AN ODE CHARACTERISATION OF MULTI-MARGINAL OPTIMAL TRANSPORT

Luca Nenna, Brendan Pass

▶ **To cite this version:**

Luca Nenna, Brendan Pass. AN ODE CHARACTERISATION OF MULTI-MARGINAL OPTIMAL TRANSPORT. 2022. hal-03912268v1

HAL Id: hal-03912268

https://hal.science/hal-03912268v1

Preprint submitted on 23 Dec 2022 (v1), last revised 5 Aug 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AN ODE CHARACTERISATION OF MULTI-MARGINAL OPTIMAL TRANSPORT

LUCA NENNA AND BRENDAN PASS

Abstract. The purpose of this paper is to introduce a new numerical method to solve multi-marginal optimal transport problems with pairwise interaction costs. The complexity of multi-marginal optimal transport generally scales exponentially in the number of marginals m. We introduce a one parameter family of cost functions that interpolates between the original and a special cost function for which the problem's complexity scales linearly in m. We then show that the solution to the original problem can be recovered by solving an ordinary differential equation in the parameter ε, whose initial condition corresponds to the solution for the special cost function mentioned above; we then present some simulations, using both explicit Euler and explicit higher order Runge-Kutta schemes to compute solutions to the ODE, and, as a result, the multi-marginal optimal transport problem.

1. Introduction

The theory of optimal transport plays an important role in many applications (see [48, 50, 52, 53]). Its generalization to the multi-marginal case consists in minimizing the functional

$$\gamma \mapsto \int_{X^1 \times \ldots \times X^m} c(x^1, \cdots, x^m) \mathrm{d}\gamma$$

among all probability measures $\gamma \in \mathcal{P}(X^1 \times \cdots \times X^m)$ having fixed $\mu^i \in \mathcal{P}(X^i)$ with $i = 1, \cdots, m$ as marginals, for a given cost function $c(x^1, ..., x^m)$. This problem has been at the center of growing interest in recent years since it arises naturally in many different areas of applications, including Economics [18], Financial Mathematics [5, 25–27], Statistics [9,17], Image Processing [49], Tomography [1], Machine Learning [32, 51], Fluid Dynamics [11] and Quantum Physics and Chemistry, in the framework of Density Functional Theory [13, 22].

The structure of solutions to the multi-marginal optimal transport problem is a notoriously delicate issue, and is still not well understood, despite substantial efforts on the part of many researchers [14, 19–21, 31, 33, 36, 37, 39, 42, 43, 45–47]; see also the surveys [44] and [24]. In many of the aforementioned applications, it is therefore pertinent to develop efficient numerical algorithms to compute solutions. This, however, represents a significant challenge, since the problem amounts to a linear (or

convex, in a popular regularized variant discussed below), yet high dimensional optimization problem: the complexity scales exponentially in the number m of marginals. For instance a crude discretization of each of 5 marginals (notice that in many applications the number of marginals could be dramatically large, e.g. in quantum mechanics where m is the number of electrons in a molecule) using 100 Dirac masses would mean that the coupling γ between the 5 marginals is supported over $100^5 = 10^{10}$ Dirac masses, rendering the problem practically intractable. There have been recently some attempts to tackle this problem by using different approaches: entropic regularization [6, 8, 40], relaxation via moment constraints approximation [3, 4], genetic column generation algorithm exploiting the existence of a sparse solution in the discrete case [28, 29], Wasserstein penalisation of the marginal constraints [38] and semidefinite relaxation [34, 35].

In many cases of interest, the cost function $c(x^1, ...x^m) = \sum_{i=1}^m w(x^i, x^j)$ is given by a sum of two marginal cost functions; when $w(x^i, x^j) = |x^i - x^j|^2$, for instance the multi-marginal problem is equivalent to the well known Wasserstein barycenter problem [2], while the Coulomb cost $w(x^i, x^j) = \frac{1}{|x^i - x^j|}$ plays a central role in the quantum chemistry applications pioneered in [22] and [13]. Here, for such pairwise interaction costs, our aim is to develop a continuation method which, by introducing a suitable one parameter family of cost functions, establishes a link between the original multi-marginal problem and a simpler one whose complexity scales linearly in the number of marginals. For discrete marginals, we show that, after the addition of an entropic regularization term, the solution of the original multi-marginal problem can be recovered by solving an ordinary differential equation (ODE) whose initial condition is the solution to the simpler problem. This method is actually inspired by the one introduced in [15] to compute the Monge solution of the two marginal problem, starting from the Knothe-Rosenblatt rearrangement; note, however, that since we apply this strategy to a regularized problem, our resulting ODE enjoys better regularity than the one in [15], which, in turn, makes it amenable to higher order numerical schemes (see the description of numerical results in Section 4 below). The above mentioned differential equation will be derived by differentiating the optimality conditions of the dual problem; in particular by penalizing the constraints with the soft-max function we will obtain a well defined ODE for which existence and uniqueness of a solution can be established.

When developing the ODE approach in Section 3 below, we restrict our attention to the case when the marginals μ^i are all identical. This has the significant advantage of reducing the Kantorovich dual problem to a maximization over a single potential function, while also capturing important applications arising in density functional theory. Though we do not pursue this direction here, our approach, with minor modifications,

will also work with distinct marginals. In this case, if each measure is discretized using N points, one would need to solve $(m-1)N$ coupled, real-valued ODEs (rather than the N coupled ODEs dealt with here in Section 3), reflecting the $m-1$ independent Kantorovich potentials needed to fully capture the solution.

The remainder of this manuscript is organized as follows. In Section 2, we recall some basic facts about multi-marginal optimal transport, as well as its entropic regularization and the duals of both problems, and prove that for a particular, simple cost function, the solutions to the regularized problem and its dual can be computing by solving $m-1$ two marginal problems. This solution will serve as the initial condition for an ODE, which is introduced, and proven to be well-posed, in Section 3. In Section 4, algorithms, based on this ODE, to compute the solution to the multi-marginal optimal transport problem are described and some resulting numerical simulations are presented.

2. Multi-marginal optimal transport and entropic regularization

Given m probability measures μ^i on bounded domains $X^i \subseteq \mathbb{R}^n$ for $i = 1, 2..., m$ and a lower semi-continuous cost function $c : X^1 \times X^2 \times ... \times X^m \to \mathbb{R} \cup \{+\infty\}$, the multi-marginal optimal transport problem consists in solving the following optimization problem

$$\inf_{\gamma \in \Pi(\mu^1, \cdots, \mu^m)} \int_{X^1 \times X^2 ... \times X^m} c(x^1, ..., x^m) \mathrm{d}\gamma \tag{1}$$

where $\Pi(\mu^1, \cdots, \mu^m)$ denotes the set of probability measures on $X^1 \times X^2 \times ... \times X^m$ whose marginals are the μ^i. One can easily show by means of the direct method of calculus of variations that this problem admits at least a solution, which will be referred as the *optimal transport plan*. It is well known that under some mild assumptions the above problem is dual to the following

$$\sup \left\{ \sum_{i=1}^{m} \int_{X^i} \varphi^i(x^i) \mathrm{d}\mu^i \mid \varphi^i \in L^1(\mu^i), \ \sum_{i=1}^{m} \varphi^i(x^i) \leq c(x^1, ..., x^m) \right\}. \tag{2}$$

We will also consider a common variant of (1), known as *entropic optimal transport* which consists in adding an entropy regularization term. For a parameter $\eta > 0$, this is to minimize

$$\inf_{\gamma \in \Pi(\mu^1, \cdots, \mu^m)} \left\{ \int_{X^1 \times X^2 ... \times X^m} c(x^1, ..., x^m) \mathrm{d}\gamma + \eta H_{\otimes_{i=1}^m \mu^i}(\gamma) \right\} \tag{3}$$

where the relative entropy $H_{\otimes_{i=1}^m \mu^i}(\gamma)$ with respect to product measure $\otimes_{i=1}^m \mu^i$ is defined by

$$H_{\otimes_{i=1}^m \mu^i}(\gamma) = \int_{X^1 \times ... \times X^m} \frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m \mu^i)} \log(\frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m \mu^i)}) \mathrm{d}(\otimes_{i=1}^m \mu^i),$$

if γ is absolutely continuous with respect to the product measure $\otimes_{i=1}^m \mu^i$ and $+\infty$ otherwise. The regularized transport is, in turn, dual to the following unconstrained optimization problem

$$\begin{aligned} \sup & \sum_{i=1}^m \int_{X^i} \varphi^i(x^i) \mathrm{d}\mu^i \\ & - \eta \int_{X^1 \times \ldots \times X^m} e^{\frac{\sum_{i=1}^m \varphi^i(x^i) - c(x^1,\ldots,x^m)}{\eta}} \mathrm{d}(\otimes_{i=1}^m \mu^i)(x^1, \cdots, x^m). \end{aligned} \tag{4}$$

The regularized problem (3) and its dual (4) arise frequently in computational work. We note in particular that (3) is the minimization of a strictly convex functional and therefore admits a unique solution. It is well known that as $\eta \to 0$, solutions of (3) and (4) converge to solutions of (1) and (2), respectively. When each X^i is a finite set (a case of particular interest in this paper), we obtain discrete versions of (1) and its dual, which amount to linear programs, taking the forms

$$\inf \left\{ \sum_{\overline{x} \in \times_{i=1}^m X^i} c(\overline{x}) \gamma_{\overline{x}} \mid \gamma \in \Pi(\mu^1, \cdots, \mu^m) \right\}, \tag{5}$$

where $\overline{x} = (x^1, \cdots, x^m) \in X^1 \times .... \times X^m$ and

$$\sup\{ \sum_{i=1}^m \sum_{x \in X^i} \varphi_x^i \mu_x^i \mid (\varphi^1, \cdots, \varphi^m) \in \mathcal{T} \} \tag{6}$$

where, if we identify functions $\varphi^i : X^i \to \mathbb{R}$ with points in $\mathbb{R}^{|X^i|}$,

$$\mathcal{T} := \{ \varphi^i \in \mathbb{R}^{|X^i|} \ \forall i = 1, \cdots, m, \sum_{i=1}^m \varphi_{x^i}^i \leq c(x^1, \cdots, x^m), \ \forall (x^1, \cdots, x^m) \in \times_{i=1}^m X^i \}.$$

Notice that, if each $|X^i| = N$, in the case of the primal problem we have to deal with N^m unknowns and mN constraints, whereas in the dual problem there are mN unknowns and N^m constraints. In both cases we have to deal with the so called "curse of dimensionality," namely the complexity of the problem increases exponentially with the number of marginals.

In this discrete setting, the entropy regularized problem (3) and its dual (4) then become finite dimensional convex optimization problems:

$$\inf \left\{ \sum_{\overline{x} \in \times_{i=1}^m X^i} c(\overline{x}) \gamma_{\overline{x}} + \eta [H(\gamma) - H(\otimes^m \mu^i)] \mid \gamma \in \Pi(\mu^1, \cdots, \mu^m), \right\} \tag{7}$$

where $\eta > 0$ and H is the entropy with respect to uniform measure on the finite set X (we suppress the subscript on H indicating the reference

measure in the finite case, as we will only deal with entropy relative to the uniform measure), $H(\gamma) = \sum_{\overline{x} \in \times_{i=1}^m X^i} h(\gamma_{\overline{x}})$, with

$$h(t) = \begin{cases} t(\log(t) - 1), \ t > 0 \\ 0, \ t = 0 \\ +\infty, \ t < 0, \end{cases}$$

and

$$\sup \left\{ \sum_{i=1}^m \sum_{x \in X^i} \varphi_x^i \mu_x^i - \eta \sum_{\overline{x} \in \times_{i=1}^m X^i} \exp\left(\frac{\sum_i \varphi_{x^i}^i - c(\overline{x})}{\eta} \right) (\otimes^m \mu^i)_{\overline{x}} \right\}. \quad (8)$$

We note in particular that (8) is an *unconstrained* finite dimensional concave maximization problem. Solutions may be computed using a multi-marginal version of the Sinkhorn algorithm [7, 23, 30, 48], and one can then recover the optimal γ in (7) from the solutions $\varphi^1, ..., \varphi^m$ to (8) via the well known formula:

$$\gamma_{\overline{x}} = \exp\left(\frac{\sum_{i=1}^m \varphi_{x^i}^i - c(\overline{x})}{\eta} \right) \mu_{x^1}^1 \mu_{x^2}^2 ... \mu_{x^m}^m$$

where $\overline{x} = (x^1, ..., x^m)$.

2.1. **Pairwise costs.** We are especially interested in this paper in cost functions $c(x^1, ..., x^m)$ involving pair-wise interactions, that is

$$c(x^1, ..., x^m) = \sum_{i<j}^m w(x^i, x^j).$$

Such costs are ubiquitous in applications: for example, for systems of interacting classical particles in [13, 22], c is a pair-wise cost, with $w(x - y) = \dfrac{1}{|x - y|}$, known as the Coulomb cost.

Let us now consider costs c_ε of the form

$$c_\varepsilon(x^1, \cdots, x^m) := \varepsilon \sum_{i=2}^m \sum_{j=i+1}^m w(x^i, x^j) + \sum_{i=2}^m w(x^1, x^i). \quad (9)$$

It is clear that when $\varepsilon = 1$ we retrieve a pair-wise cost as defined above whereas in the limit $\varepsilon \to 0$ we obtain a cost involving only the interactions between x^1 and the other x^i individually. Later on, we will develop an ordinary differential equation that governs the evolution with ε of the solutions to the regularized dual problem (8); the results below assert that the initial condition for that equation (that is, the solutions when $\varepsilon = 0$) can be recovered by solving each of the individual two marginal problems between μ^1 and μ^i.

In what follows, we will assume that each marginal μ^i is absolutely continuous with respect to a fixed based measure ν^i with density given by $\frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}$.

Proposition 1. *Assume that each marginal μ^i is absolutely continuous with respect to a fixed based measure ν^i with density given by $\frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}(x^i)$: $\mathrm{d}\mu^i(x^i) = \frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}(x^i)\mathrm{d}\nu^i(x^i)$. Consider the regularized problem (3) with limiting pairwise cost; that is, set $\varepsilon = 0$ in (9) to obtain:*

$$\min_{\gamma\in\Pi(\mu^1,\mu^2,...,\mu^m)} \int \sum_{i=2}^{m} w(x^1,x^i)\mathrm{d}\gamma + \eta H_{\otimes_{i=1}^m\mu^i}(\gamma). \tag{10}$$

Let $\frac{\mathrm{d}\overline{\pi}^i}{\mathrm{d}(\nu^1\otimes\nu^i)}$ be the density with respect to product measure $\nu^1(x^1)\otimes\nu^i(x^i)$ of the minimizer $\overline{\pi}^i = \frac{\mathrm{d}\overline{\pi}^i}{\mathrm{d}(\nu^1\otimes\nu^i)}\nu^1\otimes\nu^i$ in the regularized two marginal problem:

$$\min_{\pi^i\in\Pi(\mu^1,\mu^i)} \int w(x^1,x^i)\mathrm{d}\pi^i(x^1,x^i) + \eta H_{\mu^1\otimes\mu^i}(\pi^i).$$

Then the density $\frac{\mathrm{d}\overline{\gamma}}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}$ of the optimal $\overline{\gamma} = \frac{\mathrm{d}\overline{\gamma}}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}(\otimes_{i=1}^m\nu^i)$ in (10) is given by

$$\frac{\mathrm{d}\overline{\gamma}}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}(x^1,...,x^m) = \frac{\frac{\mathrm{d}\overline{\pi}^2}{\mathrm{d}(\nu^1\otimes\nu^2)}(x^1,x^2)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)}\frac{\frac{\mathrm{d}\overline{\pi}^3}{\mathrm{d}(\nu^1\otimes\nu^3)}(x^1,x^3)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)}...\frac{\frac{\mathrm{d}\overline{\pi}^m}{\mathrm{d}(\nu^1\otimes\nu^m)}(x^1,x^m)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)}\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1).$$

Proof. Choose any $\gamma = \frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}(\otimes_{i=1}^m\nu^i) \in \Pi(\mu^1,\mu^2,...,\mu^m)$ which is absolutely continuous with respect to $\otimes_{i=1}^m\nu^i$ and let $\pi^i(x^1,x^i) = \Big((x^1,..,x^m)\mapsto (x^1,x^i)\Big)_{\#}\gamma \in \Pi(\mu^1,\mu^i)$ be its twofold marginals. Then

$$\begin{aligned} H_{\otimes_{i=1}^m\mu^i}(\gamma) &= \int_{x^1\times...\times X^m}[\log(\frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}(x^1,...,x^m)) - \sum_{i=1}^m\log(\frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}(x^i))]\mathrm{d}\gamma(x^1,...,x^m) \\ &= \int_{X^1\times...\times X^m}[\log(\frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}(x^1,...,x^m))]\mathrm{d}\gamma(x^1,....x_N) - \sum_{i=1}^m H_{\nu^i}(\mu^i), \end{aligned} \tag{11}$$

where each $H_{\nu^i}(\mu^i) = \int_{X^i}\log(\frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}(x^i))\mathrm{d}\mu^i(x^i)$ is constant throughout $\Pi(\mu^1,...,\mu^m)$.

Now disintegrating $\gamma = \gamma_{x^1}(x^2,...,x^m)\otimes\mu^1(x^1)$ with respect to its first marginal μ^1, we note that $\gamma_{x^1}(x^2,...,x^m) = \frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m\nu^1)}(x^1,x^2,...,x^m)\frac{1}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)}(\otimes_{i=2}^m\nu^i)$ and so

$$\begin{aligned} &\int_{X^1\times...\times X^m}\log\left(\frac{\mathrm{d}\gamma}{\mathrm{d}(\otimes_{i=1}^m\nu^i)}\right)\mathrm{d}\gamma \\ =&\int_{X^1\times...\times X^m}\log\left(\frac{\mathrm{d}\gamma_{x^1}}{\mathrm{d}(\otimes_{i=2}^m\nu^i)}\right)\mathrm{d}\gamma + H_{\nu^1}(\mu^1) \\ =&\int_{X^1}\int_{X^2\times...\times X^m}\log\left(\frac{\mathrm{d}\gamma_{x^1}}{\mathrm{d}(\otimes_{i=2}^m\nu^i)}\right)\mathrm{d}\gamma_{x^1}\mathrm{d}\mu^1 + H_{\nu^1}(\mu^1) \\ =&\int_{X^1}H_{\otimes_{i=2}^m\nu_i}(\gamma_{x^1})\mathrm{d}\mu^1 + H_{\nu^1}(\mu^1) \end{aligned} \tag{12}$$

where $H_{\otimes_{i=2}^m \nu_i}(\gamma_{x^1})$ is the entropy of γ_{x^1} with respect to $\otimes_{i=2}^m \nu_i$ and $H_{\nu^1}(\mu^1)$ is the entropy of μ^1 with respect to ν^1. Now note that if we disintegrate each $\pi^i = \pi^i_{x^1}(x^i) \otimes \mu^1(x^1)$ with respect to μ^1, then for each fixed x^1, the conditional probability $\pi^i_{x^1}$ is the ith marginal of γ_{x^1} and so

$$\begin{aligned}
\int_{X^1} H_{\otimes_{i=2}^m \nu_i}(\gamma_{x^1}) \mathrm{d}\mu^1(x^1) &\geq \int_{X^1} \sum_{i=2}^m H_{\nu^i}(\pi^i_{x^1}) \mathrm{d}\mu^1(x^1) \\
&= \sum_{i=2}^m \int_{X^1} H_{\nu^i}(\pi^i_{x^1}) \mathrm{d}\mu^1(x^1) \\
&= \sum_{i=2}^m [H_{\nu^1 \otimes \nu^i}(\pi^i) - H_{\nu^1}(\mu^1)] \\
&= \sum_{i=2}^m [H_{\mu^1 \otimes \mu^i}(\pi^i) + H_{\nu^i}(\mu^i)], \qquad (13)
\end{aligned}$$

where the equality $H_{\nu^1 \otimes \nu^i}(\pi^i) = \int_{X^1} H_{\nu^i}(\pi^i_{x^1}) \mathrm{d}\mu^1(x^1) + H_{\nu^1}(\mu^1)$ in the second to last line follows very similarly to the derivation of (12) above and the equality $H_{\nu^1 \otimes \nu^i}(\pi^i) - H_{\nu^1}(\mu^1) = H_{\mu^1 \otimes \mu^i}(\pi^i) + H_{\nu^i}(\mu^i)$ in the last line follows very similarly to the derivation of (11). Therefore, combining (11), (12) and (13), we get

$$\begin{aligned}
&\int_{X^1 \times \ldots X^m} \sum_{i=1}^m w(x^1, x^i) \mathrm{d}\gamma + \eta H_{\otimes_{i=1}^m \mu_i}(\gamma) = \int_{X^1 \times X^i} \sum_{i=1}^m w(x^1, x^i) \mathrm{d}\pi^i + \eta H_{\otimes_{i=1}^m \mu_i}(\gamma) \\
&\geq \int_{X^1 \times X^i} \sum_{i=2}^m w(x^1, x^i) \mathrm{d}\pi^i + \sum_{i=2}^m [H_{\mu^1 \otimes \mu^i}(\pi^i) + H_{\nu^i}(\mu^i)] + H_{\nu^1}(\mu^1) - \sum_{i=1}^m H_{\nu^i}(\mu^i) \\
&\geq \int_{X^1 \times X^i} \sum_{i=2}^m w(x^1, x^i) \mathrm{d}\overline{\pi}^i + \sum_{i=2}^m H_{\mu^1 \otimes \mu^i}(\overline{\pi}^i)
\end{aligned}$$

by optimality of $\overline{\pi}$. We have equality in the last line if and only if $\pi^i = \overline{\pi}^i$ for each i, and equality in the line above if and only if μ^1 almost every γ_{x^1} couples the $\pi^i_{x^i}$ independently; this yields the desired result. □

Note in particular that this result allows us to recover the solution to problem (7) with cost (9), when $\varepsilon = 0$, by solving $m-1$ individual regularized two marginal optimal transport problems. In the following section, we will develop a dynamical approach to solve the dual problem (8) to (7) for cost (9) with $\varepsilon > 0$. Our initial condition will be the dual potentials when $\varepsilon = 0$, which we can obtain from the corresponding two marginal dual potentials, as the following corollary confirms.

Corollary 2. *Assume each μ^i is absolutely continuous with respect to a given reference measure ν^i. For each $i = 2, ...m$, let $\psi^i(x^1), \varphi^i(x^i)$, solve the regularized two marginal dual problem (8) between marginals μ^1 and μ^i with cost function $w(x^1, x^i)$. Then $\varphi^1(x^1), \varphi^2(x^2), ..., \varphi^m(x^m)$, with $\varphi^1(x^1) = \sum_{i=2}^m \psi^i(x^1)$ solve the regularized dual (8) with marginals $\mu^1, \mu^2, ..., \mu^m$ and cost $c(x^1, ..., x^m) = \sum_{i=2}^m w(x^1, x^i)$.*

Proof. We have that for each i, the optimizer $\overline{\pi}^i$ in the regularized two marginal primal problem satisfies

$$\frac{\mathrm{d}\overline{\pi}^i}{\mathrm{d}(\nu^1 \otimes \nu^i)}(x^1, x^i) = e^{\frac{\varphi^i(x^i)+\psi^i(x^i)-w(x^1,x^i)}{\eta}} \frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}\frac{\mathrm{d}\mu^i}{\mathrm{d}\nu^i}.$$

By the preceding proposition, the optimizer in the regularized multi-marginal problem (10) satisfies

$$\begin{aligned}
\frac{\mathrm{d}\overline{\gamma}}{\mathrm{d}(\otimes_{i=1}^m \nu^i)}(x^1, x^2, ..., x^m) &= \frac{\frac{\mathrm{d}\overline{\pi}^2}{\mathrm{d}(\nu^1\otimes\nu^2)}(x^1,x^2)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)} \frac{\frac{\mathrm{d}\overline{\pi}^3}{\mathrm{d}(\nu^1\otimes\nu^3)}(x^1,x^3)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)} ... \frac{\frac{\mathrm{d}\overline{\pi}^m}{\mathrm{d}(\nu^1\otimes\nu^m)}(x^1,x^m)}{\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1)} \frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}(x^1) \\
&= e^{\frac{\sum_{i=2}^m [\varphi^i(x^i)+\varphi^i(x^1)-w(x^1,x^i)]}{\eta}} \frac{\mathrm{d}\mu^2}{\mathrm{d}\nu^2}\frac{\mathrm{d}\mu_3}{\mathrm{d}\nu^3}...\frac{\mathrm{d}\mu^m}{\mathrm{d}\nu^m}\frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1} \\
&= e^{\frac{\sum_{i=1}^m \varphi^i(x^i)-\sum_{i=2}^m w(x^1,x^i)}{\eta}} \frac{\mathrm{d}\mu^1}{\mathrm{d}\nu^1}\frac{\mathrm{d}\mu^2}{\mathrm{d}\nu^2}\frac{\mathrm{d}\mu_3}{\mathrm{d}\nu^3}...\frac{\mathrm{d}\mu^m}{\mathrm{d}\nu^m}
\end{aligned}$$

This is exactly the first order condition identifying the regularized potentials for the multi-marginal regularized problem with cost $\sum_{i=2}^m w(x^1, x^i)$. □

3. An ODE characterisation of discrete multi-marginal optimal transport

We now turn our attention to developing an ODE for the Kantorovich potentials after discretizing the marginals. Working with the regularized discrete problem (7) and its dual (8) with pairwise cost (9), we make the following, standing assumptions throughout this section:

(1) (Equal marginals) All the marginals are equal $\mu^i = \rho = \sum_{x\in X} \rho_x \delta_x$, where X is a finite subset of $\mathbb{R}^d$,
(2) (Symmetric cost) The two body cost w is symmetric $w(x, y) = w(x, y)$.
(3) (Finite cost) The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

A motivating example of a pairwise, symmetric two body cost arises in Density Functional Theory where the cost is given by $w(x, y) = \frac{1}{|x-y|}$; in problems with this cost, the marginals are typically also identical. The cost does not satisfy the finiteness hypothesis, but one can consider a truncation $w(x, y) = \min\left(\frac{1}{|x-y|}, C\right)$ cost; it is known that the solution stays away from the diagonal, and for sufficiently large C, the solution with the truncated cost coincides with the solution for the original Coulomb cost (for instance see [12]).

Remark 3. *One could dispose of the equal marginal and symmetric cost assumptions. Analogues of the results proved below would sill hold; one could characterize the solution to the regularized dual problem* (8) *by an ODE, and prove that this ODE is well-posed. As we will see below,*

however, solving the problem numerically becomes more feasible under the hypotheses above, as the solution can be characterized by a single Kantorovich potential, and so the resulting ODE is an equation on $\mathbb{R}^N$, where N is the number of points in the support of the marginal. With unequal marginals and a non-symmetric cost, one would require $m-1$ independent Kantorovich potentials to fully characterize the solution; if each marginal is supported on N points, this would lead to an $(m-1)N$ dimensional system of ODEs.

3.1. **Formulation of the ODE problem.** Notice now that although the cost (9) at $\varepsilon = 1$ is symmetric in the variables $x^1, x^2, ..., x^m$, the one at $\varepsilon < 1$ is not. It is, however, symmetric in the variables $x^2, ..., x^m$; this means that the optimal φ^i in (8) satisfy $\varphi^i = \varphi^j = \varphi$ for $i, j \geq 2$ and so, setting $\varphi^1 = \psi$, we can rewrite (8) as

$$\inf_{\varphi,\psi:X\to\mathbb{R}} \left\{\Phi(\varphi,\psi,\varepsilon)\right\}, \tag{14}$$

where

$$\Phi(\varphi,\psi,\varepsilon) := -(m-1)\sum_{x\in X}\varphi_x\rho_x - \sum_{x\in X}\psi_x\rho_x + \eta\sum_{\overline{x}\in X^m} e^{\left(\frac{\sum_{i=2}^m \varphi_{x^i}+\psi_{x^1}-c_\varepsilon(\overline{x})}{\eta}\right)}\otimes^m\rho.$$

Remark 4 (Notation)**.** *Recall that we use the notation $\overline{x}$ to represent a point in a product space, such as $\overline{x} = (x^1, \cdots, x^m) \in X^m$, as above, or, as will often be the case in what follows, $\overline{x} = (x^1, \cdots, x^{m-1}) \in X^{m-1}$. We introduce the following notation to represent corresponding products of the densities:*

$$\tilde{\rho}_{\overline{x}} = (\otimes^{m-1}\rho)_{\overline{x}} = \otimes_{i=1}^{m-1}\rho_{x^i}$$

Since the functional $\Phi(\varphi,\psi,\varepsilon)$ is convex on the set $\{\varphi,\psi : X\to\mathbb{R}\} \approx \mathbb{R}^{2|X|}$, as the sum of a linear and an exponential function, optimal solutions (φ^*,ψ^*) can be characterized by the first order optimality conditions $\nabla_\varphi\Phi = \nabla_\psi\Phi = 0$, or (component-wise):

$$\varphi^*_z = -\eta\log\left(\sum_{\overline{x}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^{m-1}\varphi^*_{x^i}+\psi^*_{x^1}-c_\varepsilon(\overline{x},z)}{\eta}\right)\tilde{\rho}_{\overline{x}}\right)$$

and

$$\psi^*_z = -\eta\log\left(\sum_{\overline{x}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^{m}\varphi^*_{x^i}-c_\varepsilon(z,\overline{x})}{\eta}\right)\tilde{\rho}_{\overline{x}}\right). \tag{15}$$

In particular, note that (15) allows us to express the optimal ψ^* in (14) in terms of the optimal φ^*, after which (14) reduces to the following optimization problem

$$\inf_{\varphi:X\to\mathbb{R}}\left\{\tilde{\Phi}(\varphi,\varepsilon)\right\}, \tag{16}$$

where

$$\tilde{\Phi}(\varphi,\varepsilon) := -(m-1)\sum_{x\in X}\varphi_x\rho_x + \eta\sum_z \log\left(\sum_{\overline{x}\in X^{m-1}} e^{\left(\frac{\sum_{i=2}^m \varphi_{x^i} - c_\varepsilon(z,\overline{x})}{\eta}\right)}\tilde{\rho}_{\overline{x}}\right)\rho_z.$$

Remark 5 (LogSumExp and convexity)**.** *The function*

$$\varphi \mapsto \log\left(\sum_{\overline{x}\in X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi_{x^i} - c_\varepsilon(z,\overline{x})}{\eta}\right)\tilde{\rho}_{\overline{x}}\right) := LSE_{c_\varepsilon}(\varphi)_z$$

is also known as Log-Sum-Exp function (LSE). By using the Hölder inequality one can easily show that the Log-Sum-Exp is convex.

It is well known that the solution to (8) is unique up to the addition of constants $\varphi^i \mapsto \varphi^i + C^i$ adding to 0, $\sum_{i=1}^m C^i = 0$; thus, solutions to (16) are unique up to the addition of a single constant, $\varphi \mapsto \varphi + C$. We therefore impose the normalization

$$\varphi_{x_0} = 0 \tag{17}$$

for all $\varepsilon \in [0,1]$ and a fixed $x_0 \in X$.

The problem (16), restricted to φ's satisfying (17) then has a unique solution; the function $\tilde{\Phi}(\cdot,\varepsilon)$ is strictly convex when restricted to this set, and the solution $\varphi^* = \varphi(\varepsilon)$ can be characterized by the optimality condition $\nabla_\varphi\tilde{\Phi}(\varphi^*,\varepsilon) = 0$, where each component of the gradient is given by

$$\frac{\partial}{\partial\varphi_z}\tilde{\Phi} = -(m-1)\rho_z + (m-1)e^{\varphi_z/\eta}\rho_z\sum_y\sum_{\overline{x}\in X^{m-2}} e^{\left(\frac{\sum_{i=3}^m \varphi_{x^i} - c_\varepsilon(y,z,\overline{x})}{\eta}\right)}(\otimes^{m-2}\rho_{\overline{x}})\overline{\rho}_y. \tag{18}$$

where

$$\overline{\rho}_y = \frac{\rho_y}{\sum_{\overline{x}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^m \varphi_{x^i} - c_\varepsilon(y,\overline{x})}{\eta}\right)\tilde{\rho}_{\overline{x}}}.$$

Our numerical method consists then in solving an ODE for the evolution of $\varphi(\varepsilon)$ obtained by differentiating

$$\nabla_\varphi\tilde{\Phi}(\varphi(\varepsilon),\varepsilon) = 0 \tag{19}$$

with respect to ε:

$$\frac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi(\varepsilon),\varepsilon) + D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\varepsilon),\varepsilon)\frac{\mathrm{d}\varphi}{\mathrm{d}\varepsilon}(\varepsilon) = 0. \tag{20}$$

If the pure second derivatives with respect to φ as well as the mixed second derivatives with respect to φ and ε exist and are Lipschitz, and the

Hessian with respect to φ is invertible, we will obtain a characterization of φ as the solution to the following well-posed Cauchy problem:

$$\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\varepsilon}(\varepsilon) = -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\varepsilon),\varepsilon)]^{-1}\dfrac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi(\varepsilon),\varepsilon), \\ \varphi(0) = \varphi_w, \end{cases} \tag{21}$$

where, by Corollary 2, the initial value $\varphi(0)$ of φ when $\varepsilon = 0$ coincides with φ_w, the optimal potential for the two marginal optimal transport problem with cost w.

The next section is devoted to proving these properties.

3.2. **Well posedness of the ODE.** We refer the reader to appendix A for the computation of the second pure and mixed derivatives with respect to φ and the second mixed derivative with respect to φ and ε. In order to prove invertibility of $D^2_{\varphi,\varphi}\tilde{\Phi}$ and well posedness of the ODE we need some lemmas giving uniform bounds on the potential φ and the eigenvalues of $D^2_{\varphi,\varphi}\tilde{\Phi}$. We highlight that the following arguments are similar to (and largely inspired by) those in [16] (the main differences lie in the fact that we re-write the dual problem by using the Log-Sum-Exp function).

Lemma 6. *Let c_ε satisfy the boundedness assumption $\|c_\varepsilon\|_\infty \leq M$, $\forall\varepsilon \in [0,1]$ [1]. Then the maximizer $\varphi(\varepsilon)$ of* (16) *subject to the normalization constraint* (17) *satisfies*

$$\|\varphi(\varepsilon)\|_\infty \leq 4M.$$

Proof. By the first order optimality condition $\nabla_\varphi\tilde{\Phi} = 0$ for (16) we deduce that each component of $\varphi(\varepsilon)$ is given by

$$\varphi_z = -\eta\log\left(\sum_y\sum_{\overline{x}\in X^{m-2}}\exp\left(\frac{\sum_{i=3}^m\varphi_{x^i} - c_\varepsilon(y,z,\overline{x})}{\eta}\right)(\otimes^{m-2}\rho)_{\overline{x}}\overline{\rho}_y\right)$$

It is is easy to see that $\overline{\rho}_y$ can be bounded as follows

$$\frac{e^{-M/\eta}\rho_y}{\sum_{\overline{x}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^m\varphi_{x^i}}{\eta}\right)\tilde{\rho}_{\overline{x}}} \leq \overline{\rho}_y \leq \frac{e^{M/\eta}\rho_y}{\sum_{\overline{x}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^m\varphi_{x^i}}{\eta}\right)\tilde{\rho}_{\overline{x}}}.$$

[1]Note that the boundedness $\|c_\varepsilon\|_\infty \leq M$ for some $M > 0$ follows immediately from our finite cost assumption on the finite set X^m.

Since we have imposed the normalization $\varphi_{x_0} = 0$ we get

$$\begin{aligned}\varphi_z = \varphi_z - \varphi_{x_0} &\leq -\eta \log\left(e^{-2M/\eta}\frac{\sum_{\overline{x}\in X^{m-2}} \exp\left(\frac{\sum_{i=3}^m \varphi_{x^i}}{\eta}\right)(\otimes^{m-2}\rho)_{\overline{x}}}{\sum_{\overline{x}\in X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi_{x^i}}{\eta}\right)\tilde{\rho}_{\overline{x}}}\right)\\ &\quad + \eta \log\left(e^{2M/\eta}\frac{\sum_{\overline{x}\in X^{m-2}} \exp\left(\frac{\sum_{i=3}^m \varphi_{x^i}}{\eta}\right)(\otimes^{m-2}\rho)_{\overline{x}}}{\sum_{\overline{x}\in X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi_{x^i}}{\eta}\right)\tilde{\rho}_{\overline{x}}}\right),\\ &\quad \leq 4M,\end{aligned}$$

and the desired result immediately follows. □

Having established the above bounds, we aim to prove the well posedness of the Cauchy problem (21) on the set

$$U := \{\varphi : X \to \mathbb{R} \mid \varphi_{x_0} = 0,\ \|\varphi\|_\infty \leq 4M\}. \tag{22}$$

Lemma 7. *$D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi,\varepsilon)$ and $\frac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi,\varepsilon)$ are Lipschitz with respect to φ on U.*

Proof. This immediately follows from the fact that the the second pure and mixed derivatives computed in Appendix A are easily seen to be C^1, and their derivatives are all bounded on U. □

In order to prove the invertibility of $D^2_{\varphi,\varphi}\tilde{\Phi}$ we need the following lemma assuring the strong convexity of the Log-Sum-Exp function on the set U.

Lemma 8. *Let $\Psi : \tilde{U} \to \mathbb{R}$ be defined on*

$$\tilde{U}_C = \{\theta : X^{m-1} \to \mathbb{R} \mid \theta_{\overline{x}_0} = 0,\ \|\theta\|_\infty < C\}.$$

where $\overline{x}_0 = (x_0, ..., x_0) \in X^{m-1}$, by $\Psi(\theta) = \sum_{y\in X} \log\Big(\sum_{\overline{x}\in X^{m-1}} e^{\theta_{\overline{x}} - c_\varepsilon(y,\overline{x})}\tilde{\rho}_{\overline{x}}\Big)\rho_y$.

Then Ψ is β-strongly convex for some $\beta > 0$.

Proof. It is enough to show strong convexity on this set of the function

$$f_y : \theta \in \tilde{U}_C \mapsto \log\Big(\sum_{\overline{x}\in X^{m-1}} e^{\theta_{\overline{x}} - c_\varepsilon(y,\overline{x})}\tilde{\rho}_{\overline{x}}\Big) = \log\Big(e^{-c_\varepsilon(y,\overline{x}_0)}\tilde{\rho}_{\overline{x}_0} + \sum_{\overline{x}\in X^{m-1}\setminus\{\overline{x}_0\}} e^{\theta_{\overline{x}} - c_\varepsilon(y,\overline{x})}\tilde{\rho}_{\overline{x}}\Big)$$

for a fixed y.

Enumerating the set $X^{m-1} \setminus \{\overline{x}_0\}$ of independent variables as $\overline{x}_j$ for $j \in (1, ..., K)$ with $K = |X|^{m-1} - 1$, and denoting $z^j = e^{\theta_{\overline{x}} - c_\varepsilon(y,\overline{x})}\tilde{\rho}_{\overline{x}}$ the Hessian of f_y is

$$\frac{1}{\Big(e^{-c_\varepsilon(y,\overline{x}_0)}\tilde{\rho}_{\overline{x}_0} + \sum_j z^j\Big)^2}\Big(-z\otimes z + \mathrm{diag}(z)(\sum_j z^j + e^{-c_\varepsilon(y,\overline{x}_0)}\tilde{\rho}_{\overline{x}_0})\Big)$$

The first two terms together constitute a positive semi-definite matrix (see [10], p.74), while the third is positive definite, with lower bound

$$\beta = \frac{e^{-C-2M}\tilde{\rho}_{\overline{x}_0}\min_{\overline{x}}\tilde{\rho}_{\overline{x}}}{(e^{C+M}\tilde{\rho}_{\overline{x}_0}+\sum_{\overline{x}\in X^{m-1}\setminus\{\overline{x}_0\}}e^{C+M}\tilde{\rho}_{\overline{x}})^2} = e^{-4M-3C}\tilde{\rho}_{\overline{x}_0}\min_{\overline{x}}\tilde{\rho}_{\overline{x}}.$$

It follows that f_y, and therefore Ψ, is β-convex on $\tilde{U}$.

□

Lemma 9. *Let $\Lambda : U \to \tilde{U}_{4(m-1)M}$ be the linear mapping defined by $\Lambda(\varphi)_{\overline{x}} = \varphi_{x^1} + \cdots + \varphi_{x^{m-1}}$ $\forall \overline{x} \in X^{m-1}$. Then $\tilde{\Psi}(\varphi) := \Psi(\Lambda(\varphi))$ is $\alpha-$strongly convex on U.*

Proof. By the linearity of Λ, one gets that, for $\varphi \in U$, $D^2\tilde{\Psi}(\varphi)(v,v) = D^2\Psi(\Lambda(\varphi))(\Lambda(v),\Lambda(v))$ for all $v \in U$. Thus,

$$D^2\tilde{\Psi}(\varphi)(v,v) = D^2\Psi(\Lambda(\varphi))(\Lambda(v),\Lambda(v)) \geq \beta \left\|\Lambda(v)\right\|^2.$$

Since $\left\|\Lambda(v)\right\|^2 \geq \sum_{x\in X}\left\|(m-1)v_x\right\|^2$ we finally get

$$D^2\tilde{\Psi}(\varphi)(v,v) \geq \alpha \left\|v\right\|^2,$$

with $\alpha = \beta(m-1)^2 > 0$, proving the $\alpha-$strong convexity of $\tilde{\Psi}$. □

Remark 10. *The α obtained in the Lemma above is not optimal; indeed we would have obtained a better lower bound on the eigenvalues of $D^2_{\varphi,\varphi}\tilde{\Psi}$ by computing the smallest eigenvalue of $\Lambda^*\Lambda$. Moreover, in the previous lemma we take, for simplicity, $\eta = 1$, otherwise the parameter α would have taken the form $\alpha = e^{(-4C-3M)/\eta}(m-1)^2$. Notice now that α approaches to 0 as $\eta \to 0$, meaning the the condition number of the Hessian of $\tilde{\Phi}$ explodes. Namely, this will produce numerical instabilities.*

It easily follows from the previous lemma that $D^2_{\varphi,\varphi}\tilde{\Phi} = D^2_{\varphi,\varphi}\tilde{\Psi}$ is invertible on the set U; we can then state the following result on the well posedness of (21).

Theorem 11. *Let be $\varphi(\varepsilon)$ the solution to (16) for all $\varepsilon \in [0,1]$. Then $\varepsilon \mapsto \varphi(\varepsilon)$ is $\mathcal{C}^1$ and is the unique solution to the Cauchy problem*

$$\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\varepsilon}(\varepsilon) = -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\varepsilon),\varepsilon)]^{-1}\dfrac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi(\varepsilon),\varepsilon), \\ \varphi(0) = \varphi_w, \end{cases} \tag{23}$$

where φ_w is the optimal solution to (8) with cost w and two marginals equal to ρ.

Proof. As $\varphi(\varepsilon)$ minimizes $\tilde{\Phi}(\cdot,\varepsilon)$ for each fixed ε, we clearly have (19). Since $\tilde{\Phi}$ is clearly twice differentiable with respect to φ and ε and $D^2_{\varphi\varphi}\tilde{\Phi}$ is invertible by Lemma 9, the Implicit Function Theorem then implies that $\varepsilon \mapsto \varphi(\varepsilon)$ is C^1 and satisfies (20), or equivalently, (23).

Since $D^2_{\varphi,\varphi}\tilde{\Phi}$ and $\frac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}$ are Lipschitz continuous with respect to φ on U by Lemma 7 and clearly continuous with respect to ε, and since $D^2_{\varphi,\varphi}\tilde{\Phi}$ is uniformly positive definite by Lemma 9, we have that

$$(\varphi,\varepsilon)\mapsto -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\varepsilon),\varepsilon)]^{-1}\frac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi(\varepsilon),\varepsilon)$$

is Lipschitz with respect to φ and continuous with respect to ε on U. Since by Lemma 6 $\varphi(\varepsilon)\in U$ for all ε, the Cauchy-Lipschitz Theorem then implies uniqueness of the solution to (23) on $U\times[0,1]$, as desired. □

4. Algorithm and simulation

In this subsection we present some numerical simulations obtained by discretizing the above ODE. The algorithm consists in discretizing (23) by an explicit Euler scheme (notice that one could also use some high order method for the ODEs). Let h be the step size and set $\varphi(0)=\varphi_w$ the solution of a 2 marginal problem with cost w, then the φ can be defined inductively as follows.

(1) Let $\varphi^{(k)}$ the solution at step k, then compute

$$D^{(k)}:=D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi^{(k)},kh),\quad b^{(k)}:=-\frac{\partial}{\partial\varepsilon}\nabla_\varphi\tilde{\Phi}(\varphi^{(k)},kh).$$

(2) Solve the linear system $D^{(k)}z=b^{(k)}$. We denote by $z^\star$ the solution.
(3) Update the potential by setting

$$\varphi^{(k+1)}=\varphi^{(k)}+hz^\star.$$

Notice that by the regularity we have proved above, we can conclude that the Euler scheme converges linearly. Moreover, the uniform error between the discretized solution obtained via the scheme and the solution to the ODE is $O(h)$. In Figure (left plot) 1 we plot the convergence order for the Euler scheme described above. The error is computed with respect to the solution to (16) computed via a gradient descent method with backtracking. Notice that the regularity of the objective function and the boundedness of the Hessian guarantee the convergence of the method. For these simulations we have taken $m=3$, the uniform measure on $[0,1]$ uniformily discretized with 100 gridpoints and the pairwise interaction $w(x,y)=-\log(0.1+|x-y|)$. Moreover, since the RHS in (23) is regular one can try to apply an high order scheme to solve the Cauchy problem. In figure (right plot)1 we compare the convergence of the Euler method and a Runge-Kutta of order 3; notice that with 100 time steps the RK method converges to a solution with an error of order 10^{-5} and by an estimation of the slope of the two lines we obtain 3 and 1.16, respectively for RK and Euler (as expected). In Figure 3 we show the numerical result obtained with $\eta=0.006$, $h=1/1000$, $m=3$,

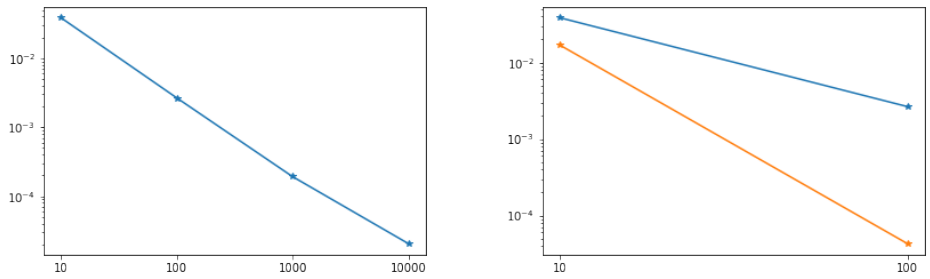

FIGURE 1. Convergence order for the ODE approach. Left: Linear convergence for an explicit Euler scheme. Right: comparison between explicit Euler (blue line) and an explicit Runge-Kutta (red line) of order 3

	ODE approach	Sinkhorn
relative error	0.0097108	0.0097085
iterations	100	820

TABLE 1. Comparison between the ODE approach and Sinkhorn

the uniform measure on $[0,1]$ uniformily discretized with 100 gridpoints and the pairwise interaction $w(x,y) = -\log(0.1 + |x-y|)$. Notice that since we have developed our continuation method by the entropic regularization of optimal transport, we can easily reconstruct the optimal plan at each k by using the potential $\varphi^{(k)}$. In Figure 2 we have compared the potential computed by the ODE with the one obtained by solving the regularized multi-marginal problem via Sinkhorn (at same η). Since in this case the optimal solution to the unregularised dual problem (as well as the primal) can be explicitly computed we compare the relative error, that is $\frac{\|\varphi - \varphi_{exact}\|_\infty}{\|\varphi_{exact}\|_\infty}$, between the ODE approach (in this case we have used an 8th order Runge-Kutta) and Sinkhorn, with respect to the number of iteration. By looking at Table 4, it is clear that both methods achieve the same relative error, but the number of iterations to reach it is smaller for the ODE approach. In Figures 2-4 we have kept the same data as before, but we have chose the negative harmonic cost, that is $w(x,y) = -|x-y|^2$. We highlight that the solution at $\varepsilon = 0$ is $-Id$ and then the final coupling is supported, as expected, on the hyperplane $x + y + z = 1.5$.

Acknowledgement. B.P. is pleased to acknowledge the support of National Sciences and Engineering Research Council of Canada Discovery Grant number 04658-2018. He is also grateful for the kind hospitality at

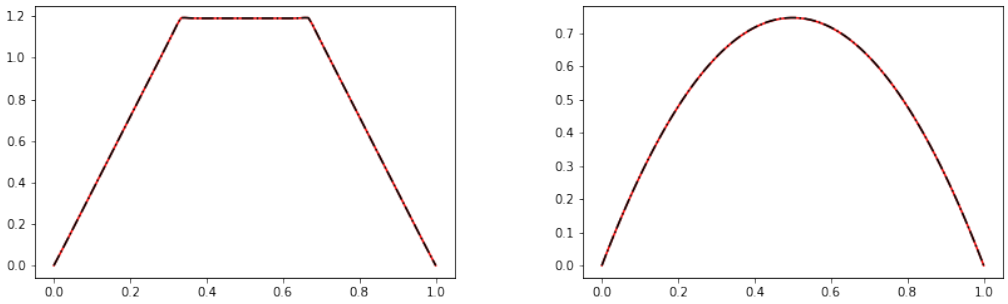

FIGURE 2. Optimal potential computed via Sinkhorn (red line). Potential computed via the ODE (black dot-dashed line). Left: Log cost. Right: Negative Harmonic cost

the Institut de Mathématiques d'Orsay, Université Paris-Saclay during his stay in November of 2019 as a missionaire scientifique invité, when this work was partially completed. L.N. was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH and from the CNRS PEPS JCJC (2022). The authors would like to thank Guillaume Carlier for all the fruitful discussions which lead to Lemma 8. A preliminary version of the results in this paper appeared in the arxiv preprint version of our paper [41], posted on January 3, 2022. Prior to that paper being accepted for publication on *Applied Mathematics and Optimization*, these results were removed and they have since been refined and expanded to form the present manuscript.

APPENDIX A. SECOND DERIVATIVES OF $\tilde{\Phi}$

In this appendix we detail the second derivatives of $\tilde{\Phi}$ with respect to φ and ε. Let us consider firstly the term (20)

$$\frac{\partial}{\partial \varepsilon} \nabla_{\varphi} \Phi(\varphi, \varepsilon),$$

notice that $\nabla_{\varphi}\Phi(\varphi,\varepsilon)$ it is a composition of an exponential with a linear function in ε, meaning that it is differentiable with respect to ε. We obtain then the following

$$\frac{\partial}{\partial \varepsilon}(\nabla_{\varphi}\Phi(\varphi,\varepsilon))_{\varphi_z} = -\frac{\exp(\varphi_z/\eta)}{\eta} \sum_{\overline{x}\in X^{m-1}} \partial_{\varepsilon}(c_{\varepsilon}(z,\overline{x})) \exp\left(\frac{\sum_{i=1}^{m-1}\varphi_{x^i} - c_{\varepsilon}(z,\overline{x})}{\eta}\right),$$

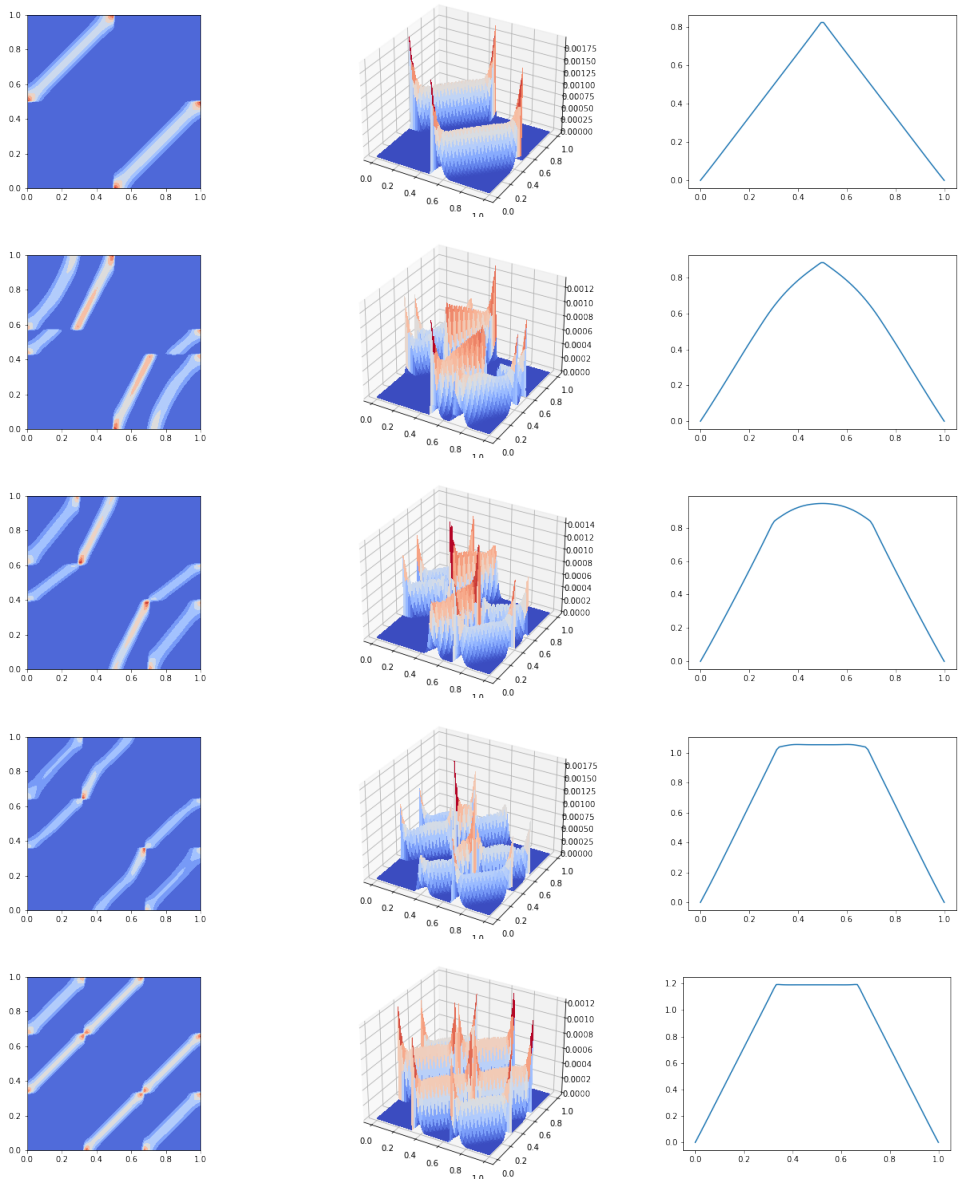

FIGURE 3. (Log cost) Left: support of the coupling $\gamma^{\varepsilon}_{1,2}$. Center: surface of the coupling $\gamma^{\varepsilon}_{1,2}$. Right: potential $\varphi(\varepsilon)$. For $\varepsilon = 0, 0.25, 0.5, 0.75, 1$

$$\frac{\partial}{\partial\varepsilon}\nabla_{\varphi}\Phi(\varphi(\varepsilon),\varepsilon) := -\frac{1}{\eta}\exp(\varphi/\eta)\rho\Bigg(\sum_{(x,\overline{y})\in X^{m-1}}\partial_{\varepsilon}c_{\varepsilon}(x,z,\overline{y})\exp\left(\frac{\sum_{i=3}^{m}\varphi_{y^i}-c_{\varepsilon}(x,z,\overline{y})}{\eta}\right)\tilde{\rho}_{\overline{y}}\overline{\rho}_x$$

$$+\sum_{(x,\overline{y})\in X^{m-1}}\exp\left(\frac{\sum_{i=3}^{m}\varphi_{y^i}-c_{\varepsilon}(x,z,\overline{y})}{\eta}\right)\tilde{\rho}_{\overline{y}}\frac{\sum_{\overline{w}\in X^{m-1}}\partial_{\varepsilon}c_{\varepsilon}(x,\overline{w})\exp\left(\frac{\sum_{i=2}^{m}\varphi_{w^i}-c_{\varepsilon}(x,\overline{w})}{\eta}\right)\tilde{\rho}_{\overline{w}}}{\sum_{\overline{w}\in X^{m-1}}\exp\left(\frac{\sum_{i=2}^{m}\varphi_{w^i}-c_{\varepsilon}(x,\overline{w})}{\eta}\right)\tilde{\rho}_{\overline{w}}}\overline{\rho}_x\Bigg)$$

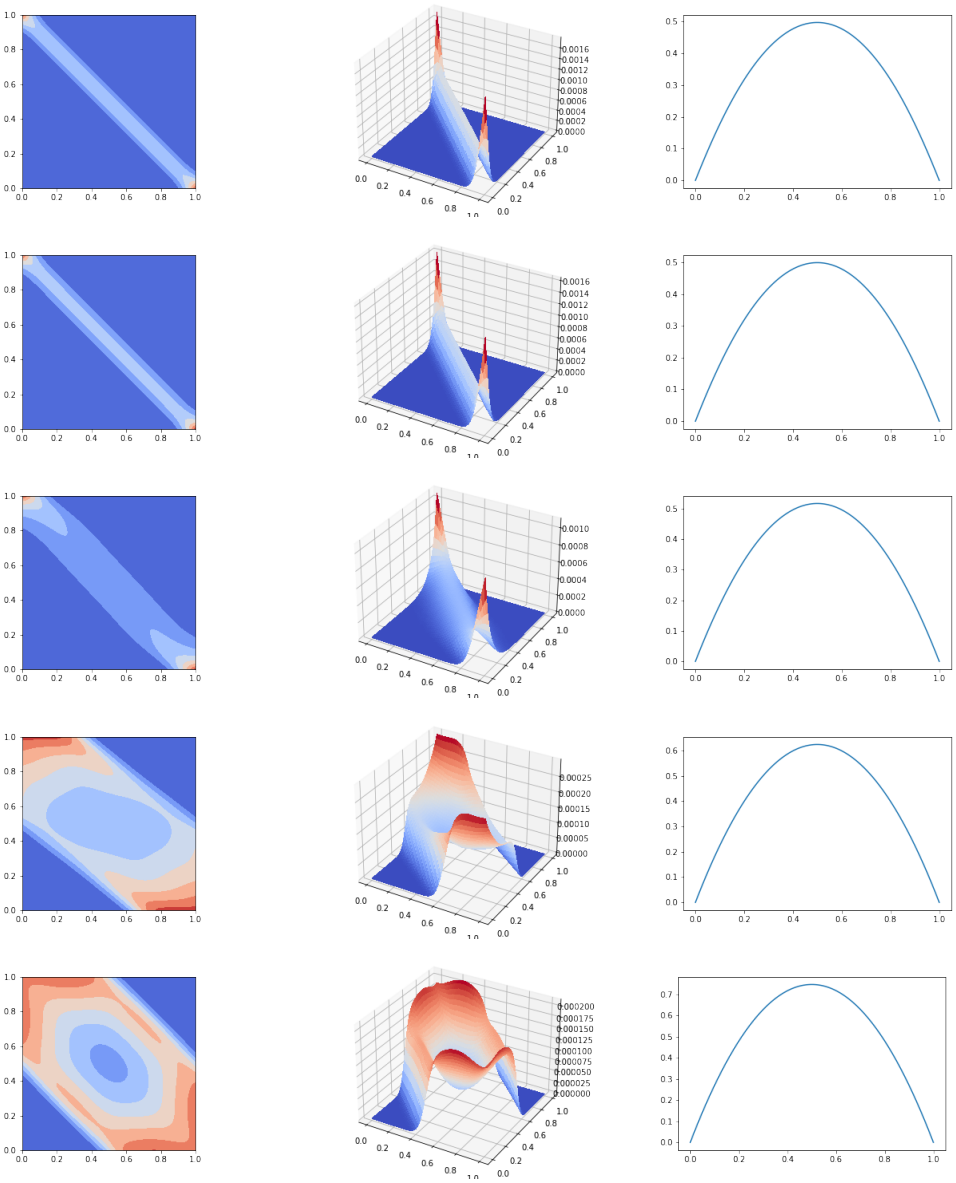

FIGURE 4. (Negative Harmonic cost) Left: support of the coupling $\gamma^{\varepsilon}_{1,2}$. Center: surface of the coupling $\gamma^{\varepsilon}_{1,2}$. Right: potential $\varphi(\varepsilon)$. For $\varepsilon = 0, 0.25, 0.5, 0.75, 1$

where

$$\partial_{\varepsilon} c_{\varepsilon}(z, \overline{x}) = \sum_{i,j=2, i \neq j}^{m-1} c_2(x^i, x^j).$$

Concerning the second derivative with respect to φ. it is again quite easy to see that Φ is twice differentiable, then we have

$$D^2_{\varphi,\varphi} \tilde{\Phi} = \frac{1}{\eta} \operatorname{diag}(I_1) + \frac{m-2}{\eta} (e^{\varphi/\eta} \rho) \otimes (e^{\varphi/\eta} \rho) I_2 - \frac{m-1}{\eta} \sum_{y \in X} I_3^y \otimes I_3^y \rho_y,$$

where

$$(I_1)_z = e^{\varphi_z/\eta}\rho_z \sum_y \sum_{\overline{x}\in X^{m-2}} \exp\left(\frac{\sum_{i=3}^m \varphi_{x^i} - c_\varepsilon(y,z,\overline{x})}{\eta}\right)(\otimes^{m-2}\rho)_{\overline{x}}\overline{\rho}_y,$$

$$(I_2)_{z,w} = \sum_y \sum_{\overline{x}\in X^{m-3}} \exp\left(\frac{\sum_{i=4}^m \varphi_{x^i} - c_\varepsilon(y,z,w,\overline{x})}{\eta}\right)(\otimes^{m-3}\rho)_{\overline{x}}\overline{\rho}_y,$$

$$(I_3^y)_z = \frac{e^{\varphi_z/\eta}\rho_z \sum_{\overline{x}\in X^{m-2}} \exp\left(\frac{\sum_{i=3}^m \varphi_{x^i} - c_\varepsilon(y,z,\overline{x})}{\eta}\right)(\otimes^{m-2}\rho)_{\overline{x}}}{\sum_{\overline{x}\in X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi_{x^i} - c_\varepsilon(y,\overline{x})}{\eta}\right)\tilde{\rho}_{\overline{x}}}$$

References

[1] Isabelle Abraham, Romain Abraham, Maıtine Bergounioux, and Guillaume Carlier. Tomographic reconstruction from a few views: a multi-marginal optimal transport approach. *Applied Mathematics & Optimization*, 75(1):55–73, 2017.

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

[3] Aurélien Alfonsi, Rafaël Coyaud, and Virginie Ehrlacher. Constrained overdamped langevin dynamics for symmetric multimarginal optimal transportation. *Mathematical Models and Methods in Applied Sciences*, 32(03):403–455, 2022.

[4] Aurélien Alfonsi, Rafaël Coyaud, Virginie Ehrlacher, and Damiano Lombardi. Approximation of optimal transport problems with marginal moments constraints. *Mathematics of Computation*, 90(328):689–737, 2021.

[5] Mathias Beiglböck, Pierre Henry-Labordere, and Friedrich Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[6] J.-D. Benamou, G. Carlier, and L. Nenna. A numerical method to solve multi-marginal optimal transport problems with coulomb cost. In *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 577–601. Springer International Publishing, 2016.

[7] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[8] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. Generalized incompressible flows, multi-marginal transport and sinkhorn algorithm. *to appear on Numerische Mathematik*, 2018.

[9] Jérémie Bigot and Thierry Klein. Characterization of barycenters in the wasserstein space by averaging optimal transport maps. *ESAIM: Probability and Statistics*, 22:35–57, 2018.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[11] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[12] Giuseppe Buttazzo, Thierry Champion, and Luigi De Pascale. Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Applied Mathematics & Optimization*, 78(1):185–200, 2018.

[13] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Physical Review A*, 85(6):062502, 2012.
[14] G. Carlier. On a class of multidimensional optimal transportation problems. *J. Convex Anal.*, 10(2):517–529, 2003.
[15] G. Carlier, A. Galichon, and F. Santambrogio. From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM J. Math. Anal.*, 41(6):2554–2576, 2009/10.
[16] Guillaume Carlier. On the linear convergence of the multi-marginal sinkhorn algorithm. 2021.
[17] Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. Vector quantile regression: an optimal transport approach. *The Annals of Statistics*, 44(3):1165–1192, 2016.
[18] Guillaume Carlier and Ivar Ekeland. Matching for teams. *Economic theory*, 42(2):397–418, 2010.
[19] Guillaume Carlier and Bruno Nazaret. Optimal transportation for the determinant. *ESAIM Control Optim. Calc. Var.*, 14(4):678–698, 2008.
[20] Maria Colombo, Luigi De Pascale, and Simone Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canad. J. Math.*, 67(2):350–368, 2015.
[21] Maria Colombo and Federico Stra. Counterexamples in multimarginal optimal transport with Coulomb cost and spherically symmetric data. *Math. Models Methods Appl. Sci.*, 26(6):1025–1049, 2016.
[22] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Density functional theory and optimal transportation with coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.
[23] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NIPS) 26*, pages 2292–2300, 2013.
[24] Simone Di Marino, Augusto Gerolin, and Luca Nenna. Optimal transportation theory with repulsive costs. In *Topological optimization and optimal transport*, volume 17 of *Radon Ser. Comput. Appl. Math.*, pages 204–256. De Gruyter, Berlin, 2017.
[25] Yan Dolinsky and H Mete Soner. Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1):391–427, 2014.
[26] Yan Dolinsky and H Mete Soner. Robust hedging with proportional transaction costs. *Finance and Stochastics*, 18(2):327–347, 2014.
[27] Hamza Ennaji, Quentin Mérigot, Luca Nenna, and Brendan Pass. Robust risk management via multi-marginal optimal transport, 2022.
[28] Gero Friesecke and Maximilian Penka. The gencol algorithm for high-dimensional optimal transport: general formulation and application to barycenters and wasserstein splines. *arXiv preprint arXiv:2209.09081*, 2022.
[29] Gero Friesecke, Andreas S Schulz, and Daniela Vögler. Genetic column generation: Fast computation of high-dimensional multimarginal optimal transport problems. *SIAM Journal on Scientific Computing*, 44(3):A1632–A1654, 2022.
[30] A. Galichon and B. Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. Technical report, Preprint SSRN-1487307, 2009.
[31] Wilfrid Gangbo and Andrzej Swiech. Optimal maps for the multidimensional monge-kantorovich problem. *Communications on pure and applied mathematics*, 51(1):23–45, 1998.

[32] Isabel Haasler, Rahul Singh, Qinsheng Zhang, Johan Karlsson, and Yongxin Chen. Multi-marginal optimal transport and probabilistic graphical models. *IEEE Transactions on Information Theory*, 67(7):4647–4668, 2021.
[33] Henri Heinich. Problème de Monge pour n probabilités. *C. R. Math. Acad. Sci. Paris*, 334(9):793–795, 2002.
[34] Yuehaw Khoo, Lin Lin, Michael Lindsey, and Lexing Ying. Semidefinite relaxation of multimarginal optimal transport for strictly correlated electrons in second quantization. *SIAM Journal on Scientific Computing*, 42(6):B1462–B1489, 2020.
[35] Yuehaw Khoo and Lexing Ying. Convex relaxation approaches for strictly correlated density functional theory. *SIAM Journal on Scientific Computing*, 41(4):B773–B795, 2019.
[36] Young-Heon Kim and Brendan Pass. A general condition for Monge solutions in the multi-marginal optimal transport problem. *SIAM J. Math. Anal.*, 46(2):1538–1550, 2014.
[37] Young-Heon Kim and Brendan Pass. Multi-marginal optimal transport on Riemannian manifolds. *Amer. J. Math.*, 137(4):1045–1060, 2015.
[38] Quentin Mérigot and Jean-Marie Mirebeau. Minimal geodesics along volume-preserving maps, through semidiscrete optimal transport. *SIAM Journal on Numerical Analysis*, 54(6):3465–3492, 2016.
[39] Abbas Moameni and Brendan Pass. Solutions to multi-marginal optimal transport problems concentrated on several graphs. *ESAIM Control Optim. Calc. Var.*, 23(2):551–567, 2017.
[40] Luca Nenna. *Numerical methods for multi-marginal optimal transportation*. PhD thesis, PSL Research University, 2016.
[41] Luca Nenna and Brendan Pass. Transport type metrics on the space of probability measures involving singular base measures. *arXiv preprint arXiv:2201.00875*, 2022.
[42] Brendan Pass. Uniqueness and Monge solutions in the multimarginal optimal transportation problem. *SIAM J. Math. Anal.*, 43(6):2758–2775, 2011.
[43] Brendan Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calc. Var. Partial Differential Equations*, 43(3-4):529–536, 2012.
[44] Brendan Pass. Multi-marginal optimal transport: theory and applications. *ESAIM Math. Model. Numer. Anal.*, 49(6):1771–1790, 2015.
[45] Brendan Pass and Adolfo Vargas-Jiménez. Monge solutions and uniqueness in multi-marginal optimal transport: weaker conditions on the cost, stronger conditions on the marginals. *arXiv preprint arXiv:2202.06783*, 2022.
[46] Brendan Pass and Adolfo Vargas-Jiménez. Monge solutions and uniqueness in multi-marginal optimal transport via graph theory. *arXiv preprint https://arxiv.org/abs/2104.09488*, 2021.
[47] Brendan W. Pass and Adolfo Vargas-Jiménez. Multi-marginal optimal transportation problem for cyclic costs. *SIAM J. Math. Anal.*, 53(4):4386–4400, 2021.
[48] Gabriel Peyré and Marco Cuturi. Computational optimal transport. Technical report, 2017.
[49] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.
[50] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[51] Nicolas Garcia Trillos, Matt Jacobs, and Jakwang Kim. The multimarginal optimal transport formulation of adversarial multiclass classification, 2022.
[52] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.
[53] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France.
Email address: luca.nenna@universite-paris-saclay.fr

Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1
Email address: pass@ualberta.ca