be Borel. Then $\gamma_\varepsilon^k := \prod_{i=1}^k (\gamma_\varepsilon)(dx_i, dy_i)$ satisfies

$$\gamma_\varepsilon^k(A) \leq e^{\frac{-\delta}{\varepsilon}}.$$

Proof. By Definition 4.2 of (c,ε)-cyclical invariance, for γ_ε^k a.e. $(x_i,y_i)_{i=1}^k \in A$ we have that

$$\prod_{i=1}^k \frac{d\gamma_\varepsilon}{\mu\otimes\nu}(x_i,y_i) \le e^{-\frac{\delta}{\varepsilon}} \prod_{i=1}^k \frac{d\gamma_\varepsilon}{\mu\otimes\nu}(x_i,y_{i+1}).$$

In one defines the set $\overline{A} := \{(x_i,y_{i+1}) : (x_i,y_i)\in A\}$, by integrating over A with respect to $\gamma_\varepsilon^k = \prod \gamma_\varepsilon(x_i,y_i) = \prod \gamma_\varepsilon(x_i,y_{i+1})$ we obtain

$$\gamma_\varepsilon^k(A) \le e^{-\frac{\delta}{\varepsilon}}\gamma_\varepsilon^k(\overline{A}) \le e^{-\frac{\delta}{\varepsilon}}.$$

□

The fact that the entropic approximation procedure selects ∞-cyclically monotone plans is then ensured by the following:

Theorem 4.4. *Under the general assumptions of Section 2, further assume that $c>0$ everywhere, and let γ_{p,ε_p} be the minimizer of J_{p,ε_p}. Then, any weak star cluster point γ_∞ as $p\to\infty$ of the family $\{\gamma_{p,\varepsilon_p}\}_{p\ge1}$ is ∞-cyclically monotone, provided*

1. *$\varepsilon_p^{\frac{1}{p}} \to 0$ as $p\to\infty$,*
2. *$\varepsilon_p = o(p(1+\lambda)^p)$ if, in addition, $c\ge 1+\lambda$ with $\lambda\ge 0$.*

Proof. Up to extracting a subsequence, let us assume that γ_{p,ε_p} weakly star converges to γ_∞. We proceed by contradiction assuming that there exist $\delta>0$ and a finite sequence of points $(x_i,y_i)_{i=1}^k$ contained in $\operatorname{spt}\gamma_\infty$, such that

$$\max_{i=1,\dots,k}\{c(x_i,y_i)\} > \max_{i=1,\dots,k}\{c(x_i,y_{i+1})\} + \delta.$$

By the continuity of the cost function c and by the uniform convergence of $\left(\sum_{i=1}^k c(x_i',y_i')^p\right)^{\frac{1}{p}}$ to $\max_{i=1,\dots,k}\{c(x_i',y_i')\}$, as $p\to+\infty$, we deduce that for every $i=1,\dots,k$ there exists an open neighborhood U_i of (x_i,y_i) and $p(\delta)>0$, such that

$$\left(\sum_{i=1}^k c(x_i',y_i')^p\right)^{\frac{1}{p}} > \left(\sum_{i=1}^k c(x_i',y_{i+1}')^p\right)^{\frac{1}{p}} + \delta,$$

for every $(x'_i, y'_i) \in U_i$ (again with the convention that $y'_{k+1} = y'_1$) and $p \geq p(\delta)$. We now observe that

$$\sum_{i=1}^{k} c(x'_i, y'_i)^p > \left(\left(\sum_{i=1}^{k} c(x'_i, y'_{i+1})^p \right)^{\frac{1}{p}} + \delta \right)^p$$
$$\geq \sum_{i=1}^{k} c(x'_i, y'_{i+1})^p + p \left(\sum_{i=1}^{k} c(x'_i, y'_{i+1})^p \right)^{\frac{p-1}{p}} \delta, \tag{4.3}$$

where the last inequality follows from the convexity of $t \mapsto t^p$, with $p > 1$. Since $c > 0$ there exists some $b > 0$ such that $c \geq b$ on each U_i, $i = 1, \ldots, k$, hence, for every $(x'_i, y'_i) \in U_i$ and $p \geq p(\delta)$

$$\sum_{i=1}^{k} c(x'_i, y'_i)^p > \sum_{i=1}^{k} c(x'_i, y'_{i+1})^p + p\delta b^{p-1}. \tag{4.4}$$

We thus have $U_1 \times \cdots \times U_k \subset A_{k,c^p}(p\delta b^{p-1})$, where $A_{k,c^p}(p\delta b^{p-1})$ is defined as in (4.2) with c replaced by c^p. Applying Lemma 4.3, we thus get:

$$\gamma_\infty^k(U_1 \times \cdots \times U_k) := \prod_{i=1}^{k} \gamma_\infty(U_i)$$
$$\leq \liminf_p \gamma_{p,\varepsilon_p}^k(U_1 \times \cdots \times U_k) := \prod_{i=1}^{k} \gamma_{p,\varepsilon_p}(U_i)$$
$$\leq \liminf_p e^{-\frac{p\delta b^{p-1}}{\varepsilon_p}} \tag{4.5}$$

so that if $\varepsilon_p^{\frac{1}{p}} \to 0$ as $p \to \infty$, for large enough p one has $\varepsilon_p \leq b^p$, which yields

$$\liminf_p e^{-\frac{p\delta b^{p-1}}{\varepsilon_p}} = 0.$$

On the other hand, since the points (x_i, y_i) belong to $\operatorname{spt} \gamma_\infty$, we have that $\gamma_\infty^k(U_1 \times \cdots \times U_k) > 0$, which yields the desired contradiction. This shows the first assertion. Now, if $c \geq (1+\lambda)$ with $\lambda \geq 0$, we can replace b by $(1+\lambda)$ in (4.5) and the same conclusion will be reached as soon as $\varepsilon_p = o(p(1+\lambda)^p)$, proving the second assertion.

□

Remark 4.5. Despite what we observed in Remark 3.3 regarding Theorem 3.1, in the proof of the second assertion of Theorem 4.4, it does not seem that the condition $c(x, y) \geq 1$ for every (x, y) can be weakened to $J_\infty \geq 1$. Note also that the condition $\varepsilon_p = o(p(1 + \lambda)^p)$ is stronger than condition (3.1) that guarantees Γ-convergence when $c \geq 1 + \lambda$.

5 Some estimates on the speed of convergence

Our aim in this Section is to give some error estimates for $v_p - v_\infty$ where

$$v_p := \min_{\gamma \in \Pi(\mu,\nu)} J_p \quad \text{and} \quad v_\infty := \min_{\gamma \in \Pi(\mu,\nu)} J_\infty, \tag{5.1}$$

where $J_p := J_{p,1}$ (i.e. for the sake of simplicity we take $\varepsilon_p = 1$ as entropic penalization parameter).

5.1 Upper bounds

Proposition 5.1 (Upper bounds on the speed of convergence)**.** *Let $c \in C^{0,\alpha}(\mathbb{R}^d \times \mathbb{R}^d)$, with $\alpha \in (0,1]$ and let us assume that $v_\infty \geq 1 + \lambda$ for some $\lambda \geq 0$. Then we have*

$$v_p - v_\infty \leq \begin{cases} O(e^{-\beta p}), \text{ with } \beta = \min\{\alpha, \log(1+\lambda)\} & \text{if } \lambda > 0 \\ O\Big(\frac{\log(\log p))}{p}\Big) & \text{if } \lambda = 0. \end{cases}$$

Proof. Let γ_∞ be a minimizer of J_∞ and γ^δ be the block approximation of γ_∞ at scale $\delta \in (0,1)$, as defined in (3.3). We observe that, by construction and by the Hölder condition on c, denoting by A the $C^{0,\alpha}$ semi-norm of c, we first have

$$||c||_{L^\infty(\gamma^\delta)} \leq ||c||_{L^\infty(\gamma_\infty)} + A\delta^\alpha.$$

Then

$$\begin{aligned} v_p &\leq \left(\int c^p d\gamma^\delta + H(\gamma^\delta | \mu \otimes \nu) \right)^{\frac{1}{p}} \leq \left(||c||^p_{L^\infty(\gamma^\delta)} + H(\gamma^\delta | \mu \otimes \nu) \right)^{\frac{1}{p}} \\ &\leq \left(||c||_{L^\infty(\gamma_\infty)} + A\delta^\alpha \right) \left(1 + \frac{H(\gamma^\delta | \mu \otimes \nu)}{(1+\lambda)^p} \right)^{\frac{1}{p}} \\ &\leq (v_\infty + A\delta^\alpha) \left(1 + \frac{d \log(L/\delta)}{(1+\lambda)^p} \right)^{\frac{1}{p}}, \end{aligned} \tag{5.2}$$

where the last inequality follows from Lemma 3.5. For $\lambda > 0$, choosing $\delta := e^{-p}$, (5.2) becomes (setting $C = d \log(L)$)

$$v_p \leq \left(v_\infty + Ae^{-\alpha p} \right) \left(1 + \frac{C + dp}{(1+\lambda)^p} \right)^{\frac{1}{p}},$$

then, we observe that for large p, one has

$$\left(1+\frac{C+dp}{(1+\lambda)^p}\right)^{\frac{1}{p}} = 1+\frac{d}{(1+\lambda)^p}+o\Big(\frac{1}{(1+\lambda)^p}\Big).$$

Therefore, for p large enough,

$$v_p \leq v_\infty + Be^{-\beta p},$$

for some $B>0$ and $\beta = \min\{\alpha, \log(1+\lambda)\}$.

Now if $\lambda = 0$, we choose $\delta = p^{-1/\alpha}$ in (5.2) which gives

$$\begin{aligned} v_p &\leq \left(v_\infty + \frac{A}{p}\right)\exp\left(\frac{1}{p}\log(1+d\log(Lp^{1/\alpha}))\right) \\ &= v_\infty + \frac{1}{\alpha}\frac{v_\infty}{p}\log(\log(p)) + o\Big(\frac{\log(\log(p)}{p}\Big) \end{aligned}$$

which ends the proof. □

5.2 Upper and lower bounds in the discrete case

Let us now consider the discrete case where there exist $x_1, \ldots, x_N$ and $y_1, \ldots, y_M$ points in $\mathbb{R}^d$ such that

$$\mu = \sum_{i=1}^N \mu_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{j=1}^M \nu_j \delta_{y_j} \tag{5.3}$$

with (strictly, without loss of generality) positive weights μ_i and ν_j summing to 1. To shorten notations let us set $c_{ij} = c(x_i, y_j) \geq 0$. In this setting, transport plans γ will simply be denoted as $N \times M$ matrices with entries γ^{ij}. We also recall that in the discrete setting $\Pi(\mu,\nu)$ is a convex polytope and the constraint $\gamma \in \Pi(\mu,\nu)$ is equivalent to

$$\gamma \mathbb{1}_M = \left(\sum_{j=1}^M \gamma^{ij}\right)_i = (\mu_i)_i \text{ and } \gamma^\intercal \mathbb{1}_N = \left(\sum_{i=1}^N \gamma^{ij}\right)_i = (\nu_j)_j.$$

In the discrete setting transport plans have a finite entropy with respect to $\mu \otimes \nu$, with the (crude) bound

$$H(\gamma|\mu\otimes\nu) \leq M := -\sum_{i=1}^N \mu_i \log(\mu_i) - \sum_{j=1}^N \nu_j \log(\nu_j)$$

for every $\gamma \in \Pi(\mu,\nu)$. So if $v_\infty \geq 1+\lambda$ with $\lambda \geq 0$, taking γ_∞ a minimizer of J_∞, we obtain

$$\begin{aligned} v_p \leq J_p(\gamma_\infty) &\leq v_\infty \Big(1+\frac{M}{(1+\lambda)^p}\Big)^{\frac{1}{p}} \\ &\leq v_\infty \Big(1+\frac{M}{p(1+\lambda)^p}+o\Big(\frac{M}{p(1+\lambda)^p}\Big)\Big) \end{aligned}$$

which gives (in a straightforward way, i.e. without using block approximation) an exponentially decaying upper bound for $v_p - v_\infty$ for $\lambda > 0$ and an algebraic upper bound $v_p - v_\infty \leq O(1/p)$ if $\lambda = 0$. The fact that $v_\infty \geq 1$ therefore ensures that $p(v_p - v_\infty)$ is bounded from above. It turns out, that in the discrete setting, this condition also guarantees that we also have an algebraically decaying lower bound for the error. To see this, we first need the following:

Lemma 5.2. *Let μ and ν be discrete measures i.e. of the form (5.3) and define*

$$F_\infty := \{\gamma \in \Pi(\mu,\nu) \ : \ J_\infty(\gamma) = v_\infty\}$$

and for every $\gamma \in F_\infty$,

$$m(\gamma) := \max\{\gamma^{ij} \ : \ \gamma^{ij} > 0,\ c_{ij} = v_\infty\}$$

then there is some $\theta > 0$ such that $m(\gamma) \geq \theta$, for every $\gamma \in F_\infty$.

Proof. Since v_∞ is the minimum of J_∞ over $\Pi(\mu,\nu)$, one can write F_∞ as the set of transport plans for which

$$\gamma^{ij} > 0 \Rightarrow c_{ij} - v_\infty \leq 0$$

or equivalently

$$l(\gamma) := \sum_{ij} \gamma^{ij}(c_{ij} - v_\infty)_+ = 0.$$

In other words, F_∞ is the facet of $\Pi(\mu,\nu)$ where the linear form l (which is nonnegative on $\Pi(\mu,\nu)$) achieves its minimum and it is therefore a convex polytope, whose extreme points belong to the (finite) set of extreme points of $\Pi(\mu,\nu)$. Let us then denote by $\{\gamma_a,\ a \in A\}$ with A a finite index set the set of extreme points of F_∞. Thanks to Minkowski's theorem, we can write any $\gamma \in F_\infty$ as

$$\gamma := \sum_{a \in A} \alpha_a \gamma_a,$$

for some weights $\alpha_a \geq 0$ summing to 1. In particular we may pick $a_0 \in A$ with $\alpha_{a_0} \geq \frac{1}{|A|}$ (with $|A|$ denoting the cardinality of A). Then we have

$$m(\gamma) \geq \frac{m(\gamma_{a_0})}{|A|} \geq \theta := \min_{a \in A} \frac{m(\gamma_a)}{|A|} > 0,$$

where the strict positivity of θ then follows from the fact that A is finite and $m(\gamma_a) > 0$ for every $a \in A$. □

We are now ready to prove the announced lower bound.

Proposition 5.3 (Lower bound on the speed of convergence, discrete case)**.** *Assume that μ and ν are discrete measures i.e. of the form (5.3) and that $v_\infty \geq 1$, then $p(v_p - v_\infty)$ is bounded from below. Hence*

$$v_p - v_\infty = O\Big(\frac{1}{p}\Big).$$

Proof. Let us argue by contradiction and assume that $p(v_p - v_\infty)$ is unbounded from below, then there is a sequence $p_n \to \infty$ as $n \to \infty$ such that

$$\lim_n p_n(v_{p_n} - v_\infty) = -\infty. \tag{5.4}$$

Letting γ_n be the minimizer of J_{p_n}, passing to a subsequence if necessary, we may assume that γ_n converges to some γ_∞ which belongs to F_∞ (as defined in Lemma 5.2) since $v_\infty \geq 1$. In particular, there exists i_0, j_0 such that

$$c_{i_0 j_0} = v_\infty \text{ and } \gamma_\infty^{i_0 j_0} \geq \theta > 0,$$

where θ is the lower bound from Lemma 5.2. Since $\gamma_n^{i_0 j_0}$ converges to $\gamma_\infty^{i_0 j_0}$ we have, for large enough n, $\gamma_n^{i_0 j_0} \geq \frac{\theta}{2}$, hence, using the fact that $c_{i_0 j_0} = v_\infty$ and again the nonnegativity of the entropy

$$\begin{aligned} v_{p_n} &\geq v_\infty \Big(\frac{\theta}{2}\Big)^{\frac{1}{p_n}} = v_\infty \exp\Big(\frac{1}{p_n} \log \frac{\theta}{2}\Big) \\ &\geq v_\infty \Big(1 + \frac{1}{p_n} \log \frac{\theta}{2}\Big) \end{aligned}$$

which is the desired contradiction to (5.4).

□

5.3 A large deviations upper bound

In this (somehow independent) paragraph, our goal is to discuss a (partial) extension of the large deviations results of [8] to the L^∞-optimal transport framework. Considering the Monge-Kantorovich problem (OT) it is well-known (see [9], [15]) that the optimality for (OT) of a plan $\gamma \in \Pi(\mu,\nu)$ is characterized by a property of c-cyclical monotonicity of its support $\Gamma := \mathrm{spt}(\gamma)$, where c-cyclical monotonicity is defined by (4.1). To analyze fine convergence properties of the entropic approximation of (OT), defined by (ε-EOT), assuming convergence (taking a subsequence if necessary) as $\varepsilon \to 0^+$, of the minimizer γ_ε of (ε-EOT) to some γ and denoting by Γ the c-cyclically monotone set $\mathrm{spt}(\gamma)$, the authors of [8] introduced

$$I(x,y) := \sup_{k\geq 2} \sup_{(x_i,y_i)_{i=2}^k \subset \Gamma} \sup_{\sigma\in\Sigma(k)} \Big\{ \sum_{i=1}^k c(x_i,y_i) - \sum_{i=1}^k c(x_i,y_{\sigma(i)}) \Big\},\ (x,y)\in\mathbb{R}^d\times\mathbb{R}^d$$

with $(x_1,y_1)=(x,y)$. They proved that I is a good rate function for the family of optimal entropic plans, $\{\gamma_\varepsilon\}_{\varepsilon>0}$ in the sense that it obeys, under very general conditions, the large deviations principle

$$\begin{aligned} &\limsup_{\varepsilon\to 0} \varepsilon \log(\gamma_\varepsilon(C)) \leq -\inf_{(x,y)\in C} I(x,y) \quad \text{and} \\ &\liminf_{\varepsilon\to 0} \varepsilon \log(\gamma_\varepsilon(U)) \geq -\inf_{(x,y)\in U} I(x,y), \end{aligned}$$

for every compact C and every open U included in $\mathrm{spt}(\mu)\times\mathrm{spt}(\nu)$. Denoting by $\gamma_{p,\varepsilon}$ the minimizer of $J_{p,\varepsilon}$, the results of [8] (using c^p instead of c) of course apply to the convergence of $\gamma_{p,\varepsilon}$ as $\varepsilon\to 0^+$ for a *fixed* exponent p. For L^∞ optimal transport, it makes more sense to rather consider the situation where $\varepsilon>0$ is fixed and p tends to ∞. More precisely, we know from Theorem 4.4, that if $c\geq 1$, $\varepsilon>0$ is fixed, the family $\{\gamma_{p,\varepsilon}\}_{p\geq 1}$ weakly star converges (again possibly after an extraction) to some γ_∞ as $p\to\infty$, $\Gamma_\infty := \mathrm{spt}(\gamma_\infty)$ is ∞-cyclically monotone. In addition to the general assumptions of Section 2, we shall further assume throughout this paragraph that

- $c\geq 1$,
- $\varepsilon>0$ being fixed, the sequence of minimizers $\{\gamma_{p,\varepsilon}\}_{p\geq 1}$ weakly star converges as $p\to\infty$ to some γ_∞, with (∞-cyclically monotone) support Γ_∞.

Let us define for every $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$

$$I_\infty(x,y) := \sup_{k\geq 2} \sup_{(x_i,y_i)_{i=2}^k\subset\Gamma_\infty} \sup_{\sigma\in\Sigma(k)} \Big\{ \max_{1\leq i\leq k}\{c(x_i,y_i)\} - \max_{1\leq i\leq k}\{c(x_i,y_{\sigma(i)})\} \Big\},$$

where $(x_1, y_1) = (x, y)$. Also define

$$\widetilde{I}_\infty(x,y) := \sup_{k\geq 2} \sup_{(x_i,y_i)_{i=2}^k \subset \Gamma_\infty} \Big\{ \max_{1\leq i\leq k}\{c(x_i,y_i)\} - \max_{1\leq i\leq k}\{c(x_i,y_{i+1})\}\Big\},$$

where $(x_1, y_1) = (x, y)$ and $y_{k+1} = y_1$. In our supremal optimal transport setting, we cannot really expect that I_∞ is a good rate function for $\{\gamma_{p,\varepsilon}\}_{p\geq 1}$; indeed, $\operatorname{argmin}_{\Pi(\mu,\nu)} J_\infty$ is unchanged when replacing c with a strictly increasing function of c, while the same does not hold for the function I_∞. However it can be interesting to have a better understanding of the function I_∞, which still provides an upper bound for the family $\{\gamma_{p,\varepsilon}\}$ (see Proposition 5.6).

Lemma 5.4. *Let I_∞ and $\widetilde{I}_\infty$ be defined as above, then*

- *I_∞ and $\widetilde{I}_\infty$ are related by $I_\infty = \max(0, \widetilde{I}_\infty)$,*
- *I_∞ and $\widetilde{I}_\infty$ are lower semicontinuous, $I_\infty \geq 0$, $I_\infty = 0$ on Γ_∞,*
- *I_∞ and $\widetilde{I}_\infty$ coincide on $(\operatorname{spt}(\mu)\times\mathbb{R}^d)\cup(\mathbb{R}^d\times\operatorname{spt}(\nu))$.*

Proof. The fact that $I_\infty \geq \max(0, \widetilde{I}_\infty)$ is obvious as well as the fact that $\widetilde{I}_\infty = 0$ on Γ_∞.
We now prove the converse inequality. Fix now $(x, y) = (x_1, y_1) \in \mathbb{R}^d \times \mathbb{R}^d$, $k \geq 2$, $(x_2, y_2), \dots (x_k, y_k)$ in Γ_∞ and $\sigma \in \Sigma(k)$. We can then partition $\{1, \dots, k\}$ into I_0 the (possibly empty) set of fixed-points of σ and disjoint (empty if σ is the identity) orbits $I_1, \dots, I_l$ on each of which σ is a cycle, this means that for $j = 1, \dots, l$, we may denote $(x_i, y_i)_{i\in I_j}$ as $(\widetilde{x}_r^j, \widetilde{y}_r^j)_{r=1,\dots,|I_j|}$ and $(x_i, y_{\sigma(i)})_{i\in I_j}$ as $(\widetilde{x}_r^j, \widetilde{y}_{r+1}^j)_{r=1,\dots,|I_j|}$ with the convention $\widetilde{y}_{|I_j|+1}^j = \widetilde{y}_1^j$. We now observe that

$$\max_{1\leq i\leq k}\{c(x_i,y_i)\} - \max_{1\leq i\leq k}\{c(x_i,y_{\sigma(i)})\} \leq \max_j \Big\{ \max_{i\in I_j} c(x_i,y_i) - \max_{i\in I_j} c(x_i, y_\sigma(i))\Big\}.$$

where the max with respect to j is taken on indices for which I_j is nonempty. To shorten notations, for such a j let us set

$$\beta_j := \max_{i\in I_j} c(x_i,y_i) - \max_{i\in I_j} c(x_i, y_\sigma(i)).$$

Of course if I_0 is nonempty, $\beta_0 = 0$, now if $j \geq 1$ and I_j is nonempty

$$\beta_j = \max_{r=1,\dots,|I_j|} c(\widetilde{x}_r^j, \widetilde{y}_r^j) - \max_{r=1,\dots,|I_j|} c(\widetilde{x}_r^j, \widetilde{y}_{r+1}^j) \leq \widetilde{I}_\infty(\widetilde{x}_1^j, \widetilde{y}_1^j).$$

So, if $(\widetilde{x}_1^j, \widetilde{y}_1^j) = (x_1, y_1)$, $\beta_j \leq \widetilde{I}_\infty(x, y)$ and if $(\widetilde{x}_1^j, \widetilde{y}_1^j) \neq (x_1, y_1)$, then $(\widetilde{x}_1^j, \widetilde{y}_1^j) \in \Gamma_\infty$, hence $\widetilde{I}_\infty(\widetilde{x}_1^j, \widetilde{y}_1^j) = 0$ by the definition of $\widetilde{I}_\infty$ and the fact that

Γ_∞ is ∞-cyclically monotone. In other words, we can bound from above each β_j by $\max(0, \widetilde{I}_\infty(x,y))$. Taking suprema with respect to k, $(x_2,y_2),\dots(x_k,y_k)$ in Γ_∞ and $\sigma \in \Sigma(k)$, we thus get $I_\infty \le \max(0,\widetilde{I}_\infty)$. Moreover, since $\widetilde{I}_\infty \le 0$ on Γ_∞, $I_\infty = \max(0,\widetilde{I}_\infty) = 0$ on Γ_∞

Lower semi continuity of I_∞ and $\widetilde{I}_\infty$ follows from the continuity of c. Finally assume that $x \in \mathrm{spt}(\mu)$ and $y \in \mathbb{R}^d$, since $\Gamma_\infty = \mathrm{spt}(\gamma_\infty)$ is compact and $\gamma_\infty \in \Pi(\mu,\nu)$, there exists $y' \in \mathbb{R}^d$ such that $(x,y') \in \Gamma_\infty$. Taking $(x_1,y_1) = (x,y)$, $(x_2,y_2) = (x,y')$ as a competitor in the definition of $\widetilde{I}_\infty(x,y)$ we see that $\widetilde{I}_\infty(x,y) \ge 0$ hence $I_\infty(x,y) = \widetilde{I}_\infty(x,y)$. The same argument shows that I_∞ and $\widetilde{I}_\infty$ coincide on $\mathbb{R}^d \times \mathrm{spt}(\nu)$. □

Lemma 5.5. *Let us fix $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$. Suppose that for some $\delta \in \mathbb{R}$, $k \in \mathbb{N}$, $k \ge 2$ and $(x_i,y_i)_{i=2}^k \subset \mathrm{spt}\,\gamma_\infty$, we have*

$$\max_{1\le i\le k}\{c(x_i,y_i)\} - \max_{1\le i\le k}\{c(x_i,y_{i+1})\} > \delta, \text{ where } (x_1,y_1) := (x,y).$$

Then there exist $\alpha > 0$, $r > 0$ and $p_0 \ge 1$ such that

$$\gamma_{p,\varepsilon}(B_r(x,y)) \le \alpha e^{\frac{-p\delta}{\varepsilon}}, \ \forall p \ge p_0,$$

where $\gamma_{p,\varepsilon}$ is the minimizer of $J_{p,\varepsilon}$.

Proof. Of course if $\delta \le 0$, one can just take $\alpha = 1$ so we may assume that $\delta > 0$. Reasoning as in the proof of Theorem 4.4 (recall that we have assumed $c \ge 1$), we know that there exist p_0 and $r > 0$ such that

$$\sum_{i=1}^k c^p(x_i',y_i') - \sum_{i=1}^k c^p(x_i',y_{i+1}') > p\delta,$$

for every $p \ge p_0$ and $(x_i',y_i')_{i=1}^k \subset B_r(x_1,y_1)\times\cdots\times B_r(x_k,y_k)$. Then $B_r(x_1,y_1)\times\cdots\times B_r(x_k,y_k) \subset A_{k,c^p}(p\delta)$ so, thanks to Lemma 4.3,

$$\gamma_{p,\varepsilon}^k(B_r(x_1,y_1)\times\cdots\times B_r(x_k,y_k)) \le e^{-\frac{p\delta}{\varepsilon}}.$$

Moreover $\liminf_{p\to\infty}\gamma_{p,\varepsilon}(B_r(x_i,y_i)) \ge \gamma_\infty(B_r(x_i,y_i)) > \beta$, for all $2 \le i \le k$, for some $\beta > 0$ since $(x_i,y_i)_{i=2}^k \subset \mathrm{spt}\,\gamma_\infty$, then

$$\gamma_{p,\varepsilon}(B_r(x,y)) \le \left(\frac{\beta}{2}\right)^{1-k} e^{-\frac{p\delta}{\varepsilon}},$$

for all $p \ge p_0$ (possibly replacing p_0 with a larger one). □

Proposition 5.6. *Under the assumptions of this paragraph, for any compact set $C \subset \mathbb{R}^d \times \mathbb{R}^d$, one has*

$$\limsup_{p\to\infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(C) \leq - \inf_{C\cap(\mathrm{spt}(\mu)\times\mathrm{spt}(\nu))} \widetilde{I}_\infty \leq - \inf_C I_\infty.$$

Proof. First note that since $\gamma_{p,\varepsilon}$ is supported on $\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)$,

$$\gamma_{p,\varepsilon}(C) = \gamma_{p,\varepsilon}(C \cap (\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)))$$

and there is noting to prove if C is disjoint from $\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)$. Therefore we can assume that $C \cap (\mathrm{spt}(\mu) \times \mathrm{spt}(\nu)) \neq \emptyset$. It then follows from Lemma 5.4 that

$$\inf_{C\cap(\mathrm{spt}(\mu)\times\mathrm{spt}(\nu))} \widetilde{I}_\infty = \inf_{C\cap(\mathrm{spt}(\mu)\times\mathrm{spt}(\nu))} I_\infty \geq \inf_C I_\infty.$$

Now let $\eta > 0$ and $(x, y) \in C \cap (\mathrm{spt}(\mu) \times \mathrm{spt}(\nu))$. By definition of $\widetilde{I}_\infty(x, y)$ there exist $k \geq 2$ and $(x_i, y_i)_{i=2}^k \subset \Gamma_\infty$, such that (setting as usual $(x_1, y_1) = (x, y)$ and $y_{k+1} = y$)

$$\max_{1\leq i\leq k}\{c(x_i, y_i)\} - \max_{1\leq i\leq k}\{c(x_i, y_{i+1})\} > \min(\eta^{-1}, \widetilde{I}_\infty(x, y)) - \eta.$$

Note that the truncation is used to handle the case where $\widetilde{I}_\infty(x, y) = +\infty$. By Lemma 5.5 we know that there exist $\alpha, r > 0$ such that

$$\gamma_{p,\varepsilon}(B_r(x, y)) \leq \alpha \exp\left(\frac{-p(\min(\eta^{-1}, \widetilde{I}_\infty(x, y)) - \eta)}{\varepsilon}\right).$$

Then

$$\limsup_{p\to\infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(B_r(x, y)) \leq - \min(\eta^{-1}, \widetilde{I}_\infty(x, y)) + \eta$$

and, by compactness of C,

$$\limsup_{p\to\infty} \frac{\varepsilon}{p} \log \gamma_{p,\varepsilon}(C) \leq - \inf_{C\cap(\mathrm{spt}(\mu)\times\mathrm{spt}(\nu))} \min(\eta^{-1}, \widetilde{I}_\infty) + \eta$$

which, letting $\eta \to 0^+$, yields the desired upper bound. □

6 Numerical results

In this section, we present several numerical examples, with the aim of illustrating the discussions and theoretical analysis of the previous sections. We shall consider discrete marginals; let $N, M \in \mathbb{N}$, with a slight abuse of notation, we will denote by μ and ν both the measures and the vectors of weights

$(\mu_i)_{i=1}^N$ and $(\nu_j)_{j=1}^M$ and γ will denote both the transport plan and the $N \times M$ matrix (γ^{ij}). For fixed $p, \varepsilon > 0$, in this discrete setting, the minimization of $J_{p,\varepsilon}$ reads

$$\min_{\Pi(\mu,\nu)} \left(\sum_{i,j} \gamma^{ij} c_{ij}^p + \varepsilon \sum_{ij} \gamma^{ij} \log\left(\frac{\gamma^{ij}}{\mu_i \nu_j}\right) \right)^{\frac{1}{p}}. \tag{6.1}$$

Raising the above cost to the power p, which does not change the minimizer, leads to a standard entropic transport problem. For such problems, we used in all our examples Sinkhorn's algorithm (see for instance Chapter 4 in [14]) to find a good approximation (with error smaller than 10^{-5}) of the solution.

If $v_\infty \geq 1$, in light of Theorem 3.1, we expect the output γ of the Sinkhorn algorithm to be, for suitable p and ε, also a good approximation of an optimal plan for the discretized L^∞- optimal transport problem

$$v_\infty := \min_{\gamma \in \Pi(\mu,\nu)} \max_{i,j} \left\{ c_{i,j} \,:\, \gamma^{ij} \neq 0 \right\}.$$

Furthermore, if $c \geq 1$, thanks to Theorem 4.4, we expect to find a plan close to an ∞-cyclically monotone one.

Remark 6.1. As the set of transport plans $\Pi(\mu, \nu)$ is a convex polytope, for any $\gamma \in \Pi(\mu, \nu)$ there exists a finite set of indices S, such that $\gamma = \sum_{s \in S} a_s \gamma_s$, with $a_s > 0$, $\sum a_s = 1$ and γ_s an extreme point of $\Pi(\mu, \nu)$. If $N = M$ and $\mu_i = \nu_j = \frac{1}{N}$, the set $\Pi(\mu, \nu)$ is the set of the so-called *bi-stochastic* matrices, whose extreme points, by Birkhoff's theorem, form the set of pemutation matrices. We observe that, by definition of $\gamma - \operatorname{ess\,sup}$, $J_\infty(\gamma) = \max_{s \in S} J_\infty(\gamma_s)$ and thus the minimum of J_∞ is attained at some permutation matrix. Therefore, if $N = M$ and $\mu_i = \nu_j = \frac{1}{N}$

$$v_\infty = \min_{\sigma \in \Sigma(N)} \max_i c_{i,\sigma(i)}.$$

This can be in principle used to compute v_∞ exactly. However this is not particularly useful in practice; regarding for instance the example on bottom of Figure 4, even if the size of μ and ν is the same, in order to calculate the exact value of v_∞ we should be able to perform 100! evaluations, which is infeasible in practice!

All the examples in this section, will be in dimension $d = 2$, μ will be represented by blue points, ν by red points and the plan will be represented by arrows: the black ones indicate that a blue point is sent to a red point with high probability, while the gray ones indicate that a blue point is sent to a red point with lower probability (but still not negligible).

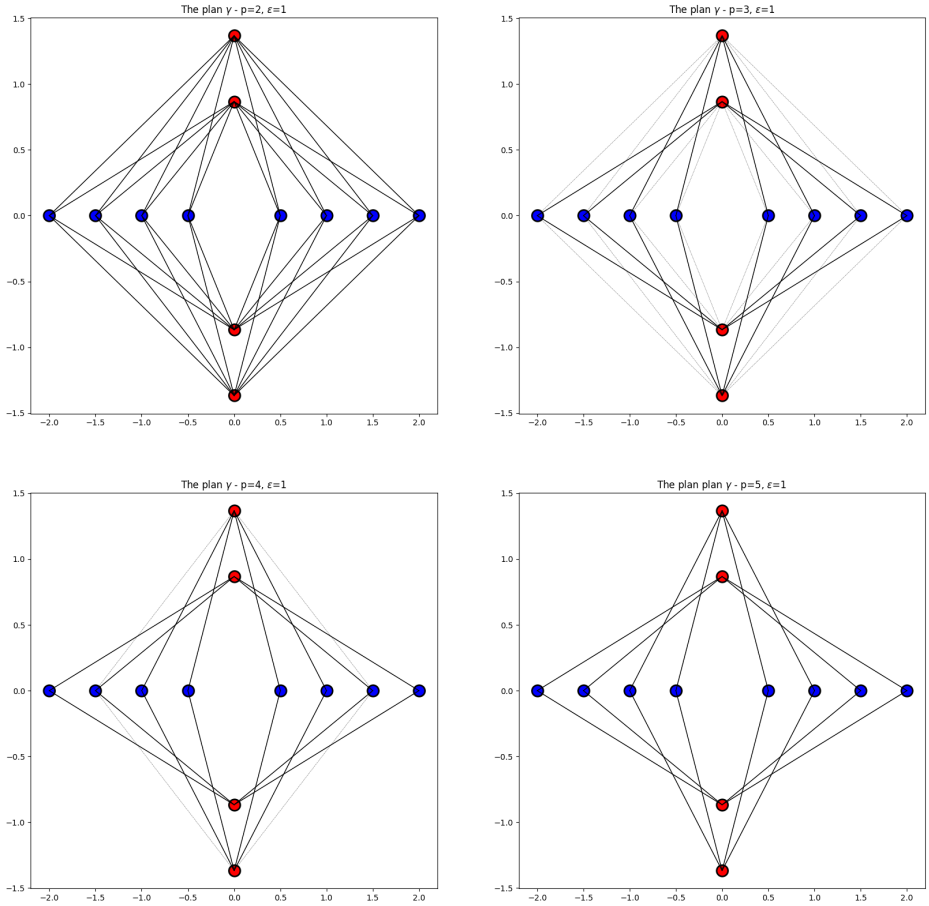

Figure 1: Example of convergence of the plan to the ∞-cm plan: $c(x, y) = |x - y|^p$, $p \in \{2, 3, 4, 5\}$, $\varepsilon = 1$ and μ and ν having orthogonal supports.

In the first example, as shown by Figure 1, we consider $c^p = |x - y|^p$, for $p \in \{2, 3, 4, 5\}$, μ which is uniformly concentrated on the blue points

$$\{(-2, 0), (-1.5, 0), (-1, 0), (-0.5, 0), (0.5, 0), (1, 0), (1.5, 0), (2, 0)\}$$

and ν on the red points

$$\{(0, -1.367), (0, -0.867), (0, 867), (0, 1.367)\}.$$

Note that with this choice of $\operatorname{spt}\mu$ and $\operatorname{spt}\nu$, $c \geq 1$ everywhere and therefore, thanks to Theorem 3.1 and Theorem 4.4, Γ-convergence and convergence of the outputs towards ∞-cm plans still hold choosing $\varepsilon = 1$. We observe that for $p = 2$, every transport plan γ is optimal. Indeed, by the orthogonality of the two supports, any plan is concentrated on a cyclically monotone set (see

(4.1)) and, as recalled in Section 5.3 (see for instance [9, 15]), this is a sufficient optimality condition. Here, since we look for a plan which minimizes the regularized problem which involves the entropy, the Sinkhorn algorithm selects the most diffuse one, as evidenced by the picture on the upper left of Figure 1. The other three pictures in Figure 1 show that convergence towards an ∞-cm plan is really fast and it occurs already for $p = 5$.

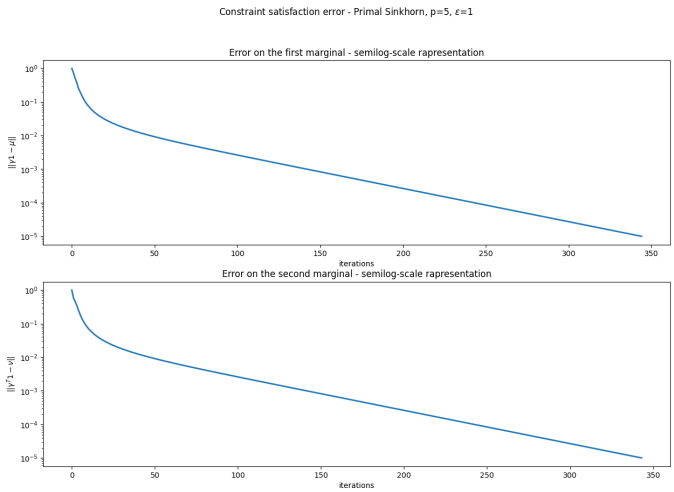

Figure 2: Error on the marginals: the first image shows the error $|\gamma \mathbb{1}_4 - \mu|$ of the output γ on the first marginal and the second one the error $|\gamma^\intercal \mathbb{1}_8 - \nu|$ on the second marginal.

Regarding the accuracy, Figure 2 shows that for $p = 5$ and $\varepsilon = 1$ the distance $|\gamma \mathbb{1}_4 - \mu|$ between the first marginal of the output γ and the distance $|\gamma^\intercal \mathbb{1}_8 - \nu|$ between the second marginal of γ and ν is of the order of 10^{-5} after only 350 iterations.

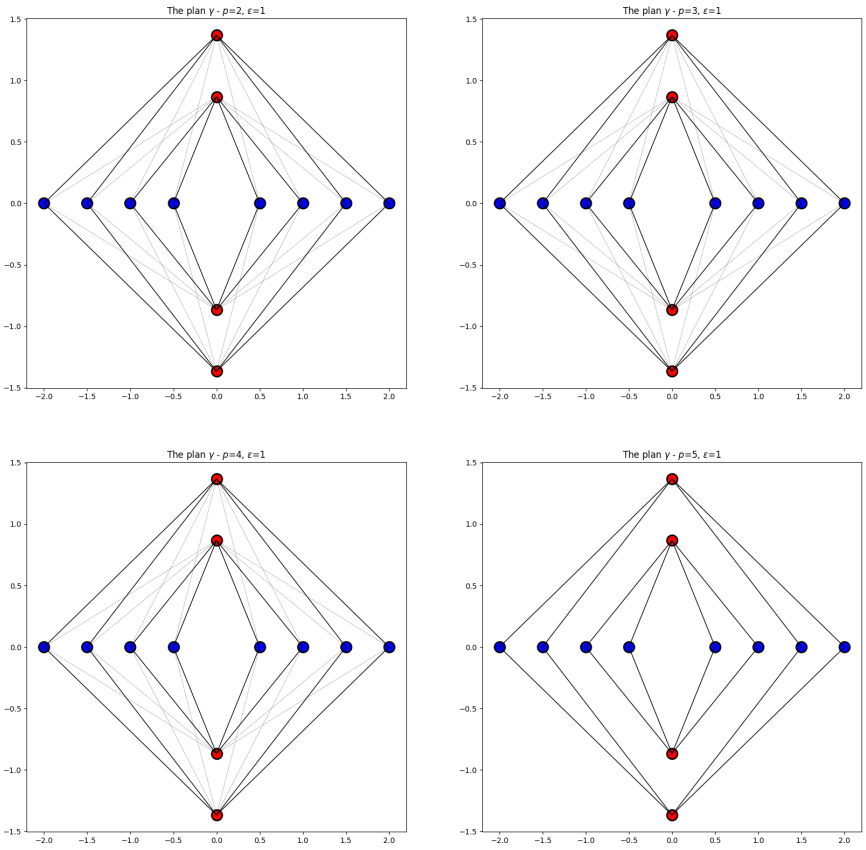

Figure 3: Example of convergence of the plan to the ∞-cm plan for $c(x,y) = (\max\{|x_1 - y_1|, |x_2 - y_2|\})^p$, for $p \in \{2,3,4,5\}$, $\varepsilon = 1$ and μ and ν having orthogonal supports.

We have also considered the same example (see Figure 3) with the cost function $c^p(x,y) := (\max\{|x_1 - y_1|, |x_2 - y_2|\})^p$. In this case the convergence is still fast and the error is small after few iterations (of order 10^{-5} after about 180 iterations).

Remark 6.2. When $c > 1$, on the one hand, we don't need ε to be small and we can even take it large as p grows (by case 2. in Theorem 3.1 we can even choose for instance $\varepsilon_p = (1+\lambda)^p$). On the other hand, we can encounter some difficulties when computing the Gibbs kernel $K_{ij} = e^{-\frac{c_{ij}^p}{\varepsilon}}$: if p is large it can happen that, for some i, j, $K_{i,j} = 0$ making impossible to perform the division in the iterations of the primal version of the Sinkhorn algorithm. Fortunately, this problem can be overcome using the Log-Domain version (see for instance Section 4.4 in [14]), as we did in the following example, represented by Figure 4.

Figure 4, which shows a comparison among three different examples, considered for $p = 2$ on the left and for $p = 15$ on the right and $\varepsilon = 1$. The two pictures on top in Figure 4 show the representation by arrows of the output when μ is uniformly concentrated on 400 points which discretize the unitary square and ν is uniformly concentrated on the points $(1, 2)$ and $(2, 1)$. This is a discretization of the case μ uniform on the square $[0, 1]^2$, where (see also Example 2.2 in [6]) every $\gamma \in \Pi(\mu, \nu)$ is optimal for the problem

$$\inf_{\gamma \in \Pi(\mu,\nu)} \gamma - \operatorname{ess\,sup} c = \|c\|_{L^\infty(\gamma)}.$$

Indeed

$$\begin{aligned} \|c\|_{L^\infty(\gamma)} &= \sup\{c(x, y) \,:\, (x, y) \in \operatorname{spt} \gamma\} \\ &= |(0, 0) - (1, 2)| = |(0, 0) - (2, 1)| = \sqrt{5}, \end{aligned}$$

for every $\gamma \in \Pi(\mu, \nu)$. Since every plan is optimal, when p is smaller, as shown in the picture on the left, the role of the entropy is more important and the algorithm selects the most diffuse plan. While increasing the value of p the entropy becomes more and more negligible and output becomes sparser: already for $p = 15$ (on the right) the output is a good approximation of the ∞-cyclically monotone plan, which in this case is unique (see Theorem 5.6 in [6]). A small variation, represented by the two figures in the middle, is to consider ν which is not uniformly concentrated on the points $(1, 2)$ and $(2, 1)$. Here we have taken $\nu = 0.1\delta_{(1,2)} + 0.9\delta_{(2,1)}$. Finally, on the bottom, we have implemented the case in which also ν is the discretization of an absolutely continuous measure. Here μ approximates the square $[-0.25, 0.25] \times [-0.25, 0.25]$ and ν the rectangle $[1.25, 1.5] \times [-0.5, 0.5]$ and both measures are supported on 100 points. As previously, one can notice that for $p = 2$ the entropy plays an important role and the algorithm selects the most diffuse plan, while, already for $p = 15$ the plan is considerably sparser.

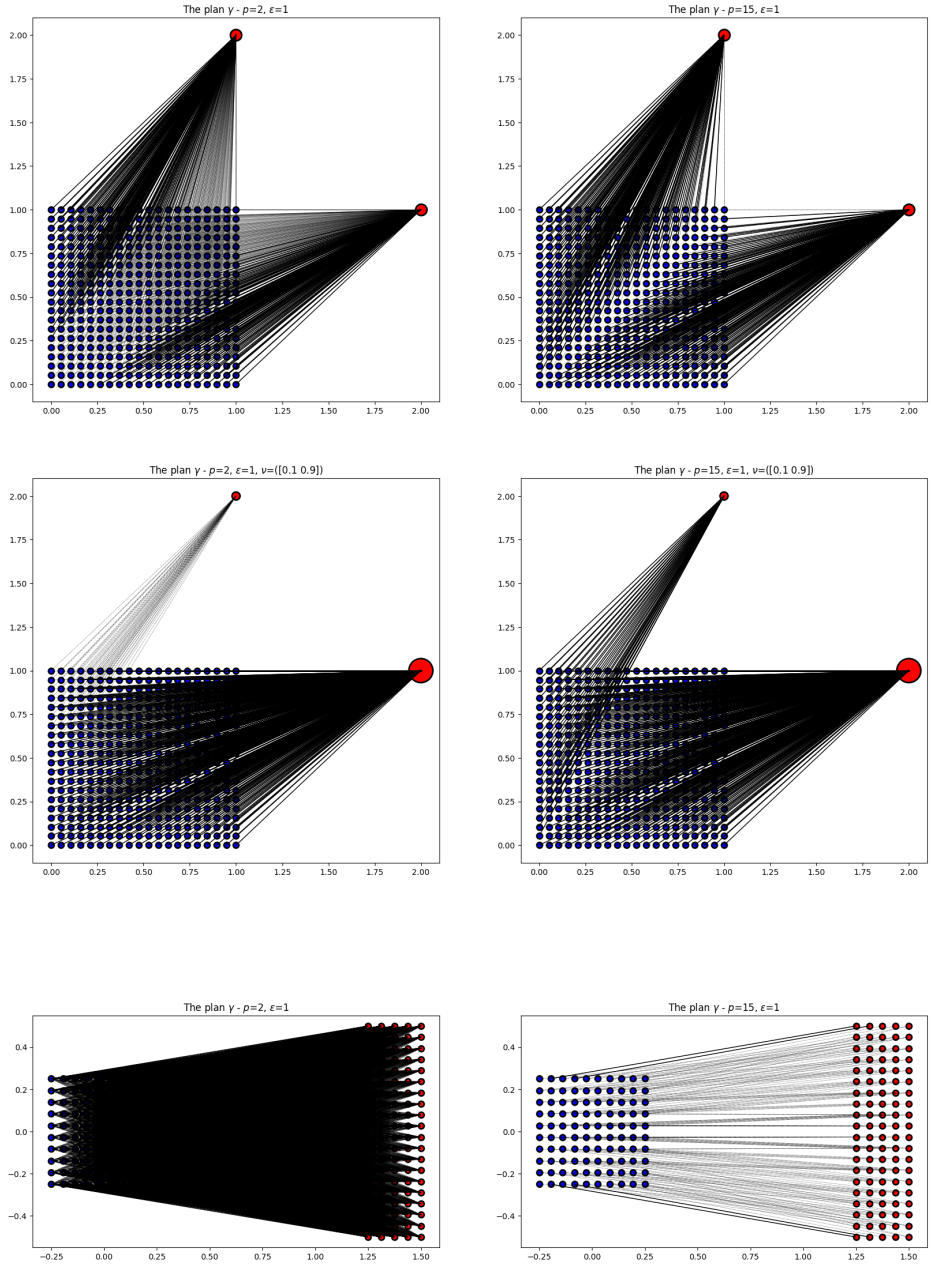

Figure 4: Comparison among three different examples: $\varepsilon = 1$, $c(x, y) = |x - y|^p$, on the left $p = 2$, on the right $p = 15$. On top: μ a uniform discretization of the unitary square and ν uniformly concentrated on the points $(1, 2)$ and $(2, 1)$. In the middle: μ the same discretization of the unitary square, $\nu = 0.1\delta_{(1,2)} + 0.9\delta_{(2,1)}$ (the point $(2, 1)$ is represented by a bigger dot). On bottom: μ a uniform discretization of the square $[-0.25, 0.25] \times [-0.25, 0.25]$ and ν of the rectangle $[1.25, 1.5] \times [-0.5, 0.5]$.

We are now interested in the asymptotic behavior of $v_p := \min_{\Pi(\mu,\nu)} J_p$ and we want to numerically represent the upper and lower bounds on the speed of convergence of v_p towards $v_\infty := \min_{\Pi(\mu,\nu)} J_\infty$ proved in Proposition 5.1 and Proposition 5.3. In order to apply Proposition 5.1 and Proposition 5.3 it is enough to assume a lower bound on v_∞ and not a pointwise one on c.

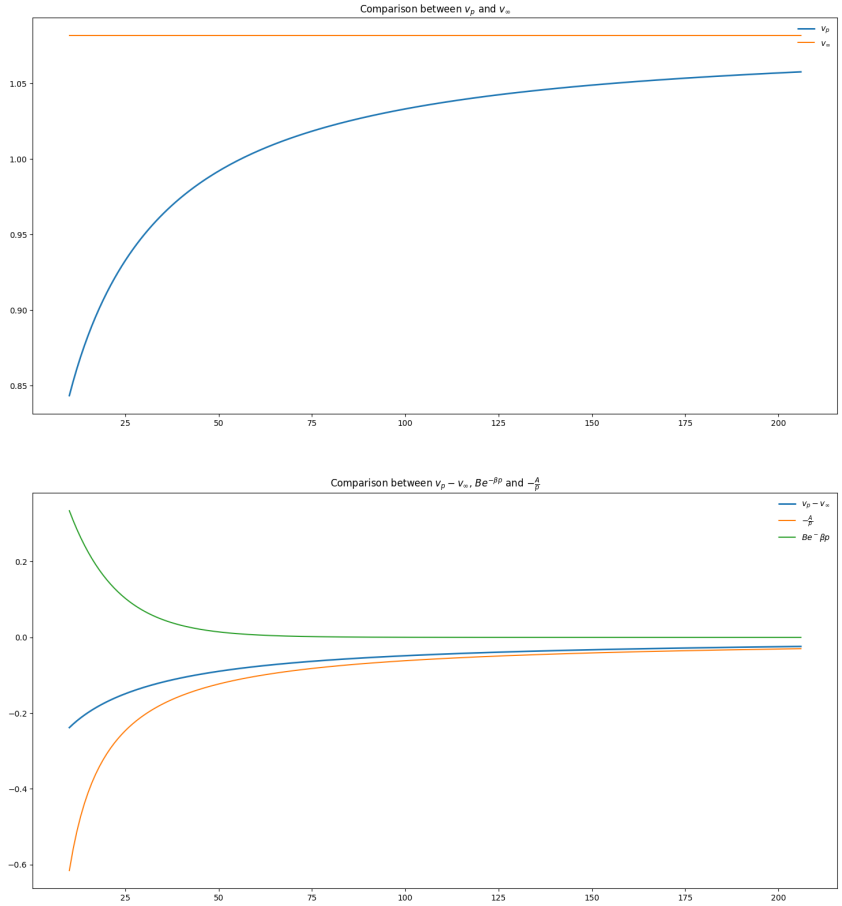

Figure 5: Comparison among the speed of convergence of $v_p - v_\infty$, $Be^{-\beta p}$ and $-\frac{A}{p}$ for $p \in [10, 206]$, μ and ν as the ones on top of Figure 4. On top: v_p in blue and v_∞ in orange. On bottom: $Be^{-\beta p}$ in green, $-\frac{A}{p}$ in orange and $v_p - v_\infty$ in blue. Here A, B are obtained by linear regression (least squares) and $\beta = \log(v_\infty)$ (the same β as in Proposition 5.1).

Figure 5 provides an example of the asymptotic behavior of v_p and of the speed of convergence in the case of μ and ν as the ones represented in the two pictures on top of Figure 4. In light of what we have just remarked, we

have re-scaled the cost c in order to have $v_\infty \simeq 1.08166$. For $p \in [10, 206]$ the image on top of Figure 5 shows in blue how v_p changes varying p, while v_∞ is constant and is represented by the orange line. On bottom of Figure 5 we have represented in blue $v_p - v_\infty$, in green the upper bound $Be^{-\beta p}$ and in orange the lower bound $-\frac{A}{p}$, where $\beta = \log(v_\infty)$ by Proposition 5.1 (indeed in this case c is Lipschitz so $\alpha = 1 > \log(v_\infty)$) and A, B have been estimated by a linear regression method (by least squares).

Finally, an example in which it is possible (even if it is really slow!) to compute v_∞ exactly (see Remark 6.1) is represented in Figure 6. Here μ is concentrated on 8 points, given by

$$\{(x_1, x_2) \,:\, x_1 = -0.25 + 0.125 \cdot i,\, i = 1, \dots, 4,\; x_2 \in \{-0.1, 0.1\}\}$$

and ν is concentrated on 8 equidistant points of the segment starting from the point $(0.625, 1.25)$ to the point $(1.25, 0)$ of the line $y_2 = -2y_1 + 2.5$. We have computed v_∞ for the cost $c(x, y) = |x - y|$ applying Remark 6.1, and we have obtained that $v_\infty \simeq 1.38647347$ and that the points which are at the minimal-maximal distance are $x_* = (-0.25, -0.1)$ and $y_* = (0.98214286, 0.53571429)$, connected by the purple segment in the picture. Regarding the speed of convergence we rescaled the cost in order to decrease further $v_\infty \simeq 1.052460609$. As shown in Figure 7, v_p is calculated varying p in the interval $[10, 172]$, with $\varepsilon = 500^2$. We observe that in this case, as shown in the picture on top, v_p is initially smaller than v_∞, then it increases becoming greater and finally it starts decreasing converging to v_∞.

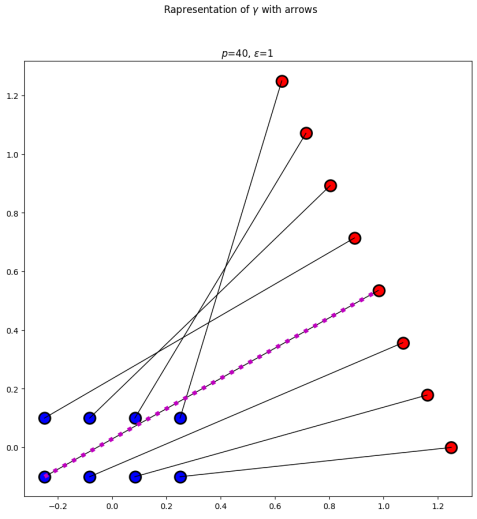

Figure 6: μ and ν uniformly distributed both concentrated on 8 points. The value of v_∞ is about 1.38647347 and it is obtained transporting mass between the two points connected by the yellow segment.

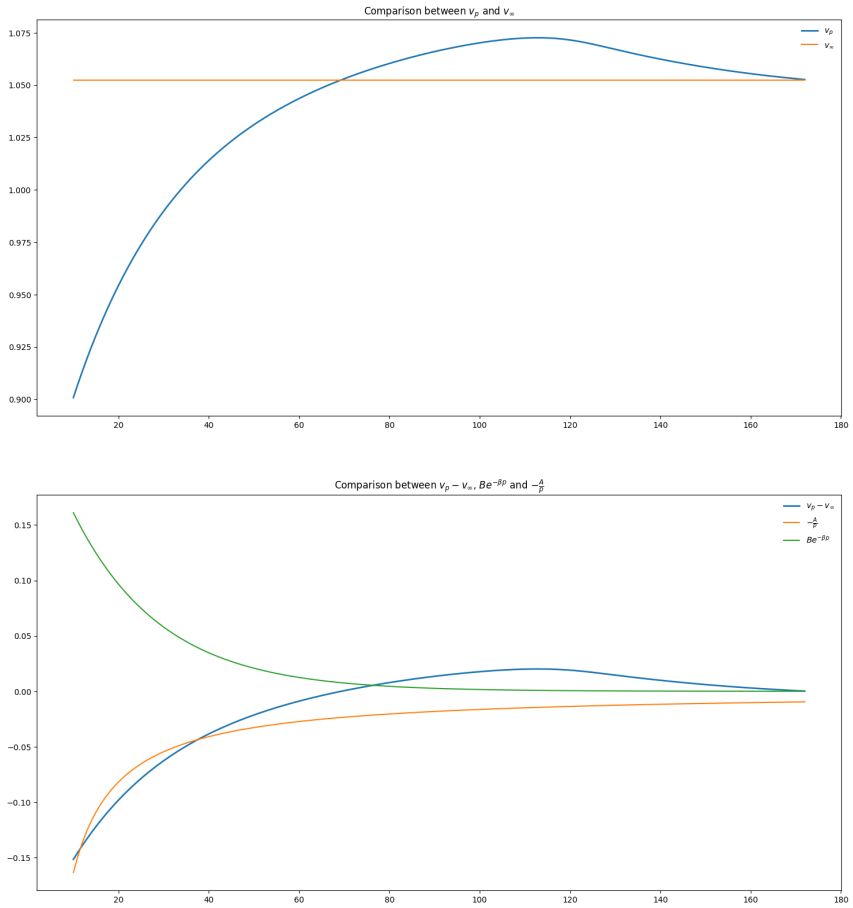

Figure 7: Comparison among the speed of convergence of $v_p - v_\infty$, $Be^{-\beta p}$ and $-\frac{A}{p}$ for $p \in [10, 172]$ and $\varepsilon = 500^2$, μ and ν as the ones in Figure 6. On top: v_p in blue and v_∞ in orange. On bottom: $Be^{-\beta p}$ in green, $-\frac{A}{p}$ in orange and $v_p - v_\infty$ in blue. Here A, B are obtained by linear regression (least squares) and $\beta = \log(v_\infty)$ (the same β as in Proposition 5.1).

Acknowledgments: G.C. acknowledges the support of the Lagrange Mathematics and Computing Research Center. The research of C.B. and L.DP is partially financed by the *"Fondi di ricerca di ateneo, ex 60 %"* of the University of Firenze and is part of the project *"Alcuni problemi di trasporto ottimo ed applicazioni"* of the GNAMPA-INDAM. C.B. and G.C. also acknowledge the support of the French Agence Nationale de la Recherche through the project MAGA (ANR-16-CE40-0014).

References

[1] Mohit Bansil and Jun Kitagawa. $\mathcal{W}_\infty$-transport with discrete target as a combinatorial matching problem. *Arch. Math. (Basel)*, 117(2):189–202, 2021.

[2] Jean-David Benamou. Optimal transportation, modelling and numerical simulation. *Acta Numer.*, 30:249–325, 2021.

[3] Camilla Brizzi, Luigi De Pascale, and Anna Kausamo. l^∞-optimal transport for a class of quasiconvex cost functions, 2021.

[4] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[5] Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs, 2022.

[6] Thierry Champion, Luigi De Pascale, and Petri Juutinen. The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40(1):1–20, 2008.

[7] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[8] Marcel Nutz Espen Bernton, Promit Ghosal. Entropic optimal transport: Geometry and large deviations. *arXiv:2102.04397*, 2021.

[9] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[10] Heikki Jylhä. The L^∞ optimal transport: infinite cyclical monotonicity and the existence of optimal transport maps. *Calc. Var. Partial Differential Equations*, 52(1-2):303–326, 2015.

[11] Christian Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.

[12] Quentin Mérigot and Boris Thibert. Optimal transport: discretization and algorithms. In *Geometric partial differential equations. Part II*, volume 22 of *Handb. Numer. Anal.*, pages 133–212. Elsevier/North-Holland, Amsterdam, [2021] ©2021.

[13] Marcel Nutz. Introduction to Entropic Optimal Transport, 2021.

[14] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[15] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[16] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[17] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[18] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.