ENTROPIC OPTIMAL TRANSPORT: CONVERGENCE OF POTENTIALS

MARCEL NUTZ AND JOHANNES WIESEL

ABSTRACT. We study the potential functions that determine the optimal density for ε-entropically regularized optimal transport, the so-called Schrödinger potentials, and their convergence to the counterparts in classical optimal transport, the Kantorovich potentials. In the limit $\varepsilon \to 0$ of vanishing regularization, strong compactness holds in L^1 and cluster points are Kantorovich potentials. In particular, the Schrödinger potentials converge in L^1 to the Kantorovich potentials as soon as the latter are unique. These results are proved for all continuous, integrable cost functions on Polish spaces. In the language of Schrödinger bridges, the limit corresponds to the small-noise regime.

1. INTRODUCTION AND MAIN RESULT

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be Polish probability spaces and $\Pi(\mu, \nu)$ the set of all couplings; i.e., probability measures π on $\mathcal{X} \times \mathcal{Y}$ with first marginal μ and second marginal ν. Moreover, let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ be continuous with

$$(1.1) \qquad \int c(x,y)\, \mu(dx)\nu(dy) < \infty.$$

Given a constant $\varepsilon > 0$, the entropic optimal transport (EOT) problem is

$$(1.2) \qquad I_\varepsilon := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y)\, \pi(dx, dy) + \varepsilon H(\pi | \mu \otimes \nu),$$

where $H(\,\cdot\,|\mu \otimes \nu)$ denotes relative entropy with respect to the product measure,

$$H(\pi|\mu \otimes \nu) := \begin{cases} \int \log(\frac{d\pi}{d(\mu \otimes \nu)})\, d\pi, & \pi \ll \mu \otimes \nu, \\ \infty, & \pi \not\ll \mu \otimes \nu. \end{cases}$$

For $\varepsilon = 0$ we recover the Monge–Kantorovich optimal transport problem, and (1.2) can be seen as its entropic regularization with parameter $\varepsilon > 0$. The minimization (1.2) admits a unique solution $\pi_\varepsilon \in \Pi(\mu, \nu)$; moreover, $\pi_\varepsilon \sim \mu \otimes \nu$ and its

1

density is of the form

$$(1.3) \qquad \frac{d\pi_\varepsilon}{d(\mu \otimes \nu)}(x, y) = \exp\left(\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x, y)}{\varepsilon}\right)$$

for two measurable functions $f_\varepsilon : \mathcal{X} \to \mathbb{R}$ and $g_\varepsilon : \mathcal{Y} \to \mathbb{R}$. We call these functions the *Schrödinger potentials*. They are unique up to normalization: any constant can be added to f_ε and subtracted from g_ε. The integrability (1.1) of c implies that $f_\varepsilon \in L^1(\mu)$ and $g_\varepsilon \in L^1(\nu)$, and we enforce the symmetric normalization

$$(1.4) \qquad \int f_\varepsilon(x)\, \mu(dx) = \int g_\varepsilon(y)\, \nu(dy)$$

to have uniqueness of the potentials in all that follows. We mention that π_ε can be characterized as the unique coupling $\pi \in \Pi(\mu, \nu)$ whose density is of the form (1.3). See, for instance, [16, Statements 3.6, 3.15, 3.19, 3.38] for existence and uniqueness, or [21] for a simple derivation including integrability under (1.1). These result heavily build on [5, 6, 11, 26], among others. Rewriting the minimization (1.2), the coupling π_ε can be interpreted as the so-called static Schrödinger bridge

$$(1.5) \qquad \pi_\varepsilon = \underset{\pi \in \Pi(\mu, \nu)}{\arg\min}\, H(\pi | R)$$

for the reference probability $dR \propto e^{-c/\varepsilon} d(\mu \otimes \nu)$ which elucidates (1.3) as the factorization property $\frac{d\pi_\varepsilon}{dR}(x, y) = e^{f_\varepsilon(x)/\varepsilon} e^{g_\varepsilon(y)/\varepsilon} =: F(x)G(y).$[1] A closely related, more analytic way to characterize the potentials are the Schrödinger equations. Writing also $C(x, y) = e^{-c(x,y)/\varepsilon}$, the fact that π_ε of (1.3) is in $\Pi(\mu, \nu)$ implies that (F, G) solves the coupled equations

(1.6)

$$F(x)^{-1} = \int G(y)C(x, y)\, \nu(dy) \quad \mu\text{-a.s.}, \quad G(x)^{-1} = \int F(y)C(x, y)\, \nu(dx) \quad \nu\text{-a.s.}$$

Conversely, we can use any solution (F, G) to define a coupling with density of the form (1.4). This coupling must coincide with π_ε by the aforementioned uniqueness, and then $(\varepsilon \log F, \varepsilon \log G)$ must be our Schrödinger potentials $(f_\varepsilon, g_\varepsilon)$, up to normalization. We refer to [4, 25] and the references therein for more on Schrödinger equations, and to [15, 19] for extensive surveys on Schrödinger bridges.

Yet another way to introduce the potentials is to consider the dual problem of (1.2) in the sense of convex analysis,

$$(1.7) \qquad S_\varepsilon := \sup_{f \in L^1(\mu), g \in L^1(\nu)} \left(\int f(x)\, \mu(dx) + \int g(y)\, \nu(dy) \right.$$
$$\left. - \varepsilon \int e^{\frac{f(x) + g(y) - c(x,y)}{\varepsilon}}\, \mu(dx)\nu(dy) + \varepsilon \right).$$

---

[1] We mention that [19] uses the term Schrödinger potentials for $f_\varepsilon/\varepsilon, g_\varepsilon/\varepsilon$ in the Schrödinger bridge context, as is natural when no parameter ε is present. On the other hand, calling $f_\varepsilon, g_\varepsilon$ potentials is more convenient in our setting, well motivated by the connection with Kantorovich potentials in Theorem 1.1, and consistent with the terminology in [17].

Then $(f_\varepsilon, g_\varepsilon)$ is the unique solution of (1.7) with the normalization (1.4). Indeed, direct arguments show the weak duality $S_\varepsilon \leq I_\varepsilon$. To see that equality is attained by $(f_\varepsilon, g_\varepsilon)$ and π_ε, we plug in (1.3) and use $\pi_\varepsilon(\mathcal{X} \times \mathcal{Y}) = 1$ to find that $S_\varepsilon \geq \int f_\varepsilon(x) \, \mu(dx) + \int g_\varepsilon(y) \, \nu(dy) \geq I_\varepsilon$. Uniqueness holds by strict concavity. See [23] and the references therein for a convex analysis perspective including (1.7).

We are interested in the relation of $(f_\varepsilon, g_\varepsilon)$ to solutions of the dual Monge–Kantorovich problem,

$$(1.8) \qquad S_0 := \sup_{f \in L^1(\mu), \, g \in L^1(\nu), \, f \oplus g \leq c} \left(\int f(x) \, \mu(dx) + \int g(y) \, \nu(dy) \right),$$

where $(f \oplus g)(x, y) := f(x) + g(y)$. It is well known that $S_0 = I_0$ and that a solution (f_0, g_0) exists [27, Theorem 5.10, Remark 5.14]. (In fact, Theorem 1.1 below yields another proof as a by-product.) There is the same ambiguity as above, and to streamline terminology, we call (f_0, g_0) Kantorovich potentials if they satisfy the normalization (1.4) for $\varepsilon = 0$. As (1.8) lacks the strict convexity of (1.7), multiple Kantorovich potentials may exist even after normalization, for instance when both marginals are discrete. Nevertheless, uniqueness of Kantorovich potentials is known to hold for most problems of interest to us, especially when c is differentiable and at least one marginal support is connected. See for instance [3, Appendix B] for sufficient conditions.

Much of the enormous recent interest in entropic optimal transport stems from the success of Sinkhorn's algorithm in high-dimensional problems, enabling data-rich applications in areas like machine learning or image processing. Popularized in this context by [12], Sinkhorn's algorithm computes the Schrödiger potentials $(f_\varepsilon, g_\varepsilon)$ by alternating projections. From a computational point of view, the Monge–Kantorovich problem is significantly harder than the entropic one; see [24] for a recent survey and numerous references. It is therefore natural to investigate $(f_\varepsilon, g_\varepsilon)$ as $\varepsilon \to 0$ to approximate Kantorovich potentials.

On the primal side, weak compactness of $\Pi(\mu, \nu)$ immediately implies that (π_ε) admits cluster points as $\varepsilon \to 0$. Moreover, any cluster point is an optimal transport, so that if uniqueness is known for the solution π_0 of the limiting optimal transport problem, then $\pi_\varepsilon \to \pi_0$. See [8, 18] for proofs by Gamma convergence, or [3] for a geometric proof assuming only continuity of c. Our aim is to establish a comparable result on the dual side. Here, compactness is not obvious (unless μ, ν are compactly supported). Our main result provides strong compactness in L^1 for (f_ε) and (g_ε) as $\varepsilon \to 0$, and moreover, that cluster points are Kantorovich potentials. In most cases of interest, the latter are unique, so that the whole sequence converges.

Theorem 1.1. *Let $(f_\varepsilon, g_\varepsilon)$ be the unique Schrödinger potentials for $\varepsilon > 0$.*

   (a) *Given $\varepsilon_n \to 0$, there is a subsequence (ε_k) such that f_{ε_k} converges in $L^1(\mu)$ and g_{ε_k} converges in $L^1(\nu)$.*

   (b) *If $\lim_n f_{\varepsilon_n} = f$ μ-a.s. and $\lim_n g_{\varepsilon_n} = g$ ν-a.s. for $\varepsilon_n \to 0$, then (f, g) are Kantorovich potentials and the convergence also holds in L^1.*

If the Kantorovich potentials (f_0, g_0) for (1.8) are unique, it follows that $\lim_\varepsilon f_\varepsilon = f_0$ in $L^1(\mu)$ and $\lim_\varepsilon g_\varepsilon = g_0$ in $L^1(\nu)$.

Applications of interest for Theorem 1.1 include costs $c(x, y) = \|x - y\|^2$ on $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ with unbounded marginal supports as in [20]; here c is continuous but not uniformly continuous. Theorem 1.1 simplifies substantially in the case of compactly supported marginals. More generally, if c is uniformly continuous, the functions $f_\varepsilon, g_\varepsilon$ inherit its modulus of continuity (uniformly in ε) and then uniform convergence on compact subsets along a subsequence follows from the Arzelà–Ascoli theorem; cf. Proposition 3.2. A result along those lines is contained in [17, Section 5] in the particular case of quadratic cost and compact marginals. We emphasize that [17] analyzes the more complex dynamic problem of approximating W_2 geodesics with entropic interpolation; the present static setting would correspond only to its marginals at times $t = 0, 1$. Given the results of [17], one may conjecture that Theorem 1.1 can be extended to interpolations and intermediate times $t \in (0, 1)$.

When $\mathcal{X}, \mathcal{Y}$ are finite sets, optimal transport is a finite-dimensional linear programming problem. For such problems, a detailed convergence analysis of entropic regularization is presented in [9]. In particular, convergence holds even when Kantorovich potentials are not unique.

Theorem 1.1 can be related to the large deviations principle (LDP) of [3] which describes the convergence of (π_ε) on the primal side (cf. Section 4 for a detailed discussion). On compact spaces, convergence of potentials is equivalent to the validity of an LDP whose rate functions includes the limiting Kantorovich potentials. On the other hand, neither result implies the other in general, and we see the results and methods as complementary. Indeed, the "easier" inequality for the present dual approach corresponds to the more delicate one in the primal approach, and vice versa. See also [10, 22] for expansions of the entropic transport cost as $\varepsilon \to 0$, which are related to the speed of convergence of (π_ε). Finally, we mention [2], studying the convergence of the discrete Sinkhorn algorithm to an optimal transport potential in the joint limit when $\varepsilon_n \to 0$ and the marginals μ, ν are approximated by discretizations μ_n, ν_n satisfying a certain density property. Beyond the aforementioned special cases and connections, Theorem 1.1 is novel, to the best of our knowledge.

Two extensions of Theorem 1.1 are obtained in the body of the text. The first one replaces c in (1.2) by a cost function c_ε that may depend on ε and converges to the continuous cost c of the Monge–Kantorovich problem as $\varepsilon \to 0$. This extension demonstrates the stability of the convergence in Theorem 1.1. In addition, it may be a natural result from the perspective of Schödinger bridges (see [19]): the corresponding reference measures R_ε in (1.5) are those with large deviations rate c. The second extension replaces the two marginals (μ, ν) by any (finite) number of marginals. The resulting "multimarginal" optimal transport problem has become a focus of attention as the primary tool to analyze Wasserstein barycenters in the sense of [1]. Its entropic regularization again admits a version of Sinkhorn's algorithm; see [7] for a very recent analysis showing linear convergence and further references. The techniques developed in the proof of Theorem 1.1 are quite versatile and extend to the multimarginal setting without effort.

The remainder of this paper is organized as follows. Section 2 collects auxiliary results for the proof of Theorem 1.1, which is carried out in Section 3 and followed

by the specialization to uniformly continuous costs. The relation with the LDP is the subject of Section 4. In Section 5 we present the extension to costs c_ε that vary with ε, and Section 6 concludes with the multimarginal case.

2. Auxiliary Results

In this section we collect a number of auxiliary results for the proof of Theorem 1.1. Anticipating the generalization in Section 5, we remark that the statements and proofs in this section hold for any measurable (but not necessarily continuous) cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ that is integrable in the sense of (1.1); further regularity is only required in Lemma 2.6, where the condition is stated explicitly.

Let $\varepsilon > 0$. We recall the Schrödinger potentials $f_\varepsilon \in L^1(\mu)$ and $g_\varepsilon \in L^1(\nu)$ from the Introduction and in particular the normalization

$$(2.1) \qquad \int f_\varepsilon(x)\, \mu(dx) = \int g_\varepsilon(y)\, \nu(dy) = S_\varepsilon/2 \geq 0.$$

The fact that π_ε of (1.3) is a probability measure with marginals μ and ν implies

$$(2.2) \quad \int e^{\frac{f_\varepsilon(x)+g_\varepsilon(y)-c(x,y)}{\varepsilon}}\, \nu(dy) = 1 \quad \mu\text{-a.s.}, \quad \int e^{\frac{f_\varepsilon(x)+g_\varepsilon(y)-c(x,y)}{\varepsilon}}\, \mu(dx) = 1 \quad \nu\text{-a.s.}$$

and hence the Schrödinger equations

$$(2.3) \qquad
\begin{aligned}
f_\varepsilon(x) &= -\varepsilon \log \int e^{\frac{g_\varepsilon(y)-c(x,y)}{\varepsilon}}\, \nu(dy) \quad \mu\text{-a.s.}, \\
g_\varepsilon(y) &= -\varepsilon \log \int e^{\frac{f_\varepsilon(x)-c(x,y)}{\varepsilon}}\, \mu(dx) \quad \nu\text{-a.s.}
\end{aligned}$$

By choosing versions of $f_\varepsilon, g_\varepsilon$ we may and will assume that these conjugacy relations hold everywhere on $\mathcal{X} \times \mathcal{Y}$. In particular, this provides canonical extensions of $f_\varepsilon, g_\varepsilon$ to the whole marginal space. The conjugacy relations can also be used to obtain a priori estimates, as has been previously exploited in [7, 14], among others.

Lemma 2.1. *For all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we have*

$$\inf_{y \in \mathcal{Y}} \left[c(x,y) - g_\varepsilon(y) \right] \leq f_\varepsilon(x) \leq \int c(x,y)\, \nu(dy),$$

$$\inf_{x \in \mathcal{X}} \left[c(x,y) - f_\varepsilon(x) \right] \leq g_\varepsilon(y) \leq \int c(x,y)\, \mu(dx).$$

Proof. Using (2.3), Jensen's inequality and (2.1),

$$(2.4) \qquad
\begin{aligned}
f_\varepsilon(x) &= -\varepsilon \log \int e^{\frac{g_\varepsilon(y)-c(x,y)}{\varepsilon}}\, \nu(dy) \\
&\leq \int [-g_\varepsilon(y) + c(x,y)]\, \nu(dy) \leq \int c(x,y)\, \nu(dy),
\end{aligned}$$

which is the upper bound. For the lower bound we note that by (2.3),

$$f_\varepsilon(x) \geq -\varepsilon \log \int e^{\frac{\sup_{y \in \mathcal{Y}}[g_\varepsilon(y) - c(x,y)]}{\varepsilon}} \nu(dy)$$

$$= -\sup_{y \in \mathcal{Y}} \big[g_\varepsilon(y) - c(x,y)\big] = \inf_{y \in \mathcal{Y}} \big[c(x,y) - g_\varepsilon(y)\big].$$

The proof for g_ε is symmetric. $\square$

Let (M, d) be a metric space. A function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ is a modulus of continuity if it is continuous at 0 with $\omega(0) = 0$. More generally, we call $\omega : M \times \mathbb{R}_+ \to \mathbb{R}_+$ a modulus of continuity if $\omega(x, \cdot)$ has those properties for each $x \in M$. A function $F : M \to \mathbb{R}$ is ω-continuous if it admits the modulus of continuity $\omega(x, \cdot)$ at $x \in M$; that is, $|F(x) - F(x')| \leq \omega(x, d(x, x'))$ for all $x, x' \in M$. To avoid ambiguity, we say that F is uniformly ω-continuous if ω can be chosen independent of x. The following generalization of the Arzelà–Ascoli theorem will be used to construct limits of f_ε and g_ε.

Lemma 2.2. *Let (M, d) be a separable metric space and let (F_n) be (arbitrary) functions on M which are pointwise bounded and satisfy*

$$(2.5) \qquad |F_n(x_1) - F_n(x_2)| \leq \omega(x_1, d(x_1, x_2)) + h_n, \quad x_1, x_2 \in M$$

for some modulus of continuity $\omega : M \times \mathbb{R}_+ \to \mathbb{R}_+$ and a sequence $h_n \to 0$ of constants. Then after passing to a subsequence, (F_n) converges uniformly on compact subsets to a ω-continuous function $F : M \to \mathbb{R}$.

Proof. Let $D \subset M$ be a countable dense set, fix $\delta > 0$ and choose $n_0 \in \mathbb{N}$ such that $h_n \leq \delta/6$ for all $n \geq n_0$. As (F_n) is pointwise bounded, a diagonal argument yields a subsequence, still denoted (F_n), converging pointwise on D. In particular, for every $x \in D$ there exists $n(x)$ such that

$$(2.6) \qquad |F_n(x) - F_m(x)| \leq \delta/3, \quad m, n \geq n(x).$$

For $x_1 \in D$, (2.5) yields an open neighborhood O_{x_1} with

$$(2.7) \quad |F_n(x_1) - F_n(x_2)| \leq \omega(x_1, d(x_1, x_2)) + h_n \leq \delta/6 + \delta/6 = \delta/3, \quad x_2 \in O_{x_1},$$

for all $n \geq n_0$. Let $K \subset M$ be compact and $D' \subseteq D$ a finite set such that $\bigcup_{x' \in D'} O_{x'}$ covers K. Choose $n_1 := \max_{x' \in D'} n(x') \vee n_0$, then as any $x \in K$ is contained in an open neighborhood $O_{x'}$ of some $x' \in D'$, we obtain from (2.6) and (2.7) that

$$|F_n(x) - F_m(x)| \leq |F_n(x) - F_n(x')| + |F_n(x') - F_m(x')| + |F_m(x') - F_m(x)| \leq \delta,$$

for all $x \in K$ and $m, n \geq n_1$. Thus (F_n) has a limit F, uniformly on compacts. Passing to the limit in (2.5) shows that F is ω-continuous. $\square$

Recall that $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ is continuous. If $\mathcal{Y}_{\mathrm{cpt}} \subset \mathcal{Y}$ is compact and $\omega(x, r) := \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}, d(x,x') \leq r} |c(x, y) - c(x', y)|$, then ω is a modulus of continuity in the above sense. That motivates the following estimates.

Lemma 2.3. *Fix $\delta \in (0,1)$ and $\varepsilon > 0$. There exist compact sets $\mathcal{X}_{\mathrm{cpt}} \subseteq \mathcal{X}, \mathcal{Y}_{\mathrm{cpt}} \subseteq \mathcal{Y}$ and measurable sets $A_\varepsilon \subseteq \mathcal{X}_{\mathrm{cpt}}$, $B_\varepsilon \subseteq \mathcal{Y}_{\mathrm{cpt}}$ with $\mu(A_\varepsilon), \nu(B_\varepsilon) \geq 1 - \delta$ such that*

$$|f_\varepsilon(x_1) - f_\varepsilon(x_2)| \leq \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}} |c(x_1, y) - c(x_2, y)| - \varepsilon \log(1 - \delta) \quad for \quad x_1, x_2 \in A_\varepsilon,$$

$$\begin{aligned}
|g_\varepsilon(y_1) - g_\varepsilon(y_2)| &\leq \sup_{x \in A_\varepsilon} |c(x, y_1) - c(x, y_2)| - \varepsilon \log(1 - \delta) \\
&\leq \sup_{x \in \mathcal{X}_{\mathrm{cpt}}} |c(x, y_1) - c(x, y_2)| - \varepsilon \log(1 - \delta) \quad for \quad y_1, y_2 \in B_\varepsilon.
\end{aligned}$$

Proof. Fix $\kappa \in (0, \delta)$, to be determined later. Choose compacts $\mathcal{X}_{\mathrm{cpt}}$ and $\mathcal{Y}_{\mathrm{cpt}}$ with $\mu(\mathcal{X}_{\mathrm{cpt}}) \geq 1 - \kappa^2/2$ and $\nu(\mathcal{Y}_{\mathrm{cpt}}) \geq 1 - \kappa^2/2$, then $\pi_\varepsilon \in \Pi(\mu, \nu)$ implies

$$(2.8) \qquad \qquad \pi_\varepsilon(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) \geq 1 - \kappa^2.$$

Consider the set

$$A_\varepsilon = \left\{ x \in \mathcal{X}_{\mathrm{cpt}} : \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy) \geq 1 - \kappa \right\};$$

we claim that its complement $A_\varepsilon^c = \mathcal{X} \setminus A_\varepsilon$ satisfies

$$(2.9) \qquad \qquad p_\varepsilon := \mu\left(A_\varepsilon^c\right) \leq \kappa.$$

Indeed, (2.2) yields

$$(2.10) \qquad \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy) \leq \int e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy) = 1$$

and thus

$$\begin{aligned}
1 - \kappa^2 &\overset{(2.8)}{\leq} \pi_\varepsilon(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) = \int_{\mathcal{X}_{\mathrm{cpt}}} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy)\mu(dx) \\
&\leq \int_{A_\varepsilon^c} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy)\mu(dx) \\
&\quad + \int_{A_\varepsilon} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy)\mu(dx) \\
&\overset{(2.10)}{\leq} (1 - \kappa)p_\varepsilon + (1 - p_\varepsilon) = 1 - p_\varepsilon\kappa,
\end{aligned}$$

which implies (2.9). Next, we observe from the definition of A_ε and (2.10) that for $x \in A_\varepsilon$,

(2.11)

$$-\varepsilon \left(\log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy) - \log(1 - \kappa) \right) \leq f_\varepsilon(x)$$

$$\leq -\varepsilon \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x,y)}{\varepsilon}} \nu(dy).$$

Let $x_1, x_2 \in A_\varepsilon$ and assume without loss of generality that $f_\varepsilon(x_1) \geq f_\varepsilon(x_2)$. Then

$$|f_\varepsilon(x_1) - f_\varepsilon(x_2)| \overset{(2.11)}{\leq} \varepsilon \left(\log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x_2, y)}{\varepsilon}} \nu(dy) - \log(1 - \kappa) \right)$$

$$- \varepsilon \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x_1, y)}{\varepsilon}} \nu(dy)$$

$$= \varepsilon \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{c(x_1, y) - c(x_2, y) + g_\varepsilon(y) - c(x_1, y)}{\varepsilon}} \nu(dy) - \varepsilon \log(1 - \kappa)$$

$$(2.12) \qquad\qquad - \varepsilon \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x_1, y)}{\varepsilon}} \nu(dy)$$

$$\leq \varepsilon \log \left(e^{\frac{\sup_{y \in \mathcal{Y}_{\mathrm{cpt}}} |c(x_1, y) - c(x_2, y)|}{\varepsilon}} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x_1, y)}{\varepsilon}} \nu(dy) \right)$$

$$- \varepsilon \log(1 - \kappa) - \varepsilon \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{g_\varepsilon(y) - c(x_1, y)}{\varepsilon}} \nu(dy)$$

$$= \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}} |c(x_1, y) - c(x_2, y)| - \varepsilon \log(1 - \kappa).$$

This concludes the proof of the first estimate in the lemma.

Turning to the second, note that by (2.8), (2.9) and the definition of A_ε,

$$\pi_\varepsilon(A_\varepsilon \times \mathcal{Y}_{\mathrm{cpt}}) \geq \pi_\varepsilon(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) - \pi_\varepsilon(\mathcal{X} \setminus A_\varepsilon \times \mathcal{Y}_{\mathrm{cpt}})$$

$$(2.13) \qquad\qquad \geq 1 - \kappa^2 - \int_{A_\varepsilon^c} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x, y)}{\varepsilon}} \nu(dy) \, \mu(dx)$$

$$\geq 1 - \kappa^2 - \kappa(1 - \kappa) = 1 - \kappa = 1 - \delta^2,$$

where we chose $\kappa := \delta^2$ (ensuring $\kappa \in (0, \delta)$, in particular). Define

$$B_\varepsilon = \left\{ y \in \mathcal{Y}_{\mathrm{cpt}} : \int_{A_\varepsilon} e^{\frac{f_\varepsilon(x) + g_\varepsilon(y) - c(x, y)}{\varepsilon}} \mu(dx) \geq 1 - \delta \right\}.$$

Arguing as for (2.9) and (2.11), now using (2.13) instead of (2.8), we see that $\nu(B_\varepsilon^c) \leq \delta$ and that for $y \in B_\varepsilon$,

$$-\varepsilon \left(\log \int_{A_\varepsilon} e^{\frac{f_\varepsilon(x) - c(x, y)}{\varepsilon}} \mu(dx) - \log(1 - \delta) \right) \leq g_\varepsilon(y)$$

$$\leq -\varepsilon \log \int_{A_\varepsilon} e^{\frac{f_\varepsilon(x) - c(x, y)}{\varepsilon}} \mu(dx).$$

We conclude the proof by arguing as in (2.12) but with f_ε, κ replaced by g_ε, δ. $\quad\square$

The following extension lemma is a variation on Kirszbraun's theorem. Recall that a pseudometric $\tilde{d}$ is defined like a metric except that $\tilde{d}(x, y) = 0$ need not imply $x = y$.

Lemma 2.4. *Let $(M, \tilde{d})$ be a pseudometric space and $A \subseteq M$. Let $F : A \to \mathbb{R}$ satisfy*

(2.14) $$|F(x_1) - F(x_2)| \leq \tilde{d}(x_1, x_2) + \gamma, \quad x_1, x_2 \in A,$$

for some $\gamma > 0$. Then the function $\tilde{F} : M \to \mathbb{R}$ defined by

$$\tilde{F}(x) := \inf_{x' \in A} \left[F(x') + \tilde{d}(x, x') + \gamma \mathbf{1}_{\{x' \neq x\}} \right], \quad x \in M$$

satisfies $\tilde{F} = F$ on A and

(2.15) $$|\tilde{F}(x_1) - \tilde{F}(x_2)| \leq \tilde{d}(x_1, x_2) + \gamma, \quad x_1, x_2 \in M.$$

Proof. Fix $x \in A$. For $x \neq x' \in A$ we have by (2.14) that

$$F(x') + \tilde{d}(x, x') + \gamma \geq F(x) - \tilde{d}(x, x') - \gamma + \tilde{d}(x, x') + \gamma = F(x).$$

It follows that $\tilde{F}(x) = F(x)$, showing the first claim. Fix $\kappa > 0$ and let $x_1, x_2 \in M$. By the definition of $\tilde{F}(x_2)$ there exists $x' \in A$ such that $\tilde{F}(x_2) \geq F(x') + \tilde{d}(x_2, x') - \kappa$, and now the definition of $\tilde{F}(x_1)$ yields

$$\tilde{F}(x_1) - \tilde{F}(x_2) \leq F(x') + \tilde{d}(x_1, x') + \gamma - F(x') - \tilde{d}(x_2, x') + \kappa$$
$$= \tilde{d}(x_1, x') - \tilde{d}(x_2, x') + \gamma + \kappa \leq \tilde{d}(x_1, x_2) + \gamma + \kappa.$$

As $\kappa > 0$ was arbitrary, (2.15) follows. $\qquad\square$

The next two lemmas show that limits of $f_\varepsilon, g_\varepsilon$ must be Kantorovich potentials.

Lemma 2.5. *Let $\varepsilon_n \to 0$ and suppose that the corresponding potentials $f_{\varepsilon_n}, g_{\varepsilon_n}$ converge a.s. Then the limits $f := \lim_n f_{\varepsilon_n}$ and $g := \lim_n g_{\varepsilon_n}$ satisfy*

$$f(x) + g(y) \leq c(x, y) \qquad \mu \otimes \nu\text{-a.s.}$$

Proof. Let $\delta > 0$. Passing to a subsequence if necessary, we may assume that $\sum_{n=1}^{\infty} e^{-\delta/\varepsilon_n} < \infty$. Define

$$A_{\delta, n} = \{(x, y) \; : \; f_{\varepsilon_n}(x) + g_{\varepsilon_n}(y) - c(x, y) \geq \delta\},$$

then

$$1 \overset{(2.2)}{=} \int e^{\frac{f_{\varepsilon_n}(x) + g_{\varepsilon_n}(y) - c(x,y)}{\varepsilon_n}} \mu(dx)\nu(dy) \geq \int_{A_{\delta,n}} e^{\frac{f_n(x) + g_n(y) - c(x,y)}{\varepsilon_n}} \mu(dx)\nu(dy)$$

$$\geq e^{\delta/\varepsilon_n}(\mu \otimes \nu)(A_{\delta,n})$$

yields $(\mu \otimes \nu)(A_{\delta,n}) \leq e^{-\delta/\varepsilon_n}$ and thus $\sum_n (\mu \otimes \nu)(A_{\delta,n}) < \infty$. The Borel–Cantelli lemma now shows that $(\mu \otimes \nu)(\limsup_n A_{\delta,n}) = 0$ and hence

$$f(x) + g(y) \leq c(x, y) + \delta \qquad \mu \otimes \nu\text{-a.s.}$$

As $\delta > 0$ was arbitrary, the claim follows. $\qquad\square$

Lemma 2.6. *Let c be upper semicontinuous and f, g measurable functions with*

$$f(x) + g(y) \leq c(x, y) \quad \mu \otimes \nu\text{-a.s.}$$

Then there exist versions $\tilde{f} = f$ μ-a.s. and $\tilde{g} = g$ ν-a.s. such that

$$\tilde{f}(x) + \tilde{g}(y) \leq c(x, y) \quad \text{for all} \quad (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Proof. Suppose first that f, g are continuous. If $f(x) + g(y) > c(x, y)$ for some $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the same inequality holds on a neighborhood $B_r(x) \times B_r(y)$, which then must be $\mu \otimes \nu$-null by the assumption. That is, $(x, y) \notin \operatorname{spt}\mu \times \operatorname{spt}\nu$. In conclusion, we can set $\tilde{f} = f$ on $\operatorname{spt}\mu$ and $\tilde{f} = -\infty$ outside $\operatorname{spt}\mu$, and similarly for $\tilde{g}$.

In general, Lusin's theorem yields an increasing sequence of closed sets $A_n \subset \mathcal{X}$ such that $f|_{A_n}$ is continuous and $\mu(A_n^c) \leq 1/n$. Let $\mu_n = \mu|_{A_n}$ and $A'_n = \operatorname{spt}\mu_n$. Defining analogously $B'_n \subset \mathcal{Y}$, the above argument shows that $f(x) + g(y) \leq c(x, y)$ on $A'_n \times B'_n$. The same inequality then holds on the product of $\cup_n A'_n$ and $\cup_n B'_n$, and these sets have full measure. It remains to set $\tilde{f} = f$ on $\cup_n A'_n$ and $\tilde{f} = -\infty$ on the complement, and similarly for $\tilde{g}$. $\qquad\square$

3. Proof of the Main Result

We can now report the proof of Theorem 1.1. To simplify the notation, let us agree that an index n always refers to an object associated with $\varepsilon = \varepsilon_n$; for instance, $f_n = f_{\varepsilon_n}$ and $g_n = g_{\varepsilon_n}$. Moreover, subsequences are not relabeled.

Steps 1–5 below establish the a.s. convergence of f_n and g_n along a subsequence. The final Step 6 shows that a.s. convergence also implies L^1-convergence, and that limits are Kantorovich potentials.

Proof of Theorem 1.1. Let $\varepsilon_n \to 0$. In addition, we fix a strictly decreasing sequence $\delta_k \to 0$ with $\delta_k < 1/2$.

Step 1. For each $k, n \in \mathbb{N}$, Lemma 2.3 yields sets

$$A_n(\delta_k) \subseteq \mathcal{X}_{\operatorname{cpt}}(\delta_k) \subseteq \mathcal{X} \quad \text{and} \quad B_n(\delta_k) \subseteq \mathcal{Y}_{\operatorname{cpt}}(\delta_k) \subseteq \mathcal{Y}$$

such that

$$\mu(A_n(\delta_k)) \geq 1 - \delta_k \quad \text{and} \quad \nu(B_n(\delta_k)) \geq 1 - \delta_k$$

as well as

(3.1)
$$\begin{aligned}
|f_n(x_1) - f_n(x_2)| &\leq \sup_{y \in \mathcal{Y}_{\operatorname{cpt}}(\delta_k)} |c(x_1, y) - c(x_2, y)| - \varepsilon_n \log(1 - \delta_k) \\
&\leq \sup_{y \in \mathcal{Y}_{\operatorname{cpt}}(\delta_k)} |c(x_1, y) - c(x_2, y)| + \varepsilon_n \log(2)
\end{aligned}$$

for $x_1, x_2 \in A_n(\delta_k)$ and similarly

$$|g_n(y_1) - g_n(y_2)| \leq \sup_{x \in \mathcal{X}_{\operatorname{cpt}}(\delta_k)} |c(x, y_1) - c(x, y_2)| + \varepsilon_n \log(2)$$

for $y_1, y_2 \in B_n(\delta_k)$. For each n, we can assume that the sequences $(\mathcal{X}_{\operatorname{cpt}}(\delta_k))_k$ and $(\mathcal{Y}_{\operatorname{cpt}}(\delta_k))_k$ are increasing, and consequently also that $(A_n(\delta_k))_k$ and $(B_n(\delta_k))_k$ are increasing.

Step 2. Define
$$\tilde{d}_k(x_1, x_2) := \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}(\delta_k)} |c(x_1, y) - c(x_2, y)|.$$

It is elementary to verify that $\tilde{d}_k$ is a pseudometric on $\mathcal{X}$. Using (3.1) and Lemma 2.4 with $\gamma = \varepsilon_n \log(2)$, there exists an extension f_n^k satisfying $f_n^k = f_n$ on $A_n(\delta_k)$ and

$$(3.2) \qquad \left| f_n^k(x_1) - f_n^k(x_2) \right| \le \tilde{d}_k(x_1, x_2) + \varepsilon_n \log(2), \quad x_1, x_2 \in \mathcal{X}.$$

Similarly, there exists an extension g_n^k for g_n with an analogous property.

Step 3. We now vary n, while still keeping k fixed, and our aim is to construct a subsequential limit $f^k = \lim_{n \to \infty} f_n^k$ μ-a.s. We first argue that $(f_n^k)_{n \in \mathbb{N}}$ is pointwise bounded from above. Indeed, after taking another subsequence if necessary, there exists $x_0 \in \mathrm{spt}\,\mu$ such that $x_0 \in A_n(\delta_k)$ for all n and $f_n(x_0) \le \int c(x_0, y)\,\nu(dy) < \infty$; cf. Lemma 2.1. Thus $f_n(x_0)^+$ is bounded uniformly in n. On the other hand, $\int f_n^+(x)\,\mu(dx) \le \int c(x, y)\,\nu(dy)\mu(dx) < \infty$, and as $\int f_n(x)\,\mu(dx) \ge 0$ by (2.1), it follows that $\int f_n^-(x)\,\mu(dx)$ is bounded. In view of (3.1), we obtain that $f_n(x_0)^-$ is bounded uniformly in n. This shows that $f_n(x_0)$ is bounded, and then so is $f_n^k(x_0)$. By (3.2), it follows that $f_n^k(x)$ is bounded uniformly in n, for any $x \in \mathcal{X}$, as claimed. Define
$$\omega_k(x, r) = \sup_{d(x,x') \le r} \tilde{d}_k(x, x') \equiv \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}(\delta_k), d(x,x') \le r} |c(x, y) - c(x', y)|.$$

Clearly $\tilde{d}(x_1, x_2) \le \omega(x_1, d(x_1, x_2))$, and ω_k is a modulus of continuity as noted above. In particular, the conditions of Lemma 2.2 are satisfied for the sequence $(f_n^k)_{n \in \mathbb{N}}$ with $\omega := \omega_k$ and $h_n := \varepsilon_n \log(2)$. After passing to a subsequence, we thus obtain an ω_k-continuous function f^k such that $f_n^k \to f^k$ uniformly on compact subsets. After passing to another subsequence, we similarly obtain a limit g^k for g_n^k.

Recall that for fixed n, the sets $A_n(\delta_k)$ are increasing in k, and $\cup_k A_n(\delta_k)$ has full μ-measure. As a consequence, $f_n^k = f_n^{k'} = f_n$ on $A_n(\delta_k)$ for all $k' \ge k$, and a diagonal argument yields a subsequence along which $\lim_{n \to \infty} f_n^k = f^k$ μ-a.s. for all k. Similarly for g_n^k, and we may assume in what follows that $\lim_{n \to \infty} f_n^k = f^k$ μ-a.s. and $\lim_{n \to \infty} g_n^k = g^k$ ν-a.s. for all k.

Step 4. In this step we show that (f^k) converges μ-a.s., after passing to a subsequence. Fix $\gamma > 0$ and choose k_0 such that $\delta_{k_0} \le \gamma$. For all $k, k' \ge k_0$ and all n, we have

$$(3.3) \quad |f^k(x) - f^{k'}(x)| \le |f^k(x) - f_n^k(x)| + |f_n^k(x) - f_n^{k'}(x)| + |f_n^{k'}(x) - f^{k'}(x)|.$$

Recalling also $f_n^k = f_n^{k'} = f_n$ on $A_n(\delta_{k_0})$ and $\mu((A_n(\delta_{k_0}))^c) \le \delta_{k_0} \le \gamma$, we deduce

$$\int [|f^k(x) - f^{k'}(x)| \wedge 1]\,\mu(dx)$$
$$\le \int_{A_n(\delta_{k_0})} \left([|f^k(x) - f_n^k(x)| \wedge 1] + [|f_n^{k'}(x) - f^{k'}(x)| \wedge 1] \right)\mu(dx) + \gamma.$$

Sending $n \to \infty$ and using the result of Step 3, dominated convergence allows us to conclude that $\int |f^k(x) - f^{k'}(x)| \wedge 1 \, \mu(dx) \leq \gamma$; that is, (f^k) is Cauchy in μ-probability. In particular, there exists a limit f in μ-probability, and after taking a subsequence, the limit also holds μ-a.s. Similarly, $\lim_k g^k = g$ ν-a.s.

Step 5. Next, we show that the potentials f_n, g_n converge a.s. to the same limits f, g, after taking another subsequence. Given $\gamma > 0$, Step 4 implies that for a.e. $x \in \mathcal{X}$ there exists $k_0(x)$ such that $|f^k(x) - f(x)| \leq \gamma/3$ and $\delta_k \leq \gamma$ for $k \geq k_0(x)$. As $\lim_n f_n^k = f^k$ μ-a.s., it follows for $k \geq k_0(x)$ and for n sufficiently large that

$$
\begin{aligned}
|f_n(x) - f(x)| &\leq |f_n(x) - f_n^k(x)| + |f_n^k(x) - f^k(x)| + |f^k(x) - f(x)| \\
&\leq |f_n(x) - f_n^k(x)| + |f_n^k(x) - f^k(x)| + \gamma/3 \\
&\leq |f_n(x) - f_n^k(x)| + \gamma/2.
\end{aligned}
$$

Recalling that $f_n(x) = f_n^k$ on $A_n(\delta_k)$, we conclude

$$
\lim_{n \to \infty} \mu\left(\{x : |f_n(x) - f(x)| \geq \gamma\}\right) \leq \limsup_{n \to \infty} \mu\left((A_n(\delta_k))^c\right) \leq \delta_k \leq \gamma;
$$

that is, $f_n \to f$ in μ-probability. Taking another subsequence, we have $\lim_n f_n = f$ μ-a.s. Similarly, we obtain $\lim_n g_n = g$. Lemmas 2.5 and 2.6 show that after modifying f, g on nullsets, we have

$$
(3.4) \qquad\qquad f(x) + g(y) \leq c(x, y), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}.
$$

Step 6. Let $C_1(x) := \int c(x, y) \, \nu(dy)$ and $C_2(y) := \int c(x, y) \, \mu(dx)$. In view of Lemma 2.1 we have

$$
(3.5) \qquad\qquad f_n, f \leq C_1 \in L^1(\mu), \qquad g_n, g \leq C_2 \in L^1(\nu).
$$

Using also $H(\pi|\mu \otimes \nu) \geq 0$, the duality $I_\varepsilon = S_\varepsilon$ from the Introduction, Fatou's lemma and (3.4), we obtain

$$
\begin{aligned}
\inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) \, \pi(dx, dy) &\leq \lim_{n \to \infty} \left(\inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) \, \pi(dx, dy) + \varepsilon_n H(\pi|\mu \otimes \nu) \right) \\
&= \lim_{n \to \infty} \left(\int f_n(x) \, \mu(dx) + \int g_n(y) \, \nu(dy) \right) \\
&\leq \int \limsup_{n \to \infty} f_n(x) \, \mu(dx) + \int \limsup_{n \to \infty} g_n(y) \, \nu(dy) \\
&= \int f(x) \, \mu(dx) + \int g(y) \, \nu(dy) \\
&= \inf_{\pi \in \Pi(\mu,\nu)} \int [f(x) + g(y)] \, \pi(dx, dy) \\
&\leq \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) \, \pi(dx, dy).
\end{aligned}
$$

In particular, $\lim_\varepsilon S_\varepsilon = \int f(x)\,\mu(dx) + \int g(y)\,\nu(dy) = S_0$. Using again (3.5), Fatou's lemma then also shows that

$$S_0/2 = \lim_{\varepsilon \to 0} S_\varepsilon/2 = \lim_{n \to \infty} \int f_n(x)\,\mu(dx) \le \int f(x)\,\mu(dx)$$

and similarly $S_0/2 \le \int g(y)\,\nu(dy)$. We conclude that

$$\int f(x)\,\mu(dx) = \int g(y)\,\nu(dy) = S_0/2$$

and hence the separate convergence

$$\lim_{n \to \infty} \int f_n(x)\,\mu(dx) = \int f(x)\,\mu(dx), \quad \lim_{n \to \infty} \int g_n(x)\,\mu(dx) = \int g(x)\,\mu(dx).$$

In view of (3.5) and the a.s. convergence $f_n \to f$, applying Scheffé's lemma to the nonpositive sequence $f_n - C_1$ allows us to conclude that $f_n \to f$ in $L^1(\mu)$. Similarly, $g_n \to g$ in $L^1(\nu)$. $\qquad \square$

The proof of Theorem 1.1 simplifies substantially if c is uniformly continuous (and in particular if $\mathcal{X}$ and $\mathcal{Y}$ are compact). Moreover, the conclusion is stronger in this case: the almost-sure convergence of $f_n \to f$ and $g_n \to g$ can be replaced by uniform convergence on compact subsets. For the remainder of this section, let $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ be a modulus of continuity as defined before Lemma 2.2.

Lemma 3.1. *Suppose that c is uniformly ω-continuous in both variables. Then the potentials $f_\varepsilon, g_\varepsilon$ are uniformly ω-continuous, for any $\varepsilon > 0$.*

Proof. Let $x_1, x_2 \in \mathcal{X}$ satisfy $f_\varepsilon(x_1) \ge f_\varepsilon(x_2)$. Then

$$|f_\varepsilon(x_1) - f_\varepsilon(x_2)|$$

$$= \varepsilon \log \int e^{\frac{g_\varepsilon(y) - c(x_2,y)}{\varepsilon}}\,\nu(dy) - \varepsilon \log \int e^{\frac{g_\varepsilon(y) - c(x_1,y)}{\varepsilon}}\,\nu(dy)$$

$$= \varepsilon \log \int e^{\frac{c(x_1,y) - c(x_2,y) + g_\varepsilon(y) - c(x_1,y)}{\varepsilon}}\,\nu(dy) - \varepsilon \log \int e^{\frac{g_\varepsilon(y) - c(x_1,y)}{\varepsilon}}\,\nu(dy)$$

$$\le \varepsilon \log \left(e^{\frac{\sup_{y \in \mathcal{Y}} |c(x_1,y) - c(x_2,y)|}{\varepsilon}} \int e^{\frac{g_\varepsilon(y) - c(x_1,y)}{\varepsilon}}\,\nu(dy) \right) - \varepsilon \log \int e^{\frac{g_\varepsilon(y) - c(x_1,y)}{\varepsilon}}\,\nu(dy)$$

$$= \sup_{y \in \mathcal{Y}} |c(x_1,y) - c(x_2,y)| \le \omega(d(x_1,x_2)).$$

The case $f_\varepsilon(x_1) \le f_\varepsilon(x_2)$ follows by symmetry and the proof for g_ε is analogous. $\quad \square$

Proposition 3.2. *Let c be uniformly ω-continuous in both variables and $\varepsilon_n \to 0$. After passing to a subsequence, $f_{\varepsilon_n} \to f$ and $g_{\varepsilon_n} \to g$ uniformly on compact subsets, for some uniformly ω-continuous Kantorovich potentials f and g.*

Proof. The functions (f_n) are w-equicontinuous by Lemma 3.1, hence (f_n) is pointwise bounded as soon as it is bounded at one point $x \in \mathcal{X}$. By Lemma 2.1, $f_n(x) \le \int c(x,y)\,\nu(dy) < \infty$ for μ-a.e. x, so that (f_n^+) is pointwise bounded. On the other hand, $\int f_n^+(x)\,\mu(dx) \le \int c(x,y)\,\nu(dy)\mu(dx) < \infty$, and as $\int f_n(x)\,\mu(dx) \ge 0$ by (2.1), it follows that $\int f_n^-(x)\,\mu(dx)$ is bounded. By equicontinuity, it follows that

(f_n^-) must be bounded at any point $x \in \mathrm{spt}\,\mu$, and then at all points. Similarly for (g_n), and now the claimed convergence to some uniformly ω-continuous functions f, g follows from the Arzelà–Ascoli theorem. To see that f, g are Kantorovich potentials, we argue as in Step 6 of the proof of Theorem 1.1. □

4. Relation to a Large Deviations Principle

In this section we discuss the connection between convergence of potentials (Theorem 1.1) and a large deviations principle (LDP) along the lines proposed in [3, Theorem 1.1]. Roughly speaking, the LDP describes the exponential rate of decay of $\pi_\varepsilon(E)$ for a set E outside the support of $\pi_0 := \lim_{\varepsilon \to 0} \pi_\varepsilon$, whereas the convergence of potentials yields the exponential rate of decay of the density $\pi_\varepsilon / d(\mu \otimes \nu)$ at points outside of $\{f + g = c\}$. Clearly, these statements are closely related, and as seen below, they are equivalent if $\mathcal{X}, \mathcal{Y}$ are compact. In the general case, however, neither result implies the other in an obvious way.

Throughout this section, we fix a sequence $\varepsilon_n \to 0$ and set $(f_n, g_n) := (f_{\varepsilon_n}, g_{\varepsilon_n})$, as in Section 3. Given a measurable function I on $\mathcal{X} \times \mathcal{Y}$, we denote by $\mathrm{ess\,inf}\,I$ the essential infimum wrt. $\mu \otimes \nu$, defined as $\mathrm{ess\,inf}\,I = \inf\{\alpha \in \mathbb{R} : (\mu \otimes \nu)\{I < \alpha\} > 0\}$.

Proposition 4.1. *Suppose that $f := \lim_n f_n$ exists in μ-probability and $g := \lim_n g_n$ exists in ν-probability. Define $I(x, y) := c(x, y) - f(x) - g(y)$, then for any measurable set $E \subset \mathcal{X} \times \mathcal{Y}$,*

$$(4.1) \qquad \liminf_{n \to \infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \geq -\mathrm{ess\,inf}_{(x,y) \in E} I(x, y).$$

If the convergence of (f_n, g_n) is a.s. uniform on E; i.e.,

$$\|(f_n, g_n)\mathbf{1}_E - (f, g)\mathbf{1}_E\|_{L^\infty(\mu \otimes \nu)} \to 0,$$

then E also satisfies the matching bound

$$(4.2) \qquad \limsup_{n \to \infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \leq -\mathrm{ess\,inf}_{(x,y) \in E} I(x, y).$$

If $\mathcal{X}, \mathcal{Y}$ are compact, that is the case for all sets E.

Proof. Let $E \subset \mathcal{X} \times \mathcal{Y}$ be measurable, $\alpha := -\mathrm{ess\,inf}_{(x,y) \in E} I(x, y)$ and $\gamma > 0$. By the definition of α, the set

$$E^\gamma := \{(x, y) \in E : f(x) + g(y) - c(x, y) \geq \alpha - 2\gamma\}$$

satisfies $\beta := (\mu \otimes \nu)(E^\gamma)/2 > 0$. In view of the assumed convergence of (f_n, g_n), there exists n_0 such that $(\mu \otimes \nu)\{|(f_n, g_n) - (f, g)| > \gamma\} \leq \beta$ for $n \geq n_0$, so that

$$E_n^\gamma := \{(x, y) \in E^\gamma : f_n(x) + g_n(y) - c(x, y) \geq \alpha - \gamma\}$$

satisfies $(\mu \otimes \nu)(E_n^\gamma) \geq \beta$ for all $n \geq n_0$. Thus

$$\pi_{\varepsilon_n}(E) \geq \pi_{\varepsilon_n}(E_n^\gamma) = \int_{E_n^\gamma} e^{\frac{f_n(x) + g_n(y) - c(x,y)}{\varepsilon_n}} \mu(dx)\,\nu(dy) \geq \beta e^{\frac{\alpha - \gamma}{\varepsilon_n}}$$

for $n \geq n_0$ and then $\liminf_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \geq \alpha - \gamma$. As $\gamma > 0$ was arbitrary, the lower bound (4.1) follows. Turning to the second claim, note that

$$
\begin{aligned}
\varepsilon_n \log \pi_{\varepsilon_n}(E) &= \varepsilon_n \log \left(\int_E e^{\frac{f_n(x) + g_n(y) - c(x,y)}{\varepsilon_n}} \, \mu(dx) \, \nu(dy) \right) \\
&\leq \operatorname*{ess\,sup}_{(x,y)\in E} \left(f_n(x) + g_n(y) - c(x,y) \right).
\end{aligned}
$$

If $\|(f_n, g_n)\mathbf{1}_E - (f,g)\mathbf{1}_E\|_\infty \to 0$, it readily follows that

$$
\lim_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \leq - \operatorname*{ess\,inf}_{(x,y)\in E} I(x,y),
$$

as desired. If $\mathcal{X}, \mathcal{Y}$ are compact, Proposition 3.2 and the assumed convergence of the potentials in probability imply that $\|(f_n, g_n) - (f,g)\|_\infty \to 0$ (without taking a subsequence), so that the above applies to any measurable set E. $\qquad\square$

Remark 4.2. (a) In Proposition 4.1 the rate is stated through an essential infimum, consistent with the fact that E can be irregular and f, g are considered as determined only up to nullsets. In many situations it is known that Kantorovich potentials admit a continuous version, for instance by c-concavity. If moreover E is suitably regular (e.g., open and contained in $\operatorname{spt}\mu \times \operatorname{spt}\nu$), the essential infimum can be written as an infimum.

(b) In the case of compactly supported marginals, an alternative proof of Proposition 4.1 can be given using Bryc's inverse to Varadhan's Integral Lemma; cf. [13, Theorem 4.4.2]. That proof, however, is longer than the direct argument given above. In connection with classical large deviations theory, we note that the sequence (π_{ε_n}) fails to be exponentially tight whenever the marginals are not compactly supported: exponential tightness implies, in particular, that any limit π_0 is compactly supported, but as $\pi_0 \in \Pi(\mu, \nu)$, the same then follows for μ, ν.

If the Kantorovich potentials (f, g) are unique, Theorem 1.1 implies that the first condition in Proposition 4.1 is satisfied.

Bounds similar to (4.1) and (4.2) are stated in [3, Theorem 1.1] for open and compact sets, respectively. While weak convergence $\pi_{\varepsilon_n} \to \pi_0$ of the couplings is assumed, it avoids any conditions on $\mathcal{X}, \mathcal{Y}$, the integrability of c, or even the finiteness of the value I_ε in (1.2). Such a setting does seem outside the scope of the methods used here.

In general, if convergence of potentials is not known a priori, Proposition 4.1 implies non-matching bounds by maximizing or minimizing over all potentials as follows. Given a family (I_λ) of measurable functions, $I^* := \operatorname{ess\,sup}_\lambda I_\lambda$ denotes the essential supremum wrt. $\mu \otimes \nu$ in the sense of probability theory.[2] Similarly, $\operatorname{ess\,inf}_\lambda I_\lambda$ is the essential infimum.

---

[2]I.e., I^* is the (a.s. unique) measurable function satisfying $I^* \geq I_\lambda$ a.s. for all λ and $I^* \leq J$ a.s. for any J satisfying $J \geq I_\lambda$ a.s. for all λ. In other words, I^* is the supremum in the lattice of measurable functions equipped with the a.s. order.

Corollary 4.3. *Define $I^* := \operatorname{ess\,sup}_{f,g} I_{f,g}$, where $I_{f,g}(x,y) := c(x,y) - f(x) - g(y)$ and the supremum is taken over all Kantorovich potentials (f,g). Similarly, define $I_* := \operatorname{ess\,inf}_{f,g} I_{f,g}$. Then*

$$(4.3) \qquad \liminf_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \geq - \operatorname*{ess\,inf}_{(x,y)\in E} I^*(x,y)$$

for any measurable set $E \subset \mathcal{X} \times \mathcal{Y}$. If $\mathcal{X}, \mathcal{Y}$ are compact, then also

$$(4.4) \qquad \limsup_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(E) \leq - \operatorname*{ess\,inf}_{(x,y)\in E} I_*(x,y).$$

Proof. Passing to a subsequence, we may assume that the $\liminf$ on the left-hand side is a limit. After passing to another subsequence, Theorem 1.1 yields that the Schrödinger potentials (f_n, g_n) converge in L^1 to some Kantorovich potentials (f,g), and then Proposition 4.1 applies to (f,g). As $\operatorname{ess\,inf}_{(x,y)\in E} I_{f,g}(x,y) \leq \operatorname{ess\,inf}_{(x,y)\in E} I^*(x,y)$, the lower bound (4.3) follows. The proof of (4.4) is analogous. $\qquad\square$

Remark 4.4. The lower bound (4.3) is quite general, and seems to be novel. Except in the case of uniqueness for the Kantorovich potentials, no analogue is stated in [3]. On the other hand, the upper bound (4.4) is similar to the bound in [3, Theorem 1.1 (a)]. The latter is stated under the condition that π_{ε_n} converges but without any conditions on $\mathcal{X}, \mathcal{Y}$.

The next result is a partial converse to Proposition 4.1. It suggests that if an LDP holds, then the Schrödinger potentials must converge (without passing to a subsequence) and the rate function must be determined by the limiting potentials. We prove this in the compact case via Varadhan's Integral Lemma, but we conjecture that the assertions remains valid in some generality.

Proposition 4.5. *Let $\mathcal{X}, \mathcal{Y}$ be compact and suppose the assertion of the LDP [3, Theorem 1.1] holds for some function $I : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$; that is,*

$$(4.5) \qquad \limsup_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(C) \leq - \inf_{(x,y)\in C} I(x,y) \quad \text{for } C \subset \mathcal{X} \times \mathcal{Y} \text{ compact,}$$

$$(4.6) \qquad \liminf_{n\to\infty} \varepsilon_n \log \pi_{\varepsilon_n}(U) \geq - \inf_{(x,y)\in U} I(x,y) \quad \text{for } U \subset \operatorname{spt}\mu \times \operatorname{spt}\nu \text{ open.}$$

Then

$$I(x,y) = c(x,y) - f(x) - g(y), \quad (x,y) \in \operatorname{spt}\mu \times \operatorname{spt}\nu$$

for some Kantorovich potentials (f,g), and

$$f = \lim_{n\to\infty} f_n \quad \text{uniformly on } \operatorname{spt}\mu, \qquad g = \lim_{n\to\infty} g_n \quad \text{uniformly on } \operatorname{spt}\nu.$$

Proof. As $\mathcal{X} \times \mathcal{Y}$ is compact, c is uniformly continuous and then f_n, g_n are uniformly equicontinuous; cf. Lemma 3.1. Fix $(x_0, y_0) \in \operatorname{spt}\mu \times \operatorname{spt}\nu$. Equicontinuity implies that given $\gamma > 0$ there exists $r, n_0 > 0$ such that for all $n \geq n_0$,

$$|I(x_0,y_0) + f_n(x_0) + g_n(y_0) - c(x_0,y_0) - J_n(r)| \leq \gamma \quad \text{for}$$

$$J_n(r) := \varepsilon_n \log \left(\int_{B_r(x_0,y_0)} e^{\frac{I(x,y) + f_n(x) + g_n(y) - c(x,y)}{\varepsilon_n}} \mu(dx)\,\nu(dy) \right).$$

To show $\lim_{n\to\infty}[f_n(x_0)+g_n(y_0)] = c(x_0,y_0)-I(x_0,y_0)$, it therefore suffices to prove for all $r > 0$ that

$$(4.7) \qquad \lim_{n\to\infty} J_n(r) = 0.$$

Next, we argue that I must be continuous. Indeed, after passing to a subsequence, Proposition 4.1 shows that I must be of the form $I = c - \tilde{f} - \tilde{g}$ on $\operatorname{spt}\mu \times \operatorname{spt}\nu$, for some (necessarily uniformly continuous) Kantorovich potentials $(\tilde{f},\tilde{g})$. Moreover, we may assume that $\mathcal{X} = \operatorname{spt}\mu$ and $\mathcal{Y} = \operatorname{spt}\nu$, by shrinking the marginal spaces if necessary. In brief, the LDP (4.5), (4.6) then holds for all closed sets C and open sets U in $\mathcal{X}\times\mathcal{Y}$ with the "good" rate function I. In this context, Varadhan's Integral Lemma [13, Theorem 4.3.1] states that

$$(4.8) \qquad \lim_{n\to\infty} \varepsilon_n \log\left(\int e^{\frac{\phi(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,dy)\right) = \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (\phi(x,y) - I(x,y))$$

for any continuous function $\phi : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ that satisfies the moment condition

$$\limsup_{n\to\infty} \varepsilon_n \log\left(\int e^{\frac{\gamma\phi(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,dy)\right) < \infty$$

for some $\gamma > 1$. As the continuous function I is bounded on the compact space $\mathcal{X}\times\mathcal{Y}$, this holds in particular for $\phi := I$, for any $\gamma > 1$. Let $(x_0,y_0) \in \mathcal{X}\times\mathcal{Y}$ and $r > 0$. Using (4.8) for $\phi = I$,

$$\limsup_{n\to\infty} J_n(r) \le \limsup_{n\to\infty} \varepsilon_n \log\left(\int e^{\frac{I(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,y)\right)$$
$$= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (I(x,y) - I(x,y)) = 0.$$

To show the converse inequality, consider a bounded continuous function ϕ with

$$\phi(x_0,y_0) = I(x_0,y_0), \qquad \phi \le I \text{ on } B_r(x_0,y_0), \qquad \phi = -1 \text{ on } B_r^c(x_0,y_0).$$

Then

$$\int_{B_r(x_0,y_0)} e^{\frac{I(x,y)+f_n(x)+g_n(y)-c(x,y)}{\varepsilon_n}} \mu(dx)\,\nu(dy) \ge \int_{B_r(x_0,y_0)} e^{\frac{\phi(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,dy)$$
$$\ge \int e^{\frac{\phi(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,dy) - e^{\frac{-1}{\varepsilon_n}}$$

and thus

$$\liminf_{n\to\infty} J_n(r) \ge \liminf_{n\to\infty} \varepsilon_n \log\left(\int e^{\frac{\phi(x,y)}{\varepsilon_n}} \pi_{\varepsilon_n}(dx,dy)\right)$$
$$= \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (\phi(x,y) - I(x,y)) = 0,$$

where we have used (4.8). This completes the proof of (4.7) and thus shows that $\lim_{n\to\infty}[f_n(x_0) + g_n(y_0)] = c(x_0,y_0) - I(x_0,y_0)$ for $(x_0,y_0) \in \operatorname{spt}\mu \times \operatorname{spt}\nu$. In view of the uniform equicontinuity, the convergence is even uniform on that set.

On the other hand, we have already shown in Proposition 3.2 that $f_n \to f$ and $g_n \to g$ uniformly, after passing to a subsequence, for some Kantorovich potentials

f, g. Thus $c - I = f + g$ on $\operatorname{spt} \mu \times \operatorname{spt} \nu$. It remains to argue that the original sequences f_n, g_n converge to f, g. Indeed, the rectangular form of $S := \operatorname{spt} \mu \times \operatorname{spt} \nu$ implies that if $f(x) + g(y) = \tilde{f}(x) + \tilde{g}(y)$ on S, then $\tilde{f} = f + a$ and $\tilde{g} = g - a$ for some $a \in \mathbb{R}$. Recalling our symmetric normalization for potentials, the claim follows. □

5. Varying Costs

In this section we extend Theorem 1.1 to cost functions that vary with ε. The continuous cost c will be used for the limiting Monge–Kantorovich transport problem, as before. In addition, we introduce a family of cost functions $c_\varepsilon : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ for the regularized problems with $\varepsilon > 0$. These functions are merely required to be measurable.

On the one hand, we are interested in the stability of Theorem 1.1 with respect to the cost function. On the other hand, this section is motivated by the large deviations perspective on Schrödinger bridges; cf. [19]. Recall that

$$(5.1) \qquad \pi_\varepsilon = \operatorname*{arg\,min}_{\pi \in \Pi(\mu, \nu)} H(\pi | R_\varepsilon) \quad \text{for} \quad \frac{dR_\varepsilon}{d(\mu \otimes \nu)} = \alpha_\varepsilon e^{-c/\varepsilon}$$

where α_ε is the normalizing constant. Theorem 1.1 and its counterparts in Section 4 can be interpreted as consequences of the large deviations of (R_ε) as $\varepsilon \to 0$, whose rate is the function c. More generally, this rate function is shared by arbitrary measures (R'_ε) with $-\varepsilon \log \frac{dR'_\varepsilon}{d(\mu \otimes \nu)} \to c$, and one may wonder if they give rise to a similar result. This convergence is equivalent to setting $\frac{dR'_\varepsilon}{d(\mu \otimes \nu)} = \alpha'_\varepsilon e^{-c_\varepsilon/\varepsilon}$ for some function c_ε with $c_\varepsilon \to c$, and returning to the language of entropic optimal transport, it corresponds to the cost c_ε under consideration.

In what follows, we assume a common bound

$$(5.2) \qquad\qquad c_\varepsilon \le \bar{c} \quad \text{for all} \quad \varepsilon > 0$$

for some function $\bar{c}(x, y) = \bar{c}_1(x) + \bar{c}_2(y)$ with $\bar{c}_1 \in L^1(\mu)$ and $\bar{c}_2 \in L^1(\nu)$, and that

$$(5.3) \qquad\qquad c_\varepsilon \to c \quad \text{uniformly on compact subsets as } \varepsilon \to 0.$$

The modified entropic optimal transport problem then reads

$$(5.4) \qquad\qquad I_\varepsilon := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c_\varepsilon(x, y)\, \pi(dx, dy) + \varepsilon H(\pi | \mu \otimes \nu).$$

As before, it has a unique solution π_ε, and we introduce the Schrödinger potentials through the formula

$$(5.5) \qquad\qquad \frac{d\pi_\varepsilon}{d(\mu \otimes \nu)}(x, y) = \exp\left(\frac{f_\varepsilon(x) + g_\varepsilon(y) - c_\varepsilon(x, y)}{\varepsilon} \right)$$

and the symmetric normalization (1.4). The Monge–Kantorovich problem and its potentials are still based on the continuous cost c. While not required for the regularized problem with $\varepsilon > 0$, continuity of costs is important for $\varepsilon = 0$, including for the validity of Theorem 1.1 (see Example 5.2).

Proposition 5.1. *Let* (5.2), (5.3) *hold. Then the assertion of Theorem 1.1 extends to the setting* (5.4), (5.5) *of variable costs* (c_ε).

Proof. We only indicate the necessary changes to the proof of Theorem 1.1. First of all, we recall that the auxiliary results in Section 2 did not require continuity. Next, we go through the steps in Section 3.

Step 1. We change (3.1) to

$$|f_n(x_1) - f_n(x_2)| \leq \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}(\delta_k)} |c_n(x_1, y) - c_n(x_2, y)| - \varepsilon_n \log(1 - \delta_k)$$

$$\leq \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}(\delta_k)} |c(x_1, y) - c(x_2, y)| + \varepsilon_n \log(2) + \eta_{n,k}$$

where, due to the uniform convergence of c_ε on the compact set $\mathcal{X}_{\mathrm{cpt}}(\delta_k) \times \mathcal{Y}_{\mathrm{cpt}}(\delta_k)$, the constant $\eta_{n,k}$ satisfies $\lim_n \eta_{n,k} = 0$ (for fixed k). The subsequent display for g_n is changed analogously.

Step 2. Instead of (3.2) we now have

$$\left| f_n^k(x_1) - f_n^k(x_2) \right| \leq \tilde{d}_k(x_1, x_2) + \varepsilon_n \log(2) + \eta_{n,k}, \quad x_1, x_2 \in \mathcal{X}.$$

Step 3. In the arguments for the pointwise boundedness, simply replace c by $\bar{c}$. In the application of Lemma 2.2, replace $h_n := \varepsilon_n \log(2)$ by $h_{n,k} := \varepsilon_n \log(2) + \eta_{n,k}$. Note that the dependence on k does not cause any difficulty, as k is fixed and $\lim_n h_{n,k} = 0$ holds for each k.

Steps 4,5. No changes are necessary in these steps; note that (3.4) is based solely on the limiting cost function c which is still assumed to be continuous.

Step 6. Define $C_1(x) := \int \bar{c}(x, y)\, \nu(dy) = \bar{c}_1(x) + \|c_2\|_{L^1(\nu)}$ and similarly $C_2(y) := \bar{c}_2(y) + \|c_1\|_{L^1(\mu)}$. Then we again have (3.5). For the subsequent display, we now need to argue that

$$(5.6) \qquad \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y)\, \pi(dx, dy) \leq \lim_{n \to \infty} \inf_{\pi \in \Pi(\mu,\nu)} \int c_n(x, y)\, \pi(dx, dy).$$

Indeed, given $\gamma > 0$, we can find a compact set $K = K_1 \times K_2 \subset \mathcal{X} \times \mathcal{Y}$ with

$$\int_{K^c} \bar{c}(x, y)\, \pi(dx, dy) \leq \int_{K_1^c} \bar{c}_1(x)\, \mu(dx) + \int_{K_2^c} \bar{c}_2(y)\, \nu(dy) < \gamma.$$

As $c_n \to c$ uniformly on K, we also have $|c - c_n| \leq \gamma$ on K for $n \geq n_0$. Thus

$$\left| \inf_{\pi \in \Pi(\mu,\nu)} \int c\, d\pi - \inf_{\pi \in \Pi(\mu,\nu)} \int c_n\, d\pi \right| \leq \sup_{\pi \in \Pi(\mu,\nu)} \int |c - c_n|\, d\pi$$

$$\leq \sup_{\pi \in \Pi(\mu,\nu)} \int_K |c - c_n|\, d\pi + \int_{K^c} \bar{c}\, d\pi \leq 2\gamma$$

for $n \geq n_0$. This implies (5.6), even with equality, and the remainder of the proof holds as stated without further changes. $\qquad \square$

The following simple example shows that continuity of c is important for the validity of Theorem 1.1.

Example 5.2. Let $\mu = \nu$ be uniform on $\mathcal{X} = \mathcal{Y} = [0, 1]$ and $c(x, y) = \mathbf{1}_{x \neq y}$. Then the Schrödinger potentials are $f_\varepsilon = g_\varepsilon \equiv 1/2$ for all $\varepsilon > 0$ but the (unique) Kantorovich potentials are $f_0 = g_0 \equiv 0$.

To put the example in a broader context, note that the entropic optimal transport problem (1.2) with $\varepsilon > 0$ remains unchanged if the cost function is altered on a $\mu \otimes \nu$-nullset, whereas the Monge–Kantorovich problem may very well change. If c is measurable and $\hat{c}$ is a continuous function with $\hat{c} = c$, Theorem 1.1 thus implies that the entropic problem (1.2) with cost c converges to the Monge–Kantorovich problem with cost $\hat{c}$ for $\varepsilon \to 0$. Example 5.2 is a particular case with $c(x, y) = \mathbf{1}_{x \neq y}$ and $\hat{c} \equiv 1$. For more general cost functions, one may conjecture that (1.2) converges to some form of upper envelope of the Monge–Kantorovich problem; we leave this question for future research.

6. Multimarginal Optimal Transport

Instead of two marginals μ and ν, we can generalize to an arbitrary number $N \in \mathbb{N}$ of marginals. Consider Polish probability spaces $(\mathcal{X}_i, \mu_i)$ for $i = 1, \ldots, N$ and let

$$\boldsymbol{\mu}(dx_1, \ldots, dx_N) := \mu_1(dx_1) \otimes \cdots \otimes \mu_N(dx_N)$$

denote the product measure. Moreover, let $c : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to \mathbb{R}_+$ be continuous with $\int c \, d\boldsymbol{\mu} < \infty$. The entropic optimal transport problem generalizes directly to the set $\pi \in \Pi(\mu_1, \ldots, \mu_N)$ of couplings,

$$(6.1) \qquad I_\varepsilon := \inf_{\pi \in \Pi(\mu_1, \ldots, \mu_N)} \int c \, \pi + \varepsilon H(\pi | \boldsymbol{\mu}),$$

and has a unique solution π_ε given by

$$(6.2) \qquad \frac{d\pi_\varepsilon}{d\boldsymbol{\mu}}(x_1, \ldots, x_N) = \exp \left(\frac{f_\varepsilon^1(x_1) + \cdots + f_\varepsilon^N(x_N) - c(x_1, \ldots, x_N)}{\varepsilon} \right)$$

with $f_\varepsilon^i \in L^1(\mu_i)$. For $\varepsilon = 0$, we again recover the multimarginal optimal transport problem, whose dual now reads

$$(6.3) \qquad S_0 := \sup_{f^i \in L^1(\mu_i), \, \sum_i f^i(x_i) \leq c(x_1, \ldots, x_N)} \sum_{i=1}^N \int f^i(x_i) \, \mu_i(dx_i).$$

We again normalize all potentials symmetrically. Extending Theorem 1.1, we have the following result.

Theorem 6.1. Let $(f_\varepsilon^1, \ldots, f_\varepsilon^N)$ be the unique Schrödinger potentials for $\varepsilon > 0$.

- (a) Given $\varepsilon_n \to 0$, there is a subsequence (ε_k) such that $f_{\varepsilon_k}^i$ converges in $L^1(\mu_i)$, for all $i = 1, \ldots, N$.
- (b) If $\lim_n f_{\varepsilon_n}^i = f^i$ μ_i-a.s. for all $i = 1, \ldots, N$, then $(f^1, \ldots, f^N)$ are Kantorovich potentials and the convergence also holds in $L^1(\mu_i)$.

If the Kantorovich potentials $(f_0^1, \ldots, f_0^N)$ for (6.3) are unique, then it follows that $\lim_{\varepsilon \to 0} f_\varepsilon^i = f_0^i$ in $L^1(\mu_i)$ for $i = 1, \ldots, N$.

Proof. The arguments are exactly the same as in the proof of Theorem 1.1, and therefore omitted. □

References

[1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.

[2] R. J. Berman. The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations. *Numer. Math.*, 145(4):771–836, 2020.

[3] E. Bernton, P. Ghosal, and M. Nutz. Entropic optimal transport: Geometry and large deviations. *Preprint arXiv:2102.04397v1*, 2021.

[4] A Beurling. An automorphism of product measures. *Ann. of Math. (2)*, 72:189–200, 1960.

[5] J. M. Borwein and A. S. Lewis. Decomposition of multivariate functions. *Canad. J. Math.*, 44(3):463–482, 1992.

[6] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum. Entropy minimization, DAD problems, and doubly stochastic kernels. *J. Funct. Anal.*, 123(2):264–307, 1994.

[7] G. Carlier. On the linear convergence of the multi-marginal Sinkhorn algorithm. *Preprint hal-03176512*, 2021.

[8] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[9] R. Cominetti and J. San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Programming*, 67(2, Ser. A):169–187, 1994.

[10] G. Conforti and L. Tamanini. A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.*, 280(11):108964, 2021.

[11] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probability*, 3:146–158, 1975.

[12] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26*, pages 2292–2300. 2013.

[13] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[14] S. Di Marino and A. Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.*, 85(2):Paper No. 27, 28, 2020.

[15] H. Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Math.*, pages 101–203. Springer, Berlin, 1988.

[16] H. Föllmer and N. Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Probab.*, 25(2):901–926, 1997.

[17] N. Gigli and L. Tamanini. Second order differentiation formula on RCD$^*(K, N)$ spaces. *J. Eur. Math. Soc. (JEMS)*, 23(5):1727–1795, 2021.

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

[19] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[20] G. Mena and J. Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems 32*, pages 4541–4551. 2019.

[21] M. Nutz. *Introduction to Entropic Optimal Transport*. Lecture notes, Columbia University, 2020. https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

[22] S. Pal. On the difference between entropic cost and the optimal transport cost. *Preprint arXiv:1905.12206v1*, 2019.

[23] T. Pennanen and A.-P. Perkkiö. Convex duality in nonlinear optimal transport. *J. Funct. Anal.*, 277(4):1029–1060, 2019.

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

[25] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I-projections. *Statist. Probab. Lett.*, 17(5):369–375, 1993.

[26] L. Rüschendorf and W. Thomsen. Closedness of sum spaces and the generalized Schrödinger problem. *Teor. Veroyatnost. i Primenen.*, 42(3):576–590, 1997.

[27] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

(MN) Departments of Statistics and Mathematics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA
  Email address: mnutz@columbia.edu

(JW) Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA
  Email address: johannes.wiesel@columbia.edu