Characterizing and computing solutions to regularized semi-discrete optimal transport via an ordinary differential equation

Luca Nenna[*], Daniyar Omarov [†] and Brendan Pass [‡]

Thursday 3rd April, 2025

Keywords. Semi-discrete optimal transport, entropic regularization, ODE, convex analysis.
2020 Mathematics Subject Classification. Primary: 49Q22; Secondary: 49N15, 94A17, 49K40.

This paper investigates the semi-discrete optimal transport (OT) problem with entropic regularization. We characterize the solution using a governing, well-posed ordinary differential equation (ODE). This naturally yields an algorithm to solve the problem numerically, which we prove has desirable properties, notably including global strong convexity of a value function whose Hessian must be inverted in the numerical scheme. Extensive numerical experiments are conducted to validate our approach. We compare the solutions obtained using the ODE method with those derived from Newton's method. Our results demonstrate that the proposed algorithm is competitive for problems involving the squared Euclidean distance and exhibits superior performance when applied to various powers of the Euclidean distance. In addition, it proves particularly effective in scenarios where the target points lie outside the support of the source measure. Finally, we note that the ODE approach yields an estimate on the rate of convergence of the solution as the regularization parameter vanishes, for a generic cost function.

[*]Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: luca.nenna@universite-paris-saclay.fr
[†]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1. email: daniyar@ualberta.ca
[‡]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1. email: pass@ualberta.ca

1. Introduction

The optimal transport (OT) problem, which was first proposed by Monge in 1781 [24] and subsequently relaxed by Kantorovich during the 1940-s [19, 20], involves identifying the most efficient way to transfer mass from one probability measure to another, all while minimizing a specified cost function. This profound problem is intricately connected many areas of mathematics, including partial differential equations [18, 26] and statistics [28, 5, 30], and its applications are vast, extending into fields such as economics [13], fluid mechanics [2, 6], image processing [27], and machine learning [29]. In recent decades, advancements in computational techniques, notably entropic regularization [9, 3, 14, 8], have facilitated the practical implementation of OT in large-scale scenarios as well as the numerical resolution of many variational problems involving Optimal Transport terms.

In this paper, we focus specifically on the semi-discrete optimal transport (OT) problem, that is

$$\max_{\gamma\in\Pi(\rho,\mu)} \int b(x,y)\mathrm{d}\gamma \ , \tag{1}$$

where the source measure $\rho(x)$ is absolutely continuous with respect to the Lebesgue measure, the target measure μ is supported on a finite set and $\Pi(\rho,\mu)$ is the set of couplings having ρ and μ as marginals. This particular variant has garnered increased interest recently due to its relevance in applications such as geometric optics [7] and mesh generation [12]. Numerous studies have discussed various numerical algorithms relevant to this problem [21, 11]; we refer the reader to [23] and the references within for an overview on this topic. Motivated by some recent works in the discrete case [25, 17], our focus here is on the entropic regularized version of the semi-discrete OT problem

$$\max_{\gamma\in\Pi(\rho,\mu)} \int_{X\times Y} b(x,y)\mathrm{d}\gamma - \varepsilon\mathrm{Ent}(\gamma \mid \rho\otimes\sigma) \ , \tag{2}$$

where σ is the counting measure on Y and $\mathrm{Ent}(\gamma \mid \rho\otimes\sigma)$ is the relative entropy with respect to the product measure. In particular by studying the dual (semi-discrete) entropic problem

$$\min_{\mathbf{v}\in\mathbb{R}^N} \int_X \varepsilon \log\left(\sum_{k=1}^N e^{\frac{b(x,y_k)-v_k}{\varepsilon}}\right)\mathrm{d}\rho + \sum_{k=1}^N v_k\mu_k + \varepsilon \ , \ \varepsilon>0 \ . \tag{3}$$

we manage to equivalently characterize the curve of solutions with respect to the regularization parameter $\varepsilon\mapsto v(\varepsilon)$ through a well-posed (even when the regularization parameter vanishes) ordinary differential equation. This extends a result in the remarkable paper by Delalande [10] where it was shown that a given curves of solutions to the entropic problem satisfy an ODE, but not the converse, that any solution of the corresponding Cauchy problem must be in fact be a curve of solutions to the entropic OT problem. Let us also emphasize that our work here applies to general cost functions, whereas [10] focuses exclusively on the quadratic cost $b(x,y)=-|x-y|^2$. Moreover, we establish uniform bound on the smallest eigenvalue of the Hessian of the dual functional when the curve

$\varepsilon \mapsto v(\varepsilon)$ approaches the solution to the unregularized problem. This bound seems to be a special feature of the semi-discrete setting, as we don't see a clear way to obtain similar estimates in fully discrete problems. It has important computational consequences, as it allows one to avoid numerical instabilities and therefore compute the solution to the fully unregularized problem (1) by numerically solving the ODE characterizing solutions to (2) up to $\varepsilon = 0$. In addition, it allows us to estimate convergence rates of the solutions as $\varepsilon \to 0$, in the spirit of the results in [1] and [10], but for a generic cost function.

Overview of the paper The paper is organized as follows: Section 2 presents theoretical results associated with the problem statement, particularly highlighting the uniform boundedness of the Hessian matrix. Next, Section 3 establishes that solutions can be characterized as solutions to an appropriate ODE. In Section 4, we provide an extensive collection of computational examples for one-dimensional, two-dimensional, and three-dimensional examples, incorporating various cost functions. Additionally, we compare the solution derived from the ODE approach with the traditional Newton's method, observing that the proposed algorithm is competitive for problems involving the squared Euclidean distance but demonstrates superior performance when dealing with different powers of the Euclidean cost function and in scenarios where the target points are not contained within the source measure.

2. Entropic optimal transport and the governing ODE

Consider X a compact convex subset of $\mathbb{R}^n$ and and a finite set $Y = \{y_1, y_2, \dots, y_N\} \subset \mathbb{R}^n$ of N points. Take two probability measures $\rho \in \mathcal{P}(X)$ and $\mu \in \mathcal{P}(Y)$ satisfying the following hypothesis

(H1) $\rho(x)$ is absolutely continuous with respect to the Lebesgue measure and bounded from above and below, that is $\exists\ \overline{m}, \underline{m} > 0$ such that $0 < \underline{m} \leq \rho(x) \leq \overline{m} < \infty$;

(H2) $\mu = \sum_{k=1}^{N} \mu_k \delta_{y_k}$ is discrete probability measure on Y, bounded from below by a positive constant, $\mu_k \geq \underline{\mu} > 0$, $\forall k$.

Then, the entropic (semi-discrete) optimal transport (OT) can be formulated as follows:

$$\max_{\gamma \in \Pi(\rho,\mu)} \int_{X\times Y} b(x,y)\mathrm{d}\gamma - \varepsilon \mathrm{Ent}(\gamma \mid \rho \otimes \sigma)\ , \tag{4}$$

where $\varepsilon \in [0, \infty)$ is a regularization parameter, $\Pi(\rho, \mu)$ is the set of probability measures on $X \times Y$ having ρ and μ as marginals, σ is the counting measure on Y, $b(x, y)$ is a cost function and $\mathrm{Ent}(\cdot|\rho \otimes \sigma)$ is the Boltzmann-Shannon relative entropy (or Kullback-Leibler divergence) w.r.t. the product measure $\rho \otimes \sigma$, defined for general probability measures p, q as

$$\mathrm{Ent}(p \,|\, q) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \eta \log(\eta)\, \mathrm{d}q & \text{if } p = \eta q, \\ +\infty & \text{otherwise.} \end{cases}$$

The fact that q is a probability measure ensures that $\mathrm{Ent}(p\,|\,q) \geq 0$. It is easy to show, see for instance [4, 16], that (4) admits the following dual formulation :

$$\min_{\mathbf{v}\in\mathbb{R}^N} \int_X \varepsilon \log\left(\sum_{k=1}^N e^{\frac{b(x,y_k)-v_k}{\varepsilon}}\right)\mathrm{d}\rho + \sum_{k=1}^N v_k\mu_k + \varepsilon \ , \ \varepsilon > 0 \ . \tag{5}$$

Remark 2.1. Notice that the term $\varepsilon \log\left(\sum_{k=1}^N e^{\frac{b(x,y_k)-v_k}{\varepsilon}}\right)$ in (5) is the so called b-soft-transform; roughly speaking this term converges to the usual b-transform of OT theory $v^b(x) := \sup_y b(x,y) - v(y)$ as $\varepsilon \to 0$.

Before going into details let us reformulate problems (4) (5) in a more convenient way: consider the following change of variables: $t = 1-\varepsilon$ with $t \in [0,1]$ and $b(x,y_i) = -tc(x,y_i)$. Then the primal problem (4) can be re-written as

$$\boxed{\min_{\gamma\in\Pi(\rho,\mu)} t\int_{X\times Y} c(x,y)\mathrm{d}\gamma + (1-t)\mathrm{Ent}(\gamma \mid \rho\otimes\sigma),\ t\in[0,1].} \tag{6}$$

Notice that (6) can be now seen as an interpolation between the case where the entropy is dominant, at $t = 0$, for which the solution is explicit, $\gamma = \rho\otimes\sigma$, and the original unregularized semi-discrete optimal transport problem, at $t = 1$. It is straightforward to then derive the new dual problem which takes the form:

$$\boxed{\min_{\psi\in\mathbb{R}^N} \Phi(\psi,t) := \int_X (1-t)\log\left[\sum_{k=1}^N e^{\frac{\psi_k - tc(x,y_k)}{1-t}}\right]\mathrm{d}\rho(x) - \sum_{k=1}^N \psi_k\mu_k - (1-t)\ .} \tag{7}$$

From now we will focus on the regularized problems (6), and (7) and their unregularized counterpart

$$\min_{\gamma\in\Pi(\rho,\mu)} \int_X c(x,y)\mathrm{d}\gamma(x,y), \quad \min_{\psi\in\mathbb{R}^N} -\sum_{k=1}^N \int_{\mathrm{Lag}_i(\psi)} (c(x,y)-\psi_i)\mathrm{d}\rho - \sum_{k=1}^N \psi_k\mu_k,$$

where $\mathrm{Lag}_i(\psi)$ is the $i-$th Laguerre cell defined below.

Definition 2.2 (Laguerre cell)**.** Given a vector $\psi \in \mathbb{R}^N$, the Laguerre cell corresponding to a target point y_i is:

$$\mathrm{Lag}_i(\psi) := \{x \in X \mid c(x,y_i) - \psi_i \leq c(x,y_k) - \psi_k,\ \forall k \neq i\}$$

Furthermore, the measure of the Laguerre cell with respect to the density $\rho(x)$ supported on X is given by $\rho(\mathrm{Lag}_i(\psi)) := \int_{\mathrm{Lag}_i(\psi)} \mathrm{d}\rho(x)$.

Definition 2.3 (Smoothed Laguerre cell)**.** Given a vector $\psi \in \mathbb{R}^N$, we define the entropic counterpart of the Laguerre cell above, that is the smoothed Laguerre cell corresponding

to a target point y_i:

$$\mathrm{RLag}_i^t(\psi) := \frac{e^{\frac{\psi_i - tc(x,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\psi_k - tc(x,y_k)}{1-t}}}$$

2.1. Preliminary results on semi-discrete entropic optimal transport and convexity of the objective function

Before introducing the governing ODE in order to characterise (7), let us state some preliminary results concerning the entropic Kantorovich functional $\Phi(\psi, t)$. We will in particular focus on some convexity properties of the entropic Kantorovich functional in two different regimes: (1) the entropy term is dominant, that is the case when $t \in [0, t^*]$; (2) the optimal transport term becomes stronger and the entropic solution $\psi(t)$ is close to the solution $\psi(1)$ of the unregularized problem $t \in [t^*, 1]$. Notice that the existence of such a t^* comes from the fact that we can easily show that $\psi(t)$ converges to $\psi(1)$ and so there exists a t^* such that for all $t \geq t^*$ we have $\|\psi(t) - \psi(1)\| \leq \delta$ for some $\delta > 0$, which will be selected later on to ensure appropriate bounds on the Hessian of Φ.

Let us start by stating some general results and then see how we can improve them in the case where $t \in [t^*, 1]$.

It is quite easy to show that a solution $\psi(t)$ to (7) is actually bounded (notice that X is compact and c is a continuous cost function, for more details see [15][Chapter 3]) by $2\|c\|_\infty$ such that we can check strong convexity of the functional Φ on the set $U := \{\psi \in \mathbb{R}^N \mid \|\psi\| \leq 2\|c\|_\infty,\ \psi \perp \mathbf{1}\}$. In particular we obtain the following results

Lemma 2.4. *Let $\psi \in \mathbb{R}^N$ be a vector such that $\sum_{i=1}^N \psi_i = 0$. Also, let $\widehat{\mu} \in \mathbb{R}^N$ be a discrete probability vector with a lower bound $\underline{\widehat{\mu}}$, $\widehat{\mu}_i \geq \underline{\widehat{\mu}} > 0$, $\forall i$. Denote by $\mathrm{Var}_{\widehat{\mu}}(\psi) :=$ the variance of ψ with respect to $\widehat{\mu}$. Then, it follows that $\mathrm{Var}_{\widehat{\mu}}(\psi) \geq \underline{\widehat{\mu}}||\psi||_2^2$.*

Proof. Let $\bar{\psi}$ denote the expectation of ψ with respect to the density $\widehat{\mu}$, $\bar{\psi} = \sum_{i=1}^N \widehat{\mu}_i \psi_i$.

$$\mathrm{Var}_{\widehat{\mu}}(\psi) = \sum_{i=1}^N \widehat{\mu}_i(\psi_i - \bar{\psi})^2 \geq \sum_{i=1}^N \underline{\widehat{\mu}}(\psi_i - \bar{\psi})^2 = \underline{\widehat{\mu}} \sum_{i=1}^N (\psi_i - \bar{\psi})^2 \geq \underline{\widehat{\mu}} \sum_{i=1}^N \psi_i^2 = \underline{\widehat{\mu}}||\psi||_2^2 \ ,$$

where

$$\sum_{i=1}^N (\psi_i - \bar{\psi})^2 = \sum_{i=1}^N (\psi_i^2 + \bar{\psi}^2 - 2\bar{\psi}\psi_i) = \sum_{i=1}^N \psi_i^2 + N\bar{\psi}^2 - 2\bar{\psi}\sum_{i=1}^N \psi_i = \sum_{i=1}^N \psi_i^2 + N\bar{\psi}^2 \geq \sum_{i=1}^N \psi_i^2 \ .$$

□

Lemma 2.5. *Let ρ a probability measure on the compact set X satisfying (H1). If $\widehat{\psi} \in U$,*

then $\exists \widehat{\underline{\mu}} > 0$*, such that* $\widehat{\mu}_i \geq \widehat{\underline{\mu}}$*,* $\forall i$*, where*

$$\widehat{\mu}_i = \int_X \frac{e^{\frac{\widehat{\psi}_i - tc(c,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(c,y_k)}{1-t}}} \mathrm{d}\rho(x) \ , t \in [0,1) \ .$$

Proof. By using the bounds on $\widehat{\psi}$, c and ρ we get

$$\widehat{\mu}_i \geq \frac{\underline{m}}{e^{2\frac{1+t}{1-t}\|c\|_\infty}}$$

□

Remark 2.6. Notice that

$$\{\nabla\Phi(\widehat{\psi},t)\}_i = \int_X \frac{e^{\frac{\widehat{\psi}_i - tc(c,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(c,y_k)}{1-t}}} \mathrm{d}\rho(x) - \widehat{\mu}_i \ , t \in [0,1) \ .$$

Theorem 2.7 (Strong convexity of Φ)**.** *Let* $\rho \in \mathcal{P}(X)$ *and* $\mu \in \mathcal{P}(Y)$ *satisfying hypothesis (H1) and (H2), respectively. If* $\widehat{\psi} \in U$ *then there exists* $\widehat{C} = \widehat{C}(t) > 0$ *such that* $\nabla^2_{\psi,\psi}\Phi(\widehat{\psi},t) \succeq \widehat{C}\mathrm{Id_N} > 0$ *for* $t \in [0,1)$*, where*

$$\nabla^2_{\psi,\psi}\Phi(\widehat{\psi},t) = \frac{1}{1-t}\big(\mathrm{diag}(\nabla\Phi(\widehat{\psi},t) - \nabla\Phi(\widehat{\psi},t) \otimes \nabla\Phi(\widehat{\psi},t)\big),$$

and

$$\{\nabla\Phi(\widehat{\psi},t)\}_i = \int_X \frac{e^{\frac{\widehat{\psi}_i - tc(x,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(x,y_k)}{1-t}}} \mathrm{d}\rho(x) - \widehat{\mu}_i \ .$$

Proof. Take a $\widehat{\psi} \in U$ we easily get that:

$$\langle v, \nabla^2\Phi(\widehat{\psi},t)v\rangle \geq \frac{1}{\tilde{C}}\mathrm{Var}_{\widehat{\mu}}(v) \ \forall v \in \mathbb{R}^N,$$

where $\tilde{C} = e^{2||c||_\infty \mathrm{diam}(X)}\frac{\overline{m}}{\underline{m}} + (1-t) > 0$, $\forall t \in [0,1)$ and $\widehat{\mu}$ defined as in Lemma 2.5. Furthermore, using Lemma 2.4, we obtain that

$$\mathrm{Var}_{\widehat{\mu}}(v) \geq \widehat{\underline{\mu}}||v||_2^2 \implies \langle v, \nabla^2\Phi(\widehat{\psi},t)v\rangle \geq \frac{1}{\tilde{C}}\widehat{\underline{\mu}}||v||_2^2 \ .$$

Finally, using Lemma 2.5, one derives the existence and bound on $\widehat{\underline{\mu}}$ and thus the lower bound for the smallest eigenvalue of $\nabla^2\Phi$:

$$\lambda_{\min}\{\nabla^2\Phi(\widehat{\psi},t)\} \geq \frac{1}{C} > 0, \ \forall t \in [0,1) \ .$$

As a result, we get that $C = \frac{m}{\tilde{C}e^{2\frac{1+t}{1-t}\|c\|_\infty}}$, and the Hessian matrix is positive semi-definite with simple zero eigenvalue and corresponding eigenvector $\mathbf{1}$. □

Notice that the result above provides strong convexity only on the interval $[0, 1)$ since the lower bound degenerates as $t \to 1$. In order to get strong convexity for all $t \in [0, 1]$ we should better study the case in which the entropic solution is close to the solution of the unregularized problem, namely we focus on the case in which $t \in [t^*, 1]$.

2.2. Strong convexity on $[t^*, 1]$

Firstly, we state some results when we consider a set of vectors ψ which are small perturbations of a vector ψ^0 such that $\mu_i = \rho(\mathrm{Lag}_i(\psi^0))$. The main results here is again the strong convexity of Φ but in the case where the lower bound for the smallest eigenvalue of $\nabla^2_{\psi,\psi}\Phi(\widehat{\psi}, t)$ does not depend on the regularization parameter.

Lemma 2.8. *Let $\mu \in \mathbb{R}^N_+$ be a discrete probability vector with a lower bound $\underline{\mu}$, $\mu_i \geq \underline{\mu} > 0$, $\forall i$. In addition, let $\psi^0 \in \mathbb{R}^N$ be a vector such that $\mu_i = \rho(Lag_i(\psi^0))$, $\forall i$. If $\widehat{\psi}$ is a small perturbation of ψ^0, $||\widehat{\psi} - \psi^0||_2^2 \leq \delta$, then $\exists \underline{\widehat{\mu}} > 0$, such that $\widehat{\mu}_i \geq \underline{\widehat{\mu}}$, $\forall i$, where*

$$\widehat{\mu}_i = \int_X \frac{e^{\frac{\widehat{\psi}_i - tc(x,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(x,y_k)}{1-t}}} \mathrm{d}\rho(x), t \in [0, 1] \ .$$

Proof. Given that $\widehat{\psi}$ is a small perturbation of ψ^0, it follows that:

$$\rho(\mathrm{Lag}_i(\widehat{\psi})) \geq \frac{1}{2}\rho(\mathrm{Lag}_i(\psi^0)) \geq \frac{1}{2}\underline{\mu} \ .$$

Next, using that $\mathrm{Lag}_i(\widehat{\psi}_i) \subset X$, $\forall i$, and $X = \cup_{i=1}^N \mathrm{Lag}_i(\widehat{\psi}_i)$, we get that:

$$\widehat{\mu}_i = \int_X \frac{e^{\frac{\widehat{\psi}_i - tc(x,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(x,y_k)}{1-t}}} \mathrm{d}\rho(x) \geq \int_{\mathrm{Lag}_i(\widehat{\psi})} \frac{e^{\frac{\widehat{\psi}_i - tc(x,y_i)}{1-t}}}{\sum_{k=1}^N e^{\frac{\widehat{\psi}_k - tc(x,y_k)}{1-t}}} \mathrm{d}\rho(x) \ .$$

From the definition of Laguerre cell it follows:

$$x \in \mathrm{Lag}_i(\widehat{\psi}) \iff c(x, y_i) - \widehat{\psi}_i \leq c(x, y_k) - \widehat{\psi_k} \iff c(x, y_i) - c(x, y_k) + \widehat{\psi_k} - \widehat{\psi}_i \leq 0 \ , \ \forall k \neq i.$$

Thus, we get:

$$\begin{aligned}\widehat{\mu}_i &\geq \int_{\mathrm{Lag}_i(\widehat{\psi})} \frac{1}{\sum_{k=1}^N e^{\frac{tc(x,y_i)-tc(x,y_k)+\widehat{\psi}_k-\widehat{\psi}_i}{1-t}}} \mathrm{d}\rho(x) \\ &\geq \int_{\mathrm{Lag}_i(\widehat{\psi})} \frac{1}{\sum_{k=1}^N e^{c(x,y_k)-c(x,y_i)}} \mathrm{d}\rho(x) \\ &\geq \int_{\mathrm{Lag}_i(\widehat{\psi})} \frac{1}{\sum_{k=1}^N e^{2\|c\|_\infty}} \mathrm{d}\rho(x) = \frac{1}{Ne^{2\|c\|_\infty}} \rho(\mathrm{Lag}_i(\widehat{\psi})) \geq \frac{1}{2Ne^{2\|c\|_\infty}} \underline{\mu} \ .\end{aligned}$$

Finally, it follows that $\widehat{\mu}_i \geq \underline{\widehat{\mu}}$, $\forall i$, where $\underline{\widehat{\mu}} = \frac{1}{2Ne^{2\|c\|_\infty}} \underline{\mu} > 0$. □

Theorem 2.9. *Let $\rho \in \mathcal{P}(X)$ and $\mu \in \mathcal{P}(Y)$ satisfying hypothesis (H1) and (H2), respectively. In addition, let $\psi^0 \in \mathbb{R}^N$ be a vector such that $\mu_i = \rho(Lag_i(\psi^0))$, $\forall i$. Then, there exists a constant $\widehat{C}$, which is independent of t, such that if $\widehat{\psi}$ is a small perturbation of ψ^0, $||\widehat{\psi} - \psi^0||_2^2 \leq \delta$, and $\widehat{\psi} \perp \mathbf{1}$, then $\lambda_{\min}\{\nabla^2_{\psi,\psi}\Phi(\psi,t)\} \geq \widehat{C} > 0$.*

Proof. Take a $\widehat{\psi}$ satisfying the hypothesis of the theorem, similarly to Theorem 3.2 of [10], we easily get that:

$$\langle v, \nabla^2_{\psi,\psi}\Phi(\widehat{\psi},t)v\rangle \geq \frac{1}{C}\mathrm{Var}_{\widehat{\mu}}(v) \ \forall v \in \mathbb{R}^N,$$

where $C = e^{L_c \mathrm{diam}(X)}\frac{\overline{m}}{\underline{m}} + 1 > 0$, L_c the Lipschitz constant of the cost and $\widehat{\mu}$ defined as in Lemma 2.8. Furthermore, using Lemma 2.4, we obtain that

$$\mathrm{Var}_{\widehat{\mu}}(v) \geq \underline{\widehat{\mu}}||v||_2^2 \implies \langle v, \nabla^2_{\psi,\psi}\Phi(\widehat{\psi},t)v\rangle \geq \frac{1}{C}\underline{\widehat{\mu}}||v||_2^2 \ .$$

Finally, using Lemma 2.8, one derives the existence and bound on $\underline{\widehat{\mu}}$:

$$\underline{\widehat{\mu}} = \frac{1}{2N}\underline{\mu} \implies \lambda_{\min}\{\nabla^2_{\psi,\psi}\Phi(\widehat{\psi},t)\} \geq \frac{1}{C}\frac{1}{2Ne^{2\|c\|_\infty}}\underline{\mu} > 0 \ .$$

As a result, we get that $\widehat{C} = \frac{1}{2NCe^{2\|c\|_\infty}}\underline{\mu}$, and the Hessian matrix is positive semi-definite with simple zero eigenvalue and corresponding eigenvector $\mathbf{1}$. □

3. An ODE characterization of semi-discrete entropic optimal transport

Notice now that since the functional is convex any minimizer of (7) is equivalently a solution to the equation $\nabla\Phi(\psi(t),t) = 0$. Thanks to the regularity of Φ one can now differentiate with respect to t and obtaining the following governing ODE:

$$\nabla^2_{\psi,\psi}\Phi(\psi(t),t)\psi'(t) + \frac{\partial}{\partial t}\nabla_\psi\Phi(\psi(t),t) = 0, \ t \in [0,1] \ , \tag{8}$$

where the gradient vector of $\Phi(\psi, t)$ from (7) is

$$\frac{\partial \Phi(\psi,t)}{\partial \psi_j} := \mu_j - \int_X \frac{e^{\frac{\psi_j - tc(x,y_j)}{1-t}}}{\sum_{k=1}^N e^{\frac{\psi_k - tc(x,y_k)}{1-t}}} \mathrm{d}\rho(x) \ , \ j = 1, 2, \ldots, N \ . \tag{9}$$

Using direct calculation, we get the derivative vector:

$$\frac{\partial}{\partial t}\left[\frac{\partial \Phi(\psi,t)}{\partial \psi_j}\right] = \int_X \frac{\sum_{k=1,k\neq j}^N A_k(x,1) e^{\frac{A_k(x,t)]}{1-t}}}{\left[(1-t)\left(1 + \sum_{k=1,k\neq j}^N e^{\frac{A_k(x,t)}{1-t}}\right)\right]^2} \mathrm{d}\rho(x) \ , \ j = 1, 2, \ldots, N \ , \tag{10}$$

where $A_k(x,t) = \psi_k - \psi_j + tc(x,y_j) - tc(x,y_k)$. Similarly, one can compute the Hessian matrix:

$$\nabla^2_{\psi\psi} \Phi(\psi(t), t) = \frac{1}{1-t} \mathbb{E}_\rho[\pi(\psi)\pi(\psi)^T - \mathrm{diag}(\pi(\psi))] \ ,$$

where $\mathbb{E}_\rho[f(x)] := \int_X f(x) \mathrm{d}\rho(x)$ and

$$\pi(\psi)_j = \frac{e^{\frac{\psi_j - tc(x,y_j)}{1-t}}}{\sum_{k=1}^N e^{\frac{\psi_k - tc(x,y_k)}{1-t}}} \ , \ j = 1, 2, \ldots, N \ ,$$

Hence we get the following Cauchy problem

Theorem 3.1. *Let $\psi(t)$ be a solution to (7) for $t \in [0,1)$. Then the trajectory $t \mapsto \psi(t)$ is smooth and is characterized on $t \in [0,1)$ as the unique solution to the Cauchy problem*

$$\boxed{\begin{cases} \nabla^2_{\psi,\psi} \Phi(\psi(t),t)\psi'(t) + \frac{\partial}{\partial t} \nabla \Phi(\psi(t),t) = 0, \ t \in [0,1), \\ \psi(0) = \log \mu - \frac{1}{N} \sum_{k=1}^N \log \mu_k. \end{cases}} \tag{11}$$

Remark 3.2. The theorem implies that one can solve the Cauchy problem (11) uniquely to obtain $\psi(t)$ for $t \in [0,1)$. As it is well known that $\lim_{t\to 1} \psi(t) = \psi(1)$, one can then take the limit to obtain $\psi(1)$ and, consequently, one can recover the solution to the unregularized problem from the ODE. In practice, one might worry that numerical instabilities could potentially arise from degeneracies of the Hessian; Theorem 2.9 ensures that this is not the case.

Under stonger conditions, one can actually go slightly further and show that the ODE is satisifed up to $t = 1$.

Proof. For any fixed $\bar{t} < 1$, we prove existence and uniqueness of a solution to the Cauchy problem on $[0, \bar{t}]$ by applying the Cauchy Lipschitz theorem on $[0, \bar{t}] \times U$, where $U := \{\psi \in \mathbb{R}^N \mid \|\psi\| \leq 2\|c\|_\infty, \ \psi \perp \mathbf{1}\}$. Since Φ is smooth, it has bounded derivative on this set, and, as noted above, it is known that the optimal ψ remains in U. Since the Hessian $\nabla^2_{\psi,\psi} \Phi(\psi, t)$ is uniformly positive definite on $[0, \bar{t}] \times U$ by Theorem 2.7, one can rewrite

the ODE as:

$$\psi'(t) = F(\psi(t), t) := -\big[\nabla^2_{\psi,\psi}\Phi(\psi(t), t)\big]^{-1}\big(\frac{\partial}{\partial t}\nabla\Phi(\psi(t), t)\big),$$

where the function F is Lipschitz on $[0, \bar{t}] \times U$. The Cauchy-Lipschitz theorem then yields well-posedness on $[0, \bar{t}]$; as $\bar{t} < 1$ is arbitrary, we get existence and uniqueness on $[0, 1)$. □

To ensure that $\psi(t)$ solves the ODE in (11) up to $t = 1$, we require stronger conditions on c, X, Y and ρ, which we define in the following. Before doing this we need three more definitions

Definition 3.3 (Twisted cost)**.** We say that c is *twisted* if for each $x \in X$ the mapping $y \mapsto \nabla_x c(x, y)$ is injective on Y.

Definition 3.4 (Y generic with respect to c)**.** We say that Y is *generic with respect to c* if for all distinct, $y_0, y_1, y_2 \in Y$ the intersection of any level sets of $x \mapsto c(x, y_0) - c(x, y_1)$ and $x \mapsto c(x, y_0) - c(x, y_2)$ has $(n-1)$ -dimensional Hausdorff measure 0.

Definition 3.5 (Y generic with respect to ∂X)**.** We say Y is *generic with respect to* ∂X if for any $y_0, y_1 \in Y$, the intersection of any level set of $x \mapsto c(x, y_0) - c(x, y_1)$ with ∂X has $(n-1)$ -dimensional Hausdorff measure 0.

(H3) The measure ρ satisfies the assumption (H1) with an $\alpha-$Hölder continuous density.

(H4) The cost function is $\mathcal{C}^2(X \times Y)$ and twisted.

Proposition 3.6. *Let c, ρ, ν verify (H1)-(H4), Y be generic with respect to both c and ∂X, and c be twisted. Then $\frac{\partial}{\partial t}\nabla\Phi(\psi, t) \to 0$ as $t \to 1$ with ψ solution to* (11).

The proof of the proposition is very similar to the one of [10][Proposition 5.1] the main difference consists in partitioning the space X, to estimate the integral in (10), by looking at what happens on the tangent space $x \mapsto \nabla_x c(x, y_j)$ for every $x \in X$ and for each component j of the derivative. For sake of completeness we have detailed the proof in this case in the Appendix A.

Proposition 3.7. *If c is twisted, Y is generic with respect to both c and ∂X and assumptions (H1)-(H4) are satsifed, the ODE in* (11) *is satisfied on* $[0, 1]$.

Proof. We need only to show that the ODE is satisfied at $t = 1$. It is well known that $\psi(t) \to \psi(1)$, the solution to the unregularized problem. Furthermore, the conditions ensure $\mathcal{C}^2$ smoothness of the Hessian $\nabla^2_{\psi\psi}\Phi(\psi, 1)$ of the unregularized problem (Theorem 47 in [23], as well as the convergence of $\nabla^2_{\psi\psi}\Phi(\psi, t)$ to it (Theorem 54 in [23]). Since the mixed derivative $\frac{\partial}{\partial t}\nabla\Phi(\psi, t) \to 0$, we can take the limit of the ODE in (11) as $t \to 1$ to obtain that the ODE is satisifed at $t = 1$. □

Remark 3.8 (Rate of convergence)**.** Notice that the results above imply that one can obtain the rate of convergence of the dual potentials, the primal solution and the entropic cost as in [10, 1], for a generic cost function. In particular, since $\psi'(1) = 0$ by Propositions 3.6 and 3.7, we have $\|\psi(1-t) - \psi(1)\| = o(1-t)$.

Remark 3.9 (Mass of the Laguerre cells)**.** In [21] the authors proved strong convexity of the hessian of the unregularized Kantorovich functional on the set $\mathcal{S}_c := \{\psi \mid \rho(\mathrm{Lag}_i(\psi)) \geq c > 0, \forall i = 1, ..., N\}$. It would have been natural then to prove the well-posedness of the ODE on the same set for some constant c. In order to do this one must assure that then for every t the vector $\psi(t)$, solution to entropic problem, belongs actually to $\mathcal{S}_c$. Unfortunately, in Figure 1 we notice that the mass of the Laguerre cells along the curve $t \mapsto \psi(t)$ can be equal to zero.

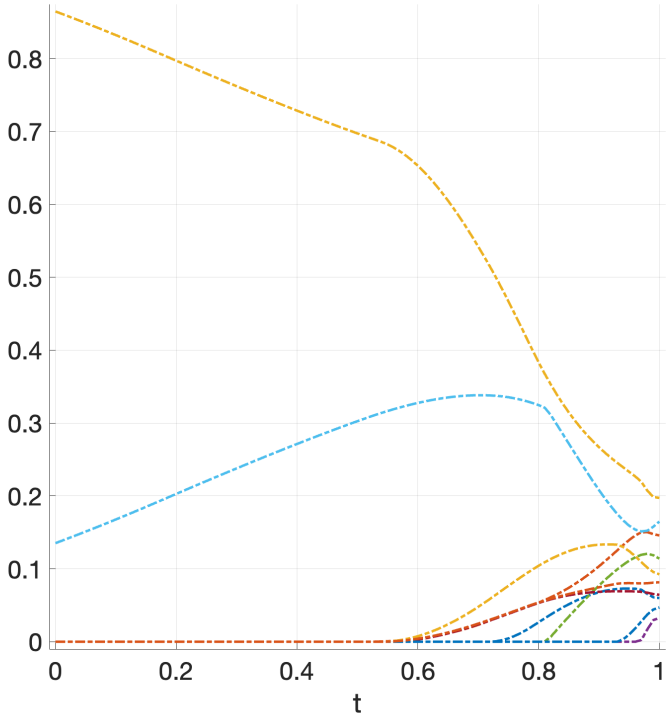


Figure 1: Mass of Laguerre cells in 2-d as a function of t

4. Computational Examples

The initial value problem (IVP) to solve is:

$$\nabla^2_{\mathbf{vv}}\Phi(\mathbf{v}(t), t)\mathbf{v}'(t) + \frac{\partial}{\partial t}\nabla\Phi(\mathbf{v}(t), t) = 0, \ \ t \in [0, 1] \ , \tag{12}$$

where $\sum_{k=1}^{N} v_k(0) = 0$. Observe that the value of the initial condition can be directly determined from (9) by setting $t = 0$ and applying the constraint $\sum_{k=1}^{N} v_k(0) = 0$. Subsequently, the initial value problem (IVP) (12) can be reformulated as follows:

$$\mathbf{v}'(t) = -[\nabla^2_{\mathbf{vv}}\Phi(\mathbf{v}(t), t)]^{\dagger}\frac{\partial}{\partial t}\nabla\Phi(\mathbf{v}(t), t),\ \mathbf{v}(0) = \log\mu - \frac{1}{N}\sum_{k=1}^{N}\log\mu,\ t \in [0, 1], \tag{13}$$

where $[\nabla^2_{\mathbf{vv}}\Phi]^{\dagger}$ denotes the pseudo-inverse of $\nabla^2_{\mathbf{vv}}\Phi$ taken in the orthogonal complement of $\ker(\nabla^2_{\mathbf{vv}}\Phi) = \mathbf{1}$. For solving (13) in all examples presented below, we will use the Runge-Kutta method from Remark 4.1 with parameters $\alpha = \frac{1}{8}$ and $\beta = \frac{1}{4}$. Furthermore, we will provide the value of the error in $\mathbf{v}$: Error $= ||\mathbf{v}(1) - \mathbf{v}_{exact}||_{\infty}$, where $\mathbf{v}_{exact}$ is the solution of the unregularized semi-discrete optimal transport problem, either determined exactly in one-dimensional examples or known a-priori for two-dimensional cases.

Remark 4.1 (Third-Order Runge-Kutta Method)**.** A family of the third-order Runge-Kutta methods with α and β parameters:

$$k_1 = f(t_n, y_n)\ ,\ k_2 = f(t_n + c_2h, y_n + a_{21}hk_1)\ ,\ k_3 = f(t_n + c_3h, y_n + h(a_{31}k_1 + a_{32}k_2)$$
$$\Longrightarrow\ y_{n+1} = y_n + h\big[b_1k_1 + b_2k_2 + b_3k_3\big]\ ,$$

where $\alpha, \beta \neq 0$, $\alpha \neq \beta$, $\alpha \neq \frac{2}{3}$, and

$$a_{21} = \alpha\ ,\ a_{31} = \frac{\beta}{\alpha}\frac{\beta - 3\alpha(1-\alpha)}{3\alpha - 2}\ ,\ a_{32} = -\frac{\beta}{\alpha}\frac{\beta - \alpha}{3\alpha - 2}\ ,$$
$$b_1 = 1 - \frac{3\alpha + 3\beta - 2}{6\alpha\beta}\ ,\ b_2 = \frac{3\beta - 2}{6\alpha(\beta - \alpha)}\ ,\ b_3 = \frac{2 - 3\alpha}{6\beta(\beta - \alpha)}\ ,\ c_1 = \alpha\ ,\ c_2 = \beta\ .$$

4.1. Problems in 1-d

Consider the following problems on $X = [0, 1]$:

$$\mathrm{d}\rho(x) = \mathrm{d}x\ ,\ y = \{0.25,\ 0.5,\ 0.75\}\ ,\ \mu = \begin{pmatrix} 0.3 \\ 0.4 \\ 0.3 \end{pmatrix}\ . \tag{E1}$$

$$\mathrm{d}\rho(x) = 1.8305e^{-10(x-0.5)^2}\mathrm{d}x\ ,\ y = \{0.25,\ 0.5,\ 0.75\}\ ,\ \mu = \begin{pmatrix} 0.3 \\ 0.4 \\ 0.3 \end{pmatrix}\ . \tag{E2}$$

$$\mathrm{d}\rho(x) = \mathrm{d}x\ ,\ y = \{-3.4584,\ -2.3668,\ 0.3374,\ 2.4005\}\ ,\ \mu = \begin{pmatrix} 0.0078 \\ 0.4920 \\ 0.4823 \\ 0.0179 \end{pmatrix}\ . \tag{E3}$$

In addition to solving IVP (13), we will also solve the above problems using Newton's method for comparison reasons. The details of Newton's method are outlined in Remark 4.2.

Remark 4.2 (Newton's Method in 1-d)**.** To solve the semi-discrete OT problem in one dimension with $c(x,y) = ||x-y||_2^p$, $p \geq 2$, one can perform the following Newton's iterations until convergence:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} - [\nabla G(\mathbf{v}^{(k)})]^{\dagger} G(\mathbf{v}^{(k)}) \ ,$$

where $[\nabla G]^{\dagger}$ represents the inverse of ∇G taken in the orthogonal complement of $\ker(\nabla G) = \mathbf{1}$,

$$\{G(\mathbf{v})\}_i = \mu_i - \rho(\mathrm{Lag}_i(\mathbf{v})) = \mu_i - \int_{\mathrm{Lag}_i(\mathbf{v})} \rho(x)dx \ , \ i = 1, 2, \ldots, N \ ,$$

$$\{\nabla G(\mathbf{v})\}_{ij} = \int_{\mathrm{Lag}(v_i)\cap\mathrm{Lag}(v_j)} \frac{\rho(x)}{|\nabla_x c(x,y_i) - \nabla_x c(x,y_j)|} ds = \frac{\rho(x_{ij})}{|\nabla_x c(x_{ij},y_i) - \nabla_x c(x_{ij},y_j)|} \ , \ i \neq j \ ,$$

$$\{\nabla G(\mathbf{v})\}_{ij} = -\sum_{j=1, j\neq i}^{N} \{\nabla G(\mathbf{v}^{(k)})\}_{ij} = -\sum_{j=1, j\neq i}^{N} \frac{\rho(x_{ij})}{|\nabla_x c(x_{ij},y_i) - \nabla_x c(x_{ij},y_j)|} \ , \ i = 1, 2, \ldots, N \ ,$$

and $x_{ij} = \mathrm{Lag}_i(\mathbf{v}) \cup \mathrm{Lag}_j(\mathbf{v})$.

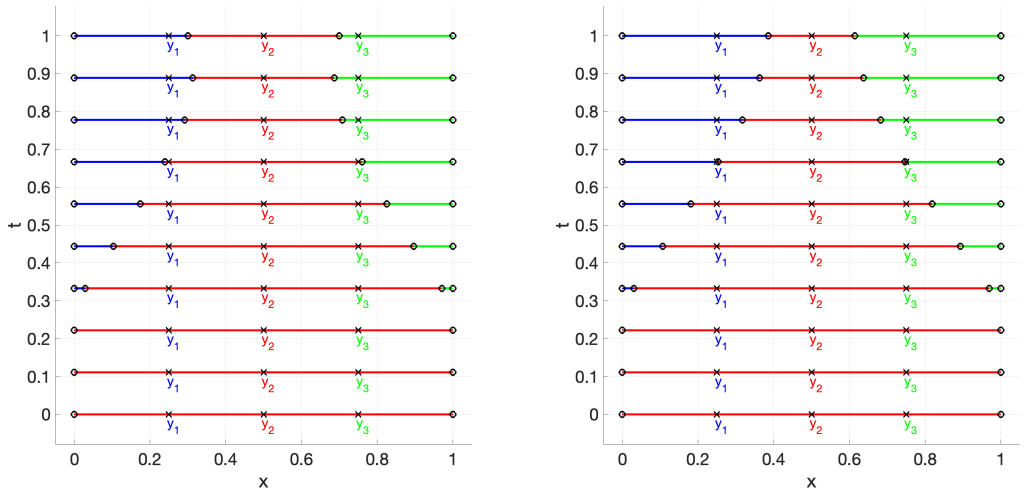
As demonstrated in Table 1, we discern a third-order convergence rate in t when $c(x,y) = ||x-y||_2^2$, as anticipated. Nevertheless, the algorithm encounters a specific mesh size where it fails, attributed to integration inaccuracies. As $t \to 1$, a high-precision integrator becomes necessary because the integrand found in the derivative term and the Hessian matrix tends to resemble a delta function. For all our experiments, we utilized the **MATLAB** integration functions: *integral* for 1-dimensional, *integral2* for 2-dimensional, and *integral3* for 3-dimensional problems. Additionally, Figure 2 illustrates the progression of Laguerre cells as we near the standard unregularized semi-discrete optimal transport problem, revealing a non-monotonic evolution.

$\Delta\mathbf{t}$	(E1)	(E2)	(E3)
10^{-1}	$1.3891 * 10^{-3}$	$3.1750 * 10^{-3}$	$1.3942 * 10^{-2}$
10^{-2}	$3.2996 * 10^{-7}$	$9.3938 * 10^{-6}$	$5.6056 * 10^{-4}$
10^{-3}	$3.9462 * 10^{-10}$	$1.1194 * 10^{-9}$	$3.1528 * 10^{-8}$
10^{-4}	$6.5607 * 10^{-13}$	$1.2871 * 10^{-12}$	NAN
10^{-5}	NAN	NAN	NAN

Table 1: Error for IVP (13) solution with $c(x,y) = ||x-y||_2^2$

Regarding Newton's method, comparison with our approach in Table 2 indicates similar error magnitudes in $\mathbf{v}$. However, Newton's method exhibits significant sensitivity to the initial conditions, leading to its failure in solving Example (E3) because the target points are outside of X, rendering the zero initial guess inadequate.

Furthermore, Table 3 presents the error convergence for $c(x,y) = ||x-y||_2^3$. Here, we observe third-order convergence once again, albeit less consistently than with the squared



(a) Example (E1) (b) Example (E2)

Figure 2: Time evolution of Laguerre cells

Initial Guess	(E1)	(E2)	(E3)
$10 * \text{rand}(N, 1)$	2.5647	2.1454	NAN
$1 * \text{rand}(N, 1)$	$3.5738 * 10^{-13}$	$1.7328 * 10^{-1}$	NAN
$0.1 * \text{rand}(N, 1)$	$5.2463 * 10^{-13}$	$4.0957 * 10^{-13}$	NAN
$0.01 * \text{rand}(N, 1)$	$7.5187 * 10^{-13}$	$1.8947 * 10^{-13}$	NAN
0	$2.7284 * 10^{-13}$	$1.8087 * 10^{-13}$	NAN

Table 2: Error for Newton's solution with $c(x, y) = ||x - y||_2^2$

Euclidean distance. In a parallel analysis in Table 4, we report the Newton's method performance under this cost. The impact of the initial guess is more pronounced; for instance, convergence to the solution in Example (E2) was only achieved when we started within 1% deviation of the zero initial guess, whereas with $c(x, y) = ||x - y||_2^2$, obtaining a solution was possible even with a 10% perturbation.

4.2. Problems in 2-d

Consider the following problems on $X = [0, 1] \times [0, 1]$:

$$\mathrm{d}\rho(x) = \mathrm{d}x_1\mathrm{d}x_2 \ , \ y = \left\{ \binom{0}{0}, \binom{0}{1}, \binom{1}{1} \right\} \ , \ \mu = \begin{pmatrix} \frac{1}{2}(1-b) \\ b \\ \frac{1}{2}(1-b) \end{pmatrix} \ , \ \mathbf{v}_{exact} = \begin{pmatrix} \frac{1}{3}(1-2\sqrt{b}) \\ -\frac{2}{3}(1-2\sqrt{b}) \\ \frac{1}{3}(1-2\sqrt{b}) \end{pmatrix} \ , \tag{E4}$$

$\Delta \mathbf{t}$	(E1)	(E2)	(E3)
10^{-1}	$7.8197 * 10^{-3}$	$7.4895 * 10^{-4}$	$7.6901 * 10^{-3}$
10^{-2}	$4.8828 * 10^{-4}$	$3.7807 * 10^{-4}$	$3.5379 * 10^{-6}$
10^{-3}	$1.7791 * 10^{-8}$	$2.3371 * 10^{-7}$	$3.5732 * 10^{-9}$
10^{-4}	$3.0193 * 10^{-11}$	$4.1936 * 10^{-11}$	NAN
10^{-5}	NAN	NAN	NAN

Table 3: Error for IVP (13) solution with $c(x, y) = ||x - y||_2^3$

Initial Guess	(E1)	(E2)	(E3)
$10 * \text{rand}(N, 1)$	2.7753	1.6201	NAN
$1 * \text{rand}(N, 1)$	$1.9808 * 10^{-1}$	$4.5294 * 10^{-1}$	NAN
$0.1 * \text{rand}(N, 1)$	$8.1341 * 10^{-13}$	$2.3304 * 10^{-1}$	NAN
$0.01 * \text{rand}(N, 1)$	$2.6465 * 10^{-13}$	$5.6829 * 10^{-13}$	NAN
0	$6.9658 * 10^{-13}$	$2.9445 * 10^{-13}$	NAN

Table 4: Error for Newton's solution with $c(x, y) = ||x - y||_2^3$

where $0 < b < 1$.

$$\mathrm{d}\rho(x) = \mathrm{d}x_1\mathrm{d}x_2 \ , \ y = \left\{ \binom{0}{0}, \binom{0}{1} \right\} \ , \ \mu = \binom{0.726759}{0.273241} \ , \ \mathbf{v}_{exact} = \begin{pmatrix} \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix} . \tag{E5}$$

$$\mathrm{d}\rho(x) = \mathrm{d}x_1\mathrm{d}x_2 \ , \ y = \left\{ \binom{0}{0}, \binom{0}{1} \right\} \ , \ \mu = \binom{0.872066}{0.127934} \ , \ \mathbf{v}_{exact} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} . \tag{E6}$$

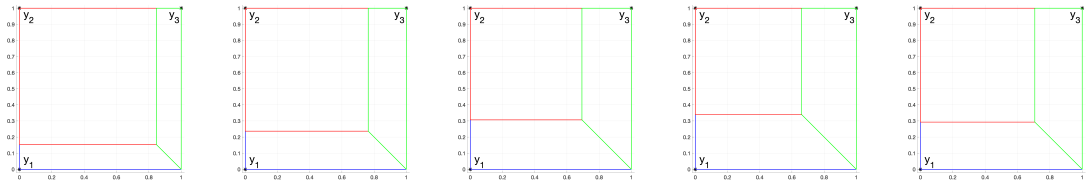
In the above examples, exact solutions are obtained for the cost function $c(x, y) = ||x - y||_2^2$ in Example (E4), and for the cost function $c(x, y) = ||x - y||_2^4$ in Examples (E5) and (E6). Analogously to the 1-d examples, we compare the solution of the IVP (13) with the results of Newton's method. When using the squared Euclidean distance as the cost, Laguerre cells can be efficiently determined using a lifting algorithm (refer to [22]). Subsequently, we calculate the Jacobian matrix using the Centered Finite Difference scheme. However, for $c(x, y) = ||x - y||_2^p$ with $p > 2$, there is no efficient computational algorithm. Consequently, we approximate the Laguerre cell on a fixed mesh and again compute the Jacobian matrix using the Centered Finite Difference scheme.

In Table 5, a third-order convergence is apparent in all examples except Example (E6). Additionally, we note that precise integration becomes increasingly critical as the method fails even with a mesh size of 10^{-3}, unlike the one-dimensional cases. Furthermore, Figure 3 illustrates the temporal progression of Laguerre partitions as $t \to 1$, exhibiting similar

non-monotonic behavior as seen in the one-dimensional scenario.

$\Delta \mathbf{t}$	(E4) with $b = 0.5$	(E4) with $b = 0.1$	(E5)	(E6)
10^{-1}	$9.5043 * 10^{-4}$ 0.41 sec.	$4.2964 * 10^{-4}$ 0.41 sec.	$2.8436 * 10^{-4}$ 0.72 sec.	$3.8182 * 10^{-4}$ 0.70 sec.
10^{-2}	$1.9683 * 10^{-8}$ 10.13 sec.	$2.1324 * 10^{-7}$ 12.46 sec.	$7.9821 * 10^{-8}$ 16.60 sec.	$1.3475 * 10^{-5}$ 14.88 sec.
10^{-3}	NAN	NAN	NAN	NAN

Table 5: Error for IVP (13) solution



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

Figure 3: Time Evolution of Laguerre cells in Example (E4) with $b = 0.5$

Next, as demonstrated in Table 6, the performance of Newton's method in terms of accuracy and time is comparable to our approach. Nonetheless, the importance of an initial guess becomes crucial in two-dimensional problems, as convergence to a solution in Example (E6) is not achieved. Overall, it is feasible to incorporate both methods by using the ODE solution as the initial guess for Newton's method.

Finally, to demonstrate robustness of ODE solver as the number of points increases. Figure 4 illustrates the convergence of the Laguerre cell measure to μ over t. In this problem, target points are randomly distributed within the unit square, both source and target measures are uniform, the cost function is the squared Euclidean distance, and the time step is $\Delta t = 10^{-2}$. The exact potential $\mathbf{v}_{\text{exact}}$ is unknown; however, we can evaluate the measure error $||\rho(\text{Lag}(\mathbf{v}(1))) - \mu||_\infty$. For instance, this error is 4.6587×10^{-4} for 10 random target points and 1.4668×10^{-3} for 25 target points. Additionally, Figures 5 and 6 illustrate the evolution of Laguerre cells with respect to t.

4.3. Problems in 3-d

This section delves into evaluating the effectiveness of the ODE solver in a three-dimensional context, comparing it to Newton's method, and examining the impact of dimensionality on computation time. Consider the following problem on $X = [0, 1] \times [0, 1] \times [0, 1]$ with

Initial Guess	(E4) with $b = 0.5$	(E4) with $b = 0.1$	(E5)	(E6)
$1 * \text{rand}(N, 1)$	$1.9160 * 10^{-8}$ 7.89 sec.	$2.1022 * 10^{-8}$ 7.36 sec.	$1.9524 * 10^{-5}$ 15.00 sec.	NAN
$0.1 * \text{rand}(N, 1)$	$1.3244 * 10^{-8}$ 6.97 sec.	$1.5153 * 10^{-8}$ 6.66 sec.	NAN	NAN
$0.01 * \text{rand}(N, 1)$	$7.5942 * 10^{-9}$ 7.10 sec.	$1.1396 * 10^{-8}$ 6.86 sec.	NAN	NAN
0	$5.7658 * 10^{-9}$ 6.70 sec.	$2.1945 * 10^{-8}$ 8.05 sec.	NAN	NAN

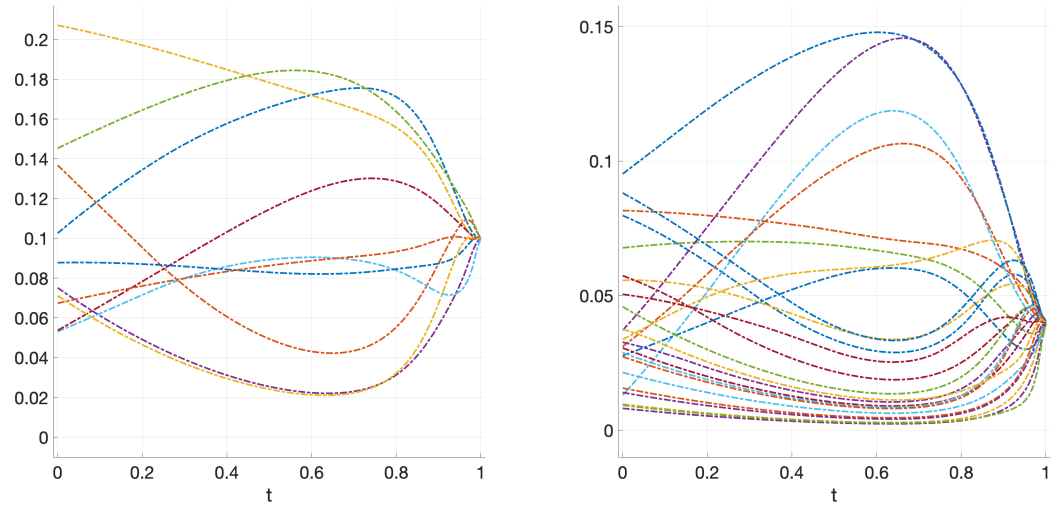
Table 6: Error for Newton's solution

$c(x, y) = ||x - y||_2^2$, and

$$\mathrm{d}\rho(x) = \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \ , \ y = \left\{ \begin{pmatrix} 0.5508 \\ 0.8963 \\ 0.0299 \end{pmatrix}, \begin{pmatrix} 0.7081 \\ 0.1256 \\ 0.4568 \end{pmatrix}, \begin{pmatrix} 0.2909 \\ 0.2072 \\ 0.6491 \end{pmatrix}, \begin{pmatrix} 0.5108 \\ 0.0515 \\ 0.2785 \end{pmatrix}, \begin{pmatrix} 0.8929 \\ 0.4408 \\ 0.6763 \end{pmatrix} \right\} \ , \ \mu = \frac{1}{5}\mathbf{1} \ . \tag{E7}$$

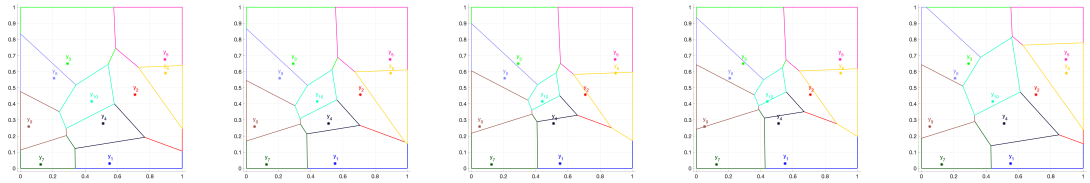
Tables 7 and 8 show the performance of the ODE method and Newton's method, respectively. For non-trivial 3-d problems, the exact solution $\mathbf{v}_{exact}$ is typically not available, rendering the calculation of Error $= ||\mathbf{v}(1) - \mathbf{v}_{exact}||_\infty$ infeasible. Nevertheless, it remains possible to compute the error in measure: Measure Error $= ||\rho(\text{Lag}(\mathbf{v}(1))) - \mu||_\infty$. As indicated in Table 7, there is first-order convergence towards μ, suggesting second-order convergence in $\mathbf{v}$ due to the integration. Moreover, Figure 7 illustrates the progression of Laguerre cells as t approaches 1.

Finally, Figure 8 provides a comparative analysis of computational time across different dimensions for the ODE solution. In this scenario, the *3-d Example* pertains to (E7), while the *1-d Example* and *2-d Example* relate to analogous problems on the unit line and unit square, respectively: five random target points, uniform target and source measures, squared Euclidean cost, and 100 time-steps are used. It is observed that calculation time escalates exponentially with increasing dimensions. Additionally, there's a rise in computation time as t nears the value of 1 in every dimension. This is attributed to the internal numerical integration process where the integrand approaches a delta function, necessitating finer spatial discretization as t converges to 1.



(a) 10 points (b) 25 points

Figure 4: Time evolution of measures



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

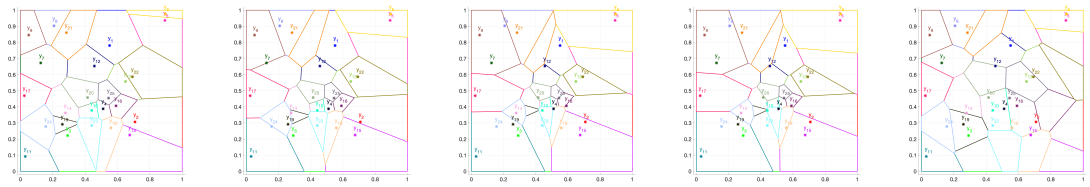
Figure 5: Time evolution of Laguerre cells with 10 random points

Acknowledgments

L.N. benefited from the support of the FMJH Program PGMO and from the ANR project GOTA (ANR-23-CE46-0001). L.N. thanks T. O. Galloüet, Q. Mérigot and P. Pegon for the fruitful discussions on semi-disrete OT and entropic regularization.
D.O. gratefully acknowledge that this research was supported in part by the Pacific Institute for the Mathematical Sciences.
B.P. is pleased to acknowledge the support of Natural Sciences and Engineering Research Council of Canada Discovery Grant number 04864-2024.

References

[1] Jason M Altschuler, Jonathan Niles-Weed, and Austin J Stromme. Asymptotics for semidiscrete entropic optimal transport. *SIAM Journal on Mathematical Analysis*,



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

Figure 6: Time evolution of Laguerre cells with 10 random points

Δt	Measure Error
10^{-1}	$6.4267 * 10^{-3}$ 70.151 sec.
10^{-2}	$1.3377 * 10^{-3}$ 3348.7 sec.
10^{-3}	NAN

Table 7: IVP (13) solution of Example (E7)

Initial Guess	Measure Error
$0.1 * \text{rand}(N, 1)$	$3.2875 * 10^{-5}$ 348.09 sec.
$0.01 * \text{rand}(N, 1)$	$1.6175 * 10^{-4}$ 238 sec.
0	$1.7587 * 10^{-4}$ 235.97 sec.

Table 8: Newton solution of Example (E7)

54(2):1718–1741, 2022.

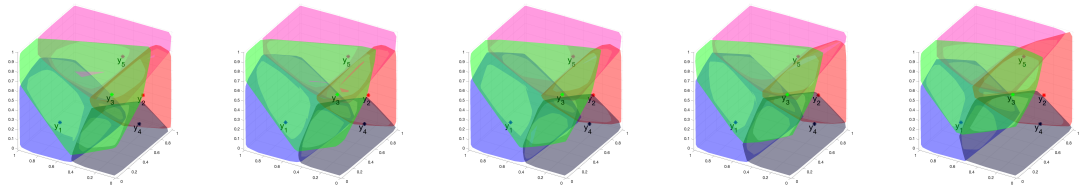
[2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[3] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[4] Bernard Bercu, Jérémie Bigot, Sébastien Gadat, and Emilia Siviero. A stochastic gauss–newton algorithm for regularized semi-discrete optimal transport. *Information and Inference: A Journal of the IMA*, 12(1):390–447, 2023.

[5] Jérémie Bigot, Elsa Cazelles, and Nicolas Papadakis. Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electronic Journal of Statistics*, 13(2):5120 – 5150, 2019.

[6] Yann Brenier. A homogenized model for vortex sheets. *Archive for Rational Mechanics and Analysis*, 138(4):319–353, 1997.



(a) $t = 0$ (b) $t = 0.25$ (c) $t = 0.5$ (d) $t = 0.75$ (e) $t = 1$

Figure 7: Time evolution of Laguerre cells in Example (E7)

[7] Pedro Machado Manhães De Castro, Quentin Mérigot, and Boris Thibert. Intersection of paraboloids and application to minkowski-type problems, 2014.

[8] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018.

[9] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pages 2292–2300. Curran Associates, Inc., 2013.

[10] Alex Delalande. Nearly tight convergence bounds for semi-discrete entropic optimal transport. In *International Conference on Artificial Intelligence and Statistics*, pages 1619–1642. PMLR, 2022.

[11] Luca Dieci and Daniyar Omarov. Solving semi-discrete optimal transport problems: star shapedeness and newton's method. *Numerical Algorithms*, pages 1–56, 2024.

[12] Qiang Du and Max Gunzburger. Grid generation and optimization based on centroidal voronoi tessellations. *Applied Mathematics and Computation*, 133(2):591–607, 2002.

[13] Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

[14] Alfred Galichon and Bernard Salanié. Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629, 2022.

[15] Aude Genevay. *Entropy-regularized optimal transport for machine learning*. PhD thesis, Université Paris sciences et lettres, 2019.

[16] Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. Stochastic optimization for large-scale optimal transport. *Advances in neural information processing systems*, 29, 2016.

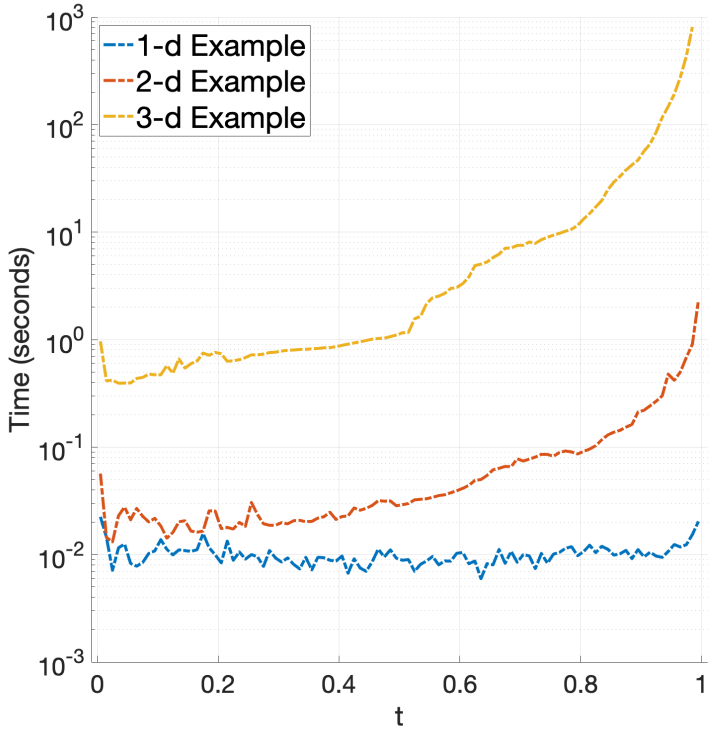


Figure 8: Comparison of time across dimensions and as $t \to 1$

[17] Joshua Zoen-Git Hiew, Luca Nenna, and Brendan Pass. An ordinary differential equation for entropic optimal transport and its linearly constrained variants. *arXiv preprint arXiv:2403.20238*, 2024.

[18] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[19] Leonid V. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[20] Leonid V. Kantorovich. On a problem of Monge. *Uspekhi Mat. Nauk*, 3:225–226, 1948.

[21] Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a Newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, April 2019.

[22] Lévy, Bruno. A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: M2AN*, 49(6):1693–1715, 2015.

[23] Quentin Mérigot and Boris Thibert. Optimal transport: discretization and algorithms. In *Geometric Partial Differential Equations - Part II*, volume 22 of *Handbook of Numerical Analysis*, pages 133–212. Elsevier, 2021.

[24] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. pages 666–704, 1781. In French.

[25] Luca Nenna and Brendan Pass. An ode characterisation of multi-marginal optimal transport with pairwise cost functions. *arXiv preprint arXiv:2212.12492*, 2022.

[26] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.

[27] Nicolas Papadakis. *Optimal transport for image processing.* PhD thesis, Université de Bordeaux, 2015.

[28] Aaditya Ramdas, Nicolás García Trillos, and Marco Cuturi. On wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2), 2017.

[29] Luis Caicedo Torres, Luiz Manella Pereira, and M Hadi Amini. A survey on optimal transport for machine learning: Theory and applications. *arXiv preprint arXiv:2106.01963*, 2021.

[30] Jonathan Weed and Quentin Berthet. Estimation of smooth densities in wasserstein distance. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 3118–3119. PMLR, 25–28 Jun 2019.

A. Proof of Proposition 3.6

Proof. The proof is based on the one of [10][Proposition 5.1] adapted to the case of a generic cost function. First of all let us re-write $\frac{\partial}{\partial t}\nabla\Phi(\psi,t)$ as follows

$$\frac{\partial}{\partial t}\nabla\Phi(\psi,t)=\int_X\sum_{j=1}^N\left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right)\pi_i(x)\pi_j(x)\mathrm{d}\rho(x),$$

where the quantity

$$\Delta_{i,j}(x,t)=\psi_i-tc(x,y_i)-(\psi_j-tc(x,y_j)),$$

can be interpreted as a duality gap and

$$\pi_i(x)=\frac{1}{\sum_{k=1}^N\exp\left(\frac{\Delta_{ki}(x,t)}{1-t}\right)}.$$

In order to estimate the integral over X we are going to make a similar partition as the one in [10]: we consider the set

$$X_{i,\eta,+} := \{x \in \mathrm{Lag}_i(\psi) \mid \forall j \neq i, \frac{\Delta_{ij}(x,1)}{||p_i^x - p_j^x||} \geq \eta\},$$

and

$$X_{i,\eta,-} := \{x \in X \mid \forall j \in \mathrm{argmax}_k(\psi_k - c(x,y_k)), \frac{\Delta_{ji}(x,1)}{||p_j^x - p_i^x||} \geq \eta\},$$

where $p_i^x := \nabla_x c(x, y_i)$. Notice that this two sets corresponds respectively to the points of $\mathrm{Lag}_i(\psi)$ and $X \setminus \mathrm{Lag}_i(\psi)$ which are far from the boundary of the Laguerre cell i. The idea behind the proof is to create then a tube around the common boundary between the Laguerre cells to have better estimates. Moreover, notice that thanks to the twisted assumption the quantity $||p_j^x - p_i^x||$ is never null. We the define for any $j \neq i$ the common boundary between $\mathrm{Lag}_i(\psi)$ and $\mathrm{Lag}_j(\psi)$ as $H_{ij} = \mathrm{Lag}_i(\psi) \cap \mathrm{Lag}_j(\psi)$ and for a parameter $\gamma > 0$ the set of points of H_{ij} that are at least a distance controlled by γ from the other Laguerre cells

$$H_{ij}^\gamma := \{x_0 \in H_{ij} \mid \forall k \neq i,j, \Delta_{ik}(x_0,1) = \Delta_{jk}(x_0,1) \geq \gamma \max(||p_i^{x_0} - p_k^{x_0}||, ||p_j^{x_0} - p_k^{x_0}||).\}$$

Denote then by

$$T_{i,\eta,\gamma} = \cup_{j \neq i}\{x_0 + s d_{ij}^{x_0}, x_0 \in H_{ij}^\gamma, s \in [-\eta||p_i^{x_0} - p_j^{x_0}||, \eta||p_i^{x_0} - p_j^{x_0}||]\}$$

with $d_{ij}^{x_0} = \frac{p_i^{x_0} - p_j^{x_0}}{||p_i^{x_0} - p_j^{x_0}||^2}$ the union of tubular sets around the common boundary without the corners and $C_{i,\eta,\gamma} = X \setminus (X_{i,\eta,+} \cup X_{i,\eta,-} \cup T_{i,\eta,\gamma})$ the neighbourhood of the corners of the Laguerre cell. Before going into the details notice that thanks to the assumptions on the cost and the set X we can easily control the term $p_i^x - p_j^x$ for all x and couple of indexes i, j. It is then quite easily to observe that on both $X_{i,\eta,+}$ and $X_{i,\eta,-}$ we get the following control on the integrand

$$\sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right)\pi_i(x)\pi_j(x) \lesssim \frac{e^{-\eta/(1-t)}}{(1-t)^2}.$$

We turn now our attention on the evaluation of the integral over $T_{i,\eta,\gamma}$. Notice first that for $x \in T_{i,\eta,\gamma}$ there exists and index j and $x_0 \in H_{ij}^\gamma$ such that $x = x_0 + s d_{ij}^{x_0}$ such that we get, by a Taylor expansion,

$$\begin{aligned}\Delta_{ij}(x,1) &= \Delta_{ij}(x_0,1) + s\langle \nabla_x \Delta(x_0,1), d_{ij}^{x_0}\rangle + \frac{s^2}{2}\langle \nabla_{xx}^2 \Delta(\xi,1) d_{ij}^{x_0}, d_{ij}^{x_0}\rangle \\ &= s + \frac{s^2}{2}\langle \nabla_{xx}^2 \Delta(\xi,1) d_{ij}^{x_0}, d_{ij}^{x_0}\rangle,\end{aligned}$$

where the quadratic term will not play an important role since we will take s depending by $1-t$ and for $t \to 1$ the first order term will be the dominant one. For any $k \neq i, j$ we have by definition of H_{ij}^{γ} and, again by a Taylor expansion,

$$\Delta_{ik}(x,t) \geq \gamma ||p_i^{x_0} - p_j^{x_0}|| - |s|C \geq \tilde{\gamma}$$

and in the same way we have $\Delta_{i,k}(x,t) \geq \tilde{\gamma}$. Now the integral on $T_{i,\eta,\gamma}$ can be written as follows

$$\sum_j \int_{x_0 \in H_{ij}^{\gamma}} \int_0^{\eta ||p_i^{x_0} - p_j^{x_0}||} (g_i(x_0 - s d_{ij}^{x_0}) + g_i(x_0 + s d_{ij}^{x_0})) \mathrm{d}t \mathrm{d}\mathcal{H}^{d-1}(x_0), \tag{E8}$$

where $g_i(x) = \sum_{j \neq i} \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x)\rho(x)$. Denote by $x_s = x_0 + s d_{ij}^{x_0}$ and $x_{-s} = x_0 - s d_{ij}^{x_0}$ then

$$\begin{aligned} g_i(x_s) &= \left(\frac{\Delta_{ij}(x_{-s},1)}{(1-t)^2})\right) \pi_i(x_s)\pi_j(x_s)\rho(x_s) + \sum_{k \neq i,j} \left(\frac{\Delta_{ik}(x_{-s},1)}{(1-t)^2})\right) \pi_i(x_s)\pi_k(x_s)\rho(x_s) \\ &\lesssim \frac{s}{(1-t)^2}\pi_i(x_s)\pi_j(x_s)\rho(x_s) + \frac{s^2}{2(1-t)^2}\pi_i(x_s)\pi_j(x_s)\rho(x_s) + \frac{1}{(1-t)^2}e^{-\tilde{\gamma}/(1-t)} \end{aligned}$$

where we have used the fact that $\pi_k(x_s) \leq e^{-\tilde{\gamma}/(1-t)}$ to get an upper bound on the second term of g. In the same way we obtain

$$g(x_{-s}) \lesssim \frac{-s}{(1-t)^2}\pi_i(x_{-s})\pi_j(x_{-s})\rho(x_{-s}) + \frac{s^2}{2(1-t)^2}\pi_i(x_{-s})\pi_j(x_{-s})\rho(x_{-s}) + \frac{1}{(1-t)^2}e^{-\tilde{\gamma}/(1-t)},$$

thus we get

$$\begin{aligned} |g_i(x_s) + g_i(x_{-s})| &\lesssim \frac{s}{(1-t)^2}|\pi_i(x_s)\pi_j(x_s)\rho(x_s) - \pi_i(x_{-s})\pi_j(x_{-s})\rho(x_{-s})| \\ &\quad + \frac{s^2}{2(1-t)^2}|\pi_i(x_s)\pi_j(x_s)\rho(x_s) + \pi_i(x_{-s})\pi_j(x_{-s})\rho(x_{-s})| \\ &\quad + \frac{1}{(1-t)^2}e^{-\tilde{\gamma}/(1-t)} \\ &\lesssim \frac{s}{(1-t)^2}|\pi_i(x_s)\pi_j(x_s) - \pi_i(x_{-s})\pi_j(x_{-s})|\rho(x_s) \\ &\quad + \frac{s}{(1-t)^2}\pi_i(x_{-s})\pi_j(x_{-s})|\rho(x_s) - \rho(x_{-s})| \\ &\quad + \frac{s^2}{2(1-t)^2}|\pi_i(x_s)\pi_j(x_s)\rho(x_s) + \pi_i(x_{-s})\pi_j(x_{-s})\rho(x_{-s})| \\ &\quad + \frac{1}{(1-t)^2}e^{-\tilde{\gamma}/(1-t)}. \end{aligned} \tag{E9}$$

First by Hölder continuity of ρ we obtain

$$|\rho(x_s) - \rho(x_{-s})| \leq C||x_s - x_{-s}||^\alpha \lesssim s^\alpha.$$

Consider now the term $\pi_i(x_s)$ and re-write it as

$$\pi_i(x_s) = \frac{1}{1 + \exp\left(\frac{\Delta_{ji}(x,t)}{1-t}\right) + \sum_{k\neq i,j}^N \exp\left(\frac{\Delta_{ki}(x,t)}{1-t}\right)},$$

now for the second term at the denominator we can use the Taylor expansion where as for the third we have the bound $\lesssim e^{-\tilde{\gamma}/(1-t)}$ so that we get

$$|\pi_i(x_s)\pi_j(x_s) - \pi_i(x_{-s})\pi_j(x_{-s})| \lesssim e^{-\tilde{\gamma}/(1-t)} e^{t(s+s^2)/(1-t)}.$$

Injecting now all these bounds into the integral on $T_{i,\eta,\gamma}$, and considering that η is going to be very small, yields

$$\left| \int_{T_{i,\eta,\gamma}} \sum_{j\neq i} \frac{\Delta_{ij}(x,t)}{(1-t)^2} \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| \lesssim \int_0^{\eta\kappa} \frac{1}{(1-t)^2} (s e^{-\tilde{\gamma}/(1-t)} e^{ts/(1-t)} + s^2 + s^{1+\alpha} + e^{\tilde{\gamma}/(1-t)}) \mathrm{d}t$$

$$\lesssim \frac{\eta^{2+\alpha}}{(1-t)^2} + \frac{\eta^3}{(1-t)^2} + \frac{e^{-\tilde{\gamma}/(1-t)}}{(1-t)^2}(\eta + (1-t)\eta e^{\eta/(1-t)} - (1-t)^2(e^{\eta/(1-t)} - 1)),$$

where $\kappa = \max_{i\neq j} \max_{x\in X} ||p_i^x - p_j^x||$. The control on $C_{i,\eta,\gamma}$ is exactly obtained as in [10] and we get that

$$\left| \int_{C_{i,\eta,\gamma}} \sum_{j\neq i} \frac{\Delta_{ij}(x,t)}{(1-t)^2} \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| \lesssim \frac{\gamma^2}{(1-t)^2}(\eta + e^{-\eta/(1-t)}).$$

Finally

$$\left| \int_X \sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| \lesssim \left| \int_{X_{i,\eta,-}} \sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| +$$

$$\left| \int_{X_{i,\eta,+}} \sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| + \left| \int_{T_{i,\eta,\gamma}} \sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| +$$

$$\left| \int_{C_{i,\eta,\gamma}} \sum_{j=1}^N \left(\frac{\Delta_{ij}(x,1)}{(1-t)^2}\right) \pi_i(x)\pi_j(x) \mathrm{d}\rho(x) \right| \lesssim \frac{e^{-\eta/(1-t)}}{(1-t)^2} + \frac{\gamma^2}{(1-t)^2}(\eta + e^{-\eta/(1-t)})$$

$$+ \frac{\eta^{2+\alpha}}{(1-t)^2} + \frac{\eta^3}{(1-t)^2} + \frac{e^{-\tilde{\gamma}/(1-t)}}{(1-t)^2}(\eta + (1-t)\eta e^{\eta/(1-t)} - (1-t)^2(e^{\eta/(1-t)} - 1))$$

and by correctly choosing γ and η we get the result. □