AUGMENTED LAGRANGIAN METHODS FOR TRANSPORT OPTIMIZATION, MEAN-FIELD GAMES AND DEGENERATE PDES

JEAN-DAVID BENAMOU AND GUILLAUME CARLIER

Abstract. Many problems from mass transport can be reformulated as variational problems under a prescribed divergence constraint (static problems) or subject to a time dependent continuity equation which again is a divergence constraint but in time and space. A large class of Mean-Field Games introduced by Lasry and Lions may also be interpreted as a generalisation of the time-dependent optimal transport problem. Following Benamou and Brenier [BB00], we show that augmented Lagrangian methods are well-suited to treat such convex but neither smooth nor strictly convex problems. It includes in particular Monge's original optimal transport problem. A finite element discretization and implementation of the method is used to provide numerical simulations and a convergence study.

1. Introduction

Context. Optimal transport theory has received a lot of attention in the last two decades and both the theory and the applications continue to develop rapidly (see the monographs of Villani [Vil03], [Vil09]). In contrast, numerical methods for optimal transport are still underdeveloped. On the one hand and independently of the transport cost, optimal transport problems are infinite-dimensional linear problem. After discretization, they can be solved using Linear Programming methods (e.g. simplex, interior points) or combinatorial methods for the assignment problem (Hungarian or auction algorithms). On the other hand, fine discretizations of continuous measures lead to large finite dimensional problems and these methods quickly become too expensive.

In many relevant applications however, the transport cost has a strong structure: it is a distance in Monge's original problem, the squared distance in the quadratic Monge-Kantorovich problem (solved by Brenier in [Bre91]) or more generally a convex function of the displacement. It is then natural to investigate if specific numerical algorithms, efficient and tractable enough to deal with measures with a large number of points in their support, can be designed for such costs.

Date: October 8, 2014.

Key words and phrases. Augmented Lagrangian, Optimal Transport, Monge Problem, Mean-Field Games, degenerate elliptic PDEs.

In [BB00], Benamou and Brenier introduced a "computational fluid dynamics (CFD)" reformulation of the quadratic optimal transport. It reformulates the squared distance cost as a time dependent kinetic energy which is minimized over solutions of the continuity equation with prescribed initial and terminal densities. The resulting variational problem is non smooth and convex and can be solved using augmented Lagrangian numerical methods. A key idea of this reformulation is to interpret the continuity equation as a divergence in time and space. The algorithm ALG2 of [FG83] was used in [BB00]. Despite the slow convergence of this iterative method, the approach has been quite successful numerically and proved very robust as an alternative to the above mentioned linear or combinatorial numerical methods for the quadratic cost.

An extension using the Augmented Lagrangian Numerical method to a mixed L^2/Wasserstein distance was proposed by Benamou in [Ben03]. More general time dependent problems were soved by Buttazo, Jimenez and Oudet in the framework of congested dynamic [BJO09] also using an Augmented Lagrangian Numerical method. More recently, Papadakis, Peyré and Oudet have shown that the Augmented Lagrangian method belongs to the larger class of proximal splitting methods [PPO14] and applied it to generalised time dependent cost functions interpolating between the quadratic cost optimal transport problem and the H^{-1} norm. Augmented Lagrangian methods have also been used for "realistic" image interpolation by Hug, Maitre and Papadakis [HMP05] but in this work the continuity equation constraint itself is modified and it breaks the convexity of the problem.

Contribution. The goal of the present paper is to push further the extension of Augmented Lagrangian methods and the numerical use of ALG2 to several other problems arising either in optimal transport or in Mean-Field Games theory. We will consider two classes of convex but neither strictly convex nor smooth variational problems. The first one is a class of time independent minimal flow problems. These consist in minimizing a convex functional among vector fields with a prescribed divergence. The Monge problem is of this type as well as many variants, including degenerate elliptic PDEs of Euler-Lagrange type

$$\sigma = \nabla H(\nabla\phi), \ -\operatorname{div}(\sigma) = f$$

where H is convex but not strictly convex (it can be 0 on a whole ball for instance which makes the previous equation even more degenerate than the p-laplacian).

The second class consists of time dependent problems where a certain energy is minimized among solutions of the continuity equation:

$$\partial_t\rho + \operatorname{div}(\rho v) = 0.$$

The Benamou-Brenier formulation of quadratic optimal optimal transport belongs to this class of problems as well as some classes of Mean-Field Games. We shall see (see remark 3.2) why Augmented Lagrangian methods are well suited to treat such problems and in particular why they ensure that mass remains nonnegative.

The paper is organized as follows. Section 2 presents various motivating examples, both in the static and the time-dependent cases. Section 3 recalls the principle of augmented Lagrangian methods, its connection with splitting methods and its convergence in finite dimensions, approximation of the infinite-dimensional problem by finite elements is also discussed. Section 4 explains the implementation of ALG2 both in our static and dynamic settings. Section 5 presents a numerical convergence study. FreeFem++ prototypes codes are available at https://team.inria.fr/mokaplan/software/

2. Variational formulation and examples

2.1. **Static problems: prescribed divergence.** The aim of this paragraph is to emphasize the role of variational problems with a prescribed divergence constraint in mass transport problems. Let us start with a convex variational problem of the form

$$\inf_{\phi\in W^{1,p}(\Omega)} J(\phi) := F(\phi) + G(\nabla\phi) \tag{1}$$

where Ω is some open subset of $\mathbb{R}^d$, $p \in (1, +\infty]$ and F and G are two convex, proper and lsc functions, $F : W^{1,p}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ and G: $L^p(\Omega)^d \to \mathbb{R} \cup \{+\infty\}$.

In mass transport, F is a linear form related to the initial and final distributions of masses and G generally lacks strict convexity and/or smoothness properties. We list below some examples to emphasize the type of singularities one has to take into account. As we will see, transport problems can be formulated in the optimal flow dual form of (1) :

$$\sup_{\sigma\in L^q(\Omega)^d} -F^*(\mathrm{div}(\sigma)) - G^*(\sigma) \tag{2}$$

where q is the dual exponent of p, $q = p/(p-1)$ if $p \neq \infty$ and $q = 1$ when $p = +\infty$ and F^* and G^* denote the Legendre transforms of F and G. Note that we are slightly abusing the terminology when $p = +\infty$ because in this case L^1 is not the dual of L^∞ but in view of applications to mass transport problems it is essential to consider this case as well. We give below some examples and numerical illustrations. A more comprehensive discussion of the numerical method follows in section 3.

Monge's problem. Given Ω a convex bounded open subset of $\mathbb{R}^d$ and two probability measures ρ_0 and ρ_1 on $\overline{\Omega}$, Monge's optimal transport problem (for the euclidean distance $|.|$) consists in finding the cheapest way to transport ρ_0 to ρ_1 for the euclidean distance. Denoting by $\gamma \in \Pi(\rho_0, \rho_1)$ the set of transport plans between ρ_0 and ρ_1 i.e. the set of probability measures on $\overline{\Omega} \times \overline{\Omega}$ having ρ_0 and ρ_1 as marginals, one thus wishes to solve the infinite-dimensional linear programming problem:

$$W_1(\rho_0,\rho_1) := \inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{\overline{\Omega}\times\overline{\Omega}} |y-x| d\gamma(x,y) \tag{3}$$

whose value $W_1(\rho_0,\rho_1)$ is by definition the 1-Wasserstein distance between ρ_0 and ρ_1. The well-known Kantorovich duality formula (see for instance [Vil03], [Vil09]) reads:

$$W_1(\rho_0,\rho_1) = \sup\{\int_{\overline{\Omega}} \phi(\rho_1-\rho_0) \ : \ \phi \text{ 1-Lipschitz } \}. \tag{4}$$

Note that since $\rho_1-\rho_0$ has zero mass, one can normalize ϕ to have mean zero. Problem (4) can be written in the standard form (1) with

$$F(\phi) := -\int_{\overline{\Omega}} \phi(\rho_1-\rho_0), \qquad G(q) := \begin{cases} 0 \text{ if } \|q\|_{L^\infty} \le 1 \\ +\infty \text{ otherwise.} \end{cases}$$

In this case, (2) reads as the minimal flow problem:

$$\sup_{\sigma\in L^1(\Omega)} \{-\int_\Omega |\sigma| \ : \ -\mathrm{div}(\sigma) = \rho_1-\rho_0, \sigma\cdot\nu = 0 \text{ on } \partial\Omega \ \} \tag{5}$$

where the divergence constraint has to be understood in the weak sense:

$$\int_\Omega \nabla u\cdot\sigma = \int_{\overline{\Omega}} u(\rho_1-\rho_0), \forall u \in C^1(\overline{\Omega}).$$

Of course, one should in general relax (5) to vector-valued measures, but it follows from the important results of De Pascale and Pratelli [DPP04] and Santambrogio [San09] that there is in fact an L^1 solution of (5) as soon as ρ_0 and ρ_1 are L^1. We wish now to explain more precisely the connections between the three optimization problems (3)-(4) and (5) (see Evans and Gangbo [EG99] or the lecture notes of Ambrosio [Amb03] for a more detailed presentation). Let ϕ be a 1-Lipschitz potential that solves (4), then by the Kantorovich duality formula a transport plan γ between ρ_0 and ρ_1 is optimal for (3) if and only if

$$\phi(y)-\phi(x) = |x-y| \ \gamma\text{-a.e.},$$

which means that the mass at x is transported along a segment on which ϕ grows at maximal rate 1, such rays whose direction is given by the gradient of ϕ are called transport rays and give the direction of optimal transportation in Monge's problem. An optimal flow field σ for (5) is formally related to an optimal ϕ in (4) by:

$$\nabla\phi = \begin{cases} \frac{\sigma}{|\sigma|} \text{ if } \sigma \neq 0 \\ \text{any vector in the unit ball if } \sigma = 0. \end{cases}$$

Hence σ also gives the direction of transport rays, moreover $|\sigma|$ is called the transport density and measures how much total mass is passing through a given point. There is also a relation between optimal flows σ's and optimal plans γ: if γ solves (3), then the vector field σ defined by, for every $X \in C_c(\Omega, \mathbb{R}^d)$:

$$\int_\Omega \sigma(x) \cdot X(x) dx := \int_{\overline{\Omega}\times\overline{\Omega}} \Big(\int_0^1 X(x + t(y-x)) \cdot (y-x) dt \Big) d\gamma(x,y)$$

actually solves (5).

Figure 1, shows Monge optimal flows computed on a 2D square ($x = (x_1, x_2) \in [0,1]^2$) in two test cases. In test case 1, we take:

$$\rho_0 = e^{-40*((x_1-0.75)^2+(x_2-0.25)^2} \quad \text{and} \quad \rho_1 = e^{-40*((x_1-0.25)^2+(x_2-0.65)^2)}$$

in the presence of an obstacle. In test case 2, ρ_0 is a constant density and ρ_1 is the sum of three concentrated Gaussians

$$\rho_1 = e^{400*((x_1-0.25)^2+(x_2-0.75)^2)} + e^{400*((x_1-0.35)^2+(x_2-0.15)^2)} + e^{400*((x_1-0.85)^2+(x_2-0.7)^2)}.$$

Variants of Monge's problem: anisotropies and heterogeneous media. There are natural variants of Monge's problem for the euclidean distance which are very simple to deal with using the Augmented Lagrangian algorithm. The first variant is when one replaces the euclidean distance by

$$d_K(x,y) := \sup_{p \in K} \langle x - y, p \rangle$$

for some (not necessarily symmetric) convex compact set K with 0 in its interior. In this case, the dual problem of (3) simply reads as

$$\sup\{ \int_{\overline{\Omega}} \phi(\rho_1 - \rho_0) \ : \ \nabla\phi \in K \text{ a.e. } \}. \tag{6}$$

Optimal transport plans are then concentrated on pairs (x,y) for which $\phi(y) - \phi(x) = d_K(x,y)$ and the optimal flow formulation is obtained by replacing the euclidean norm $|\sigma|$ by $d_K(\sigma, 0)$. Another interesting case arises when, instead of a translation-invariant distance one rather considers a Riemannian distance such as:

$$d_g(x,y) := \inf\{ \int_0^1 g(\gamma(t))|\dot\gamma(t)| dt \ : \ \gamma \in W^{1,1}([0,1], \mathbb{R}^d), \ \gamma(0) = x, \gamma(1) = y\}$$

where g is some bounded and bounded away from 0 function capturing possible media heterogeneities. For such a distance, optimal transport does not occur on straight lines anymore but on geodesics for d_g and the dual of (3) becomes

$$\sup\{ \int_{\overline{\Omega}} \phi(\rho_1 - \rho_0) \ : \ |\nabla\phi \,| \le g \text{ a.e.}\}. \tag{7}$$

We recover the standard form (1) with F as in Monge and

$$G(\nabla\phi) := \int_\Omega \mathcal{G}(x, \nabla\phi(x)) dx$$

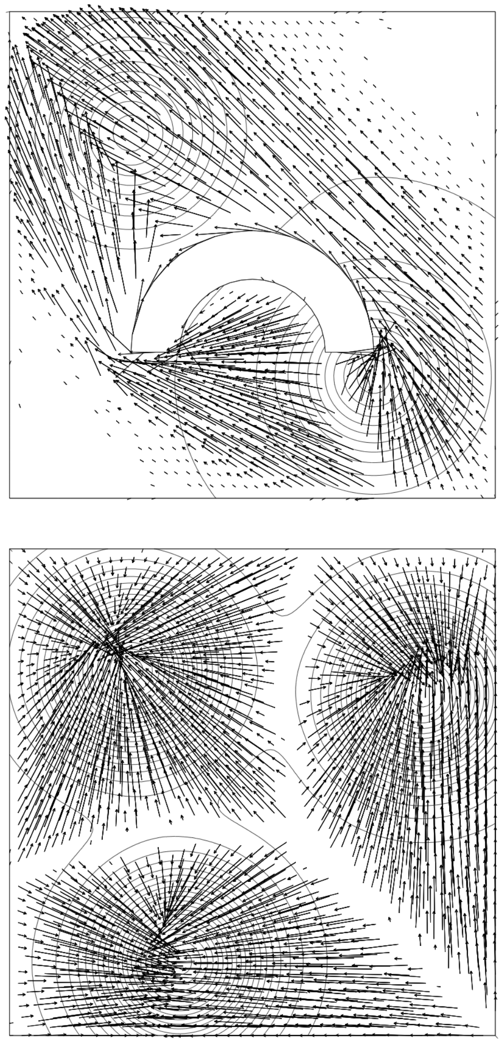

FIGURE 1. Test Case 1 and 2 : Monge problem flows, length of arrows are proportional to transport density - Level curves correspond to the right hand side density term of the divergence $\rho_1 - \rho_0$ source/sink data to be transported

with

$$\mathcal{G}(x,q) = \begin{cases} 0 \text{ if } q \leq g(x) \\ +\infty \text{ otherwise.} \end{cases}$$

Figure 2, shows such heterogeneous Monge flows between two Gaussian densities defined for $x = (x_1, x_2) \in [0,1]^2$ by:

$$\rho_0 = e^{-40*((x_1-0.75)^2+(x_2-0.25)^2} \quad \text{and} \quad \rho_1 = e^{-40*((x_1-0.25)^2+(x_2-0.65)^2)}.$$

First a concave lens - test case 3 - and then a discontinuous medium - test case 4 - are considered, corresponding to the metrics

$$g(x_1,x_2) = 3 - 2 * e^{-10*((x_1-0.5)^2+(x_2-0.5)^2)} \quad \text{or} \quad g(x_1,x_2) = 1 + 2 * \chi_{x_1>\frac{1}{2}}.$$

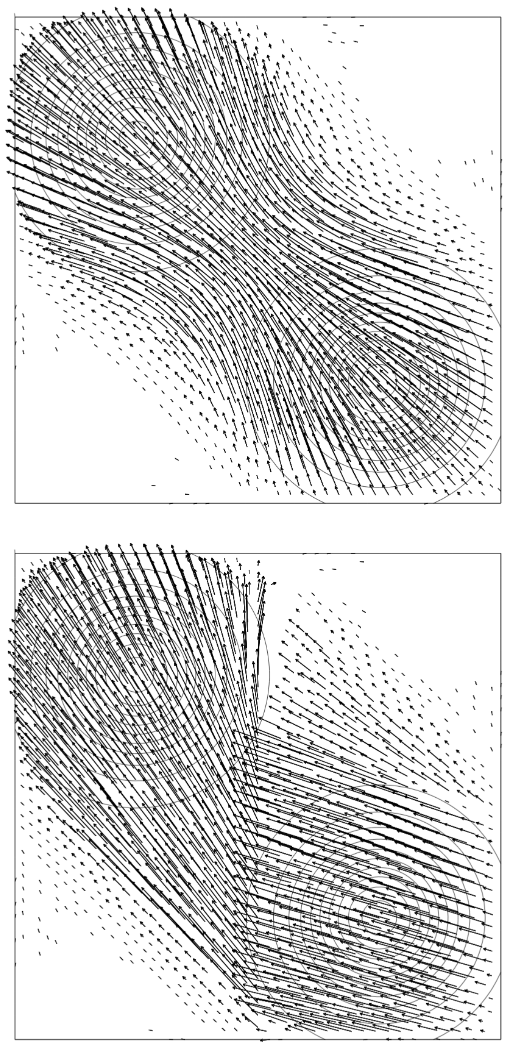

FIGURE 2. Test cases 3 and 4 : Heterogeneous medium Monge problem flows, length of arrows are proportional to transport density - Level curves correspond to the right hand side density term of the divergence $\rho_1 - \rho_0$ source/sink data to be transported

Congested transport and degenerate elliptic PDEs. In [CJS08], a variant of Monge's problem allowing for congestion effects (i.e. the fact that crossing zones of high traffic is more costly) has been proposed and in [BCS10], it has been shown that this variant leads to consider minimal flow problems with a superlinear term, a typical case being

$$\inf_{\sigma \in L^q(\Omega)} \{ \int_\Omega (\beta|\sigma| + \frac{1}{q}|\sigma|^q) \ : \ -\operatorname{div}(\sigma) = f, \sigma \cdot \nu = 0 \text{ on } \partial\Omega \ \} \tag{8}$$

where again $f = \rho_1 - \rho_0$ represents the difference between the target and source measures, ν denotes the exterior unit normal, $q > 1$ and $\beta \geq 0$. This congested variant of Monge's problem is dual to the problem of type (1):

$$\inf_{\phi \in W^{1,p}(\Omega)} \frac{1}{p} \int_\Omega (|\nabla \phi| - \beta)_+^p - \langle f, \phi \rangle \tag{9}$$

with $p = q/(q-1)$ the conjugate exponent of q. Note that G is not strictly convex since it identically vanishes whenever $|\nabla \phi| \leq \beta$. Solving (9) amounts to solve the degenerate elliptic equation

$$\begin{cases} -\operatorname{div}\left((|\nabla\phi| - \beta)_+^{p-1} \frac{\nabla\phi}{|\nabla\phi|} \right) = f, \ \text{ in } \Omega, \\ (|\nabla\phi| - \beta)_+^{p-1} \nabla\phi \cdot \nu = 0 \text{ on } \partial\Omega. \end{cases} \tag{10}$$

The optimal flow σ (which is unique) is related to ϕ solving (10) (which is not unique even with a zero mean normalization) by

$$\sigma = \Big(|\nabla\phi| - \beta \Big)_+^{p-1} \frac{\nabla\phi}{|\nabla\phi|}.$$

When $\beta = 0$, one recovers the standard p-Laplace equation, for which Glowinski and Marrocco [GM75] observed long ago that augmented Lagrangian methods are well-adapted. If on the contrary, β is very large, one recovers in (8) a regularization of the Monge flow problem (5). Another regularization which was used numerically in [BP07] consists in taking $\beta = 0$ and q very close to 1.

Figure 3, shows tests cases 5, 6, and 7 which correspond to three values of q: $q = 1.01$ (Monge like, almost no congestion), $q = 2$ (some congestion), $q = 6.66$ (more congestion), densities and obstacle are the same as in test case 1 for the classical Monge problem. Many variants (other boundary conditions, anisotropic norms, coefficients depending on x, different exponents for the different components of the flow...) can of course be considered in a similar way.

2.2. **Time-dependent problems: the continuity equation.**

Time-dependent formulation of optimal transport. The fact that Monge's optimal transport problem admits a static reformulation with a divergence constraint (5) is in fact related to the fact that the cost satisfies the triangle inequality. In contrast, for strictly convex transportation costs $c \in C^1(\mathbb{R}^d)$,

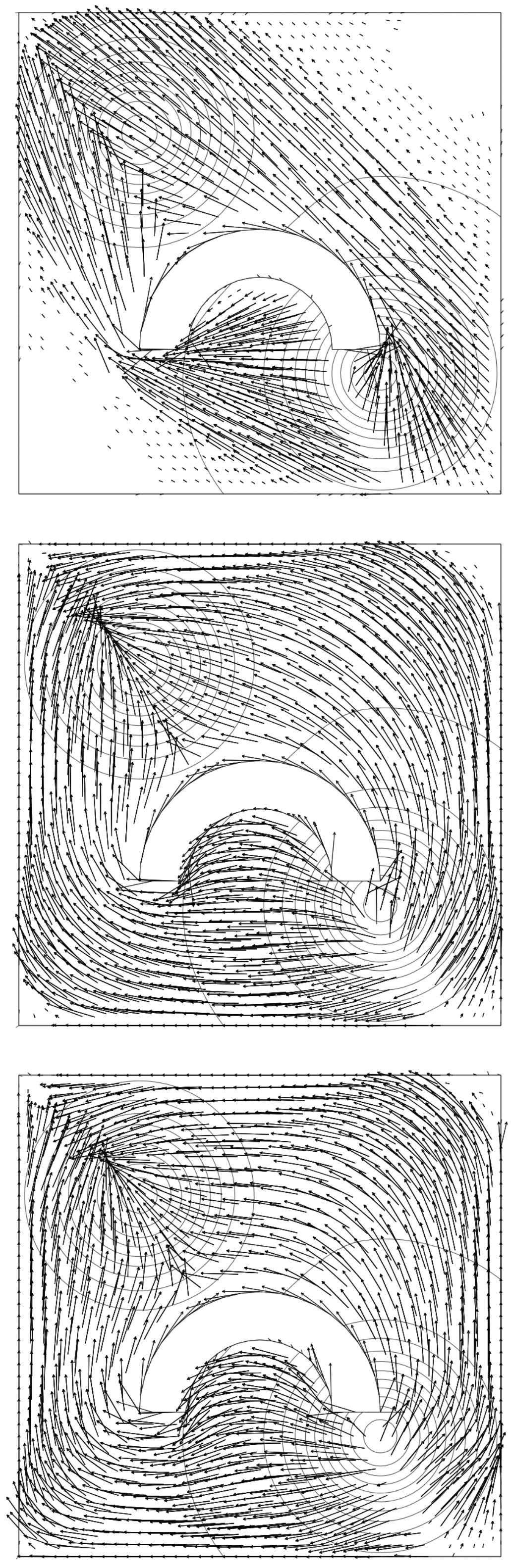

FIGURE 3. Test cases 5, 6, and 7. Congested Monge problem flows for $q = 1.01, 2, 6.66$, length of arrows are proportional to transport density - Level curves correspond to the right hand side density term of the divergence $\rho_1 - \rho_0$ source/sink data to be transported

it is necessary to introduce an extra time variable as proposed in [BB00] for the quadratic $c(y-x)=\frac{|y-x|^2}{2}$. Consider the optimal transport problem

$$W_c(\rho_0,\rho_1) := \inf_{\gamma\in\Pi(\rho_0,\rho_1)} \int_{\overline{\Omega}\times\overline{\Omega}} c(y-x)d\gamma(x,y). \tag{11}$$

Then, the dynamic formulation of (11) consists in minimizing

$$\int_0^1 \int_{\mathbb{R}^d} c(v_t(x))\rho_t(dx)dt$$

among solutions of the continuity equation

$$\partial_t\rho + \operatorname{div}(\rho v) = 0,\ \rho|_{t=0} = \rho_0,\ \rho|_{t=1} = \rho_1. \tag{12}$$

It is convenient to rewrite this problem in terms of $\sigma(t,x) = (\rho_t(x), m_t(x)) := (\rho_t(x), \rho_t(x)v_t(x)) \in \mathbb{R}^{d+1}$. Indeed, in this case, (12) simply becomes the linear constraint:

$$-\operatorname{div}_{t,x}(\sigma) = f := \delta_1\otimes\rho_1 - \delta_0\otimes\rho_0 \tag{13}$$

in the weak sense and the divergence is of course with respect to t and x. Let us then define

$$E(\sigma) = E(\rho,m) := \begin{cases} c(m/\rho)\rho \text{ if } \rho>0 \\ 0 \text{ if } \rho=0 \text{ and } m=0 \\ +\infty \text{ otherwise.}\end{cases}$$

Note that E is convex, lsc, one homogeneous and incorporates the natural constraints of the transport problem: mass is nonnegative and momentum vanishes where mass does. The time-dependent formulation of (11) then can be rewritten as:

$$\inf\{\int_0^1\int_{\mathbb{R}^d} E(\sigma_t(x))dxdt\ :\ -\operatorname{div}_{t,x}(\sigma) = f\}. \tag{14}$$

Observing that E is the support function of the closed and convex set:

$$K := \{(a,b)\in\mathbb{R}\times\mathbb{R}^d\ :\ a + c^*(b)\leq 0\}$$

problem (14) appears naturally as the dual of

$$\inf_{\phi=\phi(t,x)}\{-\langle\phi,f\rangle\ :\ \nabla_{t,x}\phi = (\partial_t\phi,\nabla\phi)\in K\} \tag{15}$$

that is the maximization of the linear form:

$$\int_{\mathbb{R}^d}\phi(1,x)d\rho_1(x) - \int_{\mathbb{R}^d}\phi(0,x)d\rho_0(x)$$

among subsolutions of the Hamilton-Jacobi equation with Hamiltonian c^*:

$$\partial_t\phi + c^*(\nabla\phi)\leq 0.$$

This is again a convex variational problem in the form (1) with linear F and a singular G, the indicatrix of the convex K, which forces the space time gradient $\nabla_{t,x}\phi$ to stay in K and captures in a dual way the constraints of the time-dependent formulation of the mass transport problem.

At least formally, the primal-dual optimality conditions for (14)-(15) reads as the Hamilton-Jacobi/continuity equation system:

$$\begin{cases} \partial_t \phi + c^*(\nabla \phi) = 0, \\ \partial_t \rho + \operatorname{div}(\rho \nabla c^*(\nabla \phi)) = 0, \\ \rho_{|t=0} = \rho_0, \ \rho_{|t=1} = \rho_1. \end{cases}$$

Deterministic Mean-Field Games. The Mean-Field Games theory of Lasry and Lions ([LL06a], [LL06b], [LL07]) naturally leads, in the deterministic case, to coupled systems of Hamilton-Jacobi and continuity equations which generalise the system above:

$$\begin{cases} \partial_t \phi + H(t, x, \nabla \phi) = \alpha(t, x, \rho), \\ \partial_t \rho + \operatorname{div}(\rho \nabla H(t, x, \nabla \phi)) = 0, \\ \rho_{|t=0} = \rho_0, \ \phi_{|t=T} = -\gamma(x, \rho_T). \end{cases} \tag{16}$$

To make things simple, we consider the periodic framework where $x \in \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ and set $\Omega := (0,1)^d$ the periodicity cell. The Hamiltonian H is given by:

$$H(t, x, p) := \sup_v \{p \cdot v - L(t, x, v)\}$$

i.e. it is associated to the (convex in v) Lagrangian L. This system characterizes equilibria for a continuum of players, each of them solving an optimal control problem with a cost which depends on the density ρ of the other players. The Hamilton-Jacobi equation captures optimality for fixed ρ and the second equation means that ρ is obtained by transporting ρ_0 by the flow of the optimal feedback $\nabla H(t, x, \nabla \phi)$. Typically both α and γ are nondecreasing in ρ which captures congestion effects.

As emphasized by Lasry and Lions in [LL06b], the MFG system above is related to a variational problem which generalizes the dynamic formulation of optimal transport i.e. problem (14). More precisely, let us define

$$A(t, x, \rho) := \begin{cases} \int_0^\rho \alpha(t, x, s) ds \text{ if } \rho \geq 0 \\ +\infty \text{ otherwise} \end{cases}$$

and

$$\Gamma(x, \rho) := \begin{cases} \int_0^\rho \gamma(t, x, s) ds \text{ if } \rho \geq 0 \\ +\infty \text{ otherwise} \end{cases}$$

and consider the variational problem

$$\inf_{\rho, v} \int_0^T \int_\Omega [L(t, x, v)\rho + A(t, x, \rho)] dx dt + \int_\Omega \Gamma(x, \rho_T) dx \tag{17}$$

subject to the constraint that ρ is again related to v through the continuity equation:

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \rho_{|t=0} = \rho_0.$$

This is again (up to the same change of variables as in Benamou-Brenier) a convex optimization problem:

(18)
$$\inf_{\sigma=(\rho,m)} \int_0^T \int_\Omega [L\Big(t,x,\frac{m}{\rho}\Big)\rho(t,x) + A(t,x,\rho(t,x))]dxdt + \int_\Omega \Gamma(x,\rho_T(x))dx$$

(where, as in the Benamou-Brenier dynamic formulation of optimal transport, the functional above is extended by $+\infty$ whenever the momentum m does not vanish where ρ does) subject to the linear constraint

$$\partial_t\rho + \mathrm{div}(m) = 0,\ \rho_{|t=0} = \rho_0 \text{ with periodic boundary conditions.}$$

The convex problem (18) is dual to

$$\inf \int_0^T \int_\Omega A^*(t,x,\partial_t\phi + H(t,x,\nabla\phi))dxdt + \int_\Omega \phi_0\rho_0 + \int_\Omega \Gamma^*(x,-\phi_T)dx. \tag{19}$$

Problem (19) can again be put into the abstract class of problems (1) at the cost of considering a slightly more general operator Λ than the gradient[1]. More precisely, one can rewrite (19) in a similar form as (1) by setting $F(\phi) = \int_\Omega \phi_0\rho_0$ and

$$G(\Lambda\phi) = \int_0^T \int_\Omega A^*(t,x,\partial_t\phi + H(t,x,\nabla\phi))dxdt + \int_\Omega \Gamma^*(x,-\phi_T)\,dx$$

with Λ the linear operator

$$\Lambda\phi = (\partial_t\phi, \nabla\phi, -\phi(T,.)) = (\nabla_{t,x}\phi, -\phi(T,.)).$$

Of course (19) is still amenable to an Augmented Lagrangian numerical resolution as the CFD formulation of optimal transport.

At least formally, the Mean-Field Game system (16) corresponds to the primal-dual optimality conditions for (18)-(19). In fact, one has to be cautious about the regularity of ϕ and the fact that the Hamilton-Jacobi is satisfied only where $\rho > 0$. We refer to Graber [Gra14], Cardaliaguet and Graber [GC14], Carlier, Cardaliaguet and Nazaret [CCN13] for precise statements.

We have tested the ALG2 algorithm for the MFG problem using the quadratic Hamiltonian $H(b) = \frac{|b|^2}{2}$ and the following form for the functions A (running cost) and Γ (terminal cost):

$$N_{q,\gamma}(\rho_1,\rho) := \begin{cases} \gamma\dfrac{|\rho-\rho_1|^q}{q} \text{ if } \rho \geq 0, \\ +\infty \text{ otherwise.} \end{cases} \tag{20}$$

[1] as recalled in section 3, it is essential for the convergence of ALG2, that the operator Λ is injective, in mass transport problems, one can normalize potentials to have zero-mean, under this normalization, the gradient operator is injective, but one cannot impose zero-mean any more in the case of MFG's this is why one has to add the terminal value as part of the operator Λ to make it injective.

In our examples $q = 1, 2$ and ρ_1 a given arbitrary positive L^1 or L^2 target function. We show 4 examples set on $[0, T] \times [0, 1]^2$ (χ_ω is the characteristic function of the set ω) :

- Test Case 8 : $\rho_0 = 1.5\,\chi_{[0.25,0.75]^2}$, $A = 0$ and $\Gamma = N_{1,1}(0.5\,\chi_{[0.25,0.75]^2})$,
- Test Case 9 : $\rho_0 = 1.5\,\chi_{[0.25,0.75]^2}$, $A = 0$ and $\Gamma = N_{2,1}(0.5\,\chi_{[0.25,0.75]^2})$,
- Test Case 10 : $\rho_0 = 1.5\,\chi_{[0.25,0.75]^2}$, $A = N_{1,0.5}(0.5\,(1 - \chi_{[0.25,0.75]^2}))$ and $\Gamma = N_{1,1}(0.5\,(1 - \chi_{[0.25,0.75]^2}))$,
- Test Case 11 : $\rho_0 = 1.5\,\chi_{[0.25,0.75]^2}$, $A = N_{2,0.5}(0.5\,(1 - \chi_{[0.25,0.75]^2}))$ and $\Gamma = N_{2,1}(0.5\,(1 - \chi_{[0.25,0.75]^2}))$.

A time slice of the space time density $\rho(t, x)$ is shown on figure 4 for cases 8 and 9. The total mass of ρ_0 is larger than the total mass of ρ_1. We see the spreading/sharpening of the excess of mass induced by the L2/L1 norms penalisation on the $t = T$ boundary. Figure 5 corresponds to cases 10 and 11. The L1/L2 norms are now set in space and time and penalise sharply the support of ρ.

There are many natural generalizations and variants of (18). In particular one can prescribe the terminal density instead of having a terminal cost (see Buttazzo, Jimenez and Oudet [BJO09] where an augmented Lagrangian method is used). Lasry and Lions more generally considered the case of Mean-Field Games with diffusion, that is the system

$$\begin{cases} \partial_t\phi + \nu\Delta\phi + H(t, x, \nabla\phi) = \alpha(t, x, \rho), \\ \partial_t\rho - \nu\Delta\rho + \operatorname{div}(\rho\nabla H(t, x, \nabla\phi)) = 0, \\ \rho_{|_{t=0}} = \rho_0,\ \phi_{|_{t=T}} = -\gamma(x, \rho_T) \end{cases}$$

for a positive diffusion parameter ν. This MFG with diffusion system is related to the minimization of the energy (17) subject to the Fokker-Planck equation:

$$\partial_t\rho - \nu\Delta\rho + \operatorname{div}(\rho v) = 0, \rho_{|_{t=0}} = \rho_0.$$

We refer to the articles of Achdou and coauthors [ACCD12], [ACD10] and [ACCD13] for finite difference schemes directly based on the MFG system with diffusion. The Augmented Lagrangian strategy may also be applied in this diffusive case. However, a bilaplacian operator appears in the algorithm and generates new numerical difficulties as discussed in [AP12].

3. The augmented Lagrangian algorithm

3.1. **On finite-element approximations.** The detailed analysis of infinite-dimensional problems of the form (1)-(2) arising in transport optimization is beyond the scope of this paper. We shall rather study finite-dimensional approximations which are easier to analyze and for which augmented Lagrangian methods are guaranteed to converge under mild assumptions that are adapted to the singularities which typically arise in transport problems.

We therefore first discuss the convergence (in the sense of Γ-convergence) of approximations by finite elements which will be used in the numerical

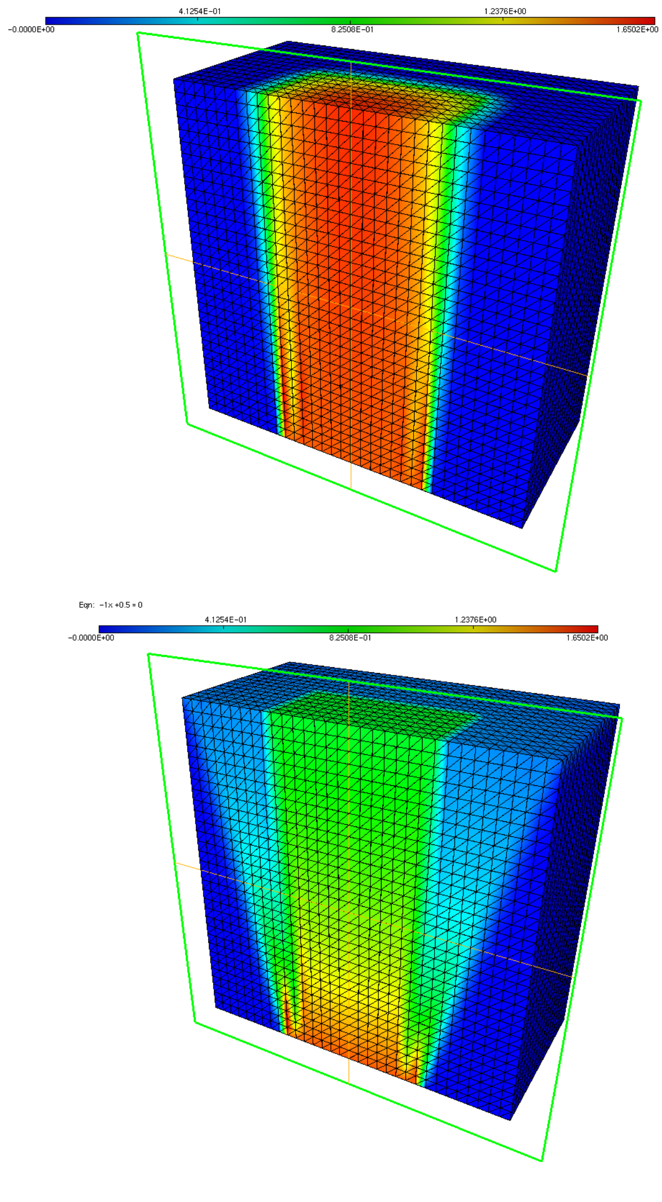

FIGURE 4. Test cases 8 and 9 : time slice of the obtained density for L1 final norm (top) and L2 final norm (bottom).

section. For the sake of simplicity, we shall restrict here ourselves to static problems with a linear F, i.e. we consider problems of the form

$$\inf_{\phi \in W^{1,p}(\Omega)} J(\phi) := G(\nabla \phi) - \langle f, \phi \rangle \tag{21}$$

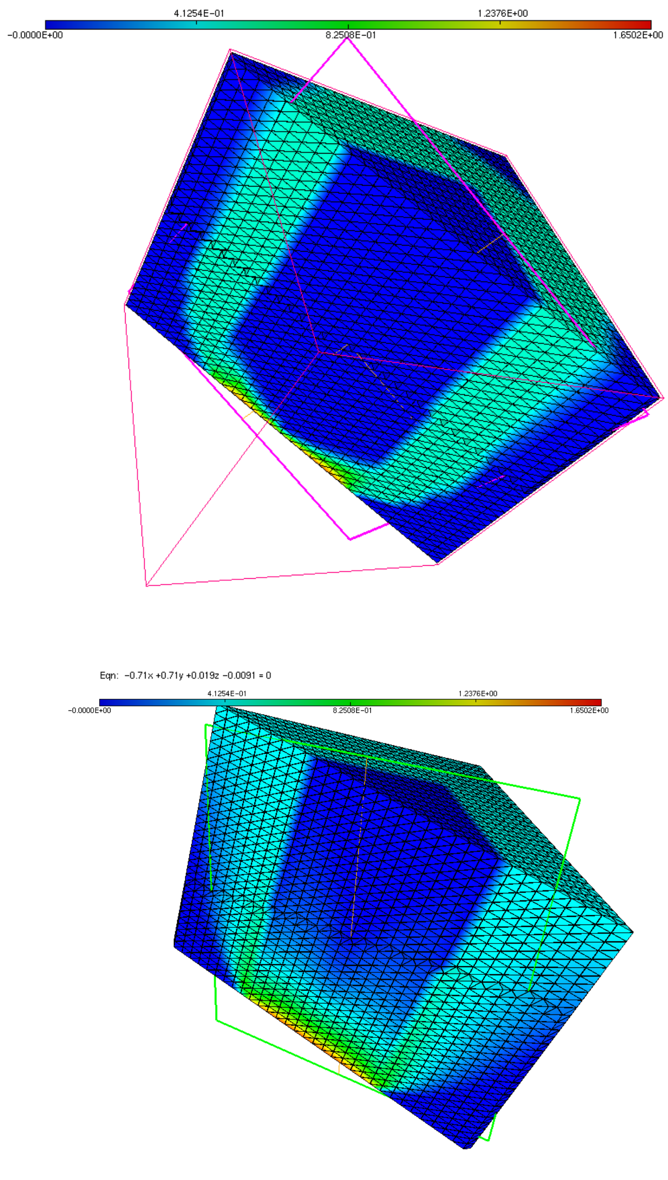

FIGURE 5. Test cases 10 and 11 : : time slice of the obtained density for L1 final norm (top) and L2 final norm (bottom).

where Ω is an bounded domain of $\mathbb{R}^d$ with Lipschitz boundary, $f = (\rho_1 - \rho_0)$ is in the dual of $W^{1,p}(\Omega)$, $\langle f, 1\rangle = 0$ (so that ϕ can be normalized so as to have zero mean) and

$$G(w) = \int_\Omega \mathcal{G}(x, w(x))dx, \ w \in L^p(\Omega)^d$$

where $\mathcal{G}$ is convex in its second argument, as in the examples of the previous paragraph.

We are now interested in discretizing (21) by finite elements (and then to solve this dicretization by an augmented Lagrangian method as we shall explain below). Given a regular triangulation of the domain with typical meshsize h, let $E_h \subset W^{1,p}(\Omega)$ be the corresponding finite-dimensional space of P_k finite elements of order k ($k \geq 1$ but in practice we'll take $k = 1$ or $k = 2$) whose generic elements are denoted ϕ_h. We approximate if necessary the linear form f by $f_h \in E_h$ (again with $\langle f_h, 1\rangle = 0$) and the nonlinear term G by a (convex) approximation G_h and consider

(22) $$\inf_{\phi_h \in E_h} J_h(\phi_h) := G_h(\nabla \phi_h) - \langle f_h, \phi_h \rangle$$

as well as its dual

(23) $$\sup_{\sigma_h \in F_h^d} \{-G_h^*(\sigma_h) \ : \ -\operatorname{div}_h(\sigma_h) = f_h\}$$

where F_h is the space of P_{k-1} finite elements and $-\operatorname{div}_h(\sigma_h)$ is of course defined by duality:

$$\langle \sigma_h, \nabla \phi_h \rangle_{F_h^d} = -\langle \operatorname{div}_h(\sigma_h), \phi_h \rangle_{E_h}.$$

As soon as:

(24) $$\frac{G_h(q_h)}{|q_h|} \to \infty \text{ as } |q_h|_{F_h^d} \to \infty$$

and the following qualification constraint holds

(25) there exists $\phi_h \in E_h$ such that G_h is continuous at $\nabla \phi_h$,

then, it follows from classical arguments that both (22) and (23) have solutions, the values of (22) and (23) coincide and solving the two problems amount to solve the primal-dual extremality relations:

(26) $$-\operatorname{div}_h(\sigma_h) = f_h, \ \sigma_h \in \partial G_h(\nabla \phi_h).$$

At this point, the natural question is whether (22) correctly approximates (21) in the sense of Γ-convergence (for the weak topology of $W^{1,p}$) which will in particular imply convergence of values and of minimizers. It is easy to see that the following assumptions (which are easy to check in the time-independent examples above, the time-dependent case being much more involved) guarantee convergence:

- density in energy of smooth functions: for every $\phi \in W^{1,p}(\Omega)$ such that $J(\phi) < +\infty$ and every $\varepsilon > 0$ there exists $\phi^\varepsilon \in C^k(\overline{\Omega})$ such that

(27) $$|J(\phi) - J(\phi^\varepsilon)| \leq \varepsilon$$

and the problem is not degenerate in the sense that there exists $\phi \in C^k(\overline{\Omega})$ such that $J(\phi) < +\infty$,

- consistency of approximation for smooth functions: for every $\phi \in C^k(\overline{\Omega})$ such that $J(\phi) < +\infty$, one has

$$J_h(I_h(\phi)) \to J(\phi) \text{ as } h \to 0, \tag{28}$$

where I_h is the usual Lagrange interpolation operator: $C^k(\overline{\Omega}) \to E_h$,

- Γ-liminf inequality: for every $\phi \in W^{1,p}(\Omega)$ and every sequence $\phi_h \in E_h$ that converges weakly to ϕ in $W^{1,p}(\Omega)$ (i.e ϕ_h converges strongly to ϕ in L^p, and $\nabla\phi_h$ converges weakly in $L^p(\Omega)^d$ to $\nabla\phi$ if $p \in (1,\infty)$, weakly $*$ in $L^\infty(\Omega)^d$ if $p = \infty$) one has

$$\liminf_{h\to 0} J_h(\phi_h) \geq J(\phi), \tag{29}$$

- equicoercivity: there exists a constant $\lambda > 0$ such that for every h and every $\phi_h \in E_h$ one has

$$J_h(\phi_h) \geq \lambda\Big(\|\nabla\phi_h\|_{L^p} - 1\Big). \tag{30}$$

We then have (also see Gabay and Mercier [GM76] for similar results and a more detailed discussion):

Proposition 3.1. *Under assumptions (24)-(25)-(27)-(28)-(29)-(30), if ϕ_h solves (22), it admits as $h \to 0$, a subsequence that converges weakly in $W^{1,p}(\Omega)$ to a ϕ that solves (21). In particular if (21) has a unique (up to an additive constant) solution, the whole sequence converges.*

3.2. **ALG2 and its convergence.** We stay in the framework of finite-dimensional convex optimization problems using the discretization by finite-elements of (1) which leads to a problem of the form :

$$\inf_{\phi\in\mathbb{R}^n} J(\phi) := F(\phi) + G(\Lambda\phi) \tag{31}$$

where F: $\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, G: $\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ are two convex lsc and proper functions and Λ is an $m \times n$ matrix with real entries. The dual of (31) then is:

$$\sup_{\sigma\in\mathbb{R}^m} -F^*(-\Lambda^T\sigma) - G^*(\sigma). \tag{32}$$

A pair $(\overline{\phi}, \overline{\sigma}) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to satisfy the primal-dual extremality relations if:

$$-\Lambda^T\overline{\sigma} \in \partial F(\overline{\phi}),\ \overline{\sigma} \in \partial G(\Lambda\overline{\phi}) \tag{33}$$

which implies that $\overline{\phi}$ solves (31) and that $\overline{\sigma}$ solves (32) as well as the fact that (31) and (32) have the same value (no duality gap). The primal-dual extremality relations are of course equivalent to finding a saddle-point of the Lagrangian

$$L(\phi, q, \sigma) := F(\phi) + G(q) + \sigma\cdot(\Lambda\phi - q),\ \ \forall(\phi, q, \sigma) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \tag{34}$$

in the sense that $(\overline{\phi}, \overline{\sigma}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfies (33) if and only if $(\overline{\phi}, \overline{q}, \overline{\sigma}) = (\overline{\phi}, \Lambda\overline{\phi}, \overline{\sigma})$ is a saddle-point of L. Now for $r > 0$, we consider the augmented Lagrangian function

$$L_r(\phi, q, \sigma) := F(\phi)+G(q)+\sigma\cdot(\Lambda\phi-q)+\frac{r}{2}|\Lambda\phi-q|^2, \ \ \forall(\phi, q, \sigma) \in \mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m \tag{35}$$

and recall (see for instance [FG83], [GM76]) that being a saddle-point of L is equivalent to being a saddle-point of L_r.

The augmented Lagrangian algorithm ALG2 splitting scheme (also known as ADMM: alternating direction method of multipliers), consists, starting from $(\phi^0, q^0, \sigma^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ to generate inductively a sequence (ϕ^k, q^k, σ^k) as follows:

- **Step 1:** minimization with respect to ϕ:

$$\phi^{k+1} := \text{argmin}_{\phi\in\mathbb{R}^n} \left\{F(\phi) + \sigma^k \cdot \Lambda\phi + \frac{r}{2}|\Lambda\phi - q^k|^2\right\} \tag{36}$$

- **Step 2:** minimization with respect to q:

$$q^{k+1} := \text{argmin}_{q\in\mathbb{R}^m} \left\{G(q) - \sigma^k \cdot q + \frac{r}{2}|\Lambda\phi^{k+1} - q|^2\right\} \tag{37}$$

- **Step 3:** update the multiplier by the gradient ascent formula

$$\sigma^{k+1} = \sigma^k + r(\Lambda\phi^{k+1} - q^{k+1}). \tag{38}$$

As emphasized by several authors (see for instance, Eckstein and Bertsekas ([EB92] or, more recently, Oudet, Papadakis and Peyré [PPO14]) ALG2 is a special case of the Douglas-Rachford splitting method for finding zeros of the sum of two maximal monotone operators. Convergence of the ALG2 iterates is guaranteed, under very general assumptions by the following result which is proved in Eckstein and Bertsekas ([EB92], Theorem 8) following contributions of the french mathematicians P.-L. Lions, Mercier, Glowinski, Gabay ([FG83], [LM79], [GM76]) to the analysis of splitting methods:

Theorem. *Let $r > 0$, assuming that Λ has full column-rank and that there exists a solution to the primal-dual extremality relations (33), then there exists an $(\overline{\phi}, \overline{\sigma}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying (33) such that the sequence (ϕ^k, q^k, σ^k) generated by the ALG2-scheme above satisfies*

$$\phi^k \to \overline{\phi}, \ q^k \to \Lambda\overline{\phi}, \ \sigma^k \to \overline{\sigma}, \quad \text{as } k \to \infty. \tag{39}$$

The use of ALG2 for transport problems was pioneered in [BB00]. The assumption that the matrix Λ has full column rank is essential, it is automatically satisfied in all our mass transport examples because in this case $\Lambda\phi$ is the gradient of ϕ and since ϕ is defined up to an additive constant we can normalize it so as to have zero mean, in which case the gradient map becomes injective, for MFG's one rather considers the injective operator $\Lambda\phi = (\nabla_{t,x}\phi, -\phi(T,.))$.

Remark 3.2. *The fact that the sequence σ^k generated by ALG2 remains in the domain of G^* follows directly from* (37) *and* (38)*. Indeed, it follows from Step 2 that G is subdifferentiable at q^{k+1} and*

$$\sigma^k + r(\Lambda\phi^{k+1} - q^{k+1}) \in \partial G(q^{k+1})$$

which with (38) *implies that $\sigma^{k+1} \in \partial G(q^{k+1})$. Since $G(q^{k+1}) < +\infty$ this gives that $G^*(\sigma^{k+1}) = \langle \sigma^{k+1}, q^{k+1}\rangle - G(q^{k+1})$ so that σ^{k+1} is in the domain of G^*. Let us recall that in the context of CFD or MFG's, the finiteness of $G^*(\sigma) = G^*(\rho, m)$ exactly means that $\rho \geq 0$ and $m = 0$ where $\rho = 0$ (see paragraph 2.2). This explains why ALG2 naturally takes into account the constraints on mass and momentum which arise in CFD and MFG's: contrary to other methods, ALG2 automatically ensures that mass remains nonnegative and that momentum vanishes where mass does. These singularities are well-known to be a problem for classical gradient descent methods when the density is not bounded away from* 0 *(see* [BB01] [AP12]*). The previous argument explains why ALG2 gives consistent results even in cases where the mass may vanish as was already observed in* [BB00]*.*

4. ALG2 steps for our problems

The continuous variational formulation of the Augmented Lagrangian and its Galerkin discretisation are convenient to interpret the three steps of the ALG2 algorithm. In what follows, the space dimension is $d = 2$.

4.1. **Static problems.** All the static variational problems we consider take the form

$$\inf_{\phi, q=(a,b)} - \int_\Omega f(x)\phi(x)\,dx + \int_\Omega \mathcal{G}(x, q(x))\,dx \tag{40}$$

subject to the constraint that $\nabla\phi = q$. The corresponding augmented Lagrangian takes the form

$$L_r(\phi, q, \sigma) := \int_\Omega \Big(- f(x)\,\phi(x) + \mathcal{G}(x, q(x)) + \langle \sigma, \nabla\phi - q\rangle + \frac{r}{2}|\nabla\phi - q|^2 \Big) dx$$

where $f = \rho_1 - \rho_0$. See section 2.1 for examples of functions $\mathcal{G}$.

- **Step 1:** can be interpreted as the variational formulation of Laplace equation :

$$- r(\Delta\phi^{k+1} - \mathrm{div}(q^k)) = f + \mathrm{div}(\sigma^k) \text{ in } \Omega \tag{41}$$

  together with the Neumann boundary condition

$$r\frac{\partial \phi^{k+1}}{\partial \nu} = rq^k \cdot \nu - \sigma^k \cdot \nu \text{ on } \partial\Omega. \tag{42}$$

Solving this problem is routine after the Galerkin discretisation of section 3.1.

- **Step 2:** For P_1 finite elements, at each vertex x_i, we have to solve the proximal problem:

$$q_i^{k+1} = \mathrm{argmin}_{(a,b)}\{\langle \mathcal{G}(x_i,(a,b)) - \langle \sigma_i^k,(a,b)\rangle + \frac{r}{2}|(\nabla\phi^k)_i - (a,b)|^2\}.$$

As detailed in the next section, the solutions are either explicit or simple projections to compute for the classes of $\mathcal{G}$ we consider.

- **Step 3:** is a straightforward update done at all vertices x_i.

4.2. **Time dependent problems.** Let us rewrite the variational problem (19) (for the sake of notational simplicity, we take H and F and G independent of (t,x)) arising in deterministic MFG's as

$$\inf_{\phi,q=(a,b,c)} \int_0^T\int_\Omega A^*(a+H(b))dxdt + \int_\Omega \phi_0\rho_0 + \int_\Omega \Gamma^*(c) \tag{43}$$

subject to the constraint $\Lambda\phi = (\partial_t\phi,\nabla\phi,-\phi_T) = q$. To the variables $(a,b,c)\in\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$, we associate the dual variables $\widetilde{\sigma} := (\rho,m,\widetilde{\rho}_T) = (\sigma,\widetilde{\rho}_T)$. The augmented Lagrangian then is:

$$\begin{aligned} L_r(\phi,q,\widetilde{\sigma}) = &\int_0^T\int_\Omega A^*(a+H(b))dxdt + \int_\Omega \Gamma^*(c) + \int_\Omega \phi_0\rho_0 \\ &+ \int_0^T\int_\Omega (\rho(\partial_t\phi - a) + m\cdot(\nabla\phi - b) - \int_\Omega \widetilde{\rho}_T(\phi_T + c) \\ &+\frac{r}{2}\Big(\|(\partial_t\phi,\nabla\phi) - (a,b)\|_{L^2}^2 + \int_\Omega |\phi_T + c|^2\Big). \end{aligned}$$

- **Step 1:** (minimization with respect to ϕ) amounts to solve the elliptic in (t,x) problem

$$-r\Delta_{t,x}\phi^{k+1} = \mathrm{div}_{t,x}(\sigma^k - r(a^k,b^k))$$

with ϕ^{k+1} periodic in x,

$$r\partial_t\phi^{k+1}(0,.) = \rho_0 - \rho^k(0,.) + ra^k(0,.)$$

and

$$r(\partial_t\phi^{k+1}(T,.) + \phi^{k+1}(T,.)) = \widetilde{\rho}_T^k(T,.) - \rho^k(T,.) + r(a^k(T,.) - c^k(T,.))$$

- **Step 2:** we have to solve two decoupled pointwise proximal subproblems:

$$\inf_c \Gamma^*(c) + \frac{r}{2}|c + \phi^{k+1}(T,.) - \frac{\widetilde{\rho}_T}{r}|^2$$

and

$$\inf_{(a,b)\in\mathbb{R}\times\mathbb{R}^d} A^*(a+H(b)) + \frac{r}{2}|a - \partial_t\phi^{k+1} - \frac{\rho^k}{r}|^2 + \frac{r}{2}|b - \nabla\phi^{k+1} - \frac{m^k}{r}|^2.$$

As in the static cases, this step translates after a P_1 discretisation into a finite number of explicit optimisation/projection at the vertices of the triangulation (see the Appendix for detailed computations in the case of the functions $N_{\gamma,q}$ used in our MFG's simulations).

- **Step 3:** is a straightforward update done at all vertices x_i.

5. Numerical convergence study

The numerical ALG2 method described in this paper has been implemented using the software FreeFem++ [2]. We use the Lagrangian finite elements and notations introduced in section 3.1, P_2 FE for ϕ_h and P_1 FE for (q_h, σ_h), $(\Lambda\phi_h)$ is the projection on P_1 of the operator Λ ($= \nabla\phi_h$ or $(\partial_t\phi_h, \nabla\phi_h, \phi_h(T, .))$). As emphasised in section 4, only the functional G and therefore the pointwise proximal/minimisation step 2 varies with our different test cases. Step 1 remains a Laplace equation in space for static problems or time and space for MFGs which can easily be implemented in FreeFem++ and step 3 is just an explicit update. This section presents the numerical convergence of ALG2 iterations indicated by the $.^k$ superscript, and the convergence of the Finite-Element discretisation indicated by the $\cdot_h$ subscript, where h is the characteristic size of the mesh elements.

5.1. **Static problems.** All static problems are computed on a triangulation of the unit square with $N = \frac{1}{h}$ element on each side. We use the following *Convergence* criteria :

- DIV_Error $= \left(\int_{\Omega_h} (\operatorname{div}\sigma_h^k + (\rho_1 - \rho_0))^2\right)^{\frac{1}{2}}$ is the L^2 error on the divergence constraint,
- BND_Error $= \left(\int_{\Gamma_h} (\sigma_h^k \cdot \nu)^2\right)^{\frac{1}{2}}$ is the $L^2(\Gamma^h)$ error on the Neumann boundary condition.
- DUAL_Error $= \max_{x_j} |\mathcal{G}^*(\sigma_h^k) + \mathcal{G}(\nabla\phi_h^k) - \nabla\phi_h^k \cdot \sigma_h^k|$, where the maximum is with respect to the vertices x_j.

The first two criteria correspond to the optimality conditions for the minimization of the Lagrangian with respect to ϕ and the third one corresponds to maximization with respect to σ.

5.1.1. *Monge's optimal transport problem.* Step 2 simply consists in

$$q_h^{k+1} = p_B\Big(\nabla\phi^{k+1} + \frac{\sigma^k}{r}\Big)$$

where p_B is the projection onto B:

$$p_B(z) = \begin{cases} z \text{ if } |z| \le 1 \\ \frac{z}{|z|} \text{ otherwise.} \end{cases}$$

[2] http://freefem.org

Figure 6 shows the convergence history for the DUAL error. The DIV and BND indicator become quickly stationary with k. The values are displayed in table 1 and show the convergence of the FE discretisation when the grid is refined and it roughly corresponds to a $O(h)$ consistency of the discretisation.

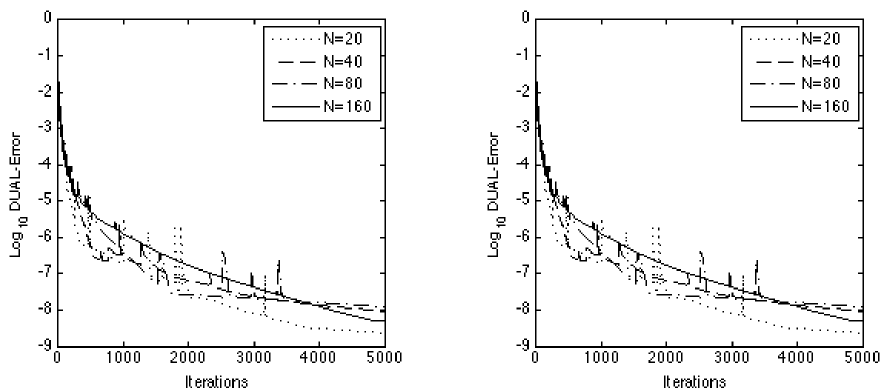

FIGURE 6. Test cases 1 (left) and 2 (right) : Convergence history of $\log_{10}$(DUAL_Error) versus ALG2 iterations and $N = 20, 40, 80, 160$.

N	DIV_Error	BND_Error	DIV_Error	BND_Error
20	2.0339e-02	6.4690e-03	6.3608e-04	5.4305e-03
40	9.2427e-03	8.8127e-04	1.5257e-04	1.7878e-03
80	2.8712e-03	4.2122e-04	3.9831e-05	7.3121e-04
160	1.0769e-03	1.3152e-04	9.5737e-06	3.9019e-04

TABLE 1. convergence of the finite element discretisation for test case 1 (left) and 2 (right).

5.1.2. *Variants of Monge's problem: heterogeneous media.* In this case, step 2 is simply modified to take into account that the projection now depends on the local weights $g_i = g(x_i)$ at the vertices x_i

$$q_{i,h}^{k+1} = p_{B_{g_i}}\left(\nabla \phi_{i,h}^{k+1} + \frac{\sigma_{i,h}^{k}}{r}\right)$$

where p_{B_g} is the projection on the ball of radius g :

$$p_{B_g} = \begin{cases} z \text{ if } |z| \leq g \\ g\,\frac{z}{|z|} \text{ otherwise.} \end{cases}$$

N	DIV_Error	BND_Error	DIV_Error	BND_Error
20	2.9528e-03	9.1262e-04	2.7689e-03	4.7240e-03
40	5.4175e-04	2.5407e-04	6.0847e-04	5.5129e-04
80	7.4324e-05	1.6041e-04	1.2336e-04	1.4319e-04
160	9.7279e-06	7.2257e-05	2.7790e-05	9.9710e-05

Table 2. convergence of the finite element discretisation for test case 3 (left) and 4 (right).

5.1.3. *Congested transport.* In the congested transport problem, the cost is of the form $\mathcal{G}(q) = \frac{1}{p}(|q| - \beta)_+^p$ so that step 2 of ALG2 requires to solve the pointwise problem

$$\inf_q \frac{1}{p}(|q| - \beta)_+^p + \frac{r}{2}|q - \widetilde{q}^k|^2$$

where $\widetilde{q}^k = \nabla\phi^{k+1} + \frac{\sigma^k}{r}$. This gives $q^{k+1} = \lambda\widetilde{q}^k$ where $\lambda \geq 0$ is the root (a dichotomy algorithm is preferred to Newton in this case) of the equation

$$(\lambda|\widetilde{q}^k| - \beta)_+^{p-1} + r\lambda|\widetilde{q}^k| = r|\widetilde{q}^k|. \tag{44}$$

We have omitted the plot of the Dual errors for test cases 6 and 7 (congested transport) as well as for test cases 3 and 4 (heterogeneous Monge) because they turn out to be significantly smaller than in test cases 1 and 2.

N	DIV_Error	BND_Error	DIV_Error	BND_Error
20	6.3333e-03	8.4775e-03	6.3333e-03	8.4775e-03
40	1.9069e-03	1.9369e-03	1.9069e-03	1.9369e-03
80	5.5455e-04	7.2514e-04	5.5455e-04	7.2514e-04
160	1.7073e-04	1.5939e-04	1.7073e-04	1.5939e-04

Table 3. convergence of the finite element discretisation for test case 6 (left) and 7 (right).

5.2. **Deterministic MFGs.** The dynamic problems are computed on a triangulation of the unit cube with $N = \frac{1}{h} = \frac{1}{dt}$ element on each side and we use periodic boundary conditions in space. We have tested our ALG2 algorithm on the MFG problem of section 2.2 with the quadratic Hamiltonian $H(b) = \frac{|b|^2}{2}$ (but other radially symmetric convex Hamiltonian can be treated as well) and as for the running and terminal costs A and Γ we took functions in the family $N_{q,\gamma}(\rho_1, .)$ defined by (20). The resolution of step 2 for the classical CFD optimal transport is well known. It is a pointwise projection on the convex $K = \{q = (a, b),\ a + \frac{1}{2}|b|^2 \leq 0\}$ and implementation details can be found in the literature. The detailed computations for step 2 of ALG2 in the MFG cases treated here involving the functions $N_{q,\gamma}(\rho_1, .)$ are detailed in the appendix.

We use the following *Convergence* criteria corresponding to the MFG optimality system (16).

- HJE_Error $= \left(\int_{\Omega_h} \rho^k (\partial_t \phi^k + H(t,x,\nabla\phi^k) - \alpha(t,x,\rho^k))^2 \right)^{\frac{1}{2}}$,
- DIV_Error $= \left(\int_{\Omega_h} \rho^k (\partial_t \rho^k + \mathrm{div}(\rho^k \nabla\phi^k))^2 \right)^{\frac{1}{2}}$,
- HJB_Error $= \left(\int_{t=1} \rho^k (\phi^k + \gamma(x,\rho^k))^2 \right)^{\frac{1}{2}}$.

We observe convergence of the FE discretisation (decrease with h of the error) and convergence (in k) of the ALG2 algorithm.

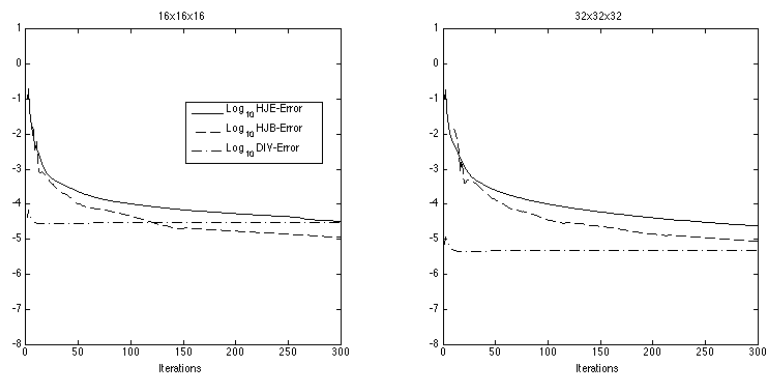

FIGURE 7. Test case 10 : Convergence history of the error indicators versus ALG2 iterations for $N = 16$ and 32.

6. Appendix : details of Step 2 for MFGs

Let $\gamma > 0$ and $\rho_1 \geq 0$ be given, then define

$$N_2(\rho) :=:= \begin{cases} \frac{\gamma}{2}(\rho-\rho_1)^2 \text{ if } \rho \geq 0 \\ +\infty \text{ otherwise} \end{cases}, \; N_1(\rho) := \begin{cases} \gamma|\rho-\rho_1| \text{ if } \rho \geq 0 \\ +\infty \text{ otherwise.} \end{cases}$$

For $N = N_1, N_2$, we have to compute: N^* as well as the two proximal operators:

- Terminal prox, given c_0, solve:

$$\inf_c \{N^*(c) + \frac{r}{2}|c-c_0|^2\}. \tag{45}$$

- Quadratic Hamiltonian prox (recall that we have taken for simplicity $H(p) = \frac{1}{2}|p|^2$ in the MFG), given $(a_0, b_0) \in \mathbb{R} \times \mathbb{R}^d$, solve

$$\inf_{(a,b)\in\mathbb{R}\times\mathbb{R}^d} \{N^*\Big(a + \frac{1}{2}|b|^2\Big) + \frac{r}{2}\Big(|a-a_0|^2 + |b-b_0|^2\Big)\}. \tag{46}$$

Proximal computations for N_2: The Legendre transform of N_2 is explicitly given by

$$N_2^*(\lambda) := \begin{cases} \frac{\lambda^2}{2\gamma} + \lambda\rho_1 \text{if } \lambda \geq -\gamma\rho_1 \\ -\gamma\frac{\rho_1^2}{2} \text{ otherwise.} \end{cases}$$

In this case, the solution of the terminal proximal-problem (45) is:

$$c = \begin{cases} c_0 \text{ if } c_0 \leq -\gamma\rho_1 \\ \frac{rc_0 - \rho_1}{r+\gamma^{-1}} \text{ otherwise} \end{cases} \tag{47}$$

Let us consider now the Hamiltonian-prox problem (46). It is convenient to formulate the optimality condition for (46) by setting

$$\lambda := (a + \frac{1}{2}|b|^2),\ r\mu := (N_2^*)'(\lambda) = (N_2^*)'\Big(a + \frac{1}{2}|b|^2\Big)$$

we then have

$$a = a_0 - \mu,\ b = \frac{b_0}{1+\mu}.$$

Defining $\lambda_0 = a_0 + \frac{1}{2}|b_0|^2$, the optimal (a, b) is given by

- case 1: $\lambda \geq -\gamma\rho_1$ then μ has to be a (nonnegative) root of the (cubic) equation

$$r\mu = \rho_1 + \frac{\lambda}{\gamma} = \rho_1 + \frac{1}{\gamma}\Big(a_0 - \mu + \frac{1}{2}\frac{|b_0|^2}{(1+\mu)^2}\Big) \tag{48}$$

  and the solvability of this equation on $\mathbb{R}_+$ is equivalent to $\lambda_0 \geq -\gamma\rho_1$.
- case 2: $\lambda_0 < -\gamma\rho_1$ then $(a, b) = (a_0, b_0)$.

Prox computations for N_1. The Legendre transform of N_1 is:

$$N_1^*(\lambda) := \begin{cases} -\gamma\rho_1 \text{ if } \lambda \leq -\gamma \\ \rho_1\lambda \text{ if } \lambda \in [-\gamma, \gamma] \\ +\infty \text{ otherwise.} \end{cases}$$

Rewriting the proximal-problem (45) as the inclusion $0 \in (c - c_0) + \frac{1}{r}\partial N_1^*(c)$ and distinguishing the different possible cases for $\partial N_1^*(c)$, one finds

$$c = \begin{cases} c_0 \text{ if } c_0 < -\gamma \\ -\gamma \text{ if } c_0 \in [-\gamma, -\gamma + \frac{\rho_1}{r}] \\ c_0 - \frac{\rho_1}{r} \text{ if } c_0 \in (-\gamma + \frac{\rho_1}{r}, \gamma + \frac{\rho_1}{r}) \\ \gamma \text{ if } c_0 \geq \gamma + \frac{\rho_1}{r}. \end{cases}$$

For the second problem (46) which corresponds to the conditions: $0 \in (a - a_0, b - b_0) + \frac{1}{r}\partial N_1^*(\lambda)(1, b)$, with $\lambda = a + \frac{1}{2}|b|^2$, we find as optimal (a, b):

$$(a, b) = \begin{cases} (a_0, b_0) \text{ if } \lambda_0 < -\gamma \\ (a(\mu), b(\mu)) \text{ with } \mu \in [0, \frac{\rho_1}{r}] \text{ solving (49) (with } - \text{ sign) if } \lambda_0^* \leq -\gamma \leq \lambda_0, \\ (a_0^*, b_0^*) \text{ if } \lambda_0^* \in (-\gamma, \gamma) \\ (a(\mu), b(\mu)) \text{ with } \mu \geq \frac{\rho_1}{r} \text{ solving (49) (with } + \text{ sign) if } \lambda_0^* \geq \gamma \end{cases}$$

where we have defined

$$a_0^* := a_0 - \frac{\rho_1}{r},\ b_0^* := \frac{rb_0}{\rho_1 + r},\ \lambda_0^* := a_0^* + \frac{1}{2}|b_0^*|^2$$

$$a(\mu) := (a_0 - \mu),\ b(\mu) := \frac{b_0}{1+\mu},$$

$$\pm\gamma = (a_0 - \mu) + \frac{1}{2}\frac{|b_0|^2}{(1+\mu)^2}. \tag{49}$$

Acknowledgements: The authors gratefully acknowledge the support of the ANR, through the project ISOTACE (ANR-12-MONU-0013) and INRIA through the "action exploratoire" MOKAPLAN.

References

[ACCD12] Yves Achdou, Fabio Camilli, and Italo Capuzzo-Dolcetta. Mean field games: numerical methods for the planning problem. *SIAM J. Control Optim.*, 50(1):77–109, 2012.

[ACCD13] Yves Achdou, Fabio Camilli, and Italo Capuzzo-Dolcetta. Mean field games: convergence of a finite difference method. *SIAM J. Numer. Anal.*, 51(5):2585–2612, 2013.

[ACD10] Yves Achdou and Italo Capuzzo-Dolcetta. Mean field games: numerical methods. *SIAM J. Numer. Anal.*, 48(3):1136–1162, 2010.

[Amb03] Luigi Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces (Funchal, 2000)*, volume 1812 of *Lecture Notes in Math.*, pages 1–52. Springer, Berlin, 2003.

[AP12] Yves Achdou and Victor Perez. Iterative strategies for solving linearized discrete mean field games systems. *Netw. Heterog. Media*, 7(2):197–217, 2012.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BB01] J. D. Benamou and Y. Brenier. Mixed L^2-Wasserstein optimal mapping between prescribed density functions. *J. Optim. Theory Appl.*, 111(2):255–271, 2001.

[BCS10] Lorenzo Brasco, Guillaume Carlier, and Filippo Santambrogio. Congested traffic dynamics, weak flows and very degenerate elliptic equations. *J. Math. Pures Appl. (9)*, 93(6):652–671, 2010.

[Ben03] Jean-David Benamou. Numerical resolution of an "unbalanced" mass transport problem. *M2AN Math. Model. Numer. Anal.*, 37(5):851–868, 2003.

[BJO09] G. Buttazzo, C. Jimenez, and E. Oudet. An optimization problem for mass transportation with congested dynamics. *SIAM J. Control Optim.*, 48(3):1961–1976, 2009.

[BP07] John W. Barrett and Leonid Prigozhin. A mixed formulation of the Monge-Kantorovich equations. *M2AN Math. Model. Numer. Anal.*, 41(6):1041–1060, 2007.

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[CCN13] P. Cardaliaguet, G. Carlier, and B. Nazaret. Geodesics for a class of distances in the space of probability measures. *Calc. Var. Partial Differential Equations*, 48(3-4):395–420, 2013.

[CJS08] G. Carlier, C. Jimenez, and F. Santambrogio. Optimal transportation with traffic congestion and Wardrop equilibria. *SIAM J. Control Optim.*, 47(3):1330–1350, 2008.

[DPP04] Luigi De Pascale and Aldo Pratelli. Sharp summability for Monge transport density via interpolation. *ESAIM Control Optim. Calc. Var.*, 10(4):549–552, 2004.

[EB92] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992.

[EG99] L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137(653):viii+66, 1999.

[FG83] Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

[GC14] Philip Jameson Graber and Pierre Cardaliaguet. Mean field games systems of first order. *ESAIM: Control, Optimisation, and Calculus of Variations*, 2014.

[GM75] R. Glowinski and A. Marrocco. Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de dirichlet non-linéaires. *Revue Francaise d'Automatique, Informatique et Recherche Opérationnelle*, 9:41–76, 1975.

[GM76] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element methods. *Computers and Mathematics with applications*, 2:17–40, 1976.

[Gra14] Philip Jameson Graber. Optimal control of first-order hamilton-jacobi equations with linearly bounded hamiltonian. *Applied Mathematics and Optimization*, 2 2014.

[HMP05] Romain Hug, Emmanuel Maitre, and Nicolas Papadakis. Multi-physics Optimal Transportation and Image Interpolation. 2014-05.

[LL06a] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.

[LL06b] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006.

[LL07] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[LM79] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[PPO14] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[San09] Filippo Santambrogio. Absolute continuity and summability of transport densities: simpler proofs and new estimates. *Calc. Var. Partial Differential Equations*, 36(3):343–354, 2009.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58. AMS Bookstore, 2003.

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.