

Optimal Partial Transport Problem with Lagrangian costs

Noureddine Igbida, van Thanh Nguyen

▶ **To cite this version:**

Noureddine Igbida, van Thanh Nguyen. Optimal Partial Transport Problem with Lagrangian costs. 2017. hal-01552075

HAL Id: hal-01552075
https://hal.science/hal-01552075

Preprint submitted on 30 Jun 2017

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.

Optimal Partial Transport Problem with Lagrangian costs

Noureddine Igbida[†] and Van Thanh Nguyen[‡]

Institut de recherche XLIM-DMI, UMR-CNRS 6172
Faculté des Sciences et Techniques
Université de Limoges, France

April 2017

Abstract. We introduce a dual dynamical formulation for the optimal partial transport problem with Lagrangian costs

$$c_L(x,y) := \inf_{\xi \in Lip([0,1];\mathbb{R}^N)} \left\{ \int_0^1 L(\xi(t), \dot{\xi}(t)) \mathrm{d}t : \xi(0) = x, \xi(1) = y \right\}$$

basing on a constrained Hamilton–Jacobi equation. Optimality condition is given that takes the form of a system of PDEs in some way similar to constrained Mean Field Games. The equivalent formulations are then used to give numerical approximations to the optimal partial transport problem via augmented Lagrangian methods. One of advantages is that the approach requires only values of L and does not need to evaluate $c_L(x,y)$, for each pair of endpoints x and y, which comes from a variational problem. This method also provides at the same time optimal active submeasures and the associated optimal transportation.

1. Introduction

Optimal transport deals with the problem to find the optimal way to move materials from a given source to a desired target in such a way to minimize the work. The problem was first proposed and studied by G. Monge in 1781 and then L. Kantorovich made fundamental contributions to the problem in the 1940s by relaxing the problem into a linear one. Since the end of the eighties, this subject has been investigated under various viewpoints with many surprising applications in partial differential equations (PDEs), differential geometry, image processing and many other areas. These cultivate a huge study with the development of intuitions and concepts for the problem from both theoretical and numerical aspects. The problem has got a lot of

Key words and phrases. Optimal transport, optimal partial transport, Fenchel–Rockafellar duality, augmented Lagrangian method.

[†]E-mail : noureddine.igbida@unilim.fr .

[‡]E-mail : van-thanh.nguyen@unilim.fr.

attention and has been also generalized in different directions. For more details on the optimal transport problem, we refer to the pedagogical books **[30]**, **[31]**, **[1]** and **[29]**.

Among generalizations of the optimal transport problem, we are interested in the so-called *optimal partial transport* which aims to study the case where only a part of the commodity (respectively, consumer demand) of total mass $\mathbf{m}$ needs to be transported (respectively, fulfilled). More precisely, let $\mu, \nu \in \mathcal{M}_b^+(\mathbb{R}^N)$ be finite Radon measures and $c : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow [0, +\infty)$ be a measurable cost function. For a prescribed total mass $\mathbf{m} \in [0, \mathbf{m}_{\max}]$ with $\mathbf{m}_{\max} := \min\left\{\mu(\mathbb{R}^N), \nu(\mathbb{R}^N)\right\}$, the optimal partial transport problem (or partial Monge-Kantorovich problem, PMK for short) reads as follows

$$\min\left\{\mathcal{K}(\gamma) := \int\limits_{\mathbb{R}^N\times\mathbb{R}^N} c(x,y)\mathrm{d}\gamma : \gamma \in \pi_{\mathbf{m}}(\mu,\nu)\right\}, \tag{1.1}$$

where

$$\pi_{\mathbf{m}}(\mu,\nu) := \left\{\gamma \in \mathcal{M}_b^+(\mathbb{R}^N \times \mathbb{R}^N) : \ \pi_x\#\gamma \le \mu, \ \pi_y\#\gamma \le \nu, \ \gamma(\mathbb{R}^N \times \mathbb{R}^N) = \mathbf{m}\right\}.$$

Here, $\pi_x\#\gamma$ and $\pi_y\#\gamma$ are marginals of γ. This generalized problem brings out new unknown quantities $\rho_0 := \pi_x\#\gamma$, $\rho_1 := \pi_y\#\gamma$ called active submeasures. Here ρ_0 and ρ_1 are the sites where the commodity is taken and the consumer demand is fulfilled, respectively. In the case where μ and ν have the same mass (i.e., $\mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$) and the desired total mass $\mathbf{m} := \mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$, the problem (1.1) is nothing else the standard optimal transport problem.

The optimal partial transport problem (1.1) was first studied theoretically by Caffarelli and McCann **[9]** (see also Figalli **[17]**) with results on the existence, uniqueness and regularity of optimal active submeasures for the case where $c(x,y) = |x-y|^2$. We also refer to the papers **[11, 13, 23]** for the regularity. On the other hand, concerning numerical approximations, the authors in **[2]** studied numerically the case $c(x,y) = |x-y|$ via an approximated PDE and Raviart–Thomas finite elements. Recently, Benamou *et al.* **[8]** introduced a general numerical framework to approximate solutions of linear programs related to optimal transport such as barycenters in Wasserstein space, multi-marginal optimal transport, optimal partial transport and optimal transport with capacity constraints. Their idea is based on an entropic regularization of the initial linear programs and Bregman–Dykstra iterations. This is a static approach to optimal transport-type problems. In this direction, we also refer to the very recent paper of Chizat *et al.* **[12]**. These approaches need to use (approximated) values of the ground cost function $c_L(x,y)$ in the definition (1.2) below.

In this paper, we are interested in the optimal partial transport problem with Lagrangian costs $c(x,y) := c_L(x,y)$, where

$$c_L(x,y) := \inf_{\xi}\left\{\int\limits_0^1 L(\xi(t),\dot\xi(t))\mathrm{d}t : \xi(0) = x,\, \xi(1) = y,\, \xi \in Lip([0,1];\Omega)\right\} \tag{1.2}$$

with a Lipschitz domain $\Omega \subset \mathbb{R}^N$. The case where $L(x,.)$ is convex, 1-homogeneous with a linear growth was treated in the recent paper **[22]**. In this case, c_L turns out to be a distance which allows to reduce the dimension of the Kantorovich-type dual problem (by ignoring the

time variable). The present article concerns the situation where $L(x,.)$ has a superlinear growth. Our main aim is to develop rigorously the variational approach to provide equivalent dynamical formulations and use them to supply numerical approximations via augmented Lagrangian methods. For the uniqueness, using basically the idea of **[26]**, we establish the uniqueness of optimal active submeasures in the case where the densities are absolutely continuous.

To introduce and comment our main results, let us take a while to focus on the typical situation where the cost is given by

$$L(x,z) := k(x)\,\frac{|z|^q}{q} \quad \text{for any } (x,z) \in \mathbb{R}^N \times \mathbb{R}^N \tag{1.3}$$

with $q > 1$ and k being a (positive) continuous function. Recall here that if $k \equiv 1$ and $q = 2$, the cost function c_L corresponds to the quadratic case:

$$c_L(x,y) = |y-x|^2 \quad \text{for any } x, y \in \mathbb{R}^N.$$

This is more or less the most studied case in the literature (cf. **[9, 11, 13, 17, 23]**). However, let us mention here that our approach is variational and goes after our program of studying the optimal partial transportation from the theoretical and numerical points of view (cf. **[21]** and **[22]**). To begin with, it is not difficult to see that using standard results concerning the Eulerian formulation of the optimal mass transport problem in the balanced case, i.e. equal mass for the source and the target, an Eulerian formulation associated with the problem (1.1)-(1.3) can be given by minimizing

$$\iint\limits_{Q} k(x)\,\frac{|\upsilon(t,x)|^q}{q}\,\mathrm{d}\rho(t,x) \tag{1.4}$$

among all the couples $(\rho,\upsilon) \in \mathcal{M}_b^+(Q) \times L^1_\rho(Q)^N$ satisfying the continuity equation

$$\partial_t \rho + \nabla \cdot (\upsilon\rho) = 0 \quad \text{in } Q := [0,1] \times \mathbb{R}^N \tag{1.5}$$

in a weak sense with $\rho(0) \le \mu$ and $\rho(1) \le \nu$ and $\rho(0)(\mathbb{R}^N) = \mathbf{m}$. However, to use the augmented Lagrangian method to solve numerically (1.1), we will prove rigorously that in fact the minimization problem of the type (1.4)-(1.5) is the Fenchel–Rockafellar dual of a new dual problem to (1.1). Indeed, using the general duality result on the optimal partial transportation of **[21]**, we prove that a dynamical formulation of the dual problem of (1.1) consists in maximizing

$$\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \tag{1.6}$$

among the couples $(\lambda,u) \in [0,\infty) \times Lip(Q)$, where u satisfies the following constrained Hamilton–Jacobi equation

$$\begin{cases} \partial_t u(t,x) + k^{-\alpha}(x)\dfrac{|\nabla_x u(t,x)|^{q'}}{q'} \le 0 \ \text{ for a.e. } (t,x) \in Q \\ \\ -\lambda \le u(0,x), \ \ u(1,x) \le 0 \ \ \forall x \in \mathbb{R}^N. \end{cases} \tag{1.7}$$

Here $q' := \dfrac{q}{q-1}$ denotes the usual conjugate of q and $\alpha = \dfrac{q'}{q}$. Then, even if the regularity of the solutions here creates an obstruction to the application of the general theory, we will prove that the minimization problem (1.4) remains to be the Fenchel–Rockafellar dual of the maximization problem (1.6)-(1.7). Using these equivalents, we overbalance the problem into the scope of augmented Lagrangian methods and give numerical approximations to the optimal partial transport problem. In particular, we will see that this approach does not need to evaluate $c_L(x,y)$, for each pair of endpoints x and y, but requires only some values of L. Also, the method provides at the same time optimal active submeasures and the associated optimal transportation.

In addition, let us mention that the Fenchel–Rockafellar duality between the maximization problem (1.6) and the minimization problem (1.4) brings up (as optimality conditions) a new type of "constrained" Mean Field Game (MFG) system. For the particular case (1.3), this system aims to find $(\rho, \upsilon) \in \mathcal{M}_b^+(Q) \times L^1_\rho(Q)^N$ satisfying both the usual MFG system associated with the cost (1.1)-(1.3):

$$\begin{cases} \partial_t u(t,x) + k^{-\alpha}(x)\dfrac{|\nabla_x u(t,x)|^{q'}}{q'} \leq 0 & \text{for a.e. } (t,x) \text{ in } Q \\ \\ \partial_t \rho + \nabla \cdot (\upsilon \rho) = 0 & \text{in } (0,1)\times\mathbb{R}^N \\ \\ \upsilon(t,x) = k(x)^{-\alpha}|\nabla_x u(t,x)|^{q'-2}\nabla_x u(t,x) & \rho\text{-a.e. } (t,x) \text{ in } Q \end{cases} \tag{1.8}$$

and the following non-standard initial boundary values:

$$\rho(0) - \mu \in \partial \mathbb{I}_{[-\lambda,+\infty)}(u(0,.)) \quad \text{and} \quad \nu - \rho(1) \in \partial \mathbb{I}_{(-\infty,0]}(u(1,.)). \tag{1.9}$$

In other words, these initial boundary values may be written as: $-\lambda \leq u(0,.), u(1,.) \leq 0$, $\rho(0) \leq \mu$, $\rho(1) \leq \nu$, $\rho(0) = \mu$ in the set $[u(0,.) > -\lambda]$ and $\rho(1) = \nu$ in the set $[u(1,.) < 0]$. In the system (1.8), λ is an arbitrary non-negative parameter and the couple $(\rho(0), \rho(1)) \in \mathcal{M}_b^+(\mathbb{R}^N) \times \mathcal{M}_b^+(\mathbb{R}^N)$ is unknown. Once the system is solved with the optimal λ for (1.4), the couple $(\rho(0), \rho(1))$ gives the optimal active submeasures and ρ gives the optimal transportation.

Actually, for a given $\lambda \geq 0$, (1.8)-(1.9) is a new type of constrained MFG system. In this direction, one can see some variant of constrained MFG systems and their connection with the Mean Field Games under congestion effects in the paper **[28]**. However, let us mention that (1.8)-(1.9) is different from the class of MFG systems introduced in **[28]**. In particular, one sees that the constraints in (1.9) focus only the state at time $t = 0$ and $t = 1$. As to the constraints in **[28]**, they are maintained on all the trajectory for every time $t \in [0,1]$ to handle some kind of congestion.

On the other hand, one sees that taking $\mathbf{m} = \mu(\mathbb{R}^N)$, the optimal partial transportation problem (1.1) can be interpreted as the projection with respect to W_L-Wasserstein distance associated with c_L onto the set of non-negative measures less or equal to ν. Recall here that the case where $k \equiv 1$ and $q = 2$, this problem appears in the study of the congestion in the modeling of crowd motion (cf. **[25]**). So, it is possible to merge our numerical approximation algorithm into the splitting process of **[25]** for the study of the crowd motion with congestion.

At last, let us mention that the main difficulty in the study of the above variational approach of the problem (1.1) for general Lagrangian L remains in the regularity of the solutions of the

optimization problems like (1.4)-(1.5) and (1.6)-(1.7). To handle this difficulty we prove some new results concerning the approximation of the solutions of general constrained Hamilton-Jacobi equation like (1.7) by regular function. Moreover, we show how to use the notion of tangential gradient to study MFG system like (1.8)-(1.9) in the general case. In particular, when $\mathbf{m} = \mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$ and $L(x, v) = L(v)$ is independent of x, this MFG system reduces to the PDE as in the work of Jimenez [**24**].

The remainder of the article is organized as follows: In the next section, we begin with the notations and assumptions. Section 3 concerns the uniqueness of optimal active submeasures. Section 4 is devoted to the rigorously theoretical study of the problem with equivalent dual formulations that would be used later for numerical approximation. Besides these, an optimality condition is also given which recovers the Monge–Kantorovich equation as a particular case. The details of our numerical approximation are discussed in Section 5. The paper ends up by some numerical examples.

2. Notations and assumptions

Let X be a topological vector space and X^* be its dual. We use $\langle x^*, x\rangle$ or $\langle x, x^*\rangle$ to indicate the value $x^*(x)$ for $x \in X$ and $x^* \in X^*$. The scalar product in $\mathbb{R}^N$ is denoted by $x \cdot y$ or $\langle x, y\rangle$ for $x, y \in \mathbb{R}^N$. We denote by $\mathbb{I}_K$ the indicator function of K and by ∂G the subdifferential of convex function $G : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ in the sense of convex analysis.

To avoid unnecessary difficulties, in our theoretical study, we will work with $\Omega = \mathbb{R}^N$ in the definition (1.2) of c_L . Throughout this article, we drop the subscript L and write simply c instead of c_L.

Assumption (A): Assume that the Lagrangian $L : \mathbb{R}^N \times \mathbb{R}^N \to [0, +\infty)$ is continuous and satisfies:

- $L(x, .)$ is convex and $L(x, 0) = 0$ for each fixed $x \in \mathbb{R}^N$;
- (Superlinearity) for any $R > 0$, there exists a function $\theta_R : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\frac{\theta_R(t)}{t} \to +\infty \text{ as } t \to +\infty \text{ and } L(x, v) \geq \theta_R(|v|) \ \ \forall x \in B(0, R).$$

For example, the function $L(x, v) := k(x)\dfrac{|v|^q}{q}, q > 1$ satisfies the above assumption whenever k is (positive) continuous.

The convex conjugate H of L is defined by

$$H(x, p) := \sup_{v\in\mathbb{R}^N} \{\langle p, v\rangle - L(x, v)\} \quad \text{for any } \ x \in \mathbb{R}^N, p \in \mathbb{R}^N.$$

Note that, under the assumption (**A**) on L, the function $H(., .)$ is continuous in both variables and that $c(., .)$ is locally Lipschitz.

We set $Q := [0, 1] \times \mathbb{R}^N$. The usual derivatives of u are denoted by $\partial_t u, \nabla_x u$, and $\nabla_{t,x} u := (\partial_t u, \nabla_x u)$. The continuity equation $\partial_t \rho + \nabla \cdot (v\rho) = 0, \rho(0) = \mu, \rho(1) = \nu$ is understood in the

sense of distribution, i.e.

$$\begin{aligned}\iint\limits_{Q} \partial_t\phi \mathrm{d}\rho + \iint\limits_{Q} \nabla_x\phi\cdot\upsilon \mathrm{d}\rho &= \int\limits_{\mathbb{R}^N} \phi(1,.)\mathrm{d}\rho(1) - \int\limits_{\mathbb{R}^N} \phi(0,.)\mathrm{d}\rho(0)\\ &= \int\limits_{\mathbb{R}^N} \phi(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} \phi(0,.)\mathrm{d}\mu\end{aligned} \tag{2.1}$$

for any compactly supported smooth function $\phi \in C_c^\infty(Q)$. For short, we denote (2.1) by

$$-\mathrm{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \nu - \delta_0 \otimes \mu.$$

We say that (ρ_0, ρ_1) is a couple of *optimal active submeasures* if $\rho_0 = \pi_x\#\gamma$ and $\rho_1 = \pi_y\#\gamma$ for some optimal plan γ of the PMK problem (1.1).

3. Uniqueness of optimal active submeasures

The existence of optimal active submeasures follows easily from the direct method (see e.g. [**9**, **21**]). The present section concerns the uniqueness.

Theorem 3.1 (Uniqueness)**.** *Assume moreover that $L(x, v) = L(v)$ is independent of x and that $L(v) = 0$ if and only if $v = 0$. If $\mu, \nu \in L^1$ and $\mathbf{m} \in [\mu \wedge \nu(\mathbb{R}^N), \mathbf{m}_{\max}]$ then there is at most one couple of optimal active submeasures.*

The idea of the proof is based on the recent paper [**26**, Proposition 5.2]. For completeness, we give here an adaptation to our case.

Lemma 3.1. *Assume that L satisfies the assumption (**A**). Let (ρ_0, ρ_1) be a couple of optimal active submeasures and $\gamma \in \pi(\rho_0, \rho_1)$ be an optimal plan. If $(x^*, y^*) \in \mathrm{supp}(\gamma)$ then $\rho_0 = \mu$ a.e. on $B_c(y^*, R) := \{t \in \mathbb{R}^N : c(t, y^*) < R\}$ and $\rho_1 = \nu$ a.e. on $B_c(x^*, R) := \{w \in \mathbb{R}^N : c(x^*, w) < R)\}$, where $R := c(x^*, y^*) = c_L(x^*, y^*)$.*

Proof. We prove that $\rho_1 = \nu$ a.e. on $B_c(x^*, R)$. If the conclusion is not true then there exists a compact set $K \Subset B_c(x^*, R)$ with a positive Lebesgue measure such that $\rho_1 < \nu$ a.e. on K. The proof consists in the construction of a better plan $\tilde{\gamma}$. Since $(x^*, y^*) \in \mathrm{supp}(\gamma)$, we have

$$0 < \gamma(B(x^*, r) \times B(y^*, r)) \leq \int\limits_{B(y^*,r)} \nu \mathrm{d}x \to 0 \text{ as } r \to 0,$$

where $B(x, r)$ is the ball w.r.t. the Euclidean norm. Now, geometrically speaking, instead of transporting mass from x^* to around y^*, we can give more mass on K. To be more precise, we construct a new plan $\tilde{\gamma}$ as follows

$$\tilde{\gamma} := \gamma - \gamma\llcorner_{B(x^*,r)\times B(y^*,r)} + \eta$$

with

$$\eta := \frac{\pi_x\#(\gamma\llcorner_{B(x^*,r)\times B(y^*,r)}) \otimes (\nu - \rho_1)\llcorner_K}{\int\limits_K (\nu - \rho_1)\mathrm{d}x}.$$

Then $\pi_x\#\tilde{\gamma} = \rho_0$ and

$$\pi_y\#\tilde{\gamma} \le \pi_y\#\gamma + \frac{\gamma(B(x^*,r)\times B(y^*,r))}{\int\limits_K (\nu-\rho_1)\mathrm{d}x}(\nu-\rho_1)\llcorner_K \le \rho_1 + (\nu-\rho_1)\llcorner_K \le \nu$$

for all sufficiently small r. It follows that $\tilde{\gamma} \in \pi_{\mathbf{m}}(\mu,\nu)$. Furthermore, we have

$$\begin{aligned}
&\int c(x,y)\mathrm{d}\tilde{\gamma}\\
&= \int c(x,y)\mathrm{d}\gamma - \int\limits_{B(x^*,r)\times B(y^*,r)} c(x,y)\mathrm{d}\gamma + \int\limits_{B(x^*,r)\times K} c(x,y)\mathrm{d}\eta\\
&\le \int c(x,y)\mathrm{d}\gamma - \left(\inf_{(x,y)\in B(x^*,r)\times B(y^*,r)} c(x,y) - \sup_{(x,y)\in B(x^*,r)\times K} c(x,y)\right)\gamma(B(x^*,r)\times B(y^*,r)).\\
&< \int c(x,y)\mathrm{d}\gamma \ \text{ for small } \ r,
\end{aligned}$$

where we used the fact that

$$\left(\inf_{(x,y)\in B(x^*,r)\times B(y^*,r)} c(x,y) - \sup_{(x,y)\in B(x^*,r)\times K} c(x,y)\right) > 0 \ \text{ for small } \ r.$$

This holds because of the definition of K and the continuity of c. □

The next lemma provides an expression for optimal active submeasures[1].

Lemma 3.2. *Under the assumptions of Theorem 3.1, let (ρ_0,ρ_1) be couple of optimal active submeasures. Then*

$$\rho_0 = \chi_{B_0^c}\mu, \quad \rho_1 = \chi_{B_1^c}\nu$$

for some measurable sets B_0, B_1.

Proof. Since $L(x,v) = L(v)$, we get $c(x,y) := c_L(x,y) = L(y-x)$. Thus $c(x,y) = 0$ if and only if $x = y$. This implies that the common mass $\mu\wedge\nu$ must belong to optimal active submeasures, i.e., $\mu\wedge\nu \le \rho_0$ and $\mu\wedge\nu \le \rho_1$. So without loss of generality, we can assume that the initial measures μ and ν are disjoint, i.e., $\mu\wedge\nu = 0$. Now, let us define

$$B_0 := \mathrm{Leb}(\mu)\cap\mathrm{Leb}(\nu)\cap\mathrm{Leb}(\rho_0)\cap\{\rho_0<\mu\}^{(1)} \ \text{ and } \ B_1 := \mathrm{Leb}(\mu)\cap\mathrm{Leb}(\nu)\cap\mathrm{Leb}(\rho_1)\cap\{\rho_1<\nu\}^{(1)}.$$

Here, $\mathrm{Leb}(g)$ is the set of Lebesgue points of g and $A^{(1)}$ is the set of points of density 1 w.r.t. A. We see that

$$B_0^c = (\mathrm{Leb}(\mu)\cap\mathrm{Leb}(\nu)\cap\mathrm{Leb}(\rho_0))^c \cup (\{\rho_0<\mu\}^{(1)})^c = Z\cup\{\rho_0=\mu\},$$

with $\mathcal{L}^N(Z) = 0$. So $\rho_0 = \mu$ a.e. on B_0^c. Next, we show that $\rho_0 = 0$ on B_0. Indeed, if $\rho_0(x) > 0$ for some $x\in B_0$ then $x\in\mathrm{supp}(\rho_0)$. Hence there exists $y\in\mathrm{supp}(\rho_1)$ such that $(x,y)\in\mathrm{supp}(\gamma)$ for some optimal plan $\gamma\in\pi(\rho_0,\rho_1)$. Since $\mu\wedge\nu = 0$, we can take $y\neq x$ and thus $R := c(x,y) = L(y-x) > 0$. Since $B_c(y,R)$ is convex, it has the cone property, i.e. there

[1]Thanks to this expression, optimal active sumeasures are completely determined by their supports, called *active regions* as used in [**9**, **17**].

is a finite cone with vertex at x contained in $B_c(y,R)$. It follows that there exists a sequence of subsets of $B_c(y,R)$ shrinking to x nicely (see e.g. [**27**, Theorem 7.10]). Using Lemma 3.1, $\rho_0 = \mu$ a.e. on $B_c(y,R)$, we obtain $\rho_0(x) = \mu(x)$, which is impossible. Consequently, the proof of the expression $\rho_0 = \chi_{B_0^c}\mu$ is completed. In much the same way, we get $\rho_1 = \chi_{B_1^c}\nu$. □

Proof of Theorem 3.1. Let (ρ_0,ρ_1) and $(\tilde\rho_0,\tilde\rho_1)$ be couples of optimal active submeasures. By Lemma 3.2, we have $\rho_0 = \chi_{B_0^c}\mu$, $\rho_1 = \chi_{B_1^c}\nu$, $\tilde\rho_0 = \chi_{\tilde B_0^c}\mu$ and $\tilde\rho_1 = \chi_{\tilde B_1^c}\nu$. By the convexity of the total cost, we see that $\frac{1}{2}(\rho_0,\rho_1) + \frac{1}{2}(\tilde\rho_0,\tilde\rho_1)$ is also an optimal couple. If $(\rho_0,\rho_1) \neq (\tilde\rho_0,\tilde\rho_1)$ then $\frac{1}{2}(\rho_0,\rho_1) + \frac{1}{2}(\tilde\rho_0,\tilde\rho_1)$ does not admit any expression as in Lemma 3.2, a contradiction. □

Remark 3.2. (i) Following the proof, Theorem 3.1 is still true for any general cost c (not necessary to be of the form c_L) if we have the following properties:

- c is continuous.
- $c(x,y) = 0$ if and only if $x = y$.
- The balls w.r.t. c defined by $B_c(y,R) := \{t \in \mathbb{R}^N : c(t,y) < R\}$ and $B_c(x,R) := \{w \in \mathbb{R}^N : c(x,w) < R)\}$ are regular in the sense that given any point on the boundary of a ball, there exists a sequence of subsets of the ball which shrinks nicely to that point.

(ii) In the case where $L(x,v) = L(v)$ is strictly convex, Figalli [**17**] studied the strict convexity of the function that associates to each $\mathbf{m} \in (\mu \wedge \nu(\mathbb{R}^N), \mathbf{m}_{\max}]$ the total Monge–Kantorovich cost to deduce the uniqueness.
(iii) When $L(x,.)$ is positively 1-homogeneous, i.e., $L(x,t\xi) = tL(x,\xi)\ \forall x \in \mathbb{R}^N, \xi \in \mathbb{R}^N, t > 0$, the uniqueness can be obtained via PDE techniques applied to the so-called obstacle Monge–Kantorovich equation (see [**21**]).

4. Equivalent formulations

In the present section, under the general assumptions of Section 2, we introduce and study the equivalent formulations for the PMK problem of the type (1.4)-(1.5) and also (1.6)-(1.7).

4.1. Dual formulation.

We start with the following Kantorovich-type dual formulation.

Theorem 4.3. *Let $\mu,\nu \in \mathcal{M}_b^+(\mathbb{R}^N)$ be compactly supported and $\mathbf{m} \in \left[0, \min\left\{\mu(\mathbb{R}^N), \nu(\mathbb{R}^N)\right\}\right]$. Suppose that L satisfies the assumption (A). Then*

$$\begin{aligned}\min_{\gamma\in\Pi_{\mathbf{m}}(\mu,\nu)} \left\{ \mathcal{K}(\gamma) := \int_{\mathbb{R}^N\times\mathbb{R}^N} c(x,y)\mathrm{d}\gamma \right\} \\ = \max_{(\lambda,u)} \left\{ \int_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \ :\ \lambda \in \mathbb{R}^+,\ u \in \mathcal{K}_c^\lambda \right\},\end{aligned} \tag{4.1}$$

where

$$\begin{aligned}\mathcal{K}_c^\lambda := \Big\{ u \in Lip(Q) : \partial_t u(t,x) + H(x,\nabla_x u(t,x)) \le 0 \ \text{ for a.e. } \ (t,x) \in Q, \\ -\lambda \le u(0,x) \ \text{ and } \ u(1,x) \le 0 \ \ \forall x \in \mathbb{R}^N \Big\}.\end{aligned}$$

To prove Theorem 4.3, we recall below the Kantorovich duality for the PMK problem. This can be found in [**21**].

Theorem 4.4. *(cf. [21]) Let $\mu, \nu \in \mathcal{M}_b^+(\mathbb{R}^N)$ be compactly supported Radon measures and $\mathbf{m} \in [0, \mathbf{m}_{\max}]$. The PMK problem has a solution $\sigma^* \in \pi_{\mathbf{m}}(\mu,\nu)$ and the Kantorovich duality turns into*

$$\mathcal{K}(\sigma^*) = \max_{(\lambda,\phi,\psi)} \left\{ \int \phi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu + \lambda \mathbf{m} \ : \ \lambda \in \mathbb{R}^+,\ (\phi,\psi) \in \mathcal{S}_c^\lambda(\mu,\nu) \right\}, \tag{4.2}$$

where

$$\mathcal{S}_c^\lambda(\mu,\nu) := \left\{ (\phi,\psi) \in L^1_\mu(\mathbb{R}^N) \times L^1_\nu(\mathbb{R}^N) : \phi \le 0,\ \psi \le 0,\ \phi(x) + \psi(y) + \lambda \le c(x,y)\ \ \forall x, y \in \mathbb{R}^N \right\}.$$

Moreover, $\sigma \in \pi_{\mathbf{m}}(\mu,\nu)$ and $(\lambda,\phi,\psi) \in \mathbb{R}^+ \times \mathcal{S}_c^\lambda(\mu,\nu)$ are solutions, respectively, if and only if

$$\phi(x) = 0 \quad \text{for } (\mu - \pi_x \# \sigma)\text{-a.e. } x \in \mathbb{R}^N, \quad \psi(y) = 0 \quad \text{for } (\nu - \pi_y \# \sigma)\text{-a.e. } y \in \mathbb{R}^N$$
$$\text{and} \quad \phi(x) + \psi(y) + \lambda = c(x,y) \quad \text{for } \sigma\text{-a.e. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N.$$

Remark 4.5. The duality (4.2) can be rewritten as

$$\mathcal{K}(\sigma^*) = \max_{(\lambda,\phi,\psi)} \left\{ \int \psi \,\mathrm{d}\nu - \int \phi \,\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \ : \ \lambda \in \mathbb{R}^+,\ (\phi,\psi) \in \Phi_c^\lambda(\mu,\nu) \right\}, \tag{4.3}$$

where

$$\Phi_c^\lambda(\mu,\nu) := \left\{ (\phi,\psi) \in L^1_\mu(\mathbb{R}^N) \times L^1_\nu(\mathbb{R}^N) : -\lambda \le \phi,\ \ \psi \le 0,\ \psi(y) - \phi(x) \le c(x,y)\ \ \forall x, y \in \mathbb{R}^N \right\}.$$

The maximization problem is called the *dual partial Monge–Kantorovich* (DPMK) problem.

Coming back to our analysis, we note that if $u \in \mathcal{K}_c^\lambda$ and u is smooth then, using the definition of convex conjugate function H, we get

$$\partial_t u(t,x) + \langle \nabla_x u(t,x), v \rangle \le L(x,v) \ \ \forall (t,x) \in Q,\ v \in \mathbb{R}^N. \tag{4.4}$$

In general, for any $u \in \mathcal{K}_c^\lambda$, we can approximate u by smooth functions satisfying a similar estimate for (4.4). This is the content of the following lemma. Although we obtain here only the estimate at the limit, this is enough for later use.

Lemma 4.3. *Fix any $u \in \mathcal{K}_c^\lambda$. There exists a sequence of smooth functions u_ε such that u_ε converges uniformly to u on Q and*

$$\limsup_{\varepsilon \to 0} \left(\partial_t u_\varepsilon(t,x) + \langle \nabla_x u_\varepsilon(t,x), v \rangle \right) \le L(x,v) \ \ \forall (t,x) \in Q, v \in \mathbb{R}^N \tag{4.5}$$

and, for all $\xi \in Lip([0,1];\mathbb{R}^N)$,

$$\limsup_{\varepsilon \to 0} \int_0^1 \left(\partial_t u_\varepsilon(t,\xi(t)) + \langle \nabla_x u_\varepsilon(t,\xi(t)), \dot\xi(t) \rangle \right) \mathrm{d}t \le \int_0^1 L(\xi(t), \dot\xi(t)) \mathrm{d}t. \tag{4.6}$$

Proof. Let $\alpha_\varepsilon, \beta_\varepsilon$ be standard mollifiers on $\mathbb{R}$ and $\mathbb{R}^N$, respectively, such that

$$\mathrm{supp}(\alpha_\varepsilon) \subset [-\varepsilon, \varepsilon]. \tag{4.7}$$

Set $\eta_\varepsilon(t,x) := \alpha_\varepsilon(t)\beta_\varepsilon(x)$. Let $\tilde{u}$ be a Lipschitz extension of u on $\mathbb{R}\times\mathbb{R}^N$. By means of convolution in both time and spacial variables, let us define

$$\tilde{u}_\varepsilon := \eta_\varepsilon \star \tilde{u},$$

and

$$u_\varepsilon(t,x) := \tilde{u}_\varepsilon(\varepsilon + (1-2\varepsilon)t, (1-2\varepsilon)x) \ \text{ for any } \ (t,x) \in \mathbb{R}\times\mathbb{R}^N.$$

Let us show that u_ε satisfies all the requirements. First, since $\tilde{u}$ is Lipschitz, u_ε converges uniformly to u on Q. It remains to check the two inequalities (4.5) and (4.6). For all $t \in [0,1]$, using (4.7), we have

$$\begin{aligned} u_\varepsilon(t,x) &= \int_{\mathbb{R}}\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)\tilde{u}(s,y)\mathrm{d}y\,\mathrm{d}s \\ &= \int_0^1\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)u(s,y)\mathrm{d}y\,\mathrm{d}s. \end{aligned}$$

Fix any $v \in \mathbb{R}^N$, for all $t \in [0,1]$, we have

$$\begin{aligned} &\partial_t u_\varepsilon(t,x) + \langle v, \nabla u_\varepsilon(t,x)\rangle \\ &= (1-2\varepsilon)\int_0^1\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)\partial_s u(s,y)\,\mathrm{d}y\,\mathrm{d}s \\ &\quad + (1-2\varepsilon)\left\langle v, \int_0^1\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)\nabla_y u(s,y)\,\mathrm{d}y\,\mathrm{d}s\right\rangle \\ &= (1-2\varepsilon)\int_0^1\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)\left(\partial_s u(s,y) + \langle v, \nabla_y u(s,y)\rangle\right)\mathrm{d}y\,\mathrm{d}s \\ &\le (1-2\varepsilon)\int_0^1\int_{\mathbb{R}^N} \alpha_\varepsilon(\varepsilon + (1-2\varepsilon)t - s)\beta_\varepsilon((1-2\varepsilon)x - y)L(y,v)\,\mathrm{d}y\,\mathrm{d}s \\ &= (1-2\varepsilon)\int_{\mathbb{R}^N} \beta_\varepsilon((1-2\varepsilon)x - y)L(y,v)\,\mathrm{d}y. \end{aligned} \tag{4.8}$$

Letting $\varepsilon \to 0$, we obtain (4.5). Finally, let us fix any $\xi \in Lip([0,1];\mathbb{R}^N)$. Using (4.5) with $x = \xi(t)$, $v = \dot{\xi}(t)$, we have

$$\limsup_{\varepsilon\to 0}\left(\partial_t u_\varepsilon(t,\xi(t)) + \langle \nabla_x u_\varepsilon(t,\xi(t)), \dot{\xi}(t)\rangle\right) \le L(\xi(t),\dot{\xi}(t)) \ \text{ for a.e. } \ t \in [0,1]. \tag{4.9}$$

Recall that (the Reverse Fatou's Lemma) if there exists an integrable function g on a measure space (X, η) such that $g_\varepsilon \le g$ for all ε, then

$$\limsup_\varepsilon \int g_\varepsilon \mathrm{d}\eta \le \int \limsup_\varepsilon g_\varepsilon \mathrm{d}\eta.$$

In our case, on $X := [0,1]$ with the Lebesgue measure, the functions $g_\varepsilon(t) := \partial_t u_\varepsilon(t, \xi(t)) + \langle \nabla_x u_\varepsilon(t, \xi(t)), \dot\xi(t)\rangle$ for a.e. $t \in [0,1]$ are bounded by a common constant depending Lipschitz constants of u and of ξ. Applying the Reverse Fatou's Lemma and (4.9), we deduce that

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \int_0^1 \Big(\partial_t u_\varepsilon(t,\xi(t)) + \langle \nabla_x u_\varepsilon(t,\xi(t)), \dot\xi(t)\rangle\Big)\,\mathrm{d}t \\
&\le \int_0^1 \limsup_{\varepsilon\to 0} \Big(\partial_t u_\varepsilon(t,\xi(t)) + \langle \nabla_x u_\varepsilon(t,\xi(t)), \dot\xi(t)\rangle\Big)\,\mathrm{d}t \\
&\le \int_0^1 L(\xi(t), \dot\xi(t))\mathrm{d}t.
\end{aligned}$$

□

Note that we can do even better in the case where $L(x,v) = L(v)$ is independent of x. Indeed, from our argument (4.8), we can choose u_ε such that

$$\partial_t u_\varepsilon(t,x) + \langle \nabla_x u_\varepsilon(t,x), v\rangle \le L(x,v) \quad \forall (t,x) \in Q, v \in \mathbb{R}^N$$

without passing ε to 0.

Now, we are ready to prove the duality (4.1). We check directly that the maximization is less than the minimum in (4.1) with the help of Lemma 4.3. For the converse inequality, we make use of the theory of Hamilton–Jacobi equations.

Proof of Theorem 4.3. Fix any $u \in \mathcal{K}_c^\lambda$. Let u_ε be the sequence of smooth functions given in Lemma 4.3. Fix any $\xi \in Lip([0,1]; \mathbb{R}^N)$ such that $\xi(0) = x, \xi(1) = y$. By (4.6), we have

$$\begin{aligned}
u(1,y) - u(0,x) &= \lim_{\varepsilon\to 0} \left(u_\varepsilon(1,\xi(1)) - u_\varepsilon(0,\xi(0))\right) \\
&= \lim_{\varepsilon\to 0} \int_0^1 \Big(\partial_t u_\varepsilon(t,\xi(t)) + \langle \nabla_x u_\varepsilon(t,\xi(t)), \dot\xi(t)\rangle\Big)\,\mathrm{d}t \\
&\le \int_0^1 L(\xi(t), \dot\xi(t))\mathrm{d}t.
\end{aligned}$$

Since ξ is arbitrary, we get

$$u(1,y) - u(0,x) \le c(x,y) \quad \forall x, y \in \mathbb{R}^N.$$

Thanks to Remark 4.5, we deduce that

$$\mathcal{K}(\sigma^*) \geq \sup_{(\lambda,u)} \left\{ \int_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) : \lambda \in \mathbb{R}^+,\ u \in \mathcal{K}_c^\lambda \right\}. \tag{4.10}$$

Conversely, let $(\phi, \psi) \in \Phi_c^\lambda(\mu, \nu)$ be a maximizer in (4.3). Set

$$\phi_1(x) := \sup_{y \in \mathrm{supp}(\nu)} (\psi(y) - c(x,y)) \ \text{ and } \ \phi^*(x) := \max\{\phi_1(x), -\lambda\} \ \text{ for } \ x \in \mathrm{supp}(\mu).$$

Since $c(.,y)$ is locally Lipschitz w.r.t. the variable x, ϕ^* is Lipschitz on the compact set $\mathrm{supp}(\mu)$. Moreover, ϕ^* is non-positive (since $\psi \leq 0$ and $c \geq 0$) and (ϕ^*, ψ) is also a maximizer of the DPMK problem. By extension, we can assume that ϕ^* is non-positive and Lipschitz on $\mathbb{R}^N$. Now, we set

$$u^*(t,x) := \inf_{\xi} \left\{ \int_0^t L(\xi(s), \dot{\xi}(s))\mathrm{d}s + \phi^*(\xi(0)) : \xi \in Lip([0,t];\mathbb{R}^N), \xi(t) = x \right\}.$$

Then (see e.g. [**16**, Chapter 10] or [**10**, Chapter 6]) u^* is Lipschitz on Q and u^* is a viscosity solution of the Hamilton–Jacobi equation

$$\partial_t u(t,x) + H(x, \nabla_x u(t,x)) = 0$$

with $u^*(0,x) = \phi^*(x)$. It is not difficult to see that $u^*(1,y) \leq \phi^*(y) \leq 0,\ u^*(0,x) = \phi^*(x) \geq -\lambda \ \ \forall x, y \in \mathbb{R}^N$ and that

$$\begin{aligned} u^*(1,y) &= \inf_{\xi} \left\{ \int_0^1 L(\xi(s), \dot{\xi}(s))\mathrm{d}s + \phi^*(\xi(0)) : \xi \in Lip([0,1];\mathbb{R}^N), \xi(1) = y \right\} \\ &\geq \inf_{x \in \mathbb{R}^N} \{c(x,y) + \phi^*(x)\} \\ &\geq \psi(y) \ \ \forall y \in \mathbb{R}^N. \end{aligned}$$

These imply that $u^* \in \mathcal{K}_c^\lambda$ and that

$$u^*(1,y) - u^*(0,x) \geq \psi(y) - \phi^*(x) \ \ \forall x, y \in \mathbb{R}^N.$$

Thus

$$\begin{aligned} \int_{\mathbb{R}^N} u^*(1,.)\mathrm{d}\nu - \int_{\mathbb{R}^N} u^*(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) &\geq \int \psi \mathrm{d}\nu - \int \phi^* \mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \\ &= \mathcal{K}(\sigma^*). \end{aligned}$$

Combing this with (4.10), the duality (4.1) holds and u^* is a solution of the maximization problem on the right hand side of (4.1). □

4.2. Eulerian formulation by Fenchel–Rockafellar duality. As we said in the introduction, the Fenchel–Rockafellar duality is an important ingredient of our analysis, especially for the numerical analysis by augmented Lagrangian method.

Theorem 4.6. *Under the assumptions of Theorem 4.3, we have*

$$\begin{aligned}&\max\left\{\int_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m}-\mu(\mathbb{R}^N)) : (\lambda,u)\in\mathbb{R}^+\times\mathcal{K}_c^\lambda\right\}\\ &= \min\left\{\iint_Q L(x,v(t,x))\mathrm{d}\rho(t,x) : (\rho,v,\theta^0,\theta^1)\in B_c\right\},\end{aligned} \tag{4.11}$$

where

$$\begin{aligned}B_c := \Big\{(\rho,v,\theta^0,\theta^1)\in\mathcal{M}_b^+(Q)\times L^1_\rho(Q)^N\times\mathcal{M}_b^+(\mathbb{R}^N)\times\mathcal{M}_b^+(\mathbb{R}^N) : \theta^0(\mathbb{R}^N)=\mu(\mathbb{R}^N)-\mathbf{m},\\ -\mathrm{div}_{t,x}(\rho,v\rho)=\delta_1\otimes(\nu-\theta^1)-\delta_0\otimes(\mu-\theta^0)\Big\}.\end{aligned}$$

Roughly speaking, the minimization in (4.11) is the Fenchel–Rockafellar dual of the maximization problem. However, the interesting point to note here is that the maximization problem in (4.11) does not satisfy the sufficient conditions to use directly the dual theory of Fenchel–Rockafellar. To overcome this difficulty, we approximate the maximization problem by a suitable supremum problem. To this end, for general Lagrangian L, we make use of the smooth approximations given in Lemma 4.3.

Proof of Theorem 4.6. Let us first show that

$$\begin{aligned}&\max\left\{\int_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m}-\mu(\mathbb{R}^N)) : (\lambda,u)\in\mathbb{R}^+\times\mathcal{K}_c^\lambda\right\}\\ &\le \inf\left\{\iint_Q L(x,v)\mathrm{d}\rho : (\rho,v,\theta^0,\theta^1)\in B_c\right\}.\end{aligned} \tag{4.12}$$

Fix any $u\in\mathcal{K}_c^\lambda$ and $(\rho,v,\theta^0,\theta^1)\in B_c$. Let u_ε be the sequence of smooth functions given in Lemma 4.3. Taking u_ε as a test function in the continuity equation

$$-\mathrm{div}_{t,x}(\rho,v\rho)=\delta_1\otimes(\nu-\theta^1)-\delta_0\otimes(\mu-\theta^0),$$

we have

$$\iint_Q \partial_t u_\varepsilon\mathrm{d}\rho + \iint_Q \nabla_x u_\varepsilon(t,x)\cdot v(t,x)\mathrm{d}\rho = \int_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}(\nu-\theta^1) - \int_{\mathbb{R}^N} u_\varepsilon(0,.)\mathrm{d}(\mu-\theta^0).$$

Since $\theta^0(\mathbb{R}^N) = \mu(\mathbb{R}^N) - \mathbf{m}$, we get

$$\begin{aligned}
&\int\limits_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u_\varepsilon(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \\
&= \int\limits_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u_\varepsilon(0,.)\mathrm{d}\mu - \lambda \int\limits_{\mathbb{R}^N} \mathrm{d}\theta^0 \\
&= \int\limits_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}(\nu - \theta^1) - \int\limits_{\mathbb{R}^N} u_\varepsilon(0,.)\mathrm{d}(\mu - \theta^0) + \int\limits_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}\theta^1 - \int \left(u_\varepsilon(0,.) + \lambda\right)\mathrm{d}\theta^0 \\
&= \iint\limits_Q \left(\partial_t u_\varepsilon + \nabla_x u_\varepsilon \cdot \upsilon\right)\mathrm{d}\rho + \int\limits_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}\theta^1 - \int \left(u_\varepsilon(0,.) + \lambda\right)\mathrm{d}\theta^0.
\end{aligned} \tag{4.13}$$

Letting $\varepsilon \to 0$, using Lemma 4.3 and the fact that $u(1,.) \leq 0$, $u(0,.) + \lambda \geq 0$, we have

$$\begin{aligned}
\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) &\leq \iint\limits_Q L(x,\upsilon(t,x))\mathrm{d}\rho + \int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\theta^1 - \int\limits_{\mathbb{R}^N} (u0,.) + \lambda)\,\mathrm{d}\theta^0 \\
&\leq \iint\limits_Q L(x,\upsilon(t,x))\mathrm{d}\rho(t,x).
\end{aligned}$$

This implies the desired inequality (4.12).
Let us now prove the converse inequality. Obviously, we have

$$\begin{aligned}
&\max\left\{\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) : (\lambda,u) \in \mathbb{R}^+ \times \mathcal{K}_c^\lambda\right\} \\
&\geq \sup\left\{\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) : (\lambda,u) \in \mathbb{R}^+ \times \mathcal{K}_c^\lambda, u \in C^{1,1}(Q)\right\}.
\end{aligned}$$

It is sufficient to show that

$$\begin{aligned}
&\sup\left\{\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) : (\lambda,u) \in \mathbb{R}^+ \times \mathcal{K}_c^\lambda, u \in C^{1,1}(Q)\right\} \\
&= \min\left\{\iint\limits_Q L(x,\upsilon(t,x))\mathrm{d}\rho : (\rho,\upsilon,\theta^0,\theta^1) \in B_c\right\}.
\end{aligned} \tag{4.14}$$

This will be proved by using the Fenchel–Rockafellar dual theory. Indeed, the supremum problem in (4.14) can be written as

$$-\inf_{(\lambda,u)\in V} \mathcal{F}(\lambda,u) + \mathcal{G}(\Lambda(\lambda,u)), \tag{4.15}$$

where

$$\mathcal{F}(\lambda,u) := -\left(\int\limits_{\mathbb{R}^N} u(1,.)\mathrm{d}\nu - \int\limits_{\mathbb{R}^N} u(0,.)\mathrm{d}\mu + \lambda(\mathbf{m}-\mu(\mathbb{R}^N))\right) \text{ for any } (\lambda,u)\in V := \mathbb{R}\times C^{1,1}(Q),$$

$$\Lambda(\lambda,u) := (\nabla_{t,x}u,\ -\lambda-u(0,.),\ u(1,.)) \in Z := C_b(Q)^{N+1}\times C_b(\mathbb{R}^N)\times C_b(\mathbb{R}^N),$$

$$\mathcal{G}(q,z,w) := \begin{cases} 0 \text{ if } z(x)\le 0, w(x)\le 0 \text{ and } q_1(t,x)+H(x,q_N(t,x))\le 0\ \forall (t,x)\in Q \\ +\infty \quad \text{otherwise} \end{cases}$$

with $q := (q_1,q_N)\in C_b(Q)\times C_b(Q)^N$ for all $(q,z,w)\in Z$. Now, using the Fenchel–Rockafellar dual theory to the problem (4.15) (see e.g. [**15**, Chapter III]), we have

$$\begin{aligned} &\inf_{(\lambda,u)\in V} \mathcal{F}(\lambda,u)+\mathcal{G}(\Lambda(\lambda,u)) \\ &= \max_{(\Phi,\theta^0,\theta^1)\in\mathcal{M}_b(Q)^{N+1}\times\mathcal{M}_b(\mathbb{R}^N)\times\mathcal{M}_b(\mathbb{R}^N)} \left(-\mathcal{F}^*(-\Lambda^*(\Phi,\theta^0,\theta^1)) - \mathcal{G}^*(\Phi,\theta^0,\theta^1)\right). \end{aligned} \tag{4.16}$$

The proof is completed by computing explicitly the quantities in this maximization problem.

- Let us compute $\mathcal{F}^*(-\Lambda^*(\Phi,\theta^0,\theta^1))$. Since $\mathcal{F}$ is linear, $\mathcal{F}^*(-\Lambda^*(\Phi,\theta^0,\theta^1))$ is finite (and is equal to 0 whenever finite) if and only if

$$\langle -\Lambda^*(\Phi,\theta^0,\theta^1),(\lambda,u)\rangle = \mathcal{F}(\lambda,u) = -\left(\int u(1,.)\,\mathrm{d}\nu - \int u(0,.)\,\mathrm{d}\mu + \lambda(\mathbf{m}-\mu(\mathbb{R}^N))\right)\forall(\lambda,u)\in V,$$

or

$$-\langle\nabla_{t,x}u,\Phi\rangle - \langle\theta^0,-\lambda-u(0,.)\rangle - \langle\theta^1,u(1,.)\rangle = -\langle u(1,.),\nu\rangle + \langle u(0,.),\mu\rangle - \lambda(\mathbf{m}-\mu(\mathbb{R}^N))\forall(\lambda,u)\in V.$$

This implies that

$$-\langle\nabla_{t,x}u,\Phi\rangle = \langle u(0,.),\mu-\theta^0\rangle - \langle u(1,.),\nu-\theta^1\rangle \text{ for all test functions } u\in C^{1,1}(Q)$$

and that

$$\theta^0(\mathbb{R}^N) = \mu(\mathbb{R}^N) - \mathbf{m}.$$

Recall that $\Phi\in\mathcal{M}_b(Q)^{N+1}$, writing $\Phi=(\rho,E)$, the above computation gives

$$-\mathrm{div}_{t,x}(\rho,E) = \delta_1\otimes(\nu-\theta^1) - \delta_0\otimes(\mu-\theta^0),$$

and

$$\theta^0(\mathbb{R}^N) = \mu(\mathbb{R}^N) - \mathbf{m}.$$

- For $\mathcal{G}^*(\Phi,\theta^0,\theta^1)$, since $H(.,.)$ is continuous, using the same arguments as in [**29**, Proposition 5.18], we have

$$\mathcal{G}^*(\Phi,\theta^0,\theta^1) = \begin{cases} \displaystyle\iint_Q L(x,\upsilon(t,x))\mathrm{d}\rho & \text{if } \theta^0\ge 0,\ \theta^1\ge 0,\ \Phi=(\rho,E), \text{ and } \rho\ge 0,\ E\ll\rho,\ E=\upsilon\rho \\ +\infty & \text{otherwise.} \end{cases}$$

Substituting $\mathcal{F}^*$ and $\mathcal{G}^*$ into (4.16), we obtain the needed equality (4.14). □

4.3. Optimality condition and constrained MFG system. To write down the optimality condition for the duality (4.11), we need to use the notion of tangential gradient to a measure. This notion is an adaptation of the usual gradient (for Lebesgue measure) to any finite Radon measure. Recall that the tangential gradient $\nabla_\eta u$ is well-defined for any Lipschitz function u and any finite Radon measure η (see e.g. [**6**, **7**, **24**]).

Optimality condition for the duality (4.11) is related to the following PDE system:

$$\begin{cases} -\operatorname{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \big(\nu - \theta^1\big) - \delta_0 \otimes \big(\mu - \theta^0\big) \\ L(x, \upsilon(t,x)) = \nabla_\rho u(t,x) \cdot (1, \upsilon(t,x)) \qquad \rho\text{-a.e. } (t,x) \text{ in } Q \\ \partial_t u(t,x) + H(x, \nabla_x u(t,x)) \le 0 \qquad \text{a.e. } (t,x) \text{ in } Q \\ -\theta^0 \in \partial \mathbb{I}_{[-\lambda, +\infty)}(u(0,.)) \\ \theta^1 \in \partial \mathbb{I}_{(-\infty, 0]}(u(1,.)) \\ (\rho, \upsilon, \theta^0, \theta^1) \in \mathcal{M}_b^+(Q) \times L^1_\rho(Q)^N \times \mathcal{M}_b^+(\mathbb{R}^N) \times \mathcal{M}_b^+(\mathbb{R}^N), \end{cases} \tag{PDE$_\lambda$}$$

where the condition $\theta^1 \in \partial \mathbb{I}_{(-\infty,0]}(u(1,.))$ means that

$$u(1,.) \le 0 \ \text{ and } \ \langle \theta^1, \phi - u(1,.) \rangle \le 0 \ \ \forall \phi \in C_b(\mathbb{R}^N), \ \ \phi \le 0,$$

or equivalently

$$u(1,.) \le 0, \ \ \theta^1 \ge 0 \ \text{ and } \int_{\mathbb{R}^N} u(1,.) \mathrm{d}\theta^1 = 0.$$

Similarly, the condition $-\theta^0 \in \partial \mathbb{I}_{[-\lambda,+\infty)}(u(0,.))$ reads as

$$u(0,.) \ge -\lambda, \ \ \theta^0 \ge 0 \ \text{ and } \int_{\mathbb{R}^N} (u(0,.) + \lambda) \, \mathrm{d}\theta^0 = 0.$$

Theorem 4.7. *Assume that $(\rho, \upsilon, \theta^0, \theta^1) \in B_c$ and $(\lambda, u) \in \mathbb{R}^+ \times \mathcal{K}_c^\lambda$ are optimal for the two problems in the duality* (4.11)*. Then $(\rho, \upsilon, \theta^0, \theta^1, u)$ satisfies the system* (PDE$_\lambda$)*. Conversely, if $(\rho, \upsilon, \theta^0, \theta^1, u)$ is a solution of* (PDE$_\lambda$)*, then $(\rho, \upsilon, \theta^0, \theta^1)$ and (λ, u) are solutions to the duality* (4.11) *w.r.t.* $\mathbf{m} = \mu(\mathbb{R}^N) - \theta^0(\mathbb{R}^N)$.

Remark 4.8. For the standard optimal transport problem, i.e., $\mathbf{m} = \mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$ (in this case $\theta^0 = \theta^1 = 0$), the optimality conditions (PDE$_\lambda$) can be reduced to the following system:

$$\begin{cases} -\operatorname{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \nu - \delta_0 \otimes \mu \\ L(x, \upsilon(t,x)) = \nabla_\rho u(t,x) \cdot (1, \upsilon(t,x)) \ \ \rho\text{-a.e. in } Q \\ \partial_t u(t,x) + H(x, \nabla_x u(t,x)) \le 0 \ \ \text{ a.e. in } Q. \end{cases} \tag{4.17}$$

In particular, if $L(x, v) = L(v)$ is independent of the variable x then the system (4.17) recovers the same PDE, called Monge–Kantorovich equation, as in the work of Jimenez [**24**].

Remark 4.9. If $\mathbf{m} = \mu(\mathbb{R}^N) = \nu(\mathbb{R}^N)$ and assume moreover that $\rho \ll \mathcal{L}^{N+1}$, then the conditions (4.17) can be rewritten as

$$\begin{cases} -\mathrm{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \nu - \delta_0 \otimes \mu \\ \partial_t u(t,x) + H(x, \nabla_x u(t,x)) \le 0 \quad \text{a.e. } (t,x) \text{ in } Q \\ \partial_t u(t,x) + H(x, \nabla_x u(t,x)) = 0 \quad \rho\text{-a.e. } (t,x) \text{ in } Q \\ \upsilon(t,x) \in \partial H(x, \nabla_x u(t,x)) \quad \rho\text{-a.e. } (t,x) \text{ in } Q, \end{cases}$$

where ∂H is the subdifferential of H w.r.t. the second variable. Indeed, the condition $L(x, \upsilon(t,x)) = \nabla_\rho u(t,x) \cdot (1, \upsilon(t,x))$ ρ-a.e. implies that

$$\begin{aligned} L(x, \upsilon(t,x)) &= \nabla_{t,x} u(t,x) \cdot (1, \upsilon(t,x)) \quad \rho\mathcal{L}^{N+1}\text{-a.e. in } Q \\ &= \partial_t u(t,x) + \upsilon(t,x) \cdot \nabla_x u(t,x) \quad \rho\mathcal{L}^{N+1}\text{-a.e. in } Q. \end{aligned}$$

This implies that

$$L(x, \upsilon(t,x)) \le \partial_t u + H(x, \nabla_x u(t,x)) + L(x, \upsilon(t,x)) \le L(x, \upsilon(t,x)) \quad \rho\mathcal{L}^{N+1}\text{-a.e. in } Q.$$

Hence $\partial_t u(t,x) + H(x, \nabla_x u(t,x)) = 0$ ρ-a.e. in Q and $\upsilon(t,x) \in \partial H(x, \nabla_x u(t,x))$ ρ-a.e. in Q.

To prove Theorem 4.7, we need a similar estimate for (4.4) for any $u \in K_c^\lambda$. Since u is not smooth in general, we will characterize the estimate (4.4) via the tangential gradient instead of the usual one.

Lemma 4.4. *Let u be a Lipschitz function on Q and $\partial_t u(t,x) + H(x, \nabla_x u(t,x)) \le 0$ a.e. $(t,x) \in Q$. For any $(\rho, \upsilon) \in \mathcal{M}_b^+(Q) \times L^1_\rho(Q)^N$ satisfying the continuity equation*

$$-\mathrm{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \rho_1 - \delta_0 \otimes \rho_0,$$

we have

$$\nabla_\rho u(t,x) \cdot (1, \upsilon(t,x)) \le L(x, \upsilon(t,x)) \ \rho\text{-a.e. } (t,x) \text{ in } Q.$$

Proof. Let u_ε be the sequence as in Lemma 4.3. Since $-\mathrm{div}_{t,x}(\rho, \upsilon\rho) = \delta_1 \otimes \rho_1 - \delta_0 \otimes \rho_0$, we see that $(1, \upsilon(t,x)) \in T_\rho(t,x)$ for ρ-a.e. (t,x), where $T_\rho(t,x)$ is the tangential space w.r.t. ρ (see **[6, 7, 24]**). Using Lemma 4.3 and the continuity of the tangential gradient (see e.g. **[24**, Proposition 4.5]), we have

$$\begin{aligned} \iint_Q \nabla_\rho u \cdot (1, \upsilon) \xi \, \mathrm{d}\rho &= \lim_{\varepsilon \to 0} \iint_Q \nabla_\rho u_\varepsilon \cdot (1, \upsilon) \xi \, \mathrm{d}\rho \\ &= \lim_{\varepsilon \to 0} \iint_Q \nabla_{t,x} u_\varepsilon \cdot (1, \upsilon) \xi \, \mathrm{d}\rho \\ &\le \iint_Q L(x, \upsilon(t,x)) \xi \, \mathrm{d}\rho \quad \forall \xi \in \mathcal{D}(\mathbb{R}^{N+1}),\ \xi \ge 0. \end{aligned}$$

Thus the result of the lemma follows. □

Proof of Theorem 4.7. Let $(\rho, \upsilon, \theta^0, \theta^1) \in B_c$ and $(\lambda, u) \in \mathbb{R}^+ \times \mathcal{K}_c^\lambda$ be admissible elements, respectively. Let u_ϵ be the sequence given by Lemma 4.3. By (4.13) and Lemma 4.4, we have

$$\begin{aligned}
&\int u(1,.)\,\mathrm{d}\nu - \int u(0,.)\,\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \\
&= \lim_{\epsilon\to 0} \int u_\epsilon\,\mathrm{d}\nu - \int u_\epsilon\,\mathrm{d}\mu + \lambda(\mathbf{m} - \mu(\mathbb{R}^N)) \\
&= \lim_{\epsilon\to 0} \left(\iint_Q (\partial_t u_\varepsilon + \nabla_x u_\varepsilon \cdot \upsilon)\,\mathrm{d}\rho + \int_{\mathbb{R}^N} u_\varepsilon(1,.)\mathrm{d}\theta^1 - \int_{\mathbb{R}^N} (u_\varepsilon(0,.) + \lambda)\,\mathrm{d}\theta^0 \right) \\
&= \lim_{\epsilon\to 0} \iint_Q (\partial_t u_\varepsilon + \nabla_x u_\varepsilon \cdot \upsilon)\,\mathrm{d}\rho + \int_{\mathbb{R}^N} u(1,.)\mathrm{d}\theta^1 - \int_{\mathbb{R}^N} (u(0,.) + \lambda)\,\mathrm{d}\theta^0 \\
&\le \lim_{\epsilon\to 0} \iint_Q \nabla_{t,x} u_\varepsilon \cdot (1,\upsilon)\mathrm{d}\rho \\
&= \iint_Q \nabla_\rho u \cdot (1,\upsilon)\mathrm{d}\rho \\
&\le \iint_Q L(x,\upsilon(t,x))\mathrm{d}\rho.
\end{aligned} \tag{4.18}$$

1. From the assumptions on optimalities and the duality (4.11), the inequalities in (4.18) become equalities. These imply that

$$\iint_Q \nabla_\rho u \cdot (1,\upsilon)\mathrm{d}\rho = \iint_Q L(x,\upsilon(t,x))\mathrm{d}\rho,$$

or equivalently

$$L(x,\upsilon(t,x)) = \nabla_\rho u(t,x) \cdot (1,\upsilon(t,x)) \ \ \rho\text{-a.e. in } Q \text{ (by Lemma 4.4)};$$

and that $\int_{\mathbb{R}^N} u(1,.)\mathrm{d}\theta^1 = 0$, $\int_{\mathbb{R}^N} (u(0,.) + \lambda)\mathrm{d}\theta^0 = 0$. These show that $(\rho, \upsilon, \theta^0, \theta^1, u)$ satisfies the system (PDE_λ).

2. Conversely, if $(\rho, \upsilon, \theta^0, \theta^1, u)$ satisfies the system (PDE_λ), the inequalities in (4.18) are equalities. Using the duality (4.11), we obtain the desired optimalities. □

Remark 4.10. A solution $(\Phi^*, \theta^{0*}, \theta^{1*})$ of the Fenchel–Rockafellar dual formulation (4.16) gives an optimal couple of inactive submeasures and therefore optimal active submeasures $\rho_0^* = \mu - \theta^{0*}$ and $\rho_1^* = \nu - \theta^{1*}$.

This remark allows us to solve the PMK problem by using numerical methods for approximation of the Fenchel–Rockafellar dual problem.

5. Numerical approximation

We will apply the ALG2 algorithm to the dual formulation on the right hand side of (4.1) in order to give numerical approximations for the optimal partial transport problem. We will solve for optimal active submeasures $\rho_0 = \mu - \theta^0$, $\rho_1 = \nu - \theta^1$ and the optimal movement of density ρ_t from ρ_0 to ρ_1.

5.1. ALG2 method. The method is to solve numerically the minimization of the form

$$\inf_{u\in V} \{\mathcal{F}(u) + \mathcal{G}(\Lambda u)\}, \tag{5.1}$$

where $\mathcal{F}, \mathcal{G}$ are l.s.c. convex functionals on Hilbert spaces V et Z; and $\Lambda \in \mathcal{L}(V, Z)$ is a linear operator. The Fenchel–Rockafellar dual problem of (5.1) reads as

$$\sup_{\sigma\in Z^*} \left(-\mathcal{F}^*(-\Lambda^*\sigma) - \mathcal{G}^*(\sigma)\right).$$

By introducing a new variable $q \in Z$ to the primal problem (5.1), we can rewrite it in the form

$$\inf_{(u,q)\in V\times Z\,:\,\Lambda u = q} \mathcal{F}(u) + \mathcal{G}(q).$$

ALG2 is a primal-dual method (i.e. it provides numerical solutions of both primal and dual problems) consisting of 3 steps. Known σ_i, q_i, the next step $(u_{i+1}, \sigma_{i+1}, q_{i+1})$ is computed as follows. Fix any parameter $r > 0$ (in practice $r = 1, 2$):

- Step 1: Compute u_{i+1}

$$\min_{u\in V} \left\{\mathcal{F}(u) + \langle \sigma_i, \Lambda u\rangle + \frac{r}{2}|\Lambda u - q_i|^2\right\}.$$

- Step 2: Compute q_{i+1}

$$\min_{q\in Z} \left\{\mathcal{G}(q) - \langle \sigma_i, q\rangle + \frac{r}{2}|\Lambda u_{i+1} - q|^2\right\}.$$

- Step 3: Update σ_{i+1}

$$\sigma_{i+1} = \sigma_i + r\left(\Lambda u_{i+1} - q_{i+1}\right).$$

For the theory of this method, its interpretation, we refer the reader to **[14, 18, 19]**. Let us recall below its convergence which is enough for our discretized problems later.

Theorem 5.11 (cf. **[14]**, Theorem 8)**.** *Fixed $r > 0$, assuming that $V = \mathbb{R}^n, Z = \mathbb{R}^m$ and that Λ has full column rank. If the primal problem* (5.1) *and the Fenchel–Rockafellar dual problem* $\sup_{\sigma\in Z^*} \left(-\mathcal{F}^*(-\Lambda^*\sigma) - \mathcal{G}^*(\sigma)\right)$ *have solutions, then $\{u_i\}$ converges to a solution of the primal problem* (5.1) *and $\{\sigma_i\}$ converges to a solution of the Fenchel–Rockafellar dual problem. Moreover, $\{q_i\}$ converges to Λu^*, where u^* is the limit of $\{u_i\}$.*

5.2. Applying ALG2 for the optimal partial transport problem. Recall that the dual maximization formulation in (4.1) can be rewritten into the form

$$\inf\left\{\mathcal{F}(\lambda,u)+\mathcal{G}(\Lambda(\lambda,u)) : (\lambda,u)\in V\right\},$$

where

$$\mathcal{F}(\lambda,u) := -\left(\int u(1,.)\,\mathrm{d}\nu - \int u(0,.)\,\mathrm{d}\mu + \lambda(\mathbf{m}-\mu(\mathbb{R}^N))\right) \text{ for any } (\lambda,u)\in V := \mathbb{R}\times C^{1,1}(Q),$$

$$\Lambda(\lambda,u) := (\nabla_{t,x}u,\, -\lambda-u(0,.),\, u(1,.)) \in Z := C_b(Q)^{N+1}\times C_b(\mathbb{R}^N)\times C_b(\mathbb{R}^N)$$

and, for all $(q,z,w)\in Z$,

$$\mathcal{G}(q,z,w) := \begin{cases} 0 & \text{if } z(x)\le 0, w(x)\le 0 \text{ and } q(t,x)=(q_1(t,x),q_N(t,x))\in K_x \quad \forall (t,x)\in Q,\\ +\infty & \text{otherwise,}\end{cases}$$

with $K_x := \left\{(a,b)\in\mathbb{R}\times\mathbb{R}^N : a+H(x,b)\le 0\right\}, x\in\mathbb{R}^N$.

Let us discuss the details of computation. Actually, in computation, we replace V, Z by finite-dimensional spaces, for example, using Lagrangian piecewise polynomials. We denote by $P_i,\ i=1,2$ the spaces of piecewise polynomials of degree i. We will use $V=(\mathbb{R},P_2)$ and $Z=(P_1^{N+1},P_2,P_2)$, where $P_1^N := (\overbrace{P_1,...,P_1}^{N \text{ times}})$. We use L^2-norm for P_1, P_2, P_1^{N+1}.

• Step 1: We split into two steps: First using λ_i to compute u_{i+1} and then using u_{i+1} to calculate λ_{i+1}.

1. For u_{i+1}, we solve

$$\min_u \Big\{-\left(\langle u(1,.),\nu\rangle - \langle u(0,.),\mu\rangle\right) + \langle(\sigma_i,\theta_i^0,\theta_i^1),(\nabla_{t,x}u,-u(0,.),u(1,.))\rangle + \frac{r}{2}|(\nabla_{t,x}u,-\lambda_i-u(0,.),u(1,.))-(q_i,z_i,w_i)|^2\Big\}.$$

This is a quadratic problem which is equivalent to a linear equation with a positive-definite coefficient matrix. So this step can be solved effectively by many solvers. The linear equation is detailed as (by taking derivative w.r.t. u)

$$\begin{aligned} & r\langle\nabla_{t,x}u_{i+1},\nabla_{t,x}\phi\rangle + r\langle u_{i+1}(1,.),\phi(1,.)\rangle + r\langle u_{i+1}(0,.),\phi(0,.)\rangle \\ = \quad & \langle\phi(1,.),\nu\rangle - \langle\phi(0,.),\mu\rangle - \langle(\sigma_i,\theta_i^0,\theta_i^1),(\nabla_{t,x}\phi,-\phi(0,.),\phi(1,.))\rangle \\ & + r\langle(q_i,z_i,w_i),(\nabla_{t,x}\phi,-\phi(0,.),\phi(1,.))\rangle - r\langle\lambda_i,\phi(0,.)\rangle \quad \text{for all } (t,\phi)\in V. \end{aligned}$$

2. For λ_{i+1},

$$\min_{\lambda\in\mathbb{R}}\left\{-\lambda(\mathbf{m}-\mu(\Omega)) + \langle(\sigma_i,\theta_i^0,\theta_i^1),(0,-\lambda,0)\rangle + \frac{r}{2}|-\lambda-u_{i+1}(0,.)-z_i|^2\right\},$$

which is equivalent to

$$\lambda_{i+1} = \frac{\mathbf{m}-\mu(\Omega)+\int\limits_\Omega\theta_i^0 - r\int\limits_\Omega(z_i+u_{i+1}(0,.))}{r\int\limits_\Omega 1}.$$

- Step 2: Since the function $\mathcal{G}(q,z,w)$ has the form of $\mathcal{G}_1(q)+\mathcal{G}_2(z)+\mathcal{G}_3(w)$, we solve separately for the variables q,z,w.

1. For z_{i+1},

$$z_{i+1} \in \arg\min_{z\in P_2}\left\{\mathbb{I}_{[z\leq 0]} - \langle\theta_i^0, z\rangle + \frac{r}{2}|-\lambda_{i+1}-u_{i+1}(0,.)-z|^2\right\}$$
$$= \mathrm{Proj}_{\{[z\in P_2: z\leq 0]\}}\left(-\lambda_{i+1}-u_{i+1}(0,.)+\frac{\theta_i^0}{r}\right).$$

This is computed in pointwise, i.e., given a grid with vertices x_j, then

$$z_{i+1}(x_j) = \mathrm{Proj}_{\{[s\in\mathbb{R}: s\leq 0]\}}\left(-\lambda_{i+1}-u_{i+1}(0,.)(x_j)+\frac{\theta_i^0(x_j)}{r}\right)$$
$$= \min\left\{-\lambda_{i+1}-u_{i+1}(0,.)(x_j)+\frac{\theta_i^0(x_j)}{r}, 0\right\}.$$

2. For w_{i+1}, similarly,

$$w_{i+1} = \mathrm{Proj}_{\{[w\in P_2: w\leq 0]\}}\left(u_{i+1}(1,.)+\frac{\theta_i^1}{r}\right).$$

3. For q_{i+1}, similarly,

$$q_{i+1} = \mathrm{Proj}_{K_x}\left(\nabla_{t,x}u_{i+1}+\frac{\sigma_i}{r}\right).$$

- Step 3: Update Lagrangian multipliers.

6. Some examples

Using the FreeFem++ software **[20]**, we test some numerical examples. In all the examples below, we work on the square $\Omega=[0,1]\times[0,1]$ in $\mathbb{R}^2$ and use the discretization size $36\times 36\times 9$ for the spatial-time variable. We test the examples for costs of the form

$$c(x,y) = \inf_{\xi}\left\{\int_0^1 L(\xi(t),\dot{\xi}(t)\mathrm{d}t : \xi\in Lip([0,1];\Omega), \xi(0)=x, \xi(1)=y\right\},$$

with $L(x,v)=k(x)|v|^2$, $k\in C(\Omega), k(x)>0$ for all $x\in\Omega, v\in\mathbb{R}^2$. For this cost, the last projection in the Step 2 (the projection on K_x) is converted to a problem on $\mathbb{R}$ and the latter is computed easily by the bisection method.

Example 6.12. The source and the target are Gaussian distributions of the same mass for the Lagrangian $L(x,v)=|v|^2$. We want to transport optimally a half of the mass. More details,

$$\mu = 10\exp(-40(x_1-0.25)^2-40(x_2-0.75)^2),$$
$$\nu = 10\exp(-40(x_1-0.75)^2-40(x_2-0.25)^2).$$

The optimal active submeasures and the optimal displacement are given in Figure 1. Timestep 0 and timestep 9 (ρ_0 and ρ_1) are optimal active submeasures of the source and the target, respectively. The intermediate timesteps show the optimal movement of density from ρ_0 to ρ_1.

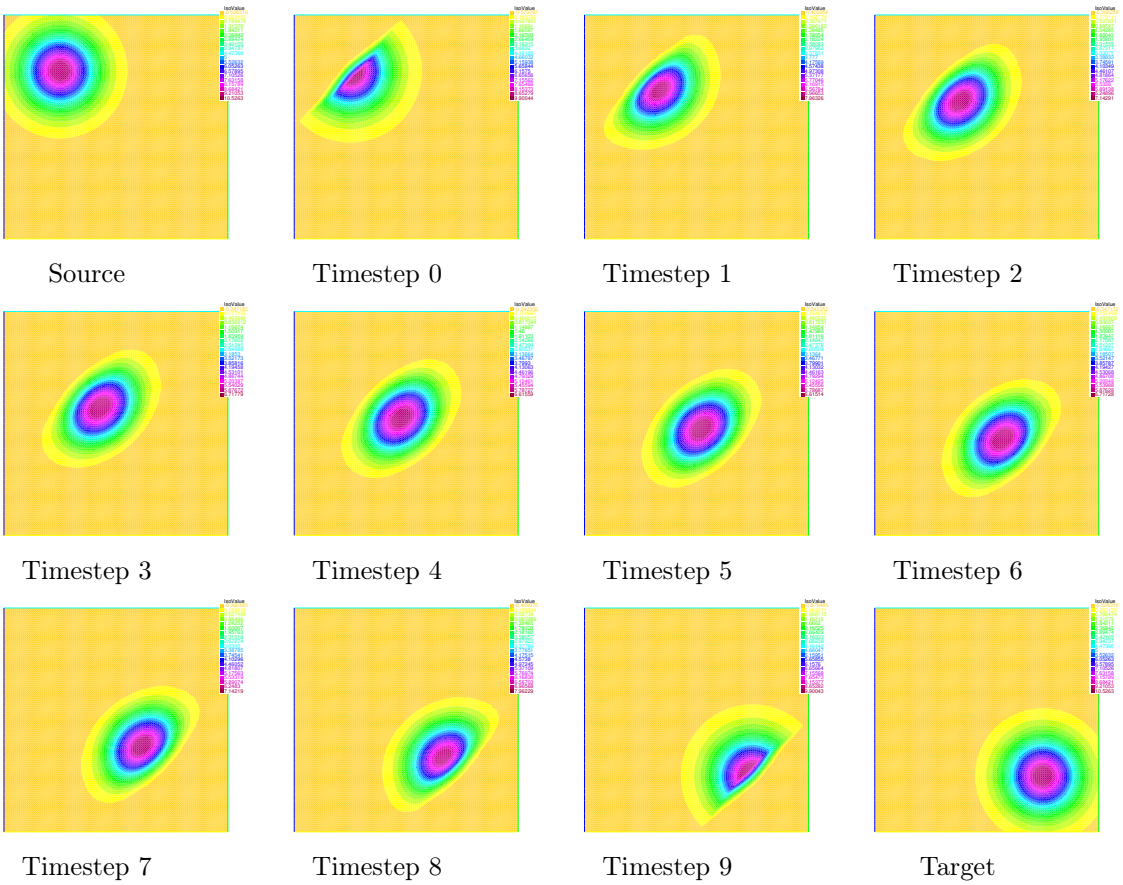


FIGURE 1. Optimal active submeasures and their displacement

Example 6.13. We take the similar data to the previous example but the source and the target are taken as the sums of two distributions,

$$\mu = 10\exp(-40(x_1-0.25)^2-40(x_2-0.25)^2)+10\exp(-40(x_1-0.75)^2-40(x_2-0.75)^2),$$
$$\nu = 10\exp(-40(x_1-0.75)^2-40(x_2-0.25)^2)+10\exp(-40(x_1-0.25)^2-40(x_2-0.75)^2).$$

The result is given in Figure 2.

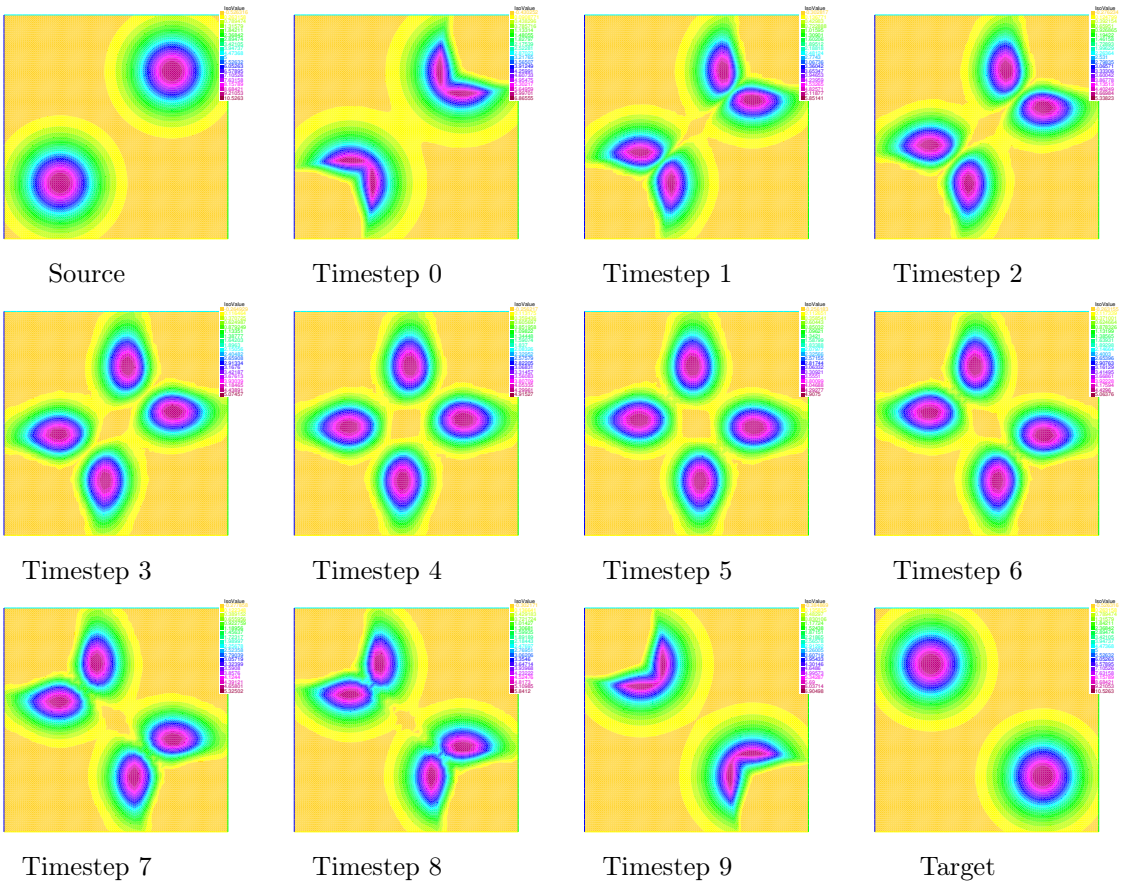


FIGURE 2. Optimal active submeasures and their displacement

Example 6.14. In this example, we take $L(x, v) = k(x)|v|^2$ with

$$k(x_1, x_2) = 1 + 15\exp(-45(x_1 - 0.5)^2 - 45(x_2 - 0.5)^2),$$
$$\mu = 20\exp(-60(x_1 - 0.2)^2 - 60(x_2 - 0.8)^2),$$
$$\nu = 20\exp(-60(x_1 - 0.8)^2 - 60(x_2 - 0.2)^2),$$

and

$$\mathbf{m} = \frac{\mathbf{m}_{\max}}{2}.$$

This cost means that we have to pay much if we transport through around $(0.5, 0.5)$ (where $k(x)$ is big). The numerical result is illustrated in Figure 3.

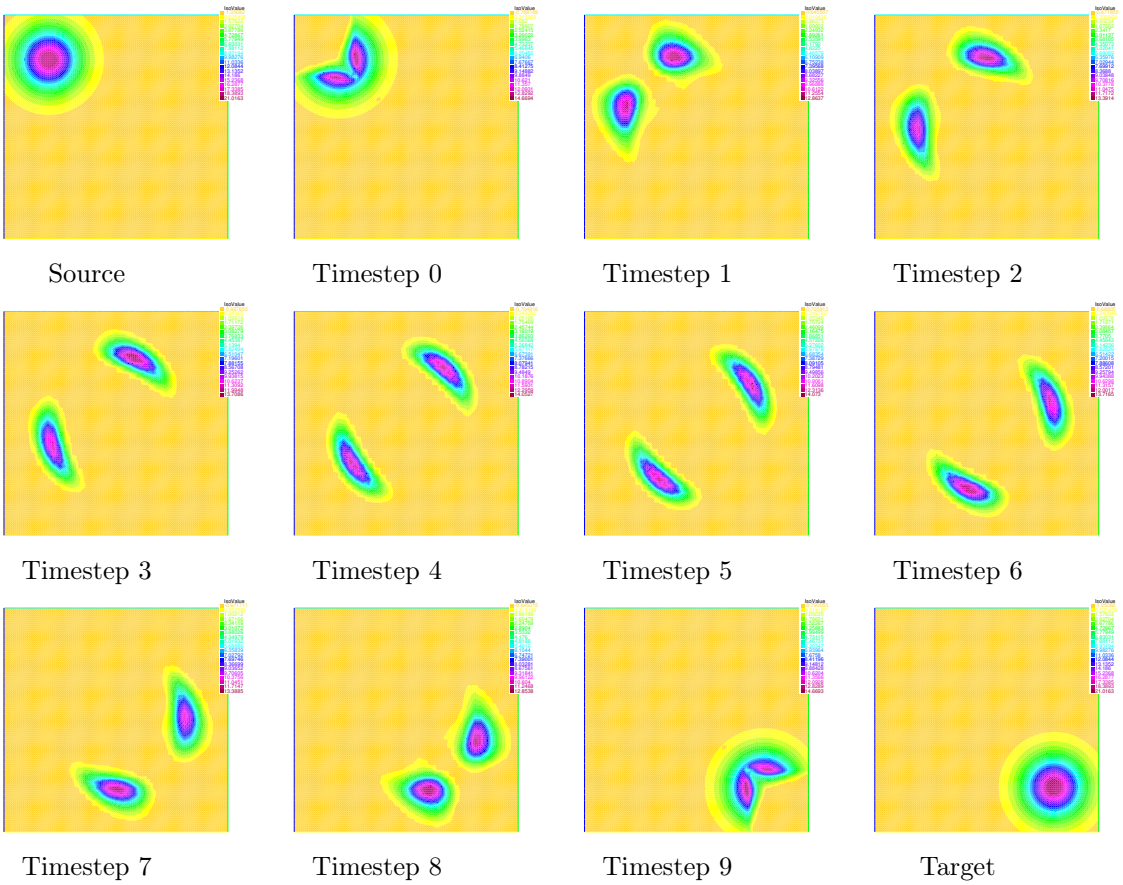


FIGURE 3. Optimal active submeasures and their displacement

Acknowledgements: The authors thank Profs J. D. Benamou and G. Carlier for many discussions.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich, Birkhäuser, 2005.

[2] J. W. Barrett and L. Prigozhin. Partial L^1 Monge–Kantorovich problem: Variational formulation and numerical approximation. *Interfaces and Free Bound.*, **11** (2009), 201–238.

[3] J. D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.*, **84** (2000), 375–393.

[4] J. D. Benamou and G. Carlier. Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations. *J Optim Theory Appl.*, **167** (2015), 1–26.

[5] J. D. Benamou, G. Carlier and R. Hatchi. A numerical solution to Monge's problem with a Finsler distance cost. *ESAIM Math. Model. Numer. Anal.*, (2017), DOI:`https://doi.org/10.1051/m2an/2016077`.

[6] G. Bouchitté, G. Buttazzo and P. Seppercher. Energy with respect to a measure and applications to low dimensional structures. *Calc. Var.*, **5** (1997), 37–54.

[7] G. Bouchitté, G. Buttazzo, and P. Seppecher. Shape optimization solutions via Monge–Kantorovich equation. *C. R. Acad. Sci. Paris Sér. I Math*, **324** (1997), 1185–1191.
[8] J. D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. PeyrÃ©. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, **37** (2015), A1111–A1138.
[9] L. Caffarelli and R.J. McCann. Free boundaries in optimal transport and Monge–Ampere obstacle problems. *Annals of Math.*, **171** (2010), 673–730.
[10] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton–Jacobi equations, and optimal control*. Vol. 58 Progress in Nonlinear Differential Equations and Their Applications, Springer, 2004.
[11] S. Chen and E. Indrei. On the regularity of the free boundary in the optimal partial transport problem for general cost functions. *J. Differential Equations*, **258** (2015), 2618–2632.
[12] L. Chizat, G. Peyré, B. Schmitzer and F. X. Vialard. Scaling algorithms for unbalanced transport problems. arXiv preprint `https://arxiv.org/abs/1607.05816` (2016).
[13] G. Davila and Y. H. Kim. Dynamics of Optimal Partial Transport. *Calc. Var.* (2016).
[14] J. Eckstein and D. P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program., **55** (1992), 293–318.
[15] I. Ekeland and R. Teman. *Convex analysis and variational problems*. Studies in Mathematics and Its Applications, North-Holland American Elsevier, 1976.
[16] L. C. Evans. *Partial Differential Equations*, second edition. Vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, 2010.
[17] A. Figalli. The Optimal Partial Transport Problem. *Arch. Ration. Mech. Anal.* **195** (2010), 533–560.
[18] M. Fortin and R. Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*. Vol. 15 of Studies in Mathematics and Its Applications, North-Holland, 1983.
[19] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*. Vol 9 of Studies in Applied and Numerical Mathematics, SIAM, 1989.
[20] F. Hecht. New development in freefem++. *J. Numer. Math.*, **20** (2012), 251-266.
[21] N. Igbida and V. T. Nguyen. Optimal partial mass transportation and obstacle Monge–Kantorovich equation. Submitted.
[22] N. Igbida and V. T. Nguyen. Augmented Lagrangian method for optimal partial transportation. *IMA J. Numer. Anal.,* (2017), DOI: `https://doi.org/10.1093/imanum/drw077`.
[23] E. Indrei. Free boundary regularity in the optimal partial transport problem. *J. Funct. Anal.*, **264** (2013), 2497–2528.
[24] C. Jimenez. Dynamic formulation of optimal transport problems. *J. Convex Anal.* **15** (2008), 593–622.
[25] B. Maury, A. Roudneff-Chupin, F. Santambrogio. A macroscopic crowd motion model of gradient flow type. *Math. Models Methods Appl. Sci.* **20** (2010), 1787–1821.
[26] G. D. Philippis, A. R. Mészáros, F. Santambrogio and B. Velichkov. BV estimates in optimal transportation and applications. *Arch. Ration. Mech. Anal.,* **219** (2016), 829–860.
[27] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co, New York, 1987.
[28] F. Santambrogio. A Modest Proposal for MFG with Density Constraints, *Netw. Heterog. Media*, **7** (2012), 337–347.
[29] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87 of Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, 2015.
[30] C. Villani. *Topics in Optimal Transportation*. Vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, 2003.
[31] C. Villani. *Optimal Transport, Old and New*. Vol. 338 of Grundlehren des Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), Springer, New York, 2009.