STABILITY OF SCHRÖDINGER POTENTIALS
AND CONVERGENCE OF SINKHORN'S ALGORITHM

MARCEL NUTZ AND JOHANNES WIESEL

ABSTRACT. We study the stability of entropically regularized optimal transport with respect to the marginals. Given marginals converging weakly, we establish a strong convergence for the Schrödinger potentials describing the density of the optimal couplings. When the marginals converge in total variation, the optimal couplings also converge in total variation. This is applied to show that Sinkhorn's algorithm converges in total variation when costs are quadratic and marginals are subgaussian, or more generally for all continuous costs satisfying an integrability condition.

1. INTRODUCTION

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be Polish probability spaces and $\Pi(\mu, \nu)$ the set of all couplings; that is, probability measures π on $\mathcal{X} \times \mathcal{Y}$ with marginals (μ, ν). Moreover, let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ be continuous. The entropic optimal transport problem with regularization parameter $\varepsilon \in (0, \infty)$ is

$$(1.1) \qquad \mathcal{C}_\varepsilon(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y)\, \pi(dx, dy) + \varepsilon H(\pi | \mu \otimes \nu),$$

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

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

Entropic optimal transport traces back to the Schrödinger bridge problem associated with Schrödinger's thought experiment [34] on the most likely evolution of a particle system, see [18, 25] for surveys. More recently, popularized by [14], the problem has received immense interest as an approximation of the (unregularized) Monge–Kantorovich optimal transport problem corresponding to $\varepsilon = 0$, especially for computing the 2-Wasserstein distance in high-dimensional applications such as machine learning, statistics, image and language processing (e.g., [1, 2, 10, 32]). As a result, the cost of principal interest is $c(x, y) = \|x - y\|^2$

on $\mathbb{R}^d \times \mathbb{R}^d$ and the convergence properties as $\varepsilon \to 0$ have been studied in detail; see [3, 6, 9, 12, 13, 21, 24, 26, 27, 29, 30, 37], among others. The main appeal of (1.1) in this computational context is that it can be solved efficiently and at large scale by Sinkhorn's algorithm; see [31] and its numerous references. The algorithm is initialized at the probability measure $\pi_{-1} \propto e^{-c/\varepsilon} d(\mu \otimes \nu)$ and its iterates are defined for $t \geq 0$ by

$$(1.2) \qquad \pi_{2t} := \underset{\Pi(*,\nu)}{\arg\min}\, H(\,\cdot\,|\pi_{2t-1}), \qquad \pi_{2t+1} := \underset{\Pi(\mu,*)}{\arg\min}\, H(\,\cdot\,|\pi_{2t}),$$

where $\Pi(*, \nu)$ is the set of measures on $\mathcal{X} \times \mathcal{Y}$ with second marginal ν (and arbitrary first marginal), and $\Pi(\mu, *)$ is defined analogously. The algorithm alternatingly "fits" the marginals μ and ν, hence is also called iterative proportional fitting procedure (IPFP). The argmin in (1.2) can be solved explicitly, and then each step of the algorithm only requires an explicit integration against $e^{-c/\varepsilon}$, cf. Section 3. The algorithm dates back as far as [16] and its convergence properties are well studied when the cost c is bounded: the convergence of π_n to the solution π_* of (1.1) holds in total variation (and in relative entropy), see [7, 9, 19, 22, 23, 33, 35, 36], among others. More precisely, [33] relaxed the boundedness condition, but introduced several other conditions, including one (see (B1) in [33]) that essentially forces c to be bounded from above in one variable and thus excludes quadratic cost with unbounded marginal supports.

One main result of this paper is the convergence $\|\pi_n - \pi_*\|_{TV} \to 0$ of Sinkhorn's algorithm (1.2) for quadratic cost and arbitrary subgaussian marginals. More generally, our result (see Corollary 3.2) holds for the continuous cost c and marginals (μ, ν) as soon as

$$(1.3) \qquad e^{\beta c} \in L^1(\mu \otimes \nu) \text{ for some } \beta > 0.$$

A fairly elementary proof of the convergence $\|\pi_n - \pi_*\|_{TV} \to 0$, following the same idea of weak-star compactness as [33], was recently given in [28] under the condition that (1.3) holds for some $\beta > \varepsilon^{-1}$. We emphasize that this condition does not capture the regime of principal interest: when approximating the 2-Wasserstein distance and marginals are standard Gaussians, say, this condition forces $\varepsilon > 1/2$, but the approximation often requires ε to be several orders of magnitude smaller (e.g., [31]). While the case $\beta > \varepsilon^{-1}$ is broadly similar to the case of bounded cost, it seems that the regime $\beta < \varepsilon^{-1}$ requires a fundamentally different line of attack—which brings us to the main theorem of this paper.

In all that follows, we focus on $\varepsilon = 1$ in (1.1) for notational simplicity; the general case is easily retrieved by replacing c with c/ε. Assuming that

$$(1.4) \qquad \mathcal{C}(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y)\, \pi(dx,dy) + H(\pi|\mu \otimes \nu)$$

is finite, there is a unique solution $\pi_* \in \Pi(\mu, \nu)$, and π_* is uniquely characterized within $\Pi(\mu, \nu)$ by having a density of the form

$$\frac{d\pi_*}{d(\mu \otimes \nu)} = e^{f \oplus g - c} \quad \mu \otimes \nu\text{-a.s.}$$

for some measurable functions $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$, where we write $f \oplus g$ for $(x, y) \mapsto f(x) + g(y)$. The functions f, g are called called *(Schrödinger) potentials* and they are integrable (under μ and ν, respectively) in the cases relevant below; a sufficient condition for integrability is $c \in L^1(\mu \otimes \nu)$. The sum $f \oplus g$ is uniquely determined $\mu \otimes \nu$-a.s., whereas the individual functions are unique up to an additive constant: clearly $(f + a, g - a)$ have the same sum for any $a \in \mathbb{R}$. Thus, they are unique after choosing a normalization removing this degree of freedom. See Appendix A for the preceding facts and further background.

Our aim is to establish stability of the potentials with respect to the marginals. Consider sequences $\mu_n \to \mu$ and $\nu_n \to \nu$ of marginals converging weakly (i.e., in the topology induced by bounded continuous functions). Denoting by (f_n, g_n) associated potentials, we want to state that $(f_n, g_n) \to (f, g)$ in a suitable sense. In view of the above characterization, a key step in this endeavor is to establish a form of compactness. In general, it is not straightforward how to formalize the convergence of potentials. E.g., in a computational context, we may be interested in discrete measures (μ_n, ν_n) approximating a continuous pair (μ, ν). Then, these measures are mutually singular and the spaces $L^p(\mu_n)$ and $L^p(\mu)$ are not immediately comparable. If $\mu \ll \mu_n$ (and similarly for ν_n), this issue is milder as convergence in μ-probability yields a natural topology. However, compactness still turns out to be an issue in the regime of interest. If c has high integrability, the weak-star topology in L^1 can be used, similarly to the arguments for Sinkhorn convergence in [33]. In some cases one can even use the Arzelà–Ascoli theorem (see Appendix B). But in the regime of interest here, where we want to cover quadratic cost and subgaussian marginals, we have not succeeded with off-the-shelf compactness concepts.

Instead, we shall build compactness through an approximation scheme and properties specific to the problem at hand, eventually using compactness of bounded sets in Euclidean space. This construction is the main technical contribution of the paper. It will also allow us to cover the case of mutually singular measures (in fact, focusing on equivalent measures would not result in a substantial simplification). The approximation scheme has the form

$$
\begin{array}{ccc}
F_n^k & \xrightarrow{\ n \to \infty\ } & F^k \\
{\scriptstyle k \to \infty}\downarrow & & \downarrow{\scriptstyle k \to \infty} \\
f_n & \dashrightarrow & f
\end{array}
$$

where for fixed n, the potential f_n is approximated by a sequence $(F_n^k)_{k \in \mathbb{N}}$ converging in μ_n-probability, with some additional uniformity in n. The functions F_n^k are piecewise constant; more precisely, they are simple functions based on a partition $(D_j^k)_{j \in \mathbb{N}}$ of $\mathcal{X}$ and have nonzero values on finitely many sets D_j^k. Using specific regularity properties of the potentials, these sets can be chosen independently of n. Therefore, $(F_n^k)_{n \in \mathbb{N}}$ can be identified with a sequence in a finite-dimensional space and compactness can be established from a priori bounds. Passing to a subsequence, this results in the (uniform) convergence $F_n^k \to F^k$ at the top of the diagram. The limiting functions F^k, in turn, are shown to form a Cauchy sequence in μ-probability,

thus yielding a limit f that is well-defined under μ. A similar construction is applied to (g_n), yielding a function g, and we shall prove that (f, g) are indeed potentials for the limiting marginals (μ, ν).

The scheme in the diagram also acts as a way to formalize a strong convergence $f_n \to f$. It implies convergence in distribution; that is, $(f_n)_{\#}\mu_n \to f_{\#}\mu$ weakly, where $f_{\#}$ denotes the pushforward under f. Convergence in distribution is a natural notion given the weak convergence setting, but it is far from strong enough to imply the desired conclusions. If $\mu \ll \mu_n$, we show that our scheme implies the convergence in μ-probability (and similarly for ν) under a fairly general condition on the Radon–Nikodym derivatives; cf. Corollary 2.4. This condition is satisfied in particular whenever the marginals convergence in total variation, allowing us to deduce via Scheffé's lemma a result of its own interest (Corollary 2.6): the optimal couplings are stable in total variation; i.e., for marginals with $\|\mu_n - \mu\|_{TV} \to 0$ and $\|\nu_n - \nu\|_{TV} \to 0$, the corresponding optimizers satisfy $\|\pi_n - \pi_*\|_{TV} \to 0$. Returning to the convergence of Sinkhorn's algorithm, we interpret each iteration of the algorithm as solving an entropic optimal transport problem with changing marginals (μ_n, ν_n). These marginals converge in total variation to (μ, ν), and we infer the convergence of the algorithm to the desired limit π_*.

Several recent works have addressed the stability of entropic optimal transport from different angles. The first result is in [8], for a setting with bounded cost and marginals equivalent to a common reference measure with densities uniformly bounded above and below. The authors show by a differential approach that the potentials are continuous in L^p relative to the marginal densities. Still with bounded cost (and some other conditions), [15] establishes uniform continuity of the potentials relative to the marginals in Wasserstein distance W_1; this result is based on the Hilbert–Birkhoff projective metric. Closer to the present unbounded setting, [20] obtains stability of the optimal couplings in weak convergence for general continuous costs. Based on the geometric approach first proposed in [4], the main restriction of the technique is that the underlying spaces need to satisfy Lebesgue's theorem on differentiation of measures which generally holds only in finite-dimensional spaces. As a by-product, the main result of the present paper yields a similar stability result for weak convergence; cf. Theorem 2.1 (i). The present result also applies in an infinite-dimensional context; the more important difference, however, is that we achieve a strong form of convergence, whereas [20] is silent about any convergence of the densities or potentials. In particular, we can infer stability in the sense of total variation convergence (Corollary 2.6) and the corresponding convergence of Sinkhorn's algorithm (Corollary 3.2). It is worth noting that these two stabilities are at opposites ends of the spectrum: in the weak topology, compactness for sets of couplings is immediate; the difficulty is to ensure that a limit is the optimal coupling for its marginals. For a limit in total variation, the latter is easy, but obtaining compactness is difficult due to the strength of the topology. In the present work, we effectively reduce the dimension by focusing on densities with a decomposition given by potentials and then obtain compactness through the potentials. A last

related work is [17], which was conducted concurrently. Here stability of the coupling in Wasserstein distance W_p is shown under certain growth and integrability conditions. Obtained by control-theoretic arguments through a transport inequality, the main strength of this result lies in being quantitative (which the present one is not). On the other hand, [17] is once again silent about the densities or potentials, and does not yield a convergence in total variation. Indeed, we are not aware of previous stability results in total variation beyond bounded settings. Finally, we would like to mention the ongoing research [11] kindly pointed out to us by Giovanni Conforti. In the setting of dynamic Schrödinger bridges satisfying a logarithmic Sobolev inequality for the underlying dynamics and marginal distributions with finite Fisher information, the authors study quantitative bounds for the relative entropy of Schrödinger bridges with different marginals and the convergence of the gradients of the potentials towards the Brenier map as $\varepsilon \to 0$.

The remainder of this paper is organized as follows. Section 2 contains the main results on stability. Section 3 details the application to Sinkhorn's algorithm. The proof of the main result, Theorem 2.1, is split into ten steps which are reported in Section 4. For convenience, Appendix A summarizes background on entropic optimal transport. Appendix B details how stability, even uniformly on compacts, can be obtained rather directly under strong integrability conditions. Lastly, Appendix C contains some proofs that we defer in the body of the text.

2. Stability

Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces endowed with their Borel σ-fields and $\mathcal{P}(\mathcal{X}), \mathcal{P}(\mathcal{Y})$ their sets of Borel probability measures. We recall that $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is continuous. In Theorem 2.1 below, we consider the entropic optimal transport problem (1.4) for marginals $(\mu_n, \nu_n) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ converging to marginals (μ, ν). The condition (2.1) of the theorem implies that $\mathcal{C}(\mu_n, \nu_n) < \infty$ and that there exist optimal couplings $\pi_n \in \Pi(\mu_n, \nu_n)$ with associated potentials (f_n, g_n); see Section A for these facts and further background.

Before stating the theorem, let us comment on the normalization chosen therein. As mentioned in the Introduction, f_n and g_n are only unique up to an additive constant. This does not affect the sum $f_n \oplus g_n$ determining the density of π_n, but in order to obtain a separate convergence for f_n and g_n, it is clearly necessarily to impose an additional condition to pin down this constant. There are many possible choices; in Theorem 2.1, we work with $\alpha_n := \int \arctan(f_n) \, d\mu_n$. As arctan is strictly increasing, fixing the value of the integral is equivalent to determining the additive constant (and conversely, there is a version of the potentials such that, e.g., $\alpha_n = 0$). Since arctan is bounded, it is clear that (α_n) always converges after passing to subsequence. Furthermore, this type of normalization is compatible with convergence in distribution.

Theorem 2.1. *For $n \in \mathbb{N}$, let $(\mu_n, \nu_n) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ satisfy $\mathcal{C}(\mu_n, \nu_n) < \infty$ and let (f_n, g_n) be corresponding potentials. Suppose that*

$$(2.1) \qquad \sup_{n \in \mathbb{N}} \int f_n^+ \, d\mu_n < \infty, \qquad \sup_{n \in \mathbb{N}} \int g_n^+ \, d\nu_n < \infty$$

and that μ_n, ν_n converge weakly to μ, ν. Then

(i)   *$\mathcal{C}(\mu, \nu) < \infty$ and the optimal couplings converge weakly: $\pi_n \to \pi_*$.*

(ii) *Suppose also that the normalizations $\alpha_n := \int \arctan(f_n) \, d\mu_n$ converge to a limit α (this always holds along a subsequence). Let $(\varepsilon_k)_{k \in \mathbb{N}} \subset (0, 1)$ satisfy $\varepsilon_k \downarrow 0$. There are measurable functions*

$$F_n^k, F^k, f : \mathcal{X} \to \mathbb{R}, \qquad G_n^k, G^k, g : \mathcal{Y} \to \mathbb{R}, \qquad k, n \in \mathbb{N}$$

and a subsequence $(n_l)_{l \in \mathbb{N}} \subset \mathbb{N}$ such that

(2.2) *$(f, g) \in L^1(\mu) \times L^1(\nu)$ are the potentials for (μ, ν) with $\int \arctan(f) \, d\mu = \alpha$,*

(2.3) *$(f_n, \mu_n) \to (f, \mu)$ in distribution; that is, $(f_n)_{\#} \mu_n \to f_{\#} \mu$,*

(2.4) $\displaystyle \lim_{k \to \infty} \sup_{n \in \mathbb{N}} \mu_n(|f_n - F_n^k| \geq \varepsilon_k) = 0$,

(2.5) $\displaystyle \lim_{l \to \infty} F_{n_l}^k = F^k$ *uniformly,   for each $k \in \mathbb{N}$,*

(2.6) $\displaystyle \lim_{k \to \infty} F^k = f$ *in μ-probability,*

and analogous properties hold for G_n^k, G^k, g.

(iii) *Suppose that $(f_n^+, g_n^+)_{n \in \mathbb{N}}$ are uniformly integrable wrt. $(\mu_n, \nu_n)_{n \in \mathbb{N}}$; that is,*

$$(2.7) \qquad \lim_{C \to \infty} \sup_{n \in \mathbb{N}} \int f_n \mathbf{1}_{f_n > C} \, d\mu_n = 0, \qquad \lim_{C \to \infty} \sup_{n \in \mathbb{N}} \int g_n \mathbf{1}_{g_n > C} \, d\nu_n = 0.$$

Then the optimal values converge: $\mathcal{C}(\mu_n, \nu_n) \to \mathcal{C}(\mu, \nu)$. If α_n, α, f, g are as in (ii), then

$$(2.8) \qquad \lim_{n \to \infty} \int f_n \, d\mu_n = \int f \, d\mu, \qquad \lim_{n \to \infty} \int g_n \, d\nu_n = \int g \, d\nu.$$

The next two results discuss the main condition (2.1) of the theorem.

Remark 2.2. Condition (2.1) implies

$$(2.9) \qquad \inf_{n \in \mathbb{N}} \int f_n \, d\mu_n > -\infty, \qquad \inf_{n \in \mathbb{N}} \int g_n \, d\nu_n > -\infty,$$

and thus (2.1) is equivalent to boundedness in L^1:

$$(2.10) \qquad \sup_{n \in \mathbb{N}} \int |f_n| \, d\mu_n < \infty, \qquad \sup_{n \in \mathbb{N}} \int |g_n| \, d\nu_n < \infty.$$

Remark 2.2 follows from the duality $\int f_n \, d\mu_n + \int g_n \, d\nu_n = \mathcal{C}(\mu_n, \nu_n) \geq 0$; cf. Proposition A.2. We can provide a sufficient condition for (2.1) in terms of the given data as follows.

Lemma 2.3. *(i)*   *Let (f_n, g_n) satisfy (2.9). The following condition is sufficient for (2.1) and (2.10):*

$$(2.11) \qquad\qquad \sup_{n \in \mathbb{N}} \int c \, d(\mu_n \otimes \nu_n) < \infty.$$

If c is uniformly integrable wrt. $(\mu_n \otimes \nu_n)_{n \in \mathbb{N}}$, then (2.7) holds as well.

(ii)   *The following condition is sufficient for (2.11):*

$$(2.12) \quad \sup_{n \in \mathbb{N}} \left[H(\mu_n | \mu) + H(\nu_n | \nu) \right] < \infty \quad and \quad e^{\beta c} \in L^1(\mu \otimes \nu) \ \ for \ some \ \beta > 0.$$

If moreover $e^{\beta \phi(c)} \in L^1(\mu \otimes \nu)$ for an increasing, superlinearly growing function ϕ, then c is uniformly integrable wrt. $(\mu_n \otimes \nu_n)_{n \in \mathbb{N}}$.

The proof is deferred to Appendix C. We remark that the assumption (2.9) is mostly a matter of normalization. Indeed, suppose that $f_n^+ \in L^1(\mu_n)$ and $g_n^+ \in L^1(\nu_n)$—which necessarily holds under (2.11), cf. Proposition A.1. Then $\int f_n \, d\mu_n + \int g_n \, d\nu_n = \mathcal{C}(\mu_n, \nu_n) \geq 0$ by duality (cf. Proposition A.2). Therefore, (2.9) always holds after choosing a suitable normalization for (f_n, g_n), for instance the centering $\int f_n \, d\mu_n = 0$.

In the remainder of the section, we discuss corollaries of Theorem 2.1 that hold when the marginals (μ_n, ν_n) have additional properties. In the general setting of Theorem 2.1, there is no natural function space where one could formulate the convergence $f_n \to f$ in a straightforward way. Whereas if $\mu \ll \mu_n$ for all $n \in \mathbb{N}$, the potentials (f_n) are well-defined μ-a.s. and one can naturally ask whether $f_n \to f$ in μ-probability. The following gives an affirmative answer under a weak condition of boundedness-in-probability on the Radon–Nikodym derivatives.

Corollary 2.4. *Let (2.1) hold and let μ_n, ν_n converge weakly to μ, ν. Suppose that $\mu \ll \mu_n$, $\nu \ll \nu_n$ and*

$$(2.13) \quad \limsup_{C \to \infty} \limsup_{n \to \infty} \mu \left(\frac{d\mu}{d\mu_n} \geq C \right) = 0, \qquad \limsup_{C \to \infty} \limsup_{n \to \infty} \nu \left(\frac{d\nu}{d\nu_n} \geq C \right) = 0.$$

Then $f_n \oplus g_n \to f \oplus g$ in $\mu \otimes \nu$-probability, where (f, g) are arbitrary potentials for (μ, ν). If α_n, α, f, g are as in Theorem 2.1 (ii), then $f_n \to f$ in μ-probability and $g_n \to g$ in ν-probability.

Proof. As $f \oplus g$ is uniquely determined and convergence in probability is metrizable, it suffices to show that any subsequence of $f_n \oplus g_n$ has a subsequence converging to $f \oplus g$. Thus we may further assume that α_n, α, f, g are as in Theorem 2.1 (ii) and show the convergence of (f_n) and (g_n). Fix $\varepsilon > 0$ and a subsequence of (f_n). By (2.6) and (2.5) in Theorem 2.1, there exist a further subsequence (not relabeled) and functions F_n^k, F^k such that along this subsequence,

$$(2.14) \qquad \lim_{k \to \infty} \mu(|F^k - f| \geq \varepsilon) = 0 \qquad and \qquad \lim_{n \to \infty} \mu(|F_n^k - F^k| \geq \varepsilon) = 0.$$

On the other hand, for any $C > 0$,

$$\mu(|f_n - F_n^k| \geq \varepsilon) \leq \mu\left(\frac{d\mu}{d\mu_n} \geq C\right) + \int \mathbf{1}_{\{|f_n - F_n^k| \geq \varepsilon\} \cap \{\frac{d\mu}{d\mu_n} \leq C\}} \frac{d\mu}{d\mu_n} \, d\mu_n$$

$$\leq \mu\left(\frac{d\mu}{d\mu_n} \geq C\right) + C\mu_n(|f_n - F_n^k| \geq \varepsilon) \leq \mu\left(\frac{d\mu}{d\mu_n} \geq C\right) + C\delta_k$$

for some $\delta_k \geq 0$ with $\lim_k \delta_k = 0$, by (2.4). Taking $n \to \infty$, then $k \to \infty$ and finally $C \to \infty$, we obtain

$$\limsup_{k \to \infty} \limsup_{n \to \infty} \mu(|f_n - F_n^k| \geq \varepsilon) = 0.$$

Together with (2.14), this yields

$$\lim_{n \to \infty} \mu(|f_n - f| \geq 3\varepsilon) \leq \limsup_{k \to \infty} \limsup_{n \to \infty} \mu(|f_n - F_n^k| \geq \varepsilon)$$

$$+ \limsup_{k \to \infty} \limsup_{n \to \infty} \mu(|F_n^k - F^k| \geq \varepsilon)$$

$$+ \limsup_{k \to \infty} \mu(|F^k - f| \geq \varepsilon) = 0.$$

The proof for g is analogous. □

Condition (2.13) holds in particular for sequences converging in total variation.

Lemma 2.5. *Suppose that $\mu \ll \mu_n$ and $\mu_n \to \mu$ in total variation. Then*

$$\limsup_{C \to \infty} \limsup_{n \to \infty} \mu\left(\frac{d\mu}{d\mu_n} \geq C\right) = 0.$$

The proof is deferred to Appendix C. Our final result is the stability of the optimal couplings in the topology of total variation, complementing the weak stability shown in Theorem 2.1 (i).

Corollary 2.6. *Let (2.1) hold and let μ_n, ν_n converge in total variation to μ, ν where $\mu \ll \mu_n$ and $\nu \ll \nu_n$. Then $\pi_n \to \pi_*$ in total variation.*

Proof. For the sake of readability, we state here the proof under the additional assumption that $\mu_n \sim \mu$ and $\nu_n \sim \nu$; the general case is deferred to Appendix C. By Corollary 2.4 and Lemma 2.5, we have $f_n \to f$ in μ-probability and $g_n \to g$ in ν-probability after passing to a subsequence. Under the additional assumption, $\frac{d\mu_n}{d\mu} \to 1$ in $L^1(\mu)$ and $\frac{d\nu_n}{d\nu} \to 1$ in $L^1(\nu)$. We see that

$$\frac{d\pi_n}{d(\mu \otimes \nu)} = \frac{d\pi_n}{d(\mu_n \otimes \nu_n)} \frac{d\mu_n}{d\mu} \frac{d\nu_n}{d\nu} = e^{f_n \oplus g_n - c} \frac{d\mu_n}{d\mu} \frac{d\nu_n}{d\nu} \to e^{f \oplus g - c} = \frac{d\pi_*}{d(\mu \otimes \nu)}$$

in $\mu \otimes \nu$-probability. But then the convergence also holds in $L^1(\mu \otimes \nu)$, by Scheffé's lemma, and we conclude that $\pi_n \to \pi_*$ in total variation. The convergence of the original sequence follows. □

3. Convergence of Sinkhorn's Algorithm

Fix marginals $(\mu, \nu) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ and a continuous cost $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$. Sinkhorn's algorithm (1.2) can be written in terms of potentials. Set $\varphi_0 := 0$ and

$$
(3.1) \qquad
\begin{aligned}
\psi_t(y) &:= -\log \int_{\mathcal{X}} e^{\varphi_t(x) - c(x,y)} \, \mu(dx), \\
\varphi_{t+1}(x) &:= -\log \int_{\mathcal{Y}} e^{\psi_t(y) - c(x,y)} \, \nu(dy)
\end{aligned}
$$

for $t \geq 0$, and define the measures

$$
d\pi(\varphi, \psi) := e^{\varphi \oplus \psi - c} \, d(\mu \otimes \nu), \qquad \pi_{2t} := \pi(\varphi_t, \psi_t), \qquad \pi_{2t-1} := \pi(\varphi_t, \psi_{t-1}),
$$

where $\psi_{-1} := 0$. One can check by direct calculation that $\pi_n \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ are the same measures as in (1.2). Denoting by (μ_n, ν_n) the marginal distributions of π_n, the following summarizes well known properties of Sinkhorn's algorithm (e.g., [28, Section 6]).

Lemma 3.1. *Let $\mathcal{C}(\mu, \nu) < \infty$. We have $\mu_n \sim \mu$ and $\nu_n \sim \nu$ for all $n \geq 0$. Moreover, $H(\mu_n | \mu) + H(\nu_n | \nu) \to 0$; in particular, $\mu_n \to \mu$ and $\nu_n \to \nu$ in total variation. For $t \geq 0$, the marginals satisfy*

$$
(3.2) \qquad \mu_{2t+1} = \mu, \qquad \nu_{2t} = \nu, \qquad \frac{d\mu_{2t}}{d\mu} = e^{\varphi_t - \varphi_{t+1}}, \qquad \frac{d\nu_{2t-1}}{d\nu} = e^{\psi_{t-1} - \psi_t}.
$$

It follows that for $n \geq 1$,

$$
\frac{d\pi_n}{d(\mu_n \otimes \nu_n)} = e^{f_n \oplus g_n - c} \qquad \mu_n \otimes \nu_n\text{-}a.s.,
$$

where

$$
(3.3) \qquad
\begin{cases}
f_n := \varphi_{t+1}, & g_n := \psi_t \qquad \text{if } n = 2t, \\
f_n := \varphi_t, & g_n := \psi_t \qquad \text{if } n = 2t - 1.
\end{cases}
$$

In brief, (f_n, g_n) are potentials for the marginals (μ_n, ν_n) which in turn converge to (μ, ν) in total variation. The stability result of Corollary 2.6 then yields the following convergence result. As emphasized in the Introduction, it covers quadratic costs with arbitrary subgaussian marginals and the problem (1.1) with arbitrary regularization parameter $\varepsilon > 0$.

Corollary 3.2. *Suppose that*

$$
e^{\beta c} \in L^1(\mu \otimes \nu) \quad \text{for some } \beta > 0.
$$

Then $\mathcal{C}(\mu, \nu) < \infty$ and the Sinkhorn iterates (π_n) converge to π_* in total variation.

Proof. As $H(\mu_n | \mu) + H(\nu_n | \nu) \to 0$ by Lemma 3.1, Lemma 2.3 yields that

$$
(3.4) \qquad \sup_{n \in \mathbb{N}} \int c \, d(\mu_n \otimes \nu_n) < \infty.
$$

In particular, $\mathcal{C}(\mu_n, \nu_n) < \infty$ and (f_n, g_n) as defined in Lemma 3.1 are potentials for the marginals (μ_n, ν_n); cf. Propositions A.1 and A.2. Next, we show that (f_n, g_n)

satisfy (2.1). In general, the Sinkhorn iterates satisfy $(\varphi_t, \psi_t) \in L^1(\mu) \times L^1(\nu)$ as well as $\int \varphi_t \, d\mu \geq 0$ and $\int \psi_t \, d\nu \geq -\log \int e^{-c} \, d(\mu \otimes \nu)$; see [28, Lemma 6.4 and its footnote]. Here, as $c \geq 0$, we have $\int \varphi_t \, d\mu \geq 0$ and $\int \psi_t \, d\nu \geq 0$.

Consider $n = 2t$, then $\nu_n = \nu$. Using (3.1), Jensen's inequality and $\int \psi_t \, d\nu \geq 0$,

$$f_n(x) = \varphi_{t+1}(x) \leq \int c(x, y) \, \nu(dy) = \int c(x, y) \, \nu_n(dy)$$

and hence $\int f_n^+ \, d\mu_n \leq \int c \, d(\mu_n \otimes \nu_n)$ which is bounded by (3.4). Similarly, (3.1) implies $g_n(y) = \psi_t(y) \leq \int c(x, y) \, \mu(dx)$ and hence

$$\int g_n^+ \, d\nu_n = \int g_n^+ \, d\nu \leq \int c \, d(\mu \otimes \nu) < \infty.$$

The argument for $n = 2t - 1$ is symmetric, so that (2.1) holds. As the marginals are equivalent and converge in total variation by Lemma 3.1, the claim follows by Corollary 2.6. $\square$

4. Proof of Theorem 2.1

The proof is structured into several steps.

Step 1: Definition of sets $A_n^k, \mathcal{X}_{\text{cpt}}^k$.

Step 2: Definition of f_n^k.

Step 3: Boundedness of $(f_n^k)_{n \in \mathbb{N}}$ on compact sets.

Step 4: Construction of F_n^k and F^k.

Step 5: Further properties related to F_n^k and F^k.

Step 6: Proof that $(F^k)_{k \in \mathbb{N}}$ is μ-Cauchy and definition of f.

Step 7: Proof that $(f_n)_{\#} \mu_n \to f_{\#} \mu$.

Step 8: Proof that f, g induce a coupling π and $\pi_n \to \pi$.

Step 9: Identification of the limit, end of proof of Theorem 2.1 (i),(ii).

Step 10: Proof of Theorem 2.1 (iii).

As noted in Appendix A, we may choose versions of the potentials satisfying the Schrödinger equations (A.2) everywhere. For brevity, we generally only detail the arguments for (f_n), the arguments for (g_n) are symmetric. Recall that a sequence $(\varepsilon_k)_{k \in \mathbb{N}} \subset (0, 1)$ with $\varepsilon_k \downarrow 0$ is given. We define $(\tilde{\delta}_k)_{k \in \mathbb{N}} \subset (0, 1/2)$ by $\tilde{\delta}_k = \varepsilon_k/2$ and another sequence $(\delta_k)_{k \in \mathbb{N}}$ by $\delta_k = 1 - e^{-\tilde{\delta}_k}$; i.e.,

$$(4.1) \qquad\qquad\qquad \tilde{\delta}_k = -\log(1 - \delta_k).$$

It follows that $0 \leq \delta_k \leq \tilde{\delta}_k$ and $\tilde{\delta}_k \downarrow 0$. For notational convenience, we set $\mu_0 := \mu$, $\nu_0 := \nu$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

Step 1 is based on the following generalization of [29, Lemma 2.3] extending that result from a single measure to a tight set of measures.

Lemma 4.1. *Let $\delta \in (0, 1)$. There are compact sets $\mathcal{X}_{\text{cpt}}(\delta) \subseteq \mathcal{X}$, $\mathcal{Y}_{\text{cpt}}(\delta) \subseteq \mathcal{Y}$ with*

$$\mu_n(\mathcal{X}_{\text{cpt}}(\delta)) \geq 1 - \delta, \qquad \nu_n(\mathcal{Y}_{\text{cpt}}(\delta)) \geq 1 - \delta, \qquad n \in \mathbb{N}_0$$

and measurable sets $A_n \subseteq \mathcal{X}_{\mathrm{cpt}}(\delta)$, $B_n \subseteq \mathcal{Y}_{\mathrm{cpt}}(\delta)$ for $n \in \mathbb{N}$ such that

$$\mu_n(A_n) \geq 1 - \delta, \qquad \nu_n(B_n) \geq 1 - \delta, \qquad n \in \mathbb{N},$$

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

$$|g_n(y_1) - g_n(y_2)| \leq \sup_{x \in \mathcal{X}_{\mathrm{cpt}}(\delta)} |c(x, y_1) - c(x, y_2)| - \log(1 - \delta) \quad \text{for} \quad y_1, y_2 \in B_n.$$

The proof of Lemma 4.1 is an adaptation of the arguments in [29]; for completeness, the details are reported in Appendix C.

Step 1. Fix $k \in \mathbb{N}$. Lemma 4.1 with $\delta = \delta_k$ yields sets

$$A_n^k \subseteq \mathcal{X}_{\mathrm{cpt}}^k \subseteq \mathcal{X} \qquad \text{and} \qquad B_n^k \subseteq \mathcal{Y}_{\mathrm{cpt}}^k \subseteq \mathcal{Y}$$

such that

$$\mu_n(A_n^k) \geq 1 - \delta_k, \qquad \nu_n(B_n^k) \geq 1 - \delta_k, \qquad n \in \mathbb{N};$$

moreover, $\mathcal{X}_{\mathrm{cpt}}^k, \mathcal{Y}_{\mathrm{cpt}}^k$ are compact and

$$\mu_n(\mathcal{X}_{\mathrm{cpt}}^k) \geq 1 - \delta_k, \qquad \nu_n(\mathcal{Y}_{\mathrm{cpt}}^k) \geq 1 - \delta_k, \qquad n \in \mathbb{N}_0.$$

Recalling (4.1), we have for all $n \in \mathbb{N}$ that

$$(4.2) \qquad |f_n(x_1) - f_n(x_2)| \leq \sup_{y \in \mathcal{Y}_{\mathrm{cpt}}^k} |c(x_1, y) - c(x_2, y)| + \tilde{\delta}_k, \qquad x_1, x_2 \in A_n^k$$

and similarly for g_n.

Step 2. Define the continuous pseudometric $\tilde{d}_k$ on $\mathcal{X}$ by

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

Let $n \in \mathbb{N}$. Using (4.2) and a version of Kirszbraun's extension theorem as detailed in [29, Lemma 2.4], there exists a function $f_n^k : \mathcal{X} \to \mathbb{R}$ satisfying

$$f_n^k = f_n \quad \text{on} \quad A_n^k,$$
$$(4.3) \qquad |f_n^k(x_1) - f_n^k(x_2)| \leq \tilde{d}_k(x_1, x_2) + \tilde{\delta}_k, \quad x_1, x_2 \in \mathcal{X}.$$

Step 3. With $k \in \mathbb{N}$ still fixed, we show that $(f_n^k)_{n \in \mathbb{N}}$ is bounded on compact sets. Suppose that $\sup_{n \in \mathbb{N}} \sup_{x \in K} |f_n^k(x)| = \infty$ for some compact $K \subset \mathcal{X}$. Then (4.3) even implies that

$$(4.4) \qquad \sup_{n \in \mathbb{N}} \inf_{x \in K'} |f_n^k(x)| = \infty \quad \text{for any compact set } \emptyset \neq K' \subset \mathcal{X},$$

because $\sup_{x_1 \in K, x_2 \in K'} \tilde{d}_k(x_1, x_2) < \infty$. We shall contradict (4.4). Indeed, by (2.10) and Markov's inequality there exists $C > 0$ such that $\mu_n(|f_n| \geq C) \leq \delta_k$ for all $n \in \mathbb{N}$. As $\mu_n(A_n^k) \geq 1 - \delta_k$, it follows that

$$\mu_n\big(\{|f_n| \leq C\} \cap A_n^k\big) \geq 1 - 2\delta_k > 0$$

and in particular $\{|f_n| \leq C\} \cap A_n^k \neq \emptyset$. As $f_n^k = f_n$ on A_n^k and $A_n^k \subseteq \mathcal{X}_{\mathrm{cpt}}^k$, it follows that $\{|f_n^k| \leq C\} \cap \mathcal{X}_{\mathrm{cpt}}^k \neq \emptyset$ for all $n \in \mathbb{N}$, contradicting (4.4) for $K' := \mathcal{X}_{\mathrm{cpt}}^k$.

As a preparation for Step 4, we record the following covering lemma.

Lemma 4.2. *Let $K \subset \mathcal{X}$ be compact and $r > 0$. There exists a measurable partition $(D_j)_{j\in\mathbb{N}}$ of $\mathcal{X}$ such that*

(i) *D_j has diameter at most r and boundary $\mu(\partial D_j) = 0$,*
(ii) *K intersects finitely many elements of $(D_j)_{j\in\mathbb{N}}$.*

Proof. Pick a dense sequence $(x_j)_{j\in\mathbb{N}}$ in $\mathcal{X}$. For each j, the boundaries $\partial B_\rho(x_j)$ are disjoint for different values of $\rho > 0$. Hence there are at most countably many values of ρ such that $\mu(\partial B_\rho(x_j)) > 0$ for some j, and we can pick $\rho \in (0, r/2]$ such that $\mu(\partial B_\rho(x_j)) = 0$ for all j.

By compactness there is a finite subset $\mathcal{N} \subset \mathbb{N}$ such that $B_\rho(x_j), j \in \mathcal{N}$ cover K; we may assume that $\mathcal{N} = \{1, \ldots, N\}$. Set $D_0 := \emptyset$ and

$$D_j = B_\rho(x_j) \setminus (D_1 \cup \cdots \cup D_{j-1}), \qquad j \geq 1.$$

The general relation $\partial(A \cap B) \subseteq \partial A \cup \partial B$ implies that $\mu(\partial D_j) = 0$, and the other requirements are satisfied by construction. $\qquad\square$

Step 4. Keeping $k \in \mathbb{N}$ fixed, our next aim is to define the functions F_n^k and F^k. As $\tilde{d}_k$ is uniformly continuous on the compact set $\mathcal{X}_{\mathrm{cpt}}^k$, there is $r \in (0, \tilde{\delta}_k)$ such that

$$(4.5) \qquad \sup_{x_1, x_2 \in \mathcal{X}_{\mathrm{cpt}}^k, \ d(x_1, x_2) \leq r} \tilde{d}_k(x_1, x_2) < \tilde{\delta}_k.$$

We apply Lemma 4.2 with $K = \mathcal{X}_{\mathrm{cpt}}^k$ to define a partition $(D_j^k)_{j\in\mathbb{N}}$ of $\mathcal{X}$ and choose a sequence $(x_{j,n}^k)_{j,n\in\mathbb{N}} \subset \mathcal{X}$ satisfying

$$x_{j,n}^k \in \begin{cases} D_j^k \cap A_n^k & \text{if } D_j^k \cap A_n^k \neq \emptyset, \\ D_j^k \cap \mathcal{X}_{\mathrm{cpt}}^k & \text{if } D_j^k \cap A_n^k = \emptyset \text{ and } D_j^k \cap \mathcal{X}_{\mathrm{cpt}}^k \neq \emptyset, \\ D_j^k & \text{otherwise.} \end{cases}$$

We then set

$$F_n^k(x) := \sum_{j\in\mathbb{N}} f_n^k(x_{j,n}^k) \mathbf{1}_{D_j^k \cap \mathcal{X}_{\mathrm{cpt}}^k}(x).$$

This is a finite sum as $\mathcal{J}^k := \{j : D_j^k \cap \mathcal{X}_{\mathrm{cpt}}^k \neq \emptyset\}$ is finite (cf. Lemma 4.2). Moreover, the points $x_{j,n}^k$ with $j \in \mathcal{J}^k$ all belong to the compact set $\mathcal{X}_{\mathrm{cpt}}^k$. In view of Step 3, it follows that the corresponding coefficients $f_n^k(x_{j,n}^k)$ are bounded uniformly in j, n.

We can now apply a diagonal sequence argument to extract a subsequence (not relabeled) along which $a_j^k := \lim_{n\to\infty} f_n^k(x_{j,n}^k)$ exists for all $j \in \mathbb{N}$. Set

$$(4.6) \qquad F^k(x) := \sum_{j\in\mathbb{N}} a_j^k \mathbf{1}_{D_j^k \cap \mathcal{X}_{\mathrm{cpt}}^k}(x),$$

then $F_n^k \to F^k$ pointwise on $\mathcal{X}$ and in particular F^k is measurable. In fact, this convergence is uniform as $\mathcal{J}^k$ is finite. By construction, F_n^k, F^k are supported on the compact $\mathcal{X}_{\mathrm{cpt}}^k$ and bounded uniformly in n. Passing to another subsequence, we achieve that $\lim_{n\to\infty} F_n^k = F^k$ holds simultaneously for all $k \in \mathbb{N}$. For the remainder

of the proof, we only consider this subsequence (denoted $(n_l)_l$ in Theorem 2.1 but not relabeled here).

Step 5. We record two more facts about the construction in Step 4 for later use. For $x \in D_j^k \cap A_n^k$, (4.2) and (4.5) yield

$$|F_n^k(x) - f_n^k(x)| = |f_n^k(x_{j,n}^k) - f_n^k(x)| \leq \tilde{d}_k(x_{j,n}^k, x) + \tilde{\delta}_k < 2\tilde{\delta}_k.$$

As $\cup_j D_j^k = \mathcal{X}$ and $f_n^k = f_n$ on A_n^k, it follows that

$$(4.7) \qquad |F_n^k(x) - f_n(x)| = |F_n^k(x) - f_n^k(x)| < 2\tilde{\delta}_k, \qquad x \in A_n^k.$$

In view of $\mu_n(A_n^k) \geq 1 - \delta_k$, this yields in particular

$$\sup_{n \in \mathbb{N}} \mu_n(|f_n - F_n^k| \geq 2\tilde{\delta}_k) \leq \delta_k$$

and hence the claim (2.4); recall $2\tilde{\delta}_k = \varepsilon_k$.

Second, we define similarly as in (4.6) the function

$$(4.8) \qquad F_{\mathrm{reg}}^k(x) := \sum_{j \in \mathcal{J}^k} a_j^k \mathbf{1}_{D_j^k}(x).$$

Then $F_{\mathrm{reg}}^k = F^k$ on $\mathcal{X}_{\mathrm{cpt}}^k$ and hence

$$(4.9) \qquad \mu_n(F_{\mathrm{reg}}^k \neq F^k) \leq \delta_k, \qquad n \in \mathbb{N}_0.$$

Clearly F_{reg}^k is continuous on the interior of the complement of $U := \cup_{j \in \mathcal{J}^k} \partial D_j^k$. As $\mathcal{J}^k$ is finite, ∂D_j^k is closed and $\mu(\partial D_j^k) = 0$, the set U is closed and μ-null. Hence its complement is open and has full μ-measure; in brief, F_{reg}^k is continuous μ-a.s. In particular, F_{reg}^k can be used in the mapping theorem [5, Theorem 2.7] for weak convergence arguments with $\mu_n \to \mu$.

Step 6. We show that $(F^k)_{k \in \mathbb{N}}$ is Cauchy in μ-probability. Fix $\varepsilon > 0$ and choose $k_0 \in \mathbb{N}$ such that $\tilde{\delta}_k \leq \varepsilon$ (hence also $\delta_k \leq \varepsilon$) for all $k \geq k_0$. Let $k, k' \geq k_0$. For $x \in A_n^k \cap A_n^{k'}$ we have $f_n^k(x) = f_n(x) = f_n^{k'}(x)$ and thus (4.7) yields

$$|F_n^k(x) - F_n^{k'}(x)| \leq |F_n^k(x) - f_n^k(x)| + |f_n^k(x) - f_n^{k'}(x)| + |f_n^{k'}(x) - F_n^{k'}(x)|$$
$$\leq 2\tilde{\delta}_k + 0 + 2\tilde{\delta}_{k'} \leq 4\varepsilon.$$

As a result,

$$|F^k(x) - F^{k'}(x)| \leq |F^k(x) - F_n^k(x)| + |F_n^k(x) - F_n^{k'}(x)| + |F_n^{k'}(x) - F^{k'}(x)|$$
$$\leq |F^k(x) - F_n^k(x)| + 4\varepsilon + |F_n^{k'}(x) - F^{k'}(x)|$$

for $x \in A_n^k \cap A_n^{k'}$. In view of $\mu_n(A_n^k \cap A_n^{k'}) \geq 1 - 2\varepsilon$, it follows that

$$\int \left[\left| F^k - F^{k'} \right| \wedge 1 \right] d\mu_n \leq \int \left[\left| F^k - F_n^k \right| \wedge 1 + \left| F_n^{k'} - F^{k'} \right| \wedge 1 \right] d\mu_n + 6\varepsilon.$$

As $F_n^k \to F^k$ and $F_n^{k'} \to F^{k'}$ uniformly (cf. Step 4), we conclude

$$(4.10) \qquad \limsup_{n \to \infty} \int \left[\left| F^k - F^{k'} \right| \wedge 1 \right] d\mu_n \leq 6\varepsilon.$$

On the other hand, using the mapping theorem and (4.9) for both $\mu \equiv \mu_0$ and μ_n,

$$\int \left[\left| F^k - F^{k'} \right| \wedge 1 \right] d\mu \leq 2\varepsilon + \int \left[\left| F_{\text{reg}}^k - F_{\text{reg}}^{k'} \right| \wedge 1 \right] d\mu$$

$$= 2\varepsilon + \lim_{n \to \infty} \int \left[\left| F_{\text{reg}}^k - F_{\text{reg}}^{k'} \right| \wedge 1 \right] d\mu_n$$

$$\leq 4\varepsilon + \limsup_{n \to \infty} \int \left[\left| F^k - F^{k'} \right| \wedge 1 \right] d\mu_n.$$

In view of (4.10), this yields

$$\int \left[\left| F^k - F^{k'} \right| \wedge 1 \right] d\mu \leq 10\varepsilon,$$

showing that $(F^k)_{k \in \mathbb{N}}$ is Cauchy in μ-probability. In particular, there exists a limit f in μ-probability.

Step 7. Let $\phi : \mathbb{R} \to [-1, 1]$ be uniformly continuous; we show that $\int \phi(f_n) \, d\mu_n$ converges to $\int \phi(f) \, d\mu$. That is, the law of f_n under μ_n converges weakly to the law of f under μ [5, Theorem 2.1], or in terms of pushforwards, $(f_n)_{\#}\mu_n \to f_{\#}\mu$. In particular, if (f_n) satisfy a normalization $\int \arctan(f_n) \, d\mu_n = \alpha_n$ and $\alpha_n \to \alpha$, then $\int \arctan(f) \, d\mu = \alpha$.

Let $\varepsilon > 0$. In view of (2.4) and (2.6), there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and all $n \in \mathbb{N}$,

$$\left| \int \phi(f_n) \, d\mu_n - \int \phi(f) \, d\mu \right| \leq \left| \int \phi(f_n) \, d\mu_n - \int \phi(F_n^k) \, d\mu_n \right|$$

$$+ \left| \int \phi(F_n^k) \, d\mu_n - \int \phi(F^k) \, d\mu \right|$$

$$+ \left| \int \phi(F^k) \, d\mu - \int \phi(f) \, d\mu \right|$$

$$\leq 2\varepsilon + \left| \int \phi(F_n^k) \, d\mu_n - \int \phi(F^k) \, d\mu \right|.$$

Fix $k \geq k_0$ such that $\delta_k \leq \varepsilon$. Then by (4.9),

$$\left| \int \phi(F_n^k) \, d\mu_n - \int \phi(F^k) \, d\mu \right|$$

$$\leq \left| \int \phi(F_n^k) \, d\mu_n - \int \phi(F^k) \, d\mu_n \right| + \left| \int \phi(F^k) \, d\mu_n - \int \phi(F^k) \, d\mu \right|$$

$$\leq 2\varepsilon + \left| \int \phi(F_n^k) \, d\mu_n - \int \phi(F^k) \, d\mu_n \right| + \left| \int \phi(F_{\text{reg}}^k) \, d\mu_n - \int \phi(F_{\text{reg}}^k) \, d\mu_n \right|.$$

As $F_n^k \to F^k$ uniformly and F_{reg}^k is continuous μ-a.s., there exists $n_0 = n_0(k)$ such that the last line is $\leq 4\varepsilon$ for all $n \geq n_0$. In summary, $|\int \phi(f_n)\, d\mu_n - \int \phi(f)\, d\mu| \leq 6\varepsilon$ for $n \geq n_0$, proving the claim.

Above, we have introduced the functions f_n^k, F_n^k, F^k, f on $\mathcal{X}$. Analogously, one constructs g_n^k, G_n^k, G^k, g on $\mathcal{Y}$. We can now detail the main step of the proof, showing that f, g are indeed potentials for a coupling $\pi \in \Pi(\mu, \nu)$. To keep track of the argument more easily, we state the technical parts as lemmas and prove them at the end.

Step 8. Recall that

$$\pi_n(dx, dy) = e^{f_n(x) + g_n(y) - c(x,y)}\, \mu(dx)\, \nu(dy) \in \Pi(\mu_n, \nu_n).$$

Consider the nonnegative measure

$$\pi(dx, dy) := e^{f(x) + g(y) - c(x,y)}\, \mu(dx)\, \nu(dy)$$

and let $S \subset \mathcal{X} \times \mathcal{Y}$ be measurable with $\pi(\partial S) = 0$; we show $\pi(S) = \lim_{n \to \infty} \pi_n(S)$. Define the auxiliary measures

$$\pi_n^C(dx, dy) = e^{f_n(x) \wedge C + g_n(y) \wedge C - c(x,y)}\, \mu_n(dx)\, \nu_n(dy),$$
$$\pi^C(dx, dy) = e^{f(x) \wedge C + g(y) \wedge C - c(x,y)}\, \mu(dx)\, \nu(dy),$$
$$\pi^{k,C}(dx, dy) = e^{F^k(x) \wedge C + G^k(y) \wedge C - c(x,y)}\, \mu(dx)\, \nu(dy).$$

Fix $\varepsilon \in (0, 1/9)$ and consider the decomposition

$$\left|\pi_n(S) - \pi(S)\right| \leq \left|\pi_n(S) - \pi_n^C(S)\right| + \left|\pi_n^C(S) - \pi^{k,C}(S)\right| + \left|\pi^{k,C}(S) - \pi^C(S)\right|$$
$$(4.11) \qquad\qquad + \left|\pi^C(S) - \pi(S)\right|.$$

We estimate separately the four terms on the right-hand side.

Lemma 4.3. *We have*

$$\lim_{C \to \infty} \sup_{n \in \mathbb{N}} \left\|\pi_n - \pi_n^C\right\|_{TV} = 0.$$

The lemma is proved at the end of Step 8. For the first term in (4.11), Lemma 4.3 shows that there exists $C > 0$ such that

$$(4.12) \qquad\qquad \sup_{n \in \mathbb{N}} \left|\pi_n(S) - \pi_n^C(S)\right| \leq \varepsilon.$$

We continue with the last term of (4.11). Since $x \mapsto e^x$ is increasing and nonnegative, an application of the monotone convergence theorem for $C \to \infty$ shows that after increasing C as necessary, we have

$$(4.13) \qquad\qquad \left|\pi^C(S) - \pi(S)\right| \leq \varepsilon \quad \text{if} \quad \pi(S) < \infty,$$
$$(4.14) \qquad\qquad \pi^C(S) \geq 2 \quad \text{if} \quad \pi(S) = \infty.$$

(The second case will be eliminated by contradiction later on.) The value of C is now fixed for the remainder of the proof.

Turning to the third term in (4.11), note that since $F^k \overset{\mu}{\to} f$ and $G^k \overset{\nu}{\to} g$,

$$e^{F^k(x)\wedge C+G^k(y)\wedge C-c(x,y)} \to e^{f(x)\wedge C+g(y)\wedge C-c(x,y)} \qquad \text{in } \mu \otimes \nu\text{-probability.}$$

As these expressions are uniformly bounded, dominated convergence implies that there exists $k_0 \in \mathbb{N}$ with

$$(4.15) \qquad \left|\pi^{k,C}(S) - \pi^C(S)\right| \leq \varepsilon \quad \text{for all} \quad k \geq k_0.$$

It remains to estimate the second term, $|\pi_n^C(S) - \pi^{k,C}(S)|$, for large n, which is the main difficulty. Choose $k_1 \in \mathbb{N}$ such that for all $k \geq k_1$ we have

$$(4.16) \qquad |e^a - e^{\tilde{a}}| \leq \varepsilon \quad \text{whenever } |a - \tilde{a}| \leq 4\tilde{\delta}_k \text{ and } a, \tilde{a} \leq 2C.$$

Moreover, choose $k_2 \in \mathbb{N}$ such that for all $k \geq k_2$ we have

$$(4.17) \qquad 2\delta_k e^{2C} \leq \varepsilon.$$

For the remainder of the proof, $k \geq \max(k_0, k_1, k_2)$ is fixed.

Lemma 4.4. *For all $k \geq \max(k_0, k_1, k_2)$,*

$$\sup_{n\in\mathbb{N}} \left|\pi_n^C(S) - \int_S e^{F_n^k(x)\wedge C+G_n^k(y)\wedge C-c(x,y)} \mu_n(dx)\,\nu_n(dy)\right| \leq 2\varepsilon.$$

Lemma 4.5. *For all $k \geq \max(k_0, k_1, k_2)$,*

$$\limsup_{n\to\infty} \left|\pi^{k,C}(S) - \int_S e^{F_n^k(x)\wedge C+G_n^k(y)\wedge C-c(x,y)} \mu_n(dx)\,\nu_n(dy)\right| \leq 4\varepsilon.$$

The lemmas are proved at the end of Step 8. Together, they show

$$\limsup_{n\to\infty} \left|\pi_n^C(S) - \pi^{k,C}(S)\right| \leq 6\varepsilon.$$

Combining this with (4.12) and (4.15) yields

$$(4.18) \qquad \limsup_{n\to\infty} \left|\pi_n(S) - \pi^C(S)\right| \leq 8\varepsilon.$$

In particular, there exists $n \in \mathbb{N}$ such that $\left|\pi_n(S) - \pi^C(S)\right| \leq 9\varepsilon$, and as π_n is a probability measure, it follows that $\pi^C(S) \leq 1 + 9\varepsilon$. This contradicts (4.14) and hence establishes that we are in the case (4.13). In view of (4.18), that yields

$$\limsup_{n\to\infty} |\pi_n(S) - \pi(S)| \leq 9\varepsilon.$$

This shows $\lim_{n\to\infty} \pi_n(S) = \pi(S)$. Thus we have proved that $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $\pi_n \to \pi$ weakly [5, Theorem 2.1]. As the marginals then also converge weakly and $\pi_n \in \Pi(\mu_n, \nu_n)$, this implies $\pi \in \Pi(\mu, \nu)$.

Proof of Lemma 4.3. Let $\kappa > 0$. By (2.10) and Markov's inequality there exists $C > 0$ such that $\mu_n(f_n \geq C) \leq \kappa$ and $\nu_n(g_n \geq C) \leq \kappa$ for all $n \in \mathbb{N}$. For any measurable set $S \subset \mathcal{X} \times \mathcal{Y}$, the definition of π_n^C then yields

$$\begin{aligned}
\pi_n(S) \geq \pi_n^C(S) &\geq \pi_n^C(S \cap \{f_n < C\} \cap \{g_n < C\}) \\
&= \pi_n(S \cap \{f_n < C\} \cap \{g_n < C\}) \\
&\geq \pi_n(S) - \mu_n(f_n \geq C) - \nu_n(g_n \geq C) \geq \pi_n(S) - 2\kappa.
\end{aligned}$$

Hence $\sup_{n \in \mathbb{N},\, S \subset \mathcal{X} \times \mathcal{Y}} |\pi_n(S) - \pi_n^C(S)| \leq 2\kappa$ and the claim follows. $\qquad\square$

Proof of Lemma 4.4. We split the integral into

$$
\left| \pi_n^C(S) - \int_S e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \mu_n(dx)\, \nu_n(dy) \right|
$$

$$
\leq \int_{S \cap (A_n^k \times B_n^k)} \left| e^{f_n(x) \wedge C + g_n(y) \wedge C - c(x,y)} - e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \right| \mu_n(dx)\, \nu_n(dy)
$$

$$
+ \int_{(A_n^k \times B_n^k)^c} \left| e^{f_n(x) \wedge C + g_n(y) \wedge C - c(x,y)} - e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \right| \mu_n(dx)\, \nu_n(dy).
$$

To estimate the first integral, recall from (4.7) that $|F_n^k - f_n| \leq 2\tilde{\delta}_k$ on A_n^k. Using also the analogue on B_n^k, (4.16) implies that the integrand is bounded by ε on $A_n^k \times B_n^k$. Regarding the second integral, the integrand is bounded by e^{2C} and

$$
(\mu_n \otimes \nu_n)\big[(A_n^k \times B_n^k)^c\big] \leq \mu_n((A_n^k)^c) + \nu_n((B_n^k)^c) \leq 2\delta_k.
$$

In summary,

$$
\left| \pi_n^C(S) - \int_S e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \mu_n(dx)\, \nu_n(dy) \right| \leq \varepsilon + 2\delta_k e^{2C} \overset{(4.17)}{\leq} 2\varepsilon
$$

as claimed. $\qquad\square$

Proof of Lemma 4.5. For brevity, denote $\tilde{\pi}_n(dx, dy) := e^{-c(x,y)} \mu_n(dx) \nu_n(dy)$ for all $n \in \mathbb{N}_0$. As $\mu_n \to \mu_0$ and $\nu_n \to \nu_0$ and $(x,y) \mapsto e^{-c(x,y)}$ is bounded and continuous, we also have $\tilde{\pi}_n \to \tilde{\pi}_0$ weakly. Recall from Step 4 (and its analogue on $\mathcal{Y}$ with self-explanatory notation) that

$$
F_n^k(x) = \sum_{j \in \mathcal{J}^k} f_n^k(x_{j,n}^k) \mathbf{1}_{D_j^k \cap \mathcal{X}_{\text{cpt}}^k}(x), \qquad G_n^k(x) = \sum_{l \in \mathcal{L}^k} g_n^k(y_{l,n}^k) \mathbf{1}_{E_l^k \cap \mathcal{Y}_{\text{cpt}}^k}(y),
$$

$$
F^k(x) = \sum_{j \in \mathcal{J}^k} a_j^k \mathbf{1}_{D_j^k \cap \mathcal{X}_{\text{cpt}}^k}(x), \qquad G^k(y) = \sum_{l \in \mathcal{L}^k} b_l^k \mathbf{1}_{E_l^k \cap \mathcal{Y}_{\text{cpt}}^k}(y),
$$

where $\mathcal{J}^k$ and $\mathcal{L}^k$ are finite sets. Thus

$$
\int_{S \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)} e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \mu_n(dx)\, \nu_n(dy)
$$

$$
= \sum_{j \in \mathcal{J}^k} \sum_{l \in \mathcal{L}^k} e^{f_n^k(x_{j,n}^k) \wedge C + g_n^k(y_{l,n}^k) \wedge C} \tilde{\pi}_n \left[S \cap (D_j^k \times E_l^k) \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k) \right]
$$

and similarly for F^k, G^k instead of F_n^k, G_n^k. By the construction in Step 4,

$$
(4.19) \qquad \lim_{n \to \infty} e^{f_n^k(x_{j,n}^k) \wedge C + g_n^k(y_{l,n}^k) \wedge C} = e^{a_j^k \wedge C + b_l^k \wedge C}.
$$

As S and $(D_j^k \times E_l^k)$ are $\mu \otimes \nu$-continuity sets, we also have

$$
(4.20) \qquad \lim_{n \to \infty} \tilde{\pi}_n \left[(D_j^k \times E_l^k) \cap S \right] = \tilde{\pi}_0 \left[(D_j^k \times E_l^k) \cap S \right].
$$

Finally, we note that for all $n \in \mathbb{N}_0$,

$$(4.21) \qquad e^{2C} \tilde{\pi}_n \big[(\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)^c \big] \leq e^{2C} (\mu_n \otimes \nu_n) \big[(\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)^c \big] \leq 2 e^{2C} \delta_k \overset{(4.17)}{\leq} \varepsilon.$$

We can now expand the difference to be estimated as

$$\left| \pi^{k,C}(S) - \int_S e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \, \mu_n(dx) \, \nu_n(dy) \right|$$

$$\leq \left| \pi^{k,C}(S \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)) - \int_{S \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)} e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \, \mu_n(dx) \, \nu_n(dy) \right|$$

$$+ e^{2C} \Big(\tilde{\pi}_0 \big[(\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)^c \big] + \tilde{\pi}_n \big[(\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k)^c \big] \Big)$$

$$\overset{(4.21)}{\leq} 2\varepsilon + \sum_{j \in \mathcal{J}^k} \sum_{l \in \mathcal{L}^k} \left| e^{a_j^k \wedge C + b_l^k \wedge C} \tilde{\pi}_0 \Big[S \cap (D_j^k \times E_l^k) \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k) \Big] \right.$$

$$\left. - e^{f_n^k(x_{j,n}^k) \wedge C + g_n^k(y_{l,n}^k) \wedge C} \tilde{\pi}_n \Big[S \cap (D_j^k \times E_l^k) \cap (\mathcal{X}_{\text{cpt}}^k \times \mathcal{Y}_{\text{cpt}}^k) \Big] \right|$$

$$\overset{(4.21)}{\leq} 4\varepsilon + \sum_{j \in \mathcal{J}^k} \sum_{l \in \mathcal{L}^k} \left| e^{a_j^k \wedge C + b_l^k \wedge C} \tilde{\pi}_0 \Big[S \cap (D_j^k \times E_l^k) \Big] \right.$$

$$\left. - e^{f_n^k(x_{j,n}^k) \wedge C + g_n^k(y_{l,n}^k) \wedge C} \tilde{\pi}_n \Big[S \cap (D_j^k \times E_l^k) \Big] \right|.$$

In view of (4.19) and (4.20), taking $n \to \infty$ yields

$$\limsup_{n \to \infty} \left| \pi^{k,C}(S) - \int_S e^{F_n^k(x) \wedge C + G_n^k(y) \wedge C - c(x,y)} \, \mu_n(dx) \, \nu_n(dy) \right| \leq 4\varepsilon. \qquad \square$$

Step 9. In Step 8 we have shown that

$$\pi(dx, dy) := e^{f(x) + g(y) - c(x,y)} \, \mu(dx) \, \nu(dy)$$

defines a coupling of μ, ν. Moreover, Step 7 and (2.1) imply that $f^+ \in L^1(\mu)$ and $g^+ \in L^1(\nu)$. By the general verification result in Proposition A.2, the form of π with $(f \oplus g)^+ \in L^1(\mu \otimes \nu)$ implies that $(f, g) \in L^1(\mu) \times L^1(\nu)$, that $\mathcal{C}(\mu, \nu) < \infty$ and that $\pi = \pi_*$ is the unique minimizer for the entropic optimal transport problem (1.4). It follows that $\pi_n \to \pi_*$ also holds along the original sequence, completing the proof of Theorem 2.1 (i).

The potentials corresponding to π_* are a.s. uniquely determined up to an additive constant (Proposition A.1) and we have seen in Step 7 that $\alpha_n = \int \arctan(f_n) \, d\mu_n \to \int \arctan(f) \, d\mu$. As $\int \arctan(f + a) \, d\mu$ is strictly increasing in $a \in \mathbb{R}$, it follows that (f, g) are the unique potentials with normalization $\int \arctan(f) \, d\mu = \alpha$. In particular, (f, g) do not depend on the subsequence chosen in Step 4 and the proof of Theorem 2.1 (ii) is complete.

Step 10. It remains to prove Theorem 2.1 (iii). Let (2.7) hold. Passing to a subsequence, we may assume that $\alpha_n \to \alpha$ and f, g are as in Theorem 2.1 (ii). We first

show the upper semicontinuity

$$(4.22) \qquad \limsup_{n\to\infty} \int f_n \, d\mu_n \le \int f \, d\mu, \qquad \limsup_{n\to\infty} \int g_n \, d\nu_n \le \int g \, d\mu.$$

Indeed, the weak convergence (2.3) and the uniform integrability (2.7) imply that $\lim_{n\to\infty} \int f_n^+ \, d\mu_n = \int f^+ \, d\mu$. Together with Portmanteau's theorem for (f_n^-), the first part of (4.22) follows, and similarly for the second.

Next, we argue the lower semicontinuity of the sum,

$$(4.23) \qquad \liminf_{n\to\infty} \left(\int f_n \, d\mu_n + \int g_n \, d\nu_n \right) \ge \int f \, d\mu + \int g \, d\nu.$$

By (2.1) and Proposition A.2, we have the duality

$$\inf_{\pi\in\Pi(\mu_n,\nu_n)} \int c(x,y) \, \pi(dx,dy) + H(\pi|\mu_n \otimes \nu_n) = \int f_n \, d\mu_n + \int g_n \, d\nu_n.$$

Similarly, $(f,g) \in L^1(\mu) \times L^1(\nu)$ implies the duality for the limiting problem. As a consequence, it suffices to argue lower semicontinuity in the primal problem. Let $\pi_n = \arg\min_{\pi\in\Pi(\mu_n,\nu_n)} \int c(x,y) \, \pi(dx,dy) + H(\pi|\mu_n \otimes \nu_n)$. Then $(\pi_n)_{n\in\mathbb{N}}$ is tight and any weak cluster point belongs to $\Pi(\mu,\nu)$. Using the lower semicontinuity of $(\pi',\mu',\nu') \mapsto \int c \, d\pi' + H(\pi'|\mu' \otimes \nu')$, cf. [28, Lemma 1.3], we deduce

$$\liminf_{n\to\infty} \left(\int f_n \, d\mu_n + \int g_n \, d\nu_n \right) = \liminf_{n\to\infty} \left(\inf_{\pi\in\Pi(\mu_n,\nu_n)} \int c \, d\pi + H(\pi|\mu_n \otimes \nu_n) \right)$$

$$= \liminf_{n\to\infty} \int c \, d\pi_n + H(\pi_n|\mu_n \otimes \nu_n)$$

$$\ge \inf_{\pi\in\Pi(\mu,\nu)} \int c \, d\pi + H(\pi|\mu \otimes \nu)$$

$$= \int f \, d\mu + \int g \, d\nu.$$

Together, the lower semicontinuity (4.23) of the sum and the separate upper semicontinuity (4.22) imply (2.8). This completes the proof of Theorem 2.1. □

Appendix A. Background on Entropic Optimal Transport

Let $(\mu,\nu) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ and let $c : \mathcal{X} \times \mathcal{Y} \to [0,\infty)$ be measurable. We have the following result on existence and uniqueness for the entropic optimal transport problem (1.4).

Proposition A.1. *If $\mathcal{C}(\mu,\nu) < \infty$, there is a unique minimizer $\pi_* \in \Pi(\mu,\nu)$ for (1.4). Moreover, $\pi_* \sim \mu \otimes \nu$ and there are measurable functions $f : \mathcal{X} \to \mathbb{R}$, $g : \mathcal{Y} \to \mathbb{R}$, called potentials, such that*

$$\frac{d\pi_*}{d(\mu \otimes \nu)} = e^{f\oplus g - c} \quad \mu \otimes \nu\text{-a.s.}$$

The potentials are a.s. unique up to an additive constant: if f', g' are potentials, then $f' = f + a$ μ-a.s. and $g' = g - a$ ν-a.s. for some $a \in \mathbb{R}$.
If $c \in L^1(\mu \otimes \nu)$, then $\mathcal{C}(\mu,\nu) < \infty$ and $(f,g) \in L^1(\mu) \times L^1(\nu)$.

See [28, Theorem 4.2] for a proof. Conversely, the next result shows that the form of the density characterizes the minimizer. We also include the duality relation.

Proposition A.2. *Let $\pi_0 \in \Pi(\mu, \nu)$ admit a density of the form*

$$\frac{d\pi_0}{d(\mu \otimes \nu)} = e^{f_0 \oplus g_0 - c} \quad \mu \otimes \nu\text{-}a.s.$$

for some measurable functions $f_0 : \mathcal{X} \to \mathbb{R}$ and $g_0 : \mathcal{Y} \to \mathbb{R}$.

(a) *If $\mathcal{C}(\mu, \nu) < \infty$, then π_0 is the minimizer π_* and f_0, g_0 are its potentials.*

(b) *If $(f_0 \oplus g_0)^+ \in L^1(\mu \otimes \nu)$, then necessarily $(f_0, g_0) \in L^1(\mu) \times L^1(\nu)$ and*

$$(A.1) \qquad \mathcal{C}(\mu, \nu) = \int f_0 \, d\mu + \int g_0 \, d\nu.$$

In particular, $\mathcal{C}(\mu, \nu) < \infty$ and (a) applies.

See [28, Theorem 4.2, Theorem 4.7, Remark 4.8]. When f, g are potentials as in Proposition A.1, the fact that $\pi_* \in \Pi(\mu, \nu)$ implies the so-called Schrödinger equations

$$(A.2) \qquad \begin{aligned} f(x) &= -\log \int_{\mathcal{Y}} e^{g(y) - c(x,y)} \, \nu(dy) \quad \mu\text{-}a.s., \\ g(y) &= -\log \int_{\mathcal{X}} e^{f(x) - c(x,y)} \, \mu(dx) \quad \nu\text{-}a.s. \end{aligned}$$

We may choose versions of the potentials such that these relations hold without exceptional sets.

Appendix B. Uniform Stability under Strong Integrability

The following result shows that stability of the potentials, even uniformly on compacts, can be obtained quite easily when the cost is sufficiently integrable. As discussed in the Introduction, the integrability condition is *not* satisfied in the regime of principal interest. For the statement, we choose versions of the potentials such that (A.2) holds without exceptional sets.

Proposition B.1. *For $n \in \mathbb{N}$, let $(\mu_n, \nu_n) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ satisfy $\mathcal{C}(\mu_n, \nu_n) < \infty$ and let (f_n, g_n) be corresponding potentials. Suppose that μ_n, ν_n converge weakly to μ, ν and*

$$(B.1) \qquad \sup_{n \in \mathbb{N}} \int e^{\beta f_n} \, d\mu_n < \infty, \qquad \sup_{n \in \mathbb{N}} \int e^{\beta g_n} \, d\nu_n < \infty \qquad \text{for some } \beta > 1.$$

Then $f_n \oplus g_n \to f \oplus g$ uniformly on compacts, where (f, g) are arbitrary potentials for (μ, ν), and the corresponding optimal couplings converge weakly. If α_n, α, f, g are as in Theorem 2.1 (ii), then also $f_n \to f$ and $g_n \to g$, uniformly on compacts.

Similarly as in Lemma 2.3, a sufficient condition for (B.1) is that

$$(B.2) \qquad \sup_{n \in \mathbb{N}} \int e^{\beta c} \, d(\mu_n \otimes \nu_n) < \infty$$

and (f_n, g_n) are normalized such that (2.9) holds, for instance $\int f_n \, d\mu_n = 0$.

Proof of Proposition B.1. We first show that $(f_n)_{n\in\mathbb{N}}$ is pointwise bounded. Let $C = \sup_n (\int e^{\beta g_n}\, d\nu_n)^{1/\beta}$. As $c \geq 0$, we have by (A.2) that

$$e^{-f_n(x)} = \int e^{g_n(y)-c(x,y)}\, \nu_n(dy) \leq \int e^{g_n(y)}\, \nu_n(dy) \leq C.$$

This yields the uniform lower bound $f_n \geq -\log C$, and similarly $g_n \geq -\log C$. Fix $x \in \mathcal{X}$. Using (A.2) and the lower bound,

$$e^{-f_n(x)} = \int e^{g_n(y)-c(x,y)}\, \nu_n(dy) \geq C^{-1} \int e^{-c(x,y)}\, \nu_n(dy).$$

As $e^{-c(x,\cdot)}$ is bounded and continuous and $\nu_n \to \nu$ weakly, $\int e^{-c(x,y)}\, \nu_n(dy)$ converges to $\int e^{-c(x,y)}\, \nu(dy) > 0$. We conclude that $\sup_{n\in\mathbb{N}} f_n(x) < \infty$, completing the proof that (f_n) is pointwise bounded.

Next, we show that (f_n) is equicontinuous. On the strength of the pointwise boundedness, it suffices to show that e^{-f_n} is equicontinuous. Fix a compatible metric on $\mathcal{X}$, let $x_0 \in \mathcal{X}$ and $\varepsilon > 0$. Using tightness, choose a compact $K \subset \mathcal{Y}$ with $\nu_n(K^c) < \varepsilon$ for all n. Let $q \in (1,\infty)$ satisfy $1/\beta + 1/q = 1$. By Hölder's inequality,

$$\begin{aligned}
|e^{-f_n(x_0)} - e^{-f_n(x)}| &= \left| \int e^{g_n(y)-c(x_0,y)}\, \nu_n(dy) - \int e^{g_n(y)-c(x,y)}\, \nu_n(dy) \right| \\
&\leq \left(\int e^{\beta g_n(y)}\, \nu_n(dy) \right)^{1/\beta} \left(\int \left| e^{-c(x_0,y)} - e^{-c(x,y)} \right|^q \nu_n(dy) \right)^{1/q} \\
&\leq C \left(\int \left| e^{-c(x_0,y)} - e^{-c(x,y)} \right|^q \nu_n(dy) \right)^{1/q}.
\end{aligned}$$

Recalling that $e^{-c} \leq 1$, the integral can be estimated by

$$\int \left| e^{-c(x_0,y)} - e^{-c(x,y)} \right|^q \nu_n(dy) \leq \int_K \left| e^{-c(x_0,y)} - e^{-c(x,y)} \right|^q \nu_n(dy) + 2\nu_n(K^c).$$

As c is continuous and K is compact, $\sup_{y\in K} \left| e^{-c(x_0,y)} - e^{-c(\cdot,y)} \right|$ is continuous. Therefore, $\int_K |e^{-c(x_0,y)} - e^{-c(x,y)}|^q \nu_n(dy) < \varepsilon$ for $x \in B_\delta(x_0)$, for $\delta > 0$ sufficiently small, and we obtain the desired equicontinuity,

$$|e^{-f_n(x_0)} - e^{-f_n(x)}| \leq C(\varepsilon + 2\varepsilon)^{1/q}, \qquad x \in B_\delta(x_0).$$

We have shown that (f_n) is equicontinuous and pointwise bounded, and the same arguments hold for (g_n). Passing to a subsequence, the Arzelà–Ascoli theorem shows that $f_n \to f$ and $g_n \to g$ uniformly on compacts. Note that $(e^{f_n\oplus g_n-c})_n$ is $(\mu_n\otimes\nu_n)_n$-uniformly integrable by (B.1). As $e^{f_n\oplus g_n-c} \to e^{f\oplus g-c}$ uniformly on compacts and $\mu_n \otimes \nu_n \to \mu \otimes \nu$ weakly, it follows for the optimal couplings π_n that

$$\pi_n = e^{f_n\oplus g_n-c}\, d(\mu_n \otimes \nu_n) \to e^{f\oplus g-c}\, d(\mu \otimes \nu) =: \pi_0 \qquad \text{weakly.}$$

In particular, $\pi_0 \in \Pi(\mu,\nu)$. Proposition A.2 now shows that π_0 is the optimal coupling for (μ,ν) and (f,g) are corresponding potentials. As $f \oplus g$ is unique (Proposition A.1), the claim for the original sequence follows. $\square$

Remark B.2. Proposition B.1 extends to the boundary case $\beta = 1$ if e^{-c} is uniformly continuous. This holds in particular when $c(x, y) = \|x - y\|^2$ on $\mathbb{R}^d \times \mathbb{R}^d$, or more generally whenever c is coercive on $\mathcal{X} \times \mathcal{Y}$.

Following the above proof, the argument for pointwise boundedness still applies, and the argument for equicontinuity is even simpler under the additional hypothesis. The argument using uniform integrability may no longer be clear, but we can instead use Theorem 2.1 to conclude that (f, g) must be potentials for (μ, ν).

APPENDIX C. OMITTED PROOFS

Proof of Lemma 2.3. (i) Let $\inf_{n \in \mathbb{N}} \int g_n \, d\nu_n \geq -C$ for $C \geq 0$. By (A.2) and Jensen's inequality,

$$f_n(x) = -\log \left(\int e^{g_n(y) - c(x,y)} \, \nu_n(dy) \right)$$

$$\leq \int [c(x, y) - g_n(y)] \, \nu_n(dy) \leq C + \int c(x, y) \, \nu_n(dy).$$

Recalling $c \geq 0$, this shows

(C.1) $$f_n^+(x) \leq C + \int c(x, y) \, \nu_n(dy),$$

and now Tonelli's theorem yields $\int f_n^+ \, \mu_n \leq C + \int c \, d(\mu_n \otimes \nu_n)$. The analogue holds for g_n. Thus (2.11) implies (2.1), and via (2.9) also (2.10).

More generally, given a measurable set $A \subset \mathcal{X}$, (C.1) also implies

$$\int_A f_n^+ \, \mu_n \leq C\mu_n(A) + \int_{A \times \mathcal{Y}} c \, d(\mu_n \otimes \nu_n).$$

If $\mu_n(A) \leq \delta$, then $(\mu_n \otimes \nu_n)(A \times \mathcal{Y}) \leq \delta$, so that the ε–δ characterization of uniform integrability yields the claim.

(ii) The variational representation of relative entropy [28, Lemma 1.3] shows that

$$H(\mu_n \otimes \nu_n | \mu \otimes \nu) \geq \int \psi \, d(\mu_n \otimes \nu_n) - \log \int e^\psi \, d(\mu \otimes \nu)$$

for any measurable function ψ bounded from below. Choosing $\psi = \beta c$, we deduce

$$\int \beta c \, d(\mu_n \otimes \nu_n) \leq H(\mu_n \otimes \nu_n | \mu \otimes \nu) + \log \int e^{\beta c} \, d(\mu \otimes \nu).$$

Noting that $H(\mu_n \otimes \nu_n | \mu \otimes \nu) = H(\mu_n | \mu) + H(\nu_n | \nu)$, the right-hand side is bounded under (2.12), so that (2.11) applies. To obtain the last claim, we replace c with $\phi(c)$ in the preceding argument and apply the la Vallée–Poussin theorem. $\square$

Proof of Lemma 2.5. Let $C > 0$ and $A_n = \{\frac{d\mu}{d\mu_n} \geq C\}$. Then $\mu(A_n) = \int_{A_n} \frac{d\mu}{d\mu_n} \, d\mu_n \geq C\mu_n(A_n)$ and thus $\mu_n(A_n) \leq C^{-1}$ for all $n \in \mathbb{N}$. Writing

$$\mu(A_n) = \mu(A_n) - \mu_n(A_n) + \mu_n(A_n) \leq \|\mu - \mu_n\|_{TV} + C^{-1}$$

then shows that $\limsup_n \mu(A_n) \leq C^{-1}$. The claim follows. $\square$

Proof of Corollary 2.6. By Corollary 2.4 and Lemma 2.5, we have $f_n \to f$ in μ-probability and $g_n \to g$ in ν-probability after passing to a subsequence. Consider the Lebesgue decomposition $\mu_n = \mu'_n + \mu''_n$ into $\mu'_n \ll \mu$ and $\mu''_n \perp \mu$. Then $\mu'_n \to \mu$ in total variation and hence $\frac{d\mu'_n}{d\mu} \to 1$ in $L^1(\mu)$. This implies the convergence in μ-probability of the reciprocal, and as $\frac{d\mu}{d\mu'_n} = \frac{d\mu}{d\mu_n}$ μ-a.s., that $\frac{d\mu}{d\mu_n} \to 1$ in μ-probability. Similarly, $\frac{d\nu}{d\nu_n} \to 1$ in ν-probability. Following the proof of the particular case in Section 2 but writing the reciprocals,

$$\frac{d(\mu \otimes \nu)}{d\pi_n} = \frac{d(\mu_n \otimes \nu_n)}{d\pi_n} \frac{d\mu}{d\mu_n} \frac{d\nu}{d\nu_n} = e^{-(f_n \oplus g_n - c)} \frac{d\mu}{d\mu_n} \frac{d\nu}{d\nu_n} \to e^{-(f \oplus g - c)} = \frac{d(\mu \otimes \nu)}{d\pi_*}$$

in $\mu \otimes \nu$-probability. Consider the Lebesgue decomposition $\pi_n = \pi'_n + \pi''_n$ into $\pi'_n \ll \mu \otimes \nu$ and $\pi''_n \perp \mu \otimes \nu$. Then it follows that

$$\frac{d\pi'_n}{d(\mu \otimes \nu)} = \left(\frac{d(\mu \otimes \nu)}{d\pi_n} \right)^{-1} \mathbf{1}_{\left\{ \frac{d(\mu \otimes \nu)}{d\pi_n} \neq 0 \right\}} \to \frac{d\pi_*}{d(\mu \otimes \nu)}$$

in $\mu \otimes \nu$-probability, which by Scheffé's lemma implies $\pi'_n \to \pi_*$ in total variation. As π_n, π_* are probability measures, it follows that $\pi''_n(\mathcal{X} \times \mathcal{Y}) \to 0$ and finally $\pi_n \to \pi_*$ in total variation. The convergence of the original sequence follows. $\square$

Proof of Lemma 4.1. Fix $\varepsilon \in (0, \delta)$, to be determined later. By Prokhorov's theorem, we can find compacts $\mathcal{X}_{\mathrm{cpt}}$ and $\mathcal{Y}_{\mathrm{cpt}}$ with $\inf_{n \in \mathbb{N}_0} \mu_n(\mathcal{X}_{\mathrm{cpt}}) \geq 1 - \varepsilon^2/2$ and $\inf_{n \in \mathbb{N}_0} \nu_n(\mathcal{Y}_{\mathrm{cpt}}) \geq 1 - \varepsilon^2/2$. As $\pi_n \in \Pi(\mu_n, \nu_n)$, it follows that

$$(\mathrm{C.2}) \qquad \inf_{n \in \mathbb{N}} \pi_n(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) \geq 1 - \varepsilon^2.$$

Fix $n \in \mathbb{N}$ and consider the set

$$A_n = \left\{ x \in \mathcal{X}_{\mathrm{cpt}} : \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x) + g_n(y) - c(x,y)} \nu_n(dy) \geq 1 - \varepsilon \right\};$$

we claim that its complement satisfies

$$(\mathrm{C.3}) \qquad p_n := \mu_n(A_n^c) \leq \varepsilon.$$

Indeed, (A.2) yields

$$(\mathrm{C.4}) \qquad \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x) + g_n(y) - c(x,y)} \nu_n(dy) \leq \int e^{f_n(x) + g_n(y) - c(x,y)} \nu_n(dy) = 1$$

and thus

$$1 - \varepsilon^2 \overset{(C.2)}{\leq} \pi_n(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) = \int_{\mathcal{X}_{\mathrm{cpt}}} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x)+g_n(y)-c(x,y)} \, \nu_n(dy)\mu_n(dx)$$

$$\leq \int_{A_n^c} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x)+g_n(y)-c(x,y)} \, \nu_n(dy)\mu_n(dx)$$

$$+ \int_{A_n} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x)+g_n(y)-c(x,y)} \, \nu_n(dy)\mu_n(dx)$$

$$\overset{(C.4)}{\leq} (1-\varepsilon)p_n + (1-p_n) = 1 - p_n\varepsilon,$$

which implies (C.3). Next, we observe from the definition of A_n and (C.4) that for $x \in A_n$,

$$-\left(\log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x,y)} \, \nu_n(dy) - \log(1-\varepsilon) \right) \leq f_n(x)$$

$$(C.5) \hspace{5cm} \leq - \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x,y)} \, \nu_n(dy).$$

Let $x_1, x_2 \in A_n$ and assume without loss of generality that $f_n(x_1) \geq f_n(x_2)$. Then

$$|f_n(x_1) - f_n(x_2)| \overset{(C.5)}{\leq} \left(\log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x_2,y)} \, \nu_n(dy) - \log(1-\varepsilon) \right)$$

$$- \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x_1,y)} \, \nu_n(dy)$$

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

$$(C.6) \hspace{3cm} - \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x_1,y)} \, \nu_n(dy)$$

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

$$- \log(1-\varepsilon) - \log \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{g_n(y)-c(x_1,y)} \, \nu_n(dy)$$

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

This concludes the proof of the first estimate in the lemma. Turning to the second, note that by (C.2), (C.3) and the definition of A_n,

$$\pi_n(A_n \times \mathcal{Y}_{\mathrm{cpt}}) \geq \pi_n(\mathcal{X}_{\mathrm{cpt}} \times \mathcal{Y}_{\mathrm{cpt}}) - \pi_n(\mathcal{X} \setminus A_n \times \mathcal{Y}_{\mathrm{cpt}})$$

$$(C.7) \hspace{2.5cm} \geq 1 - \varepsilon^2 - \int_{A_n^c} \int_{\mathcal{Y}_{\mathrm{cpt}}} e^{f_n(x)+g_n(y)-c(x,y)} \, \nu_n(dy) \, \mu_n(dx)$$

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

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

$$B_n = \left\{ y \in \mathcal{Y}_{\mathrm{cpt}} : \int_{A_n} e^{f_n(x)+g_n(y)-c(x,y)} \, \mu_n(dx) \geq 1 - \delta \right\}.$$

Arguing as for (C.3) and (C.5), now using (C.7) instead of (C.2), we see that $\nu_n(B_n^c) \leq \delta$ and that for $y \in B_n$,

$$-\left(\log \int_{A_n} e^{f_n(x)-c(x,y)} \, \mu_n(dx) - \log(1-\delta) \right) \leq g_n(y)$$

$$\leq -\log \int_{A_n} e^{f_n(x)-c(x,y)} \, \mu_n(dx).$$

We conclude the proof by arguing as in (C.6) but with f_n, ε replaced by g_n, δ. $\qquad \square$

References

[1] D. Alvarez-Melis and T. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890, 2018.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. volume 70 of *Proceedings of Machine Learning Research*, pages 214–223, 2017.

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

[4] E. Bernton, P. Ghosal, and M. Nutz. Entropic optimal transport: Geometry and large deviations. *Duke Math. J., to appear*, 2021. arXiv:2102.04397.

[5] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

[6] J. Blanchet, A. Jambulapati, C. Kent, and A. Sidford. Towards optimal running times for optimal transport. *Preprint arXiv:1810.07717v1*, 2018.

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

[8] G. Carlier and M. Laborde. A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.*, 52(1):709–717, 2020.

[9] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.*, 169(2):671–691, 2016.

[10] V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, 45(1):223–256, 2017.

[11] A. Chiarini, G. Conforti, G. Greco, and L. Tamanini. *In preparation.*

[12] 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.

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

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

[15] G. Deligiannidis, V. De Bortoli, and A. Doucet. Quantitative uniform stability of the iterative proportional fitting procedure. *Preprint arXiv:2108.08129v1*, 2021.

[16] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statistics*, 11:427–444, 1940.

[17] S. Eckstein and M. Nutz. Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *Preprint arXiv:2110.06798v2*, 2021.

[18] 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.

[19] J. Franklin and J. Lorenz. On the scaling of multidimensional matrices. *Linear Algebra Appl.*, 114/115:717–735, 1989.

[20] P. Ghosal, M. Nutz, and E. Bernton. Stability of entropic optimal transport and Schrödinger bridges. *Preprint arXiv:2106.03670v1*, 2021.

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

[22] C. T. Ireland and S. Kullback. Contingency tables with given marginals. *Biometrika*, 55:179–188, 1968.

[23] S. Kullback. Probability densities with given marginals. *Ann. Math. Statist.*, 39:1236–1243, 1968.

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

[25] 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.

[26] T. Mikami. Optimal control for absolutely continuous stochastic processes and the mass transportation problem. *Electron. Comm. Probab.*, 7:199–213, 2002.

[27] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields*, 129(2):245–260, 2004.

[28] M. Nutz. *Introduction to Entropic Optimal Transport.* Lecture notes, Columbia University, 2021. https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

[29] M. Nutz and J. Wiesel. Entropic optimal transport: Convergence of potentials. *Probab. Theory Related Fields, to appear.* arXiv:2104.11720v2.

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

[31] 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.

[32] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.*, 40:99–121, 2000.

[33] L. Rüschendorf. Convergence of the iterative proportional fitting procedure. *Ann. Statist.*, 23(4):1160–1174, 1995.

[34] E. Schrödinger. Über die Umkehrung der Naturgesetze. Sitzungsberichte Preuss. Akad. Wiss. *Akad. Wiss., Berlin. Phys. Math.*, 144:144–153, 1931.

[35] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.

[36] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, 21:343–348, 1967.

[37] J. Weed. An explicit analysis of the entropic penalty in linear programming. volume 75 of *Proceedings of Machine Learning Research*, pages 1841–1855, 2018.

(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