# Asymptotics for semi-discrete entropic optimal transport*

Jason M. Altschuler†, Jonathan Niles-Weed‡, and Austin J. Stromme†

**Abstract.** We compute exact second-order asymptotics for the cost of an optimal solution to the entropic optimal transport problem in the continuous-to-discrete, or *semi-discrete*, setting. In contrast to the discrete-discrete or continuous-continuous case, we show that the first-order term in this expansion vanishes but the second-order term does not, so that in the semi-discrete setting the difference in cost between the unregularized and regularized solution is *quadratic* in the inverse regularization parameter, with a leading constant that depends explicitly on the value of the density at the points of discontinuity of the optimal unregularized map between the measures. We develop these results by proving new pointwise convergence rates of the solutions to the dual problem, which may be of independent interest.

**Key words.** Optimal transport, entropic optimal transport, semi-discrete optimal transport, second-order asymptotics

**AMS subject classifications.** 41A60, 58E30, 49N15

## 1. Introduction.

The entropically regularized optimal transportation problem, originally inspired by a thought experiment of Schrödinger [52] and the subject of a great deal of recent interest in probability [26, 40], statistics [13, 27, 39, 50] and machine learning [19, 28], is an optimization problem which seeks a coupling between two probability measures that minimizes the transport cost between them, subject to an additional entropic penalty. Specifically, given Borel probability measures $\mu$ and $\nu$ on $\mathbb{R}^d$ with finite second moment and $\eta > 0$, the problem reads

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi \,\|\, \mu \otimes \nu)\,, \tag{1.1}$$

where $\Pi(\mu,\nu)$ denotes the set of couplings of $\mu$ and $\nu$ and $\mathrm{KL}(\cdot\,\|\,\cdot)$ denotes the Kullback–Leibler divergence or relative entropy, defined by

$$\mathrm{KL}(\pi \,\|\, \rho) := \begin{cases} \int \log \frac{d\pi}{d\rho}(x) d\pi(x) & \pi \ll \rho \\ +\infty & \text{otherwise.} \end{cases}$$

Recent interest in (1.1) has been driven by the fact that, as $\eta \to \infty$, the solution $\pi_\eta$ to (1.1) approaches the solution $\pi^*$ to the unregularized optimal transport problem [10, 35],

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[\|x-y\|^2]\,, \tag{1.2}$$

*Submitted to the editors August 11, 2021.

**Funding:** This work was partially supported by NSF Graduate Research Fellowship 1122374, a TwoSigma PhD Fellowship, NSF grant DMS-2015291, and NDSEG Fellowship F-6749924378.

†Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA, 02139. (jasonalt@mit.edu, astromme@mit.edu)

‡Courant Institute of Mathematical Sciences and the Center for Data Science, New York University, New York, NY, 10003. (jnw@cims.nyu.edu)

which defines the squared Wasserstein distance $W_2^2(\mu,\nu)$ [57]. In statistics and machine learning applications, it has been recognized that (1.1) represents a computationally and statistically attractive proxy for (1.2). Statistically, the entropically regularized problem offers improved sample complexity [27] and cleaner limit laws [39] than its unregularized counterpart; computationally, the strict convexity of (1.1) opens the door to much faster algorithms [1, 19].

The importance of the $\eta \to \infty$ limit has spurred a line of work which seeks to quantify the speed of convergence of $\pi_\eta \to \pi^*$ and to develop higher-order asymptotics in the $\eta \to \infty$ regime. Of particular interest is the *suboptimality* of the entropically regularized solution:

$$\mathbb{E}_{\pi_\eta}[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2]\,.$$

This quantity measures the suitability of $\pi_\eta$ as an approximation for $\pi^*$, and giving precise bounds is essential for statistical and computational applications.

Two cases are well understood, with vastly different rates: when $\mu$ and $\nu$ are both finitely supported, then it is known that the difference in cost approaches zero exponentially fast as $\eta \to \infty$ [14, 58]. On the other hand, when $\mu$ and $\nu$ are absolutely continuous measures with bounded, compactly supported densities, then precise asymptotics to second order are known for the cost *including the entropic term* [13, 15, 25, 47]: as $\eta \to \infty$,

$$\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\,\mathrm{KL}(\pi_\eta \,\|\, \mu\otimes\nu) = W_2^2(\mu,\nu) - \frac{d}{2\eta}\log\left(\frac{\pi}{\eta}\right) + \frac{1}{2\eta}(h(\mu)+h(\nu)) + \frac{1}{16\eta^2}I(\mu,\nu) + o(\eta^{-2})\,, \tag{1.3}$$

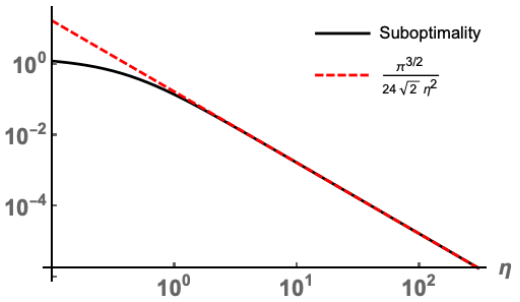
where for a probability measure $\mu$ with density $\mu(\cdot)$ with respect to the Lebesgue measure we write

$$h(\mu) := -\int \log(\mu(x))\mu(x)dx$$

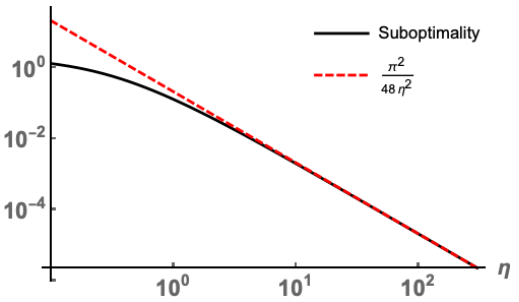
for the entropy relative to the Lebesgue measure, and where $I$ is the integrated Fisher information along the Wasserstein geodesic connecting $\mu$ to $\nu$. It does not seem possible to extract asymptotics for the cost $\mathbb{E}_{\pi_\eta}[\|x-y\|^2]$ directly from (1.3); however, it is easy to show that in general for absolutely continuous $\mu$ and $\nu$, the suboptimality is linear in $\eta^{-1}$. For example, when $\mu$ and $\nu$ are Gaussian measures on $\mathbb{R}$, it can be checked directly that

$$\mathbb{E}_{\pi_\eta}[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2] = \frac{1}{2\eta} + o(\eta^{-1})\,.$$

The large gulf between these convergence rates—exponential for finitely supported measures, linear in $\eta^{-1}$ for absolutely continuous measures—raises the question of which of the two behaviors should be expected in general. As a first step towards understanding this question, we study a situation between these two extremes: the *semi-discrete* case, in which one measure is absolutely continuous and the other is finitely supported. This setting is important for both theoretical and practical reasons, but prior work gives no hint of how the suboptimality in the semi-discrete case should behave. Should one expect to recover the exponential rate or the linear rate?




(a) $\mu$ is standard Gaussian distribution, i.e., has density $\mu(x) = e^{-x^2/2}/\sqrt{2\pi}$.




(b) $\mu$ is standard Laplacian distribution, i.e., has density $\mu(x) = e^{-|x|}/2$.

Figure 1: For two toy examples in one dimension, numerics show that the suboptimality scales quadratically in $\eta^{-1}$, and that the leading constant is an explicit function of the value of the density at 0. Our main result, Theorem 1.1, extends this to the general setting. The agreement between the predicted limiting value and the numerical results is precise.

A numerical computation in one dimension, where all the quantities are explicit, shows, perhaps surprisingly, that the rate in the semi-discrete case is something else entirely. Figure 1 plots the suboptimality for two different one-dimensional examples as $\eta$ varies, one where $\mu$ is the Gaussian density, and the other where $\mu$ is the Laplacian density. For both experiments, we take $\nu$ to be a discrete measure, uniform on $\{-1, +1\}$. The apparent result is that in both cases, the suboptimality is neither linear nor exponential but *quadratic* in $\eta^{-1}$. Moreover, the very careful reader will note that the asymptotic suboptimality appears to agree with $\frac{\pi^2\mu(0)}{24}\eta^{-2}$, where $\mu(0)$ is the value of the density $\mu$ at the origin, which is also the point at which the optimal *unregularized* map from $\mu$ to $\nu$ changes value from $-1$ to $+1$. We give a full exposition of this example in section 3.

Our main theorem shows that this phenomenon is completely general: in any dimension, if $\nu$ is discrete and $\mu$ has sufficiently regular density with respect to the Lebesgue measure, then the suboptimality scales as $\eta^{-2}$, with leading constant given by the value of $\mu$'s density on the hyperplanes on which the optimal map changes value.

Theorem 1.1. *Suppose $\mu$ and $\nu$ are Borel probability measures on $\mathbb{R}^d$ such that $\nu$ is finitely supported on $y_1, \ldots, y_n$, and $\mu$ is absolutely continuous and compactly supported, with positive, continuous density on the interior of its connected support. Then*

$$\mathbb{E}_{\pi_\eta}[\|x - y\|^2] = W_2^2(\mu, \nu) + \frac{\zeta(2)}{2\eta^2} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|} + o(\eta^{-2}), \tag{1.4}$$

*where $w_{ij}$ is the $(d-1)$-dimensional integral of $\mu(x)$ on $\overline{T^{-1}(y_i)} \cap \overline{T^{-1}(y_j)}$ for the optimal map $T$ transporting $\mu$ to $\nu$ (see subsection 2.3), and where $\zeta(2) = \frac{\pi^2}{6}$.*

See section 5 for a precise statement and proof of this result. The assumption that $\mu$ is compactly supported is mostly for convenience and can be substantially weakened; see Assump-

tion 2.9 and Proposition 2.10. By contrast, the continuity and positivity of $\mu$ are essential: in the absence of these assumptions, the convergence rate is no faster than $O(\eta^{-1})$ in general.

As an intermediate result, we also obtain an exact second-order expression for the cost with the entropic term. In what follows, we write $H(\nu) = -\sum_{i=1}^n \nu_i \log \nu_i$ to denote the Shannon entropy of a discrete distribution $\nu$ with weights $\nu_1, \dots, \nu_n$ on its atoms.

Theorem 1.2. *Suppose $\mu, \nu$ are as in Theorem 1.1. Then*

$$\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi_\eta \,\|\, \mu \otimes \nu) = W_2^2(\mu,\nu) + \frac{1}{\eta}H(\nu) - \frac{\zeta(2)}{2\eta^2}\sum_{i<j}\frac{w_{ij}}{\|y_i - y_j\|} + o(\eta^{-2})\,. \tag{1.5}$$

It would be interesting to find a heuristic argument to relate (1.5) to (1.3). In any case, the fact that the right side is $O(\eta^{-1})$ rather than $O(\eta^{-1}\log\eta)$ is a manifestation of the fact that the unregularized optimal coupling $\pi^*$ has finite relative entropy with respect to the product measure $\mu \otimes \nu$ [43].

Our proof techniques differ from prior work on the asymptotics of entropically regularized optimal transport. Prior work in the continuous setting has exploited a dynamical formulation [11, 29, 30] analogous to the celebrated Benamou–Brenier formula from the theory of unregularized optimal transport [5]. Instead, we take a different approach that, similar in spirit to the one employed in the analysis of the discrete problem [14], focuses on the convex dual of (1.1). However, our proof techniques depart substantially from those available in the discrete case, where finite-dimensional considerations make the analysis of the dual problem more tractable. In particular, the quadratic terms in Theorems 1.1 and 1.2 come from showing that the discrepancy of the quadratic cost between $\pi_\eta$ and $\pi^*$ localizes around the boundaries of the power diagram that determines the optimal unregularized transport map $\pi^*$, and then explicitly computing the resulting integrals up to low-order terms.

Our main technical result, which is of possible independent interest, gives first-order asymptotics for the convergence of solutions of the convex dual of (1.1) to solutions of the dual of (1.2), showing that this convergence happens faster than $\eta^{-1}$.

Theorem 1.3. *Suppose $\mu, \nu$ are as in Theorem 1.1. Let $(f_\eta, g_\eta)$ and $(f^*, g^*)$ solve the dual problems to (1.1) and (1.2), respectively with appropriate normalization constraints. (See Definitions 2.3 and 2.6.) Then*

$$\eta(f_\eta - f^*) \to 0$$
$$\eta(g_\eta - g^*) \to 0$$

*pointwise, with the latter convergence uniform.*

**1.1. Related work.** The study of optimal transport dates back to the fundamental contributions of Monge in the 18th century [41] and Kantorovich in the 20th [32]. Later in the 20th century, significant progress was made on the qualitative nature of optimal transport solutions, with many independent discoveries of a fundamental characterization of optimal transport solutions (Theorem 2.1) [9, 17, 18, 33, 51]. Around the turn of the 21st century, it was recognized that optimal transport gives a deep geometric perspective on the space of probability distributions [38, 45]. This discovery led to new functional inequalities, stable notions of curvature for metric measure spaces, and especially new means of analyzing difficult PDEs [22, 23, 37, 46, 55].

In parallel to these theoretical developments, major effort was devoted to practical algorithms for computing optimal transport maps, particularly in the discrete-discrete case. Standard linear programming methods work quite effectively when the supports of each distribution are discrete with up to several thousand atoms [19,24]. However, for larger datasets linear programming methods become prohibitively slow, and approximations are required. The entropic regularization approach is the most popular approximation, first considered algorithmically by Sinkhorn [53] and Sinkhorn and Knopp [54] in the 1960s. These works gave fast algorithms based off iterative matrix scaling for computing the approximate optimal coupling. Cuturi introduced this work to the machine learning community in 2013 [19], which led to an explosion of interest in optimal transport for applications [48]. Subsequently, the entropic penalty has been applied to variants of the optimal transport problem, where it also leads to fast and practical algorithms [2,6,7,12].

Apart from its algorithmic implications, the entropic penalty has an interesting probabilistic interpretation dating back to Schrödinger. In Schrödinger's original motivation, (1.1) represents a formalization of the following *hot gas experiment*. Consider a collection of particles evolving according to Brownian motion, and suppose their initial and final distribution approximately coincide with the measures $\mu$ and $\nu$, respectively. Schrödinger asked for a description of the "most likely paths" of each particle, conditional on this starting and ending configuration. The entropically regularized optimal transport problem gives a way of making mathematical sense of this problem: the path measure governing the evolution of the particles can be obtained by convolving the optimal coupling $\pi_\eta$ given by the solution to (1.1) with a Brownian bridge [26]. This interpretation has led to a fruitful line of work understanding (1.1) through the lens of large-deviations principles, which also has helped to clarify the nature of the convergence of (1.1) to (1.2) as $\eta \to \infty$ [35].

Obtaining an asymptotic expansion of the cost $\mathbb{E}_{\pi_\eta}[\|x-y\|^2]$ or the entropic cost $\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi \,\|\, \mu \otimes \nu)$ in the $\eta \to \infty$ limit is the subject of a great deal of recent interest. In the discrete-discrete case, this question was first investigated in the broader context of entropically regularized linear programs by Cominetti and San Martín [14], who showed that the suboptimality converges to zero exponentially fast as $\eta \to \infty$.

In the continuous-continuous case, asymptotics have been computed to second order for the entropic cost, under regularity assumptions (see [15] and references therein). To our knowledge, however, no general asymptotics for the suboptimality (without the entropic term) are known, but examples—such as the Gaussian case mentioned above—show that the rate $\Theta(\eta^{-1})$ is typical.

Recently, Bernton et al. [8] developed a structural characterization of $\pi_\eta$ which allows them to establish a large-deviations principle for the convergence of $\pi_\eta$ to $\pi^*$, but they do not extract asymptotics for the cost. Our results in section 2 develop a similar structural characterization for semi-discrete couplings by a direct argument.

The semi-discrete case, which is the central focus of this work, is important both for theoretical and practical reasons. For instance, it reflects the practical situation of the statistician who has access to an empirical distribution $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ of samples from an unknown measure, and wishes to compare these samples to an absolutely continuous reference measure $\mu$. From a theoretical perspective, the semi-discrete setting is closely connected to the *optimal quantization* problem [21,31,49], which seeks the best approximation of an absolutely continu-

ous measure by a measure with finite support. The study of the structure of optimal couplings for semi-discrete problems has a long history in computational geometry, where such couplings are known as "power diagrams" [3,4]. We draw extensively on the properties of such diagrams in our geometrical results of section 2.

We emphasize that throughout this paper, the optimal transport problem is defined with respect to the quadratic cost $\|x-y\|^2$. This corresponds to the 2-Wasserstein distance over Euclidean space, which plays a preponderant role in both the theory and application of optimal transport [48, 56, 57]. It is an interesting question how the results we develop change if the transportation cost is different—e.g., if one considers a non-Euclidean ground metric or $p$-Wasserstein distances, $p \neq 2$. We suspect that the analysis techniques we develop may still be useful, but new challenges arise. For example, in the case of general costs, the optimal unregularized map is given by a much more complicated partition of $\mathbb{R}^d$ than a power diagram, and this raises new challenges for developing the integration identities that we establish in this paper for the quadratic cost.

**1.2. Organization of the remainder of the paper.** In section 2, we formalize important definitions and establish basic geometrical results on the structure of the optimal regularized and unregularized couplings. To illustrate our ideas, in section 3 we develop the one-dimensional example mentioned above, and give a preview of the argument that will follow in the general case. Section 4 contains the proof of our main technical result, Theorem 1.3, which is at the heart of our arguments. In section 5, we apply this convergence result to prove Theorem 1.1. Finally, section 6 contains necessary background information on the dilogarithm and zeta functions, as well as several intermediate integration lemmas needed for the proofs of our main theorems. It also contains the proofs of two technical results from section 2.

**2. Background on semi-discrete OT and Sinkhorn problems.** In this section we recall relevant background on semi-discrete OT and Sinkhorn problems, as well as provide several useful propositions and intuitions for the work that comes. For further background we refer the reader to the standard textbooks [48, 56, 57], as well as to the detailed treatment of the semi-discrete setting in [42, section 4].

**2.1. Semi-discrete optimal transport.** The foundational observation in optimal transport theory declares the existence, uniqueness, and structure of the optimal coupling in the transport problem. For a proof, see e.g., [56, Theorem 2.12].

Theorem 2.1. *Suppose $\mu, \nu$ are probability measures with finite second moment. Then there is an optimal coupling $\pi^* \in \Pi(\mu,\nu)$ such that*

$$W_2^2(\mu,\nu) = \mathbb{E}_{\pi^*}[\|x-y\|^2].$$

*Moreover, we have the following form of strong duality:*

$$W_2^2(\mu,\nu) = \sup_{(f,g)\in L^1(\mu)\times L^1(\nu)\,:\, f(x)+g(y)\le \|x-y\|^2} \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g]. \tag{2.1}$$

*If $\mu$ has a density with respect to the Lebesgue measure, then in fact there is a unique optimal $\pi^*$, it is supported on the graph of a function $T\colon \mathbb{R}^d \to \mathbb{R}^d$, and $T$ is the gradient of a (proper,*

*lower semi-continuous) convex function. We shall usually write* $T = T_{\mu\to\nu} = \nabla\phi_{\mu\to\nu}$. *In this case the supremum in the dual problem* (2.1) *is attained by*

$$(f, g) = (\|x\|^2 - 2\phi_{\mu\to\nu}, \|y\|^2 - 2\phi^c_{\mu\to\nu})$$

*where we are using the Legendre conjugate*

$$\phi^c_{\mu\to\nu}(y) := \sup_x \langle x, y\rangle - \phi_{\mu\to\nu}(x).$$

The optimal $f$ and $g$ are typically not unique. However, the following assumptions guarantee that, up to an additive shift, $f$ and $g$ are unique $\mu$ (respectively, $\nu$) almost surely [8,20].

Assumption 2.2. *The measure* $\nu$ *is finitely supported and* $\mu$ *is absolutely continuous with finite second moment. The interior of the support of* $\mu$ *is connected, the boundary of the support has zero Lebesgue measure, and* $\mu$ *has positive density on the interior of its support.*

Under Assumption 2.2, we can therefore uniquely identify a pair of optimal dual solutions.

Definition 2.3 (Optimal unregularized potentials). *We denote by* $(f^*, g^*)$ *optimal solutions to* (2.1) *subject to the additional normalization constraint that* $\mathbb{E}_\nu[g^*] = 0$.

Using Theorem 2.1, we can completely characterize the optimal transport maps in the semi-discrete case. In what follows, we identify $\mu$ with its Lebesgue density $\mu(\cdot)$, and write $\{y_i\}_{i=1}^n$ for the support of $\nu$.

Theorem 2.4 ( [4]). *Adopt Assumption 2.2. Then,* $\mu$*-almost surely,*

$$T_{\mu\to\nu}(x) = \operatorname*{arg\,min}_{y_i \in \mathrm{supp}(\nu)} (\|x - y_i\|^2 - g^*(y_i)).$$

*Proof.* For ease of notation, write $\phi := \phi_{\mu\to\nu}$. Since $\phi$ is convex and closed, we know that $\phi = (\phi^c)^c$, where $(\cdot)^c$ denotes Legendre conjugation. Therefore,

$$\phi(x) = \max_{y_i} \langle x, y_i\rangle - \phi^c(y_i).$$

Since $\mu$ is absolutely continuous, there is a unique maximizer for $\mu$-almost every $x$, and if $y_i$ is the unique maximizer for such an $x$, then $\nabla\phi(x) = y_i$, and

$$\|x - y_i\|^2 - \|y_i\|^2 + 2\phi^c(y_i) < \|x - y_j\|^2 - \|y_j\|^2 + 2\phi^c(y_j) \quad \forall j \neq i\,.$$

Therefore we have shown that $\mu$-almost everywhere,

$$T(x) = \operatorname*{arg\,min}_{y_i} (\|x - y_i\|^2 - (\|y_i\|^2 - 2\phi^c(y_i))).$$

This yields the result by the characterization in Theorem 2.1. ∎

In view of this result, the next definition is natural.

Definition 2.5 ( [3]). *We define the* power cells *with respect to the optimal dual potential* $g^*$ *by*

$$S_i := \{x \in \mathbb{R}^d : \forall j\ \|x - y_i\|^2 - g^*(y_i) \le \|x - y_j\|^2 - g^*(y_j)\}, \quad i = 1, \dots, n.$$

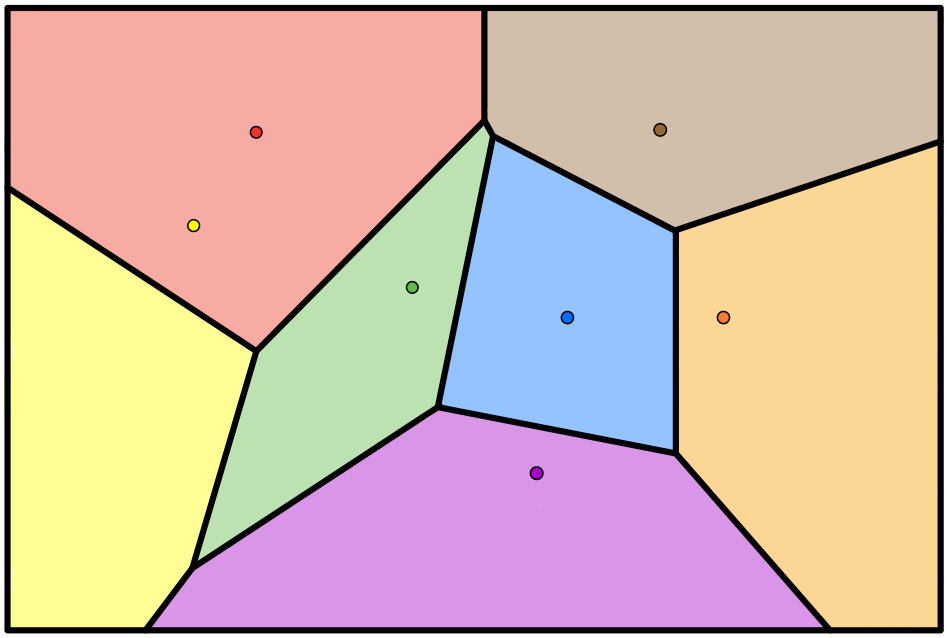


Figure 2: Illustration of a power diagram, or equivalently the optimal coupling for a semi-discrete OT problem. Each shaded region is a power cell $S_i$ corresponding to the point $y_i$ with the same color.

The significance of the power cells $S_i$ is that they are precisely the pull-back of $y_i$ under $T_{\mu\to\nu}$:

$$S_i = T_{\mu\to\nu}^{-1}(y_i).$$

The power cells for $\pi^*$ form a convex polyhedral partition of $\mathbb{R}^d$. In Figure 2 we show an example of an optimal mapping between a measure on the larger rectangle and a finitely supported measure. Note that a point $y_i$ in the support of $\nu$ can lie in the power cell $S_j$ corresponding to a different point $y_j \neq y_i$. For example, this occurs if $\mu$ is supported on $(-\infty, -2]$ and $\nu = (1/2)\delta_{-1} + (1/2)\delta_1$.

**2.2. Semi-discrete entropic optimal transport.** In this subsection, we discuss the entropy regularized version of the semi-discrete optimal transport problem. Denote by $\rho$ the counting measure on the support of $\nu$. We first note that for any $\pi \in \Pi(\mu, \nu)$, we have

$$\mathrm{KL}(\pi \,\|\, \mu \otimes \nu) = \mathrm{KL}(\pi \,\|\, \mu \otimes \rho) + H(\nu)\,. \tag{2.2}$$

The regularized optimal transport problem (1.1) is therefore equivalent to

$$\inf_{\pi \in \Pi(\mu,\nu)} \mathbb{E}_\pi[\|x - y\|^2] + \frac{1}{\eta}\,\mathrm{KL}(\pi \,\|\, \mu \otimes \rho)\,. \tag{2.3}$$

The benefit of the formulation (2.3) is that under Assumption 2.2,

$$\mathrm{KL}(\pi^* \,\|\, \mu \otimes \rho) = 0\,,$$

which leads to a simplification in some of the formulas appearing in what follows.

Csiszár's theory of "I-projection" [16] implies that as long as $\mu$ and $\nu$ have finite second moment, the value of (2.3) equals the value of the dual problem

$$\sup_{(f,g)\in L^1(\mu)\times L^1(\nu)} \mathbb{E}_\mu[f] + \mathbb{E}_\nu[g] - \frac{1}{\eta}\sum_{j=1}^{n}\int_{\mathbb{R}^d} e^{-\eta(\|x-y_j\|^2 - f(x) - g(y_j))}\mu(x)dx + \frac{1}{\eta}\,. \tag{2.4}$$

Moreover, the optimal solution to (2.3) satisfies

$$\frac{d\pi_\eta}{d(\mu\otimes\rho)}(x,y) = e^{-\eta(\|x-y\|^2 - f_\eta(x) - g_\eta(y))}\,, \tag{2.5}$$

where $f_\eta$ and $g_\eta$ solve (2.4).

The strict convexity of (2.4) implies that $f_\eta$ and $g_\eta$ are unique up to an additive shift; as above, we therefore fix a unique optimal pair by adding an additional constraint.

Definition 2.6 (Optimal regularized potentials). *We denote by $(f_\eta, g_\eta)$ solutions to (2.4), subject to the additional normalization constraint $\mathbb{E}_\nu[g_\eta] = 0$.*

**2.3. Useful geometric notions.** The power cell decomposition of Definition 2.5 gives us a useful way to separate the subproblems arising in our proof into individual problems over the cells $S_i$. In the service of analyzing these problems, we will focus on the distance of a point $x \in S_i$, from each of the hyperplanes defining $S_i$. We call these quantities the *slacks*, in reference to the fact that they represent the slack in the dual feasibility constraints in (2.1).

Definition 2.7 (Slack). *Let $i, j \in [n]$. The $j$-th slack at point $x \in S_i$ is*

$$\Delta_{ij}(x) := \|x - y_j\|^2 - f^*(x) - g^*(y_j). \tag{2.6}$$

We establish several basic properties of this slack operator.

Lemma 2.8 (Properties of slack). *For $i, j \in [n]$ and $x \in S_i$,*

- *Nonnegativity.* $\Delta_{ij}(x) \geq 0$, *with strict inequality $\mu$-almost everywhere if $i \neq j$.*
- *Diagonals vanish.* $\Delta_{ij}(x) = 0$ *if $i = j$.*
- *Expression via $g^*$.* $\Delta_{ij}(x) = 2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g^*(y_j) + g^*(y_i)$.

*Proof.* Nonnegativity follows by feasibility of $(f^*, g^*)$ for the dual OT problem (2.1), with strict inequality following from the fact that $\|x-y_i\|^2 - g^*(y_i) < \|x-y_j\|^2 - g^*(y_j)$ in the interior of $S_i$. The vanishing $\Delta_{ii} \equiv 0$ follows from the fact that $\|x - y\|^2 - f^*(x) - g^*(y) = 0$ $\pi^*$-almost surely, by strong duality. For the final item, observe that

$$\Delta_{ij}(x) = \|x - y_j\|^2 - f^*(x) - g^*(y_j) = \|x - y_j\|^2 - \|x - y_i\|^2 + g^*(y_i) - g^*(y_j)$$

where the second step is because $\|x - y_i\|^2 = f^*(x) + g_i^*$ by the previous item $\Delta_{ii}(x) = 0$. Now expand the square. ∎

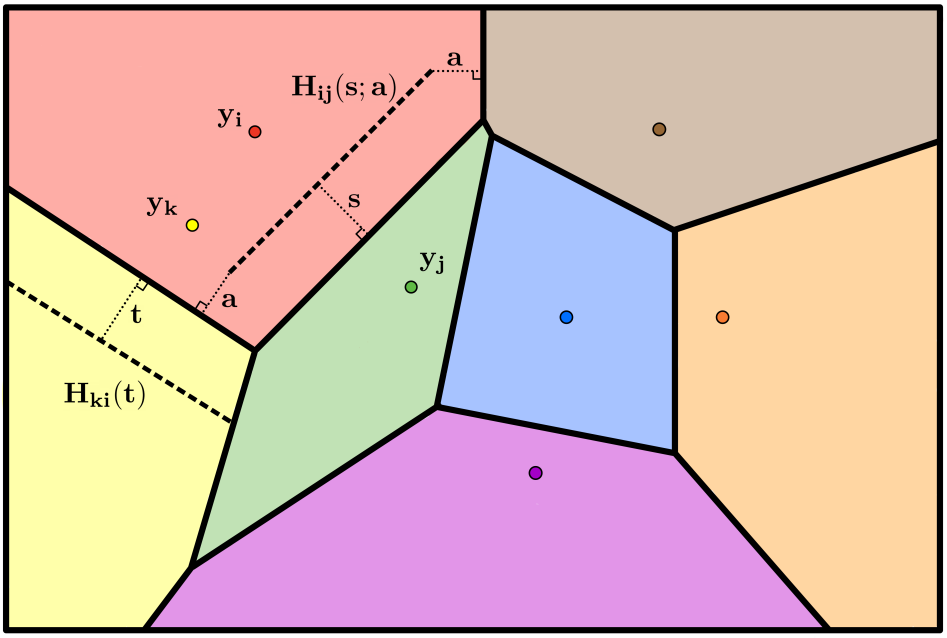



Figure 3: Power diagram with our $H_{ki}(t)$, $H_{ij}(s;a)$ notation depicted. The region $S_{ij}(a)$ defined in (2.7) is the subset of $S_i$ obtained by pushing in the hyperplanes separating $S_i$ from all neighboring cells other than $S_j$.

Our second main assumption on the measure $\mu$ relates to the regularity of the density along level sets defined by the slacks. We require several definitions. For $i \neq j$ and $a \geq 0$, set

$$S_{ij}(a) := \{x \in \mathbb{R}^d : \|x - y_i\|^2 - g_i^* \leq \|x - y_k\|^2 - g_k^* - a\mathbf{1}_{k \neq i,j}, \quad \forall k \in [n]\} \tag{2.7}$$
$$= \{x \in S_i : \Delta_{ik}(x) \geq a, \quad \forall k \neq i, j\}.$$

When $a = 0$, $S_{ij}(0) = S_i$. Also, for $t \geq 0$, let $H_{ij}(t;a) = \{x \in S_{ij}(a) : \Delta_{ij}(x) = t\}$ be the intersection of this set with a hyperplane parallel to the boundary between $S_i$ and $S_j$. See Figure 3 for an illustration.

Since $\mathbf{1}[x \in S_{ij}(a)]\mu(x)$ is in $L^1(\mathbb{R}^d)$, we can define

$$h_{ij}(t;a) := \int_{H_{ij}(t;a)} \mu(x) d\mathcal{H}_{d-1}(x) \in L^1(\mathbb{R}) \tag{2.8}$$

where $\mathcal{H}_{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure on $H_{ij}(t;a)$. When $a = 0$, we abbreviate $H_{ij}(t;a)$ and $h_{ij}(t;a)$ by $H_{ij}(t)$ and $h_{ij}(t)$, respectively.

The benefit of this definition is that it gives us a convenient way to integrate functions that depend only on the slacks; indeed, the coarea formula implies that for any nonnegative $\phi : \mathbb{R} \to \mathbb{R}$,

$$\int_{S_{ij}(a)} \phi(\Delta_{ij}(x))\mu(x)dx = \frac{1}{2\|y_i - y_j\|} \int_0^\infty \phi(t) h_{ij}(t;a) dt. \tag{2.9}$$

We require the following crucial condition on the measure $\mu$.

Assumption 2.9. *For all $i \neq j$ and $a \geq 0$ sufficiently small, the functions $t \mapsto h_{ij}(t;a)$ and $a \mapsto h_{ij}(0;a)$ are continuous at 0.*

Assumption 2.9 is a strong requirement on the regularity of $\mu$ along hyperplanes, and it is essential for our results. As alluded to in the statement of Theorem 1.1, it is possible to verify Assumption 2.9 under easy conditions on $\mu$. Say that $\mu$ is *dominated along hyperplanes* if for any affine hyperplane $H$ orthogonal to a vector $v$ there exists a nonnegative $\psi : \mathbb{R}^{d-1} \to \mathbb{R}$, integrable with respect to the Lebesgue measure, and an affine isometry $P : H \to \mathbb{R}^{d-1}$ such that

$$\mu(x + tv) \leq \psi(Px) \quad \forall t \in \mathbb{R}, x \in H\,.$$

If $\mu$ is pointwise bounded and compactly supported, then it is dominated along hyperplanes; however, some non-compactly supported measures, such as the standard Gaussian measure on $\mathbb{R}^d$ also enjoy this property.

Proposition 2.10. *If $\mu$ is continuous and dominated along hyperplanes, then Assumption 2.9 holds.*

Finally, we record a simple consequence of the connectedness of the support of $\mu$, which we will rely on extensively in section 4.

Lemma 2.11. *Under Assumption 2.2, we have $h_{ij}(0) = h_{ji}(0)$ for all $i \neq j$, and the graph on $[n]$ with edge set $\{(i,j) : h_{ij}(0) > 0\}$ is connected.*

The proofs of Proposition 2.10 and Lemma 2.11 appear in section 6.

**3. Case study: symmetric one-dimensional measures.** In order to provide intuition for our main result, we consider here a toy example which, despite its simplicity, illustrates many of the key underlying phenomena. Specifically, in this section we explicitly compute the suboptimality in the case where $\mu$ has a symmetric density on $\mathbb{R}$ and $\nu$ is the discrete distribution $\nu = (1/2)\delta_{-1} + (1/2)\delta_1$. The symmetry of both distributions around 0 allows us to compute closed-form expressions for $\pi^*$ and $\pi_\eta$, and hence also for the suboptimality. These closed-form expressions hold for any $\eta > 0$ and facilitate understanding our assumptions and main techniques.

**Unregularized optimal transport plan $\pi^*$.** By symmetry of $\mu$, the optimal coupling $\pi^*$ is supported on the graph a function that sends $x \in \mathrm{supp}(\mu)$ to $\mathrm{sgn}(x)$. That is,

$$\pi^*(x,y) = \mathbf{1}[y = \mathrm{sgn}(x)] \cdot \mu(x).$$

**Regularized optimal transport plan $\pi_\eta$.** Let us compute the dual potentials $f_\eta, g_\eta$ from Definition 2.6. Symmetry of the distributions around 0 implies

$$\pi_\eta(x,y) = \pi_\eta(-x,-y).$$

Using (2.5) and solving, this means $f_\eta(x) - f_\eta(-x) = g_\eta(-y) - g_\eta(y)$ for all $x \in \mathrm{supp}(\mu)$ and $y \in \mathrm{supp}(\nu)$. Replacing $x$ with $-x$, we see that both $f_\eta$ and $g_\eta$ must be even functions. By our convention in Definition 2.6, it follows that $g_\eta(1) = g_\eta(-1) = 0$.

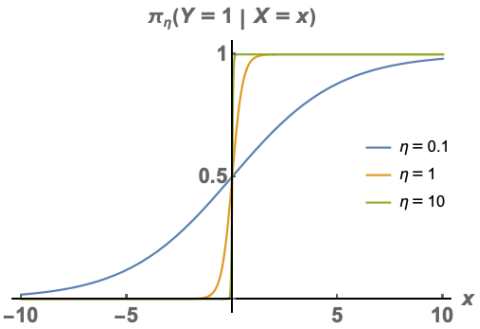



Figure 4: For the toy example in section 3, the conditional distribution $\pi_\eta(Y = 1|X = x)$ of the regularized plan $\pi_\eta$ is the sigmoid function $1/(1 + e^{-4\eta x})$ by (3.1). As $\eta \to \infty$, this converges to the conditional distribution $\pi^*(Y = 1|X = x) = \mathbf{1}[\mathrm{sign}(x) = 1]$ of the unregularized plan $\pi^*$. The convergence is exponential in $\eta$ at any $x \neq 0$. There is a symmetric region around the origin of width $\Theta(1/\eta)$ on which $\pi_\eta(Y = 1|X = x)$ is bounded away from 0 and 1.

We can now solve for $f_\eta$ using the marginal constraint $\mu(x) = \pi_\eta(x, 1)+\pi_\eta(x, -1)$. Plugging in the optimality conditions (2.5) for $\pi_\eta$ and simplifying implies

$$e^{\eta f_\eta(x)} = \frac{1}{e^{-\eta(x-1)^2} + e^{-\eta(x+1)^2}}.$$

Rearranging, we conclude that

$$\pi_\eta(x, y) = \frac{e^{-\eta(x-y)^2}}{e^{-\eta(x-1)^2} + e^{-\eta(x+1)^2}}\mu(x) = \frac{\mu(x)}{e^{2\eta x(1-y)} + e^{-2\eta x(1+y)}}. \tag{3.1}$$

See Figure 4 for an intuitive interpretation of $\pi_\eta$ as a smoothed version of $\pi^*$.

**Explicit evaluation of suboptimality.** By symmetry, marginal constraints, and the formula (3.1), we find

$$\begin{aligned}
\mathbb{E}_{\pi_\eta}[(x-y)^2] - \mathbb{E}_{\pi^*}[(x-y)^2] &= 2\int_0^\infty \Big((x-1)^2(\pi_\eta(x,1)-1) + (x+1)^2\pi_\eta(x,-1)\Big)\,dx \\
&= 2\int_0^\infty ((x+1)^2 - (x-1)^2)\pi_\eta(x,-1)dx \\
&= 8\int_0^\infty \frac{x}{1+e^{4\eta x}}\mu(x)dx. \qquad (3.2)
\end{aligned}$$

The dominant part of (3.2) as $\eta \to \infty$ is at $x = 0$, and if $\mu$ is continuous it can be shown that it is valid to replace $\mu(x)$ by $\mu(0)$ to obtain

$$\mathbb{E}_{\pi_\eta}[(x-y)^2] - \mathbb{E}_{\pi^*}[(x-y)^2] \approx 8\int_0^\infty \frac{x}{1+e^{4\eta x}}\mu(0)dx = -\frac{\mathrm{Li}_2(-1)\mu(0)}{2\eta^2} = \frac{\pi^2\mu(0)}{24\eta^2}.$$

Here, $\mathrm{Li}_2$ is the *dilogarithm* function, which will play a central role in our argument. More details about this function—as well as the so-called Fermi-Dirac integral identity used above—can be found in section 6.

**Necessity of assumptions.** If $\mu$ fails to be continuous at zero, convergence to 0 may be slower than quadratic. Consider $\mu(x) = c_p|x|^{-p}$ on $[-1,1]$ for $p < 1$ and normalizing constant $c_p = (1-p)/2$. The analysis above holds unchanged up to Equation 3.2. However, the following step, in which we approximated the integral by replacing $\mu(x)$ with $\mu(0)$, does not hold here since $\mu$ is not continuous at 0. Specifically,

$$\mathbb{E}_{\pi_\eta}[(x-y)^2] - \mathbb{E}_{\pi^*}[(x-y)^2] = 8c_p \int_0^1 \frac{x^{1-p}}{1+e^{4\eta x}} dx = \frac{2c_p}{\eta^{2-p}} \int_0^{4\eta} \frac{u^{1-p}}{1+e^u} du = \Theta\left(\frac{1}{\eta^{2-p}}\right).$$

This shows that in fact any polynomial rate faster than $1/\eta$ is achievable when our assumptions are violated. Morever, taking $\mu$ supported away from 0 shows that an exponential rate can be obtained when $\mu$ is not supported at the decision boundary.

**4. Convergence of dual potentials.** In this section, we develop an asymptotic expansion for the solution $g_\eta$ of (2.4) around the optimal solution $g^*$ to the unregularized problem (2.1). Recall that Assumption 2.2 implies that $g^*$ is unique, and also [44] that under this assumption $g_\eta$ converges to $g^*$. The main result of this section is a more precise result, showing that this convergence happens at the rate $o(\eta^{-1})$.

We prove the following.

Theorem 4.1. *Under Assumptions 2.2 and 2.9, the following convergence holds:*

$$\lim_{\eta\to\infty} \|\eta(g_\eta - g^*)\|_\infty = 0\,.$$

A consequence of Theorem 4.1 is that $\eta(f_\eta - f^*) \to 0$ pointwise, though we stress that this convergence is not uniform. Together, these results establish Theorem 1.3.

Corollary 4.2. *Under Assumptions 2.2 and 2.9, the following pointwise convergence holds:*

$$\lim_{\eta\to\infty} \eta(f_\eta - f^*) = 0\,.$$

From the general theory of entropic optimal transport, these results Theorem 4.1 and Corollary 4.2 are unexpected, and they reflect particular features of the semi-discrete setting. For instance, when $\mu$ and $\nu$ are both discrete, the quantities $\eta(g_\eta - g^*)$ and $\eta(f_\eta - f^*)$ both converge to positive limits in general. Moreover, Assumption 2.2 is essential: if $\mu$ is not positive on the interior of its support, it is possible for $\eta(g_\eta - g^*)$ to diverge.[1]

The proof of Theorem 4.1 also yields the following corollary on the difference between the Wasserstein distance and the entropic cost, which gives Theorem 1.2.

Corollary 4.3. *Under Assumptions 2.2 and 2.9,*

$$\lim_{\eta\to\infty} \eta^2 \left(\mathbb{E}_{\pi^*}[\|x-y\|^2] - (\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi_\eta \,\|\, \mu\otimes\rho))\right) = \frac{\zeta(2)}{2} \sum_{i<j} \frac{h_{ij}(0)}{\|y_i - y_j\|}\,.$$

[1] This occurs, for instance, when $\mu$ decays to zero at different rates on opposite sides of one of the hyperplane boundaries $H_{ij}$.

*Equivalently,*

$$\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi_\eta \,\|\, \mu\otimes\nu) = W_2^2(\mu,\nu) + \frac{1}{\eta}H(\nu) - \frac{\zeta(2)}{2\eta^2}\sum_{i<j}\frac{h_{ij}(0)}{\|y_i-y_j\|} + o(\eta^{-2})\,. \tag{4.1}$$

Below, we prove Theorem 4.1 in subsection 4.1, and then we show how Corollaries 4.2 and 4.3 follow from this in subsection 4.2.

**4.1. Proof of Theorem 4.1.** To prove Theorem 4.1, we define the function

$$d_\eta := \eta(g_\eta - g^*)\,.$$

We will show that $d_\eta$ is the unique solution to an auxiliary convex optimization problem whose solution gives the first-order difference between the Wasserstein distance $W_2^2(\mu,\nu)$ and the entropic cost $\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi_\eta \,\|\, \mu\otimes\rho)$. By showing that the zero function is an approximate optimizer of this auxiliary problem and establishing a form of strong convexity around 0 in the limit, we obtain that $d_\eta \to 0$, proving the claim.

We begin by defining these auxiliary optimization problems.

Proposition 4.4. *The function $d_\eta$ is the unique solution of*

$$\min_{d\in L_1(\nu):\mathbb{E}_\nu d=0}\sum_{i=1}^n\int_{S_i}\log(1+\sum_{j\neq i}e^{d(y_j)-d(y_i)-\eta\Delta_{ij}(x)})\mu(x)dx\,. \tag{4.2}$$

*Moreover, if we denote by $\Phi(\eta)$ the value of (4.2), then*

$$\Phi(\eta) = \eta\left(\mathbb{E}_{\pi^*}[\|x-y\|^2] - (\mathbb{E}_{\pi_\eta}[\|x-y\|^2] + \frac{1}{\eta}\mathrm{KL}(\pi_\eta\,\|\,\mu\otimes\rho))\right), \tag{4.3}$$

*and $\pi_\eta$ satisfies*

$$\frac{d\pi_\eta}{d(\mu\otimes\rho)}(x,y_j) = \frac{e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \qquad \forall x\in S_i, i\in[n]\,. \tag{4.4}$$

*Proof.* Recall that $f_\eta$ and $g_\eta$ are the unique solutions to (2.4) subject to the constraint $\mathbb{E}_\nu[g_\eta]=0$, so they also uniquely solve

$$\eta\cdot\min_{\substack{(f,g)\in L_1(\mu)\times L_1(\nu)\\ \mathbb{E}_\nu[g]=0}}\mathbb{E}_\mu[f^*]+\mathbb{E}_\nu[g^*]-\mathbb{E}_\mu[f]-\mathbb{E}_\nu[g]+\frac{1}{\eta}\sum_{j=1}^n\int_{\mathbb{R}^d}e^{-\eta(\|x-y_j\|^2-f(x)-g(y_j))}\mu(x)dx-\frac{1}{\eta}\,.$$

By duality, the optimal value of this program is exactly (4.3). Decomposing the integrals over the cells $S_i$ and recalling (2.6), we obtain that $f_\eta$ and $g_\eta$ are the unique solutions to

$$\min_{\substack{(f,g)\in L_1(\mu)\times L_1(\nu)\\ \mathbb{E}_\nu[g]=0}}\sum_{i=1}^n\int_{S_i}\Big(\eta(f^*(x)-f(x))+\eta(g^*(y_i)-g(y_i)) + \sum_{j=1}^n e^{-\eta(\Delta_{ij}(x)+f^*(x)-f(x)+g^*(y_j)-g(y_j))}\Big)\mu(x)dx - 1\,. \tag{4.5}$$

Reparametrizing in terms of $\delta_f = \eta(f - f^*)$ and $\delta_g = \eta(g - g^*)$ yields the equivalent representation

$$\min_{\substack{(\delta_f,\delta_g)\in L_1(\mu)\times L_1(\nu)\\ \mathbb{E}_\nu[\delta_g]=0}} \sum_{i=1}^n \int_{S_i} \Big( -1 - \delta_f(x) - \delta_g(y_i) + \sum_{j=1}^n e^{\delta_f(x)+\delta_g(y_j)-\eta\Delta_{ij}(x)}\Big)\mu(x)dx\,,$$

with optimal solutions $\eta(f_\eta - f^*)$ and $\eta(g_\eta - g^*)$. Fixing $\delta_g$ and minimizing this expression with respect to $\delta_f$ yields that the optimal solutions $\delta_f$ and $\delta_g$ are related by

$$\delta_f(x) = -\log\left(\sum_{j=1}^n e^{\delta_g(y_j)-\eta\Delta_{ij}(x)}\right) \tag{4.6}$$

for $\mu$-almost every $x \in S_i$. Plugging in this expression gives

$$\begin{aligned}\min_{\delta_g\in L_1(\nu):\,\mathbb{E}_\nu[\delta_g]=0} &\sum_{i=1}^n \int_{S_i} \Big( \log\big(\sum_{j=1}^n e^{\delta_g(y_j)-\eta\Delta_{ij}(x)}\big) - \delta_g(y_i)\Big)\mu(x)dx\\ &= \min_{\delta_g\in L_1(\nu):\mathbb{E}_\nu[\delta_g]=0} \sum_{i=1}^n \int_{S_i} \log\big(1 + \sum_{j\neq i} e^{\delta_g(y_j)-\delta_g(y_i)-\eta\Delta_{ij}(x)}\big)\mu(x)dx\,.\end{aligned}$$

Writing $d$ for $\delta_g$ yields (4.2).

Finally, applying the same argument to (2.5) yields

$$\begin{aligned}\frac{d\pi_\eta}{d(\mu\otimes\rho)}(x,y_j) &= e^{-\eta(\|x-y_j\|^2 - f_\eta(x) - g_\eta(y_j)}\\ &= e^{\delta_f(x)+\delta_g(y_j)-\eta\Delta_{ij}(x)}\\ &= \frac{e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}}\end{aligned}$$

for all $x \in S_i$ and $i \in [n]$, as desired. ∎

To prove the theorem, we require two intermediate results. First, we obtain an upper bound on $\Phi$ by comparing it to the value of (4.2) at $d = 0$. Though crude, this comparison will turn out to be accurate to first order.

Lemma 4.5.

$$\limsup_{\eta\to\infty} \eta\Phi(\eta) \le \frac{\zeta(2)}{4}\sum_{i\neq j}\frac{h_{ij}(0)}{\|y_i - y_j\|}\,.$$

*Proof.* Choose $d = 0$ in (4.2). The subadditivity of the function $\alpha \mapsto \log(1+\alpha)$ for $\alpha > 0$ and the optimality of $d(\eta)$ then imply

$$\begin{aligned}\Phi(\eta) &\le \sum_{i=1}^n \int_{S_i} \log\left(1 + \sum_{j\neq i} e^{-\eta\Delta_{ij}(x)}\right)\mu(x)dx\\ &\le \sum_{i=1}^n \sum_{j\neq i} \int_{S_i} \log(1 + e^{-\eta\Delta_{ij}(x)})\mu(x)dx\,.\end{aligned}$$

Multiplying by $\eta$, taking the limit, and applying Lemma 6.3 (with $S_{ij}(0) = S_i$) yields the claim. ∎

Next, we use Lemma 4.5 to show that the solutions to (4.2) remain bounded.

Proposition 4.6. *Under Assumption 2.2, $d_\eta$ is bounded as $\eta \to \infty$.*

*Proof.* The claim is obvious if $n = 1$, so assume $n \geq 2$. Fix $(i,j)$ for which $h_{ij}(0) > 0$. (Such a pair exists by Lemma 2.11.) Then by Proposition 4.4,

$$\begin{aligned}\eta\Phi(\eta) &= \eta \sum_{i=1}^n \int_{S_i} \log(1 + \sum_{j \neq i} e^{d_\eta(y_j) - d_\eta(y_i) - \eta\Delta_{ij}(x)})\mu(x)dx \\ &\geq \eta \int_{S_i} \log(1 + e^{d_\eta(y_j) - d_\eta(y_i) - \eta\Delta_{ij}(x)})\mu(x)dx.\end{aligned}$$

To bound this integral, we require the following lemma.

Lemma 4.7. *For any $a \geq 0$ and $b \in [0,1]$,*

$$\log(1 + ab) \geq \log(1 + a)\log(1 + b)\,. \tag{4.7}$$

*Proof.* Fix $b \in [0,1]$. Then (4.7) holds for $a = 0$, and the derivative of the left side in $a$ is $b/(1+ab) \geq b/(1+a)$, whereas the derivative of the right side in $a$ is $(\log(1+b))/(1+a) \leq b/(1+a)$. We obtain that (4.7) therefore holds for all $a \geq 0$. ∎

With this lemma in hand, we obtain

$$\eta\Phi(\eta) \geq \log(1 + e^{d_\eta(y_j) - d_\eta(y_i)}) \cdot \eta \int_{S_i} \log(1 + e^{-\eta\Delta_{ij}(x)})\mu(x)dx.$$

Taking the limit of both sides and using the change-of-variables (2.9) and Lemmas 4.5 and 6.2, we obtain

$$\sum_{i' \neq j'} \frac{h_{i'j'}(0)}{\|y_{i'} - y_{j'}\|} \geq \limsup_{\eta \to \infty} \log(1 + e^{d_\eta(y_j) - d_\eta(y_i)}) \frac{h_{ij}(0)}{\|y_i - y_j\|}\,,$$

showing that $d_\eta(y_j) - d_\eta(y_i)$ is bounded above for all $(i,j)$ for which $h_{ij}(0) > 0$. Now by Lemma 2.11, the graph on $[n]$ with edge set $\{(i,j) : h_{ij}(0) > 0\}$ is connected, so for any $(i,j) \in [n]^2$ we may find a path $(k_l)_{l=1}^L$ such that $k_1 = i$ and $k_L = j$, and $d_\eta(y_{k_{l+1}}) - d_\eta(y_{k_l})$ is bounded above for all $l = 1, \ldots, L-1$; as a result, we conclude that in fact $d_\eta(y_j) - d_\eta(y_i)$ is bounded for all $(i,j) \in [n]^2$. Finally, since $\mathbb{E}_\nu d_\eta = 0$, we conclude that $d_\eta$ is bounded. ∎

We now turn to the proof of the theorem. The boundedness of $d_\eta$ allows us to extract a convergent subsequence, and by passing to the limit we obtain strong convexity of (4.2) in the limit around 0.

*Proof of Theorem 4.1.* As above, we may assume $n \geq 2$. We will show that for any sequence $(\eta_s)_{s \geq 1}$, there exists a subsequence along which $d_\eta \to 0$. Let us fix such a sequence.

Since $d_\eta$ is bounded, by passing to a subsequence—which we again denote by $\eta_s$—we may assume that $d_\eta$ tends to a limit $d_\infty$.

Now, fix an $\varepsilon > 0$. Recall from section 2 that $S_{ij}(\varepsilon)$ is the subset of $S_i$ on which $\Delta_{ik} \geq \varepsilon$ for all $k \neq i, j$. By definition, then, the sets $S_{ij}(\varepsilon) \cap \{x \in S_i : \Delta_{ij} < \varepsilon\}$ for $j \neq i$ are disjoint subsets of $S_i$. We can therefore decompose the integral over $S_i$ into these sets to obtain

$$\begin{aligned}\Phi(\eta) &= \sum_{i=1}^{n} \int_{S_i} \log(1 + \sum_{k\neq i} e^{d_\eta(y_k) - d_\eta(y_i) - \eta\Delta_{ik}(x)})\mu(x)dx \\ &\geq \sum_{i=1}^{n}\sum_{j\neq i} \int_{S_{ij}(\varepsilon)\cap\{x\in S_i : \Delta_{ij}<\varepsilon\}} \log(1 + \sum_{k\neq i} e^{d_\eta(y_k) - d_\eta(y_i) - \eta\Delta_{ik}(x)})\mu(x)dx \\ &\geq \sum_{i=1}^{n}\sum_{j\neq i} \int_{S_{ij}(\varepsilon)\cap\{x\in S_i : \Delta_{ij}<\varepsilon\}} \log(1 + e^{d_\eta(y_j) - d_\eta(y_i) - \eta\Delta_{ij}(x)})\mu(x)dx\,.\end{aligned}$$

Multiplying by $\eta$ and taking the limit using Lemma 6.3 yields for $\varepsilon$ sufficiently small

$$\liminf_{s\to\infty} \eta_s \Phi(\eta_s) \geq \sum_{i\neq j} -\mathrm{Li}_2(-e^{d_\infty(y_j) - d_\infty(y_i)}) \frac{h_{ij}(0;\varepsilon)}{2\|y_i - y_j\|}\,.$$

Letting $\varepsilon \to 0$ and applying Assumption 2.9, we obtain

$$\liminf_{s\to\infty} \eta_s \Phi(\eta_s) \geq \sum_{i\neq j} -\mathrm{Li}_2(-e^{d_\infty(y_j) - d_\infty(y_i)}) \frac{h_{ij}(0)}{2\|y_i - y_j\|}\,.$$

Since $h_{ij}(0) = h_{ji}(0)$ by Lemma 2.11, we may symmetrize this sum to obtain

$$\liminf_{s\to\infty} \eta_s \Phi(\eta_s) \geq \sum_{i\neq j} \frac{1}{2}\left[-\mathrm{Li}_2(-e^{d_\infty(y_j) - d_\infty(y_i)}) - \mathrm{Li}_2(-e^{d_\infty(y_i) - d_\infty(y_j)})\right] \frac{h_{ij}(0)}{2\|y_i - y_j\|}\,.$$

By the inversion formula for the dilogarithm function [36, A.2.1(5)],

$$\frac{1}{2}\left[-\mathrm{Li}_2(-e^{d_\infty(y_j) - d_\infty(y_i)}) - \mathrm{Li}_2(-e^{d_\infty(y_i) - d_\infty(y_j)})\right] = \frac{\zeta(2)}{2} + \frac{1}{4}(d_\infty(y_j) - d_\infty(y_i))^2\,.$$

Combined with Lemma 4.5, we conclude

$$\begin{aligned}\frac{\zeta(2)}{4}\sum_{i\neq j}\frac{h_{ij}(0)}{\|y_i - y_j\|} &\geq \limsup_{s\to\infty} \eta_s\Phi(\eta_s) \\ &\geq \liminf_{s\to\infty} \eta_s\Phi(\eta_s) \\ &\geq \frac{\zeta(2)}{4}\sum_{i\neq j}\frac{h_{ij}(0)}{\|y_i - y_j\|} + \frac{1}{8}\sum_{i\neq j}(d_\infty(y_j) - d_\infty(y_i))^2\frac{h_{ij}(0)}{\|y_i - y_j\|}\,,\end{aligned}$$

implying that $d_\infty(y_j) = d_\infty(y_i)$ if $h_{ij}(0) \neq 0$, and that

$$\lim_{\eta\to\infty} \eta\Phi(\eta) = \frac{\zeta(2)}{4}\sum_{i\neq j}\frac{h_{ij}(0)}{\|y_i - y_j\|} = \frac{\zeta(2)}{2}\sum_{i<j}\frac{h_{ij}(0)}{\|y_i - y_j\|}\,. \tag{4.8}$$

We conclude as in the proof of Proposition 4.6. ∎

### 4.2. Proof of Corollaries 4.2 and 4.3.

*Proof of Corollary 4.2.* For shorthand, denote $\delta_f := \eta(f_\eta - f^*)$ and $\delta_g := \eta(g_\eta - g^*)$. By (4.6), these functions are related by the identity

$$\delta_f(x) = \log\left(\sum_{j=1}^n e^{\delta_g(y_j) - \eta\Delta_{ij}(x)}\right)$$

for $\mu$-almost every $x \in S_i$. Now by Theorem 4.1, $\lim_{\eta\to\infty} \delta_g(y_j) = 0$ for all $j \in [n]$. And by Lemma 2.8, for $\mu$-almost every $x \in S_i$, the slack $\Delta_{ij}(x)$ is zero for $j = i$ and otherwise is strictly positive for $j \neq i$. Thus, for $\mu$-almost every $x$,

$$\lim_{\eta\to\infty} \delta_f(x) = \log 1 = 0.$$

■

*Proof of Corollary 4.3.* This is immediate in light of (4.8), (4.3), and (2.2). ■

**5. Convergence of the suboptimality.** In this section we prove our main result, from which Theorem 1.1 follows.

Theorem 5.1. *Under Assumptions 2.2 and 2.9,*

$$\lim_{\eta\to\infty} \eta^2(\mathbb{E}_{\pi_\eta}[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2]) = \frac{\zeta(2)}{2}\sum_{i<j}\frac{h_{ij}(0)}{\|y_i - y_j\|}.$$

The proof uses two lemmas. The first lemma decomposes the suboptimality of an arbitrary coupling $\pi \in \Pi(\mu,\nu)$ into a sum of nonnegative terms involving the slack operators $\Delta_{ij}$.

Lemma 5.2 (Suboptimality decomposition). *For any $\pi \in \Pi(\mu,\nu)$,*

$$\mathbb{E}_\pi[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2] = \sum_{i\neq j}\int_{S_i}\Delta_{ij}(x)d\pi(x,y_j). \tag{5.1}$$

*Proof.* By strong duality and the fact that $\pi \in \Pi(\mu,\nu)$,

$$\mathbb{E}_{\pi^*}[\|x-y\|^2] = \mathbb{E}_\mu f^* + \mathbb{E}_\nu g^* = \mathbb{E}_\pi[f^*(x) + g^*(y)].$$

Therefore

$$\begin{aligned}\mathbb{E}_\pi[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2] &= \mathbb{E}_\pi[\|x-y\|^2 - f^*(x) - g^*(y)]\\ &= \sum_{i,j}\int_{S_i}[\|x-y_j\|^2 - f^*(x) - g^*(y_j)]d\pi(x,y_j)\\ &= \sum_{i,j}\int_{S_i}\Delta_{ij}(x)d\pi(x,y_j),\end{aligned}$$

where the last step uses the definition of $\Delta_{ij}$ (2.7). Since $\Delta_{ij}(x) = 0$ if $i = j$, the diagonal terms vanish, proving the claim. ■

The second lemma explicitly computes the integrals that result from using this decomposition on the coupling $\pi_\eta$. We recall the notation $d_\eta = \eta(g_\eta - g^*)$ from section 4.

Lemma 5.3 (Sigmoid slack integrals). *Under Assumptions 2.2 and 2.9, for any $i \neq j$,*

$$\lim_{\eta\to\infty} \eta^2 \int_{S_i} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \mu(x)dx = \frac{\zeta(2)h_{ij}(0)}{4\|y_i - y_j\|}. \tag{5.2}$$

*Proof.* First,

$$\lim_{\eta\to\infty} \eta^2 \int_{S_i} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \mu(x)dx \le \lim_{\eta\to\infty} \eta^2 \int_{S_i} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-d_\eta(y_i)-\eta\Delta_{ij}(x)}}{1 + e^{d_\eta(y_j)-d_\eta(y_i)-\eta\Delta_{ij}(x)}} \mu(x)dx\,,$$

and since $d_\eta \to 0$ by Theorem 4.1, we can apply Lemma 6.4 to conclude that the limit is bounded above by

$$-\mathrm{Li}_2(-1)\frac{h_{ij}(0)}{2\|y_i - y_j\|} = \frac{\zeta(2)h_{ij}(0)}{4\|y_i - y_j\|}\,.$$

On the other hand, for any $\varepsilon > 0$ and $c > 1$, we have

$$\begin{aligned}
\lim_{\eta\to\infty} \eta^2 \int_{S_i} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \mu(x)dx &\ge \lim_{\eta\to\infty} \eta^2 \int_{S_{ij}(\varepsilon)} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \mu(x)dx \\
&\ge \lim_{\eta\to\infty} \eta^2 \int_{S_{ij}(\varepsilon)} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{e^{d_\eta(y_i)} + (n-2)e^{2\|d_\eta\|_\infty - \eta\varepsilon} + e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}} \mu(x)dx \\
&\ge \lim_{\eta\to\infty} \eta^2 \int_{S_{ij}(\varepsilon)} \frac{\Delta_{ij}(x) e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{c + e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}} \mu(x)dx\,,
\end{aligned}$$

where we have used the fact that $d_\eta \to 0$, so that $e^{d_\eta(y_i)} + (n-2)e^{2\|d_\eta\|_\infty - \eta\varepsilon} < c$ for all $\eta$ sufficiently large. By Lemma 6.4, for $\varepsilon$ sufficiently small, this limit is

$$-\mathrm{Li}_2(-1/c)\frac{h_{ij}(0;\varepsilon)}{2\|y_i - y_j\|}\,,$$

and taking $c \to 1$ and $\varepsilon \to 0$ and applying Assumption 2.9, we obtain that the limit is also bounded below by

$$\frac{\zeta(2)h_{ij}(0)}{4\|y_i - y_j\|}\,,$$

completing the proof. ∎

With these two lemmas in hand, the proof of Theorem 1.1 follows readily.

*Proof of Theorem 1.1.* By Lemma 5.2 and (4.4),

$$\lim_{\eta\to\infty} \eta^2(\mathbb{E}_{\pi_\eta}[\|x-y\|^2] - \mathbb{E}_{\pi^*}[\|x-y\|^2]) = \lim_{\eta\to\infty} \sum_{i\neq j} \eta^2 \int_{S_i} \Delta_{ij}(x) \frac{e^{d_\eta(y_j)-\eta\Delta_{ij}(x)}}{\sum_k e^{d_\eta(y_k)-\eta\Delta_{ik}(x)}} \mu(x)dx\,.$$

By Lemma 5.3, this is equal to

$$\frac{\zeta(2)}{4}\sum_{i\neq j}\frac{h_{ij}(0)}{\|y_i-y_j\|}.$$

Summing over $i \neq j$ and using the symmmetry

$$\frac{h_{ij}(0)}{\|y_i-y_j\|} = \frac{h_{ji}(0)}{\|y_j-y_i\|}$$

finishes the proof. ∎

**6. Supplementary results.** This section collects several supplementary lemmas relating to the integration of relevant quantities depending on the slacks in the cell $S_i$, as well as the proofs of two technical claims from section 2.

**6.1. The dilogarithm function.** The properties of our asymptotic expansion—including the presence of the constant $\zeta(2)/2$—rely on several classical properties of the dilogarithm function. The claims below appear in [36].

Definition 6.1. *The* dilogarithm *function is given by*

$$\mathrm{Li}_2(z) = \sum_{s=1}^{\infty}\frac{z^s}{s^2} \quad |z|\leq 1$$

*and extended to* $\mathbb{C}\setminus(1,\infty)$ *by analytic continuation.*

An immediate consequence of this definition is the special value

$$\mathrm{Li}_2(-1) = \sum_{s=1}^{\infty}\frac{(-1)^s}{s^2} = -\frac{\zeta(2)}{2} = -\frac{\pi^2}{12}. \tag{6.1}$$

Moreover, the analyticity of $\mathrm{Li}_2$ away from the branch cut implies in particular that it is continuous on the negative reals.

The appearance of the dilogarithm in our proofs follows directly from two of its integral representations, which arise naturally from the solutions of the entropic optimal transport problem in the semi-discrete setting studied in this paper. These integral identities are often called Fermi-Dirac integrals in the mathematical physics literature.

Lemma 6.2 ( [36]). *The dilogarithm satisfies*

$$-\mathrm{Li}_2(-1/c) = \int_0^\infty \frac{te^{-t}}{c+e^{-t}}\,dt = \int_0^\infty \log(1+e^{-t}/c)\,dt$$

*for all* $c>0$. *In particular,*

$$-\mathrm{Li}_2(-1) = \int_0^\infty \frac{te^{-t}}{1+e^{-t}}\,dt = \int_0^\infty \log(1+e^{-t})\,dt = \frac{\zeta(2)}{2}.$$

Rather than using Lemma 6.2 directly, we will typically be integrating with respect to the measure $\mu$ over a power cell. However, as the following lemmas show, in the large-$\eta$ limit we can still employ the integral identities of Lemma 6.2 to obtain explicit expressions in terms of the dilogarithm.

Lemma 6.3. *Let $M_\eta$ be such that $\lim_{\eta\to\infty} M_\eta = M > 0$, and let $a > 0$ be small enough that Assumption 2.9 holds. Then*

$$\lim_{\eta\to\infty} \eta \int_{S_{ij}(a)} \log(1 + M_\eta e^{-\eta\Delta_{ij}(x)})\mu(x)dx = -\mathrm{Li}_2(-M)\frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\,.$$

*The same claim holds if $S_{ij}(a)$ is replaced by $S_{ij}(a) \cap \{x \in S_i : \Delta_{ij}(x) < a\}$.*

*Proof.* By a change of variables, we can write

$$\eta \int_{S_{ij}(a)} \log(1 + M_\eta e^{-\eta\Delta_{ij}(x)})\mu(x)dx = \frac{\eta}{2\|y_i - y_j\|}\int_0^\infty \log(1 + M_\eta e^{-\eta t})h_{ij}(t;a)dt\,.$$

Since $M_\eta$ tends to a limit, it is bounded, and so for any $\varepsilon > 0$ the function $\eta\log(1 + M_\eta e^{-\eta t})$ tends uniformly to 0 on $[\varepsilon, \infty)$. Since $h_{ij}(t;a) \in L_1$, this implies that

$$\lim_{\eta\to\infty} \eta \int_\varepsilon^\infty \log(1 + M_\eta e^{-\eta t})h_{ij}(t;a)dt = 0\,.$$

The integral therefore only depends on an interval near zero; in particular, replacing the set $S_{ij}(a)$ by $S_{ij}(a) \cap \{x \in S_i : \Delta_{ij}(x) < a\}$, which has the effect of integrating from 0 to $a$ instead of 0 to $\infty$, does not affect the value of the limit.

A second change of variables gives

$$\lim_{\eta\to\infty} \frac{\eta}{2\|y_i - y_j\|}\int_0^\varepsilon \log(1+M_\eta e^{-\eta t})h_{ij}(t;a)dt = \lim_{\eta\to\infty} \frac{1}{2\|y_i - y_j\|}\int_0^{\eta\varepsilon} \log(1+M_\eta e^{-t})h_{ij}(\eta^{-1}t;a)dt\,.$$

Let us first consider replacing $h_{ij}(\eta^{-1}t;a)$ by $h_{ij}(0;a)$. Dominated convergence and Lemma 6.2 then imply

$$\lim_{\eta\to\infty} \frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\int_0^{\eta\varepsilon} \log(1 + M_\eta e^{-t})dt = \frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\int_0^\infty \log(1 + Me^{-t})dt = -\mathrm{Li}_2(-M)\frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\,,$$

which is the desired limit.

It therefore suffices to show that replacing $h_{ij}(\eta^{-1}t;a)$ by $h_{ij}(0,a)$ is justified. If we make this replacement, we incur an error of size at most

$$\sup_{\delta\le\varepsilon} |h_{ij}(\delta;a) - h_{ij}(0,a)|\frac{1}{2\|y_i - y_j\|}\int_0^{\eta\varepsilon} \log(1 + M_\eta e^{-t})dt\,.$$

Since the integral is bounded and $h_{ij}(t;a)$ is continuous at $t = 0$ (Assumption 2.9), this error vanishes as $\varepsilon \to 0$, completing the proof. ∎

Lemma 6.4. *Let $M_\eta$ be such that $\lim_{\eta\to\infty} M_\eta = M > 0$, let $a \ge 0$ be small enough that Assumption 2.9 holds, and let $c > 0$ be arbitrary. Then*

$$\lim_{\eta\to\infty} \eta^2 \int_{S_{ij}(a)} \frac{\Delta_{ij}(x)M_\eta e^{-\eta\Delta_{ij}(x)}}{c + M_\eta e^{-\eta\Delta_{ij}(x)}}\mu(x)dx = -\mathrm{Li}_2(-M/c)\frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\,.$$

*Proof.* The proof is exactly analogous to that of Lemma 6.3. Fix $\varepsilon > 0$. First, by change of variables and the uniform convergence of $\frac{\eta^2 t M_\eta e^{-\eta t}}{c + M_\eta e^{-\eta t}}$ to 0 on $[\varepsilon, \infty)$, it suffices to evaluate

$$\lim_{\eta\to\infty} \frac{1}{2\|y_i - y_j\|}\eta^2 \int_0^\varepsilon \frac{tM_\eta e^{-\eta t}}{c + M_\eta e^{-\eta t}} h_{ij}(t;a)dt = \lim_{\eta\to\infty} \frac{1}{2\|y_i - y_j\|} \int_0^{\varepsilon\eta} \frac{tM_\eta e^{-t}}{c + M_\eta e^{-t}} h_{ij}(\eta^{-1}t;a)dt\,.$$

As above, replacing $h_{ij}(\eta^{-1}t;a)$ by $h_{ij}(0;a)$ incurs error that vanishes as $\varepsilon \to 0$. We obtain that the desired limit is

$$\lim_{\eta\to\infty} \frac{h_{ij}(0;a)}{2\|y_i - y_j\|} \int_0^{\varepsilon\eta} \frac{tM_\eta e^{-t}}{c + M_\eta e^{-t}} dt\,.$$

By dominated convergence and Lemma 6.2, this is

$$-\mathrm{Li}_2(-M/c)\frac{h_{ij}(0;a)}{2\|y_i - y_j\|}\,,$$

as desired. ∎

**6.2. Proof of Proposition 2.10.** The proof is inspired by [42, Lemma 46]. For any $i \neq j$, define the hyperplane

$$H_{ij} = \{x \in \mathbb{R}^d : 2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* = 0\}\,.$$

We require the following lemma.

Lemma 6.5. *If $g^*$ is optimal, then $H_{ij} \neq H_{ik}$ for all $j \neq k$.*

*Proof.* Suppose that $H_{ik}$ and $H_{ij}$ coincide for some $j \neq k$. Then the definition of $H_{jk}$ implies that it coincides with $H_{ik}$ and $H_{ij}$ as well. The cells $S_i$, $S_j$, and $S_k$ are convex sets with positive $\mu$ (and hence positive Lebesgue) measure; therefore, because the boundary of a convex set has zero Lebesgue measure (e.g., [34, Theorem 1]), it follows that the cells $S_i$, $S_j$, and $S_k$ have non-empty interiors. If we consider the two open halfspaces defined by the hyperplane $H_{ij} = H_{ik} = H_{jk}$, then there exist two of the cells—say, $S_i$ and $S_j$—whose interiors lie in the same open halfspace. But this contradicts the fact that $2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* > 0$ for all $x \in \mathrm{int}(S_i)$, and $2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* < 0$ for all $x \in \mathrm{int}(S_j)$. So $H_{ik}$ and $H_{ij}$ cannot coincide, as claimed. ∎

Let us fix an $a \geq 0$ sufficiently small and prove the continuity of $t \mapsto h_{ij}(t;a)$. Given a nonnegative sequence $t_n \to 0$, consider

$$\begin{aligned} h_{ij}(t_n;a) - h_{ij}(0;a) &= \int_{H_{ij}(t_n;a)} \mu(x)d\mathcal{H}_{d-1}(x) - \int_{H_{ij}(0;a)} \mu(x)d\mathcal{H}_{d-1}(x) \\ &= \int_{H_{ij}} (\mathbf{1}[x + t_n v \in S_{ij}(a)]\mu(x + t_n v) - \mathbf{1}[x \in S_{ij}(a)]\mu(x))d\mathcal{H}_{d-1}(x). \end{aligned}$$

Here $v$ denotes the vector $(y_i - y_j)/(2\|y_i - y_j\|^2)$, c.f., Lemma 2.8. Continuity of $\mu$ implies that $\mu(x + t_n) \to \mu(x)$ pointwise. We will now show that $\mathbf{1}[x + t_n v \in S_{ij}(a)] \to \mathbf{1}[x \in S_{ij}(a)]$ for

$\mathcal{H}_{d-1}$-almost every $x$. First, since $S_{ij}(a)$ is closed, if $x \notin S_{ij}(a)$ then $x \notin S_{ij}(a) - t_n v$ for all $t_n$ sufficiently close to 0. Thus, $\limsup_{n\to\infty} \mathbf{1}[x + t_n v \in S_{ij}(a)] \le \mathbf{1}[x \in S_{ij}(a)]$.

On the other hand, the set $S_{ij}(a)$ is a convex set defined by the constraints

$$\begin{aligned} &2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* \ge 0 \\ &2\langle x, y_i - y_k\rangle - \|y_i\|^2 + \|y_k\|^2 - g_k^* + g_i^* \ge a \qquad \forall k \ne i, j\,. \end{aligned}$$

By Lemma 6.5, $H_{ij} \ne H_{ik}$ for all $k \ne i, j$. It follows that for all $k \ne i, j$ and all $a \ge 0$ sufficiently small, the intersection of $H_{ij}$ and $\{x \in \mathbb{R}^d : 2\langle x, y_i - y_k\rangle - \|y_i\|^2 + \|y_k\|^2 - g_k^* + g_i^* = a\}$ has codimension at least 2. Therefore, for $\mathcal{H}_{d-1}$-almost every $x \in S_{ij}(a) \cap H_{ij}$,

$$\begin{aligned} &2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* = 0 \\ &2\langle x, y_i - y_k\rangle - \|y_i\|^2 + \|y_k\|^2 - g_k^* + g_i^* > a \qquad \forall k \ne i, j\,. \end{aligned}$$

For such $x$, we therefore have that $x + t_n v \in S_{ij}(a)$ for $t_n$ sufficiently close to 0, and $\liminf_{n\to\infty} \mathbf{1}[x + t_n v \in S_{ij}(a)] \ge \mathbf{1}[x \in S_{ij}(a)]$. Therefore, $\mathbf{1}[x + t_n v \in S_{ij}(a)] \to \mathbf{1}[x \in S_{ij}(a)]$ for $\mathcal{H}_{d-1}$-almost every $x$.

Since $\mu$ is dominated along hyperplanes, $\mathbf{1}[x + t_n v \in S_{ij}(a)]\mu(x + t_n v) - \mathbf{1}[x \in S_{ij}(a)]\mu(x)$ is dominated by an integrable function on $H_{ij}$, and the claim follows.

The second argument is simpler: given a sequence $a_n \to 0$, we have

$$h_{ij}(0; a_n) - h_{ij}(0; 0) = \int_{H_{ij}} (\mathbf{1}[x \in S_{ij}(a_n)] - \mathbf{1}[x \in S_i])\mu(x) d\mathcal{H}_{d-1}(x)\,.$$

Since $S_{ij}(a_n) \subseteq S_i$, it is clear that $\limsup_{n\to\infty} \mathbf{1}[x \in S_{ij}(a_n)] \le \mathbf{1}[x \in S_i]$. And as above, $\mathcal{H}_{d-1}$-almost every $x \in S_{ij} \cap H_{ij}$ satisfies

$$\begin{aligned} &2\langle x, y_i - y_j\rangle - \|y_i\|^2 + \|y_j\|^2 - g_j^* + g_i^* = 0 \\ &2\langle x, y_i - y_k\rangle - \|y_i\|^2 + \|y_k\|^2 - g_k^* + g_i^* > 0 \qquad \forall k \ne i, j\,. \end{aligned}$$

and for these $x$, $\liminf_{n\to\infty} \mathbf{1}[x \in S_{ij}(a_n)] \ge \mathbf{1}[x \in S_i]$. This proves the claim.

**6.3. Proof of Lemma 2.11.** That $h_{ij}(0) = h_{ji}(0)$ follows from the fact that $H_{ij}(0) = H_{ji}(0) = S_i \cap S_j$.

Now, we show that the graph with edge set $\{(i, j) : h_{ij}(0) > 0\}$ is connected. Since $\mu$ is positive on the interior of its support, if $h_{ij}(0) = 0$, then $\mathrm{int}(\mathrm{supp}(\mu)) \cap (S_i \cap S_j)$ has zero $\mathcal{H}_{d-1}$ measure. By [42, Lemma 49], this implies that the set

$$Z := \mathrm{int}(\mathrm{supp}(\mu)) \setminus \left( \bigcup_{ij : h_{ij}(0) = 0} S_i \cap S_j \right)$$

is path connected.

Now, suppose that the graph has $K$ connected components. For each component $C_k \subseteq [n]$, let

$$Z_k = \bigcup_{i \in C_k} (Z \cap S_i)\,.$$

Since each cell $S_i$ is closed and has positive $\mu$ mass, each $Z_k$ is nonempty and closed in the subspace topology on $Z$. Moreover, they are disjoint by the definition of $Z$. Therefore the $Z_k$ form a non-empty, closed partition of the connected set $Z$, so $K = 1$.

## REFERENCES


[1] J. Altschuler, J. Niles-Weed, and P. Rigollet, *Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration*, in Advances in Neural Information Processing Systems, 2017, pp. 1964–1974.

[2] J. M. Altschuler and P. A. Parrilo, *Approximating Min-Mean-Cycle for low-diameter graphs in near-optimal time and memory*, arXiv preprint arXiv:2004.03114, (2020).

[3] F. Aurenhammer, *Power diagrams: properties, algorithms and applications*, SIAM Journal on Computing, 16 (1987), pp. 78–96.

[4] F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica, 20 (1998), pp. 61–76.

[5] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[6] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM Journal on Scientific Computing, 37 (2015), pp. A1111–A1138.

[7] J.-D. Benamou, W. Ijzerman, and G. Rukhaia, *An entropic optimal transport numerical approach to the reflector problem*, (2020).

[8] E. Bernton, P. Ghosal, and M. Nutz, *Entropic optimal transport: geometry and large deviations*, arXiv preprint arXiv:2102.04397, (2021).

[9] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, CR Acad. Sci. Paris Sér. I Math., 305 (1987), pp. 805–808.

[10] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM Journal on Mathematical Analysis, 49 (2017), pp. 1385–1418.

[11] Y. Chen, T. T. Georgiou, and M. Pavon, *On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint*, Journal of Optimization Theory and Applications, 169 (2016), pp. 671–691.

[12] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling algorithms for unbalanced optimal transport problems*, Mathematics of Computation, 87 (2018), pp. 2563–2609.

[13] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard, and G. Peyré, *Faster Wasserstein distance estimation with the Sinkhorn divergence*, Advances in Neural Information Processing Systems, 33 (2020).

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

[15] G. Conforti and L. Tamanini, *A formula for the time derivative of the entropic cost and applications*, Journal of Functional Analysis, 280 (2021), p. 108964.

[16] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, The Annals of Probability, (1975), pp. 146–158.

[17] J. A. Cuesta and C. Matrán, *Notes on the Wasserstein metric in Hilbert spaces*, The Annals of Probability, 17 (1989), pp. 1264–1276.

[18] M. J. Cullen and R. J. Purser, *An extended lagrangian theory of semi-geostrophic frontogenesis*, Journal of the Atmospheric Sciences, 41 (1984), pp. 1477–1497.

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

[20] E. del Barrio, A. González-Sanz, and J.-M. Loubes, *Central limit theorems for general transportation costs*, arXiv preprint arXiv:2102.06379, (2021).

[21] S. Dereich, M. Scheutzow, and R. Schottstedt, *Constructive quantization: Approximation by empirical measures*, in Annales de l'IHP Probabilités et statistiques, vol. 49, 2013, pp. 1183–1203.

[22] L. Desvillettes and C. Villani, *On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: The linear Fokker-Planck equation*, Communications on Pure and Applied Mathematics, 54 (2001), pp. 1–42.

[23] L. Desvillettes and C. Villani, *On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation*, Inventiones Mathematicae, 159 (2005), pp. 245–316.

[24] Y. Dong, Y. Gao, R. Peng, I. Razenshteyn, and S. Sawlani, *A study of performance of optimal transport*, arXiv preprint arXiv:2005.01182, (2020).

[25] M. Erbar, J. Maas, M. Renger, et al., *From large deviations to Wasserstein gradient flows in multiple dimensions*, Electronic Communications in Probability, 20 (2015).

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

[27] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré, *Sample complexity of Sinkhorn divergences*, in International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 1574–1583.

[28] A. Genevay, G. Peyré, and M. Cuturi, *Learning generative models with Sinkhorn divergences*, in International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 1608–1617.

[29] I. Gentil, C. Léonard, and L. Ripani, *About the analogy between optimal transport and minimal entropy*, in Annales de la Faculté des sciences de Toulouse: Mathématiques, vol. 26, 2017, pp. 569–600.

[30] N. Gigli and L. Tamanini, *Benamou-Brenier and duality formulas for the entropic cost on* $\mathsf{RCD}^*(K,N)$ *spaces*, Probability Theory and Related Fields, 176 (2020), pp. 1–34.

[31] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Springer, 2007.

[32] L. V. Kantorovich, *Mathematical methods of organizing and planning production*, Management Science, 6 (1960), pp. 366–422. Translation. Originally published by Leningrad University in 1939.

[33] M. Knott and C. S. Smith, *On the optimal mapping of distributions*, Journal of Optimization Theory and Applications, 43 (1984), pp. 39–49.

[34] R. Lang, *A note on the measurability of convex sets*, Archiv der Mathematik, 47 (1986), pp. 90–92.

[35] C. Léonard, *From the Schrödinger problem to the Monge–Kantorovich problem*, Journal of Functional Analysis, 262 (2012), pp. 1879–1920.

[36] L. Lewin, *Polylogarithms and associated functions*, North Holland, 1981.

[37] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics, (2009), pp. 903–991.

[38] R. J. McCann, *A convexity principle for interacting gases*, Advances in Mathematics, 128 (1997), pp. 153–179.

[39] 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, 2019.

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

[41] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris, (1781).

[42] Q. Mérigot and B. Thibert, *Chapter 2 -optimal transport: discretization and algorithms*, in Geometric Partial Differential Equations - Part II, A. Bonito and R. H. Nochetto, eds., vol. 22 of Handbook of Numerical Analysis, Elsevier, 2021, pp. 133–212.

[43] M. Nutz, *Lectures on entropic optimal transport*. Lecture Notes, Columbia University., 2020.

[44] M. Nutz and J. Wiesel, *Entropic optimal transport: Convergence of potentials*, arXiv preprint arXiv:2104.11720, (2021).

[45] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations, 26 (2001), pp. 101–174.

[46] F. Otto and C. Villani, *Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality*, Journal of Functional Analysis, 173 (2000), pp. 361–400.

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

[48] G. Peyré and M. Cuturi, *Computational optimal transport: with applications to data science*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607.

743 [49] D. Pollard, *Quantization and the method of k-means*, IEEE Transactions on Information theory, 28
744 (1982), pp. 199–205.
745 [50] P. Rigollet and J. Weed, *Entropic optimal transport is maximum-likelihood deconvolution*, Comptes
746 Rendus Mathematique, 356 (2018), pp. 1228–1235.
747 [51] L. Rüschendorf and S. T. Rachev, *A characterization of random variables with minimum l2-distance*,
748 Journal of multivariate analysis, 32 (1990), pp. 48–54.
749 [52] E. Schrödinger, *Über die Umkehrung der Naturgesetze.*, Angewandte Chemie, 44 (1931), pp. 636–636.
750 [53] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, The
751 Annals of Mathematical Statistics, 35 (1964), pp. 876–879.
752 [54] R. Sinkhorn and P. Knopp, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific
753 Journal of Mathematics, 21 (1967), pp. 343–348.
754 [55] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta mathematica, 196 (2006), pp. 65–131.
755 [56] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American
756 Mathematical Society, Providence, RI, 2003.
757 [57] C. Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.
758 [58] J. Weed, *An explicit analysis of the entropic penalty in linear programming*, in Conference On Learning
759 Theory, 2018, pp. 1841–1855.