HAL
open science

An augmented Lagrangian approach to Wasserstein gradient flows and applications

Jean-David Benamou, Guillaume Carlier, Maxime Laborde

To cite this version:

Jean-David Benamou, Guillaume Carlier, Maxime Laborde. An augmented Lagrangian approach to Wasserstein gradient flows and applications. ESAIM: Proceedings and Surveys, 2019. hal-01245184

HAL Id: hal-01245184

https://hal.science/hal-01245184v1

Submitted on 16 Dec 2015

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

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

An augmented Lagrangian approach to Wasserstein gradient flows and applications.

Jean-David Benamou[*†] , Guillaume Carlier [†*] , Maxime Laborde[‡†*]

December 16, 2015

Abstract

Taking advantage of the Benamou-Brenier dynamic formulation of optimal transport, we propose a convex formulation for each step of the JKO scheme for Wasserstein gradient flows which can be attacked by an augmented Lagrangian method which we call the ALG2-JKO scheme. We test the algorithm in particular on the porous medium equation. We also consider a semi implicit variant which enables us to treat nonlocal interactions as well as systems of interacting species. Regarding systems, we can also use the ALG2-JKO scheme for the simulation of crowd motion models with several species.

Keywords: Benamou-Brenier formulation, augmented Lagrangian, nonlinear diffusion, granular media, systems, crowd motions, Wasserstein gradient flows, semi-implicit JKO scheme.

MS Classification 2010: 49M29, 65M60.

1 Introduction

It is well-known since the seminal work of Jordan Kinderlehrer and Otto [25] that the Fokker-Planck equation

$$\partial_t \rho = \Delta \rho + \mathrm{div}(\rho \nabla V),\ \rho|_{t=0} = \rho_0 \tag{1.1}$$

[*]INRIA Rocquencourt, Mokaplan, Domaine de Voluceau, 78153 Le Chesnay, FRANCE. Jean-David.Benamou@inria.fr.

[†]CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE carlier@ceremade.dauphine.fr

[‡]laborde@ceremade.dauphine.fr.

where the initial condition ρ_0 is a probability density may be viewed as the Wasserstein gradient flow of the (relative) entropy functional

$$S_V(\rho) := \int_{\mathbb{R}^d} \rho(x) \log\left(\frac{\rho(x)}{e^{-V(x)}}\right) \mathrm{d}x. \tag{1.2}$$

More generally, given an internal energy E, a potential V and an interaction potential W, evolution equations of the form

$$\partial_t \rho = \mathrm{div}(\rho \nabla(E'(\rho) + V + W \star \rho)),\ \rho|_{t=0} = \rho_0 \tag{1.3}$$

is the Wasserstein gradient flow of the energy

$$\mathcal{E}(\rho) := \int_{\mathbb{R}^d} E(\rho(x))\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\rho(x)\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y.$$

For instance, if $E(\rho) = \frac{1}{m-1}\rho^m$ and $V = W = 0$ one in particular recovers the porous medium equation $\partial_t \rho = \Delta \rho^m$, see the seminal work of Otto [34]. Convolution terms $\nabla W \star \rho$ in (1.3) arise naturally in aggregation equations [16] and models of granular media [18], [19].

The celebrated Jordan-Kinderlehrer-Otto (henceforth JKO) scheme consists, given a time-step $\tau > 0$ in constructing inductively, starting from ρ_0 a sequence of probability measures ρ^k by the implicit Euler scheme:

$$\rho^{k+1} \in \mathrm{argmin}_{\rho\in\mathcal{P}_2}\left\{\frac{1}{2\tau}W_2^2(\rho,\rho^k) + \mathcal{E}(\rho)\right\} \tag{1.4}$$

where $\mathcal{P}_2$ denotes the set of probability measures on $\mathbb{R}^d$ having finite second moments and W_2^2 is the squared 2-Wasserstein distance defined for every $(\rho,\nu) \in \mathcal{P}_2 \times \mathcal{P}_2$ by

$$W_2(\rho,\nu) := \inf_{\gamma\in\Pi(\rho,\nu)}\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \mathrm{d}\gamma(x,y)\right\}^{\frac{1}{2}}$$

where $\Pi(\mu,\nu)$ is the set of transport plans between ρ and ν i.e. the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as marginals. When $\mathcal{E} = S_V$ is given by (1.2), Jordan, Kinderlehrer and Otto [25] proved that one recovers the solution of the Fokker-Planck equation (1.1) by letting τ tend to 0 in the JKO scheme. Similar convergence results hold for the more general equation (1.3) under suitable assumptions on E, V and W. The theory of Wasserstein gradient flows is by now well-developed and it is detailed in the textbooks of Ambrosio, Gigli and Savaré [1], Villani [38], [39] and Santambrogio [37].

We remark that the JKO scheme (1.4) is constructive and it is very natural and tempting to try to apply it for numerical purposes. The positivity,

mass conservation and energy dissipation are inbuilt in the JKO scheme and non trivial to preserve with non-linear finite-difference or finite volume schemes (see [14] and references therein). Also some JKO gradient flows, like congested crowd motions [31], cannot be formulated as nonlinear PDEs and the JKO semi-discretisation is the only numerical option.

A serious difficulty with this approach is in the Wasserstein term which involves solving a costly optimal transport problem at each step. In dimension one, this is not really an issue since the optimal transport is essentially a rearrangement problem, and in fact, this 1-D numerical approach was proposed in the early work of Kinderlehrer and Walkington [26] and was used repeatedly. See in particular the recent work of Osberger and Matthes [33] for application to fourth-order evolution PDEs of thin films type. In higher dimensions, the optimisation problem in (1.4) is much more complicated because the optimal transport is given by Brenier's map, the gradient of a convex potential which solves some Monge-Ampère equation.

At least three categories of approaches have been followed to solve (1.4) numerically.

A first "Lagrangian" strategy based on Brenier's Theorem is to formulate the problem in terms of the transport map or its potential instead of the density ρ to avoid dealing with the positivity and mass constraints. It also allows a more consistent discretisation of the mass when the density concentrates or dilates. This is done for instance in Carrillo and Moll [17] who proposed a Lagrangian scheme, based on a gradient flow for evolving diffeomorphisms (not necessarily the optimal transport maps) related to a system of evolution equations which is very nonlinear since it involves cofactors. Osberger and Matthes [30] introduce a Galerkin discretisation of the potential. As illustrated in [24], where another Lagrangian method is introduced, a difficulty with the Lagrangian approach is the construction of a discrete density to be used in the internal energy. A semi-discrete solution to this problem has been proposed in [3] based on optimal maps, a discretisation of the Monge-Ampère operator and techniques of computational geometry. This method is provably convergent and enables one to use of a Newton method. Note that using monotone finite difference Monge-Ampère solvers as introduced in [10], [9] could be an option but it does not seem to have been tried.

A second strategy, which is Eulerian, is to use the Monge-Kantorovich linear relaxation of the Wasserstein distance in (1.4). The size of the discretisation is very limited by the linear programming approach. However, Peyré [36] recently generalised entropic regularisation techniques that are computationally efficient in optimal transport [8] to treat JKO gradient flows.

In the present paper we investigate a third approach, also Eulerian, based on replacing the Wasserstein distance with the Benamou-Brenier formulation

[5]. This idea has already been used in [11], [12], [2], [4] either for JKO steps or in optimisation problems where the Wasserstein distance intervenes. Our original contribution, initiated in [7] on a general class of Optimal Transport problems and variational Mean Field Games [29], is to extend this convex reformulation and its augmented Lagrangian numerical resolution based on the algorithm ALG2 of Glowinski and Fortin [22] to solve a succession of problems of the form (1.4), we will call this method ALG2-JKO. We also show that the method can be adapted to treat systems (which are not necessarily gradient flows) using the relaxation introduced in [20] by Di Francesco and Fagioli and extended to the diffusive case in [15].

The Benamou-Brenier formulation induces an extra time dimension in each of the JKO steps. The resulting extra cost because of the discretisation of the inner (Benamou-Brenier) time dimension is usually considered a draw back. In the ALG2-JKO scheme however, since the successive JKO density time snapshots are close only a very few inner timesteps are needed in practice. The ALG2 augmented Lagrangian method is very robust, can deal with non-smooth energies but remains a proximal splitting first order method and converges slowly [35].

The paper is organized as follows. In section 2, we describe the ALG2-JKO scheme. In section 3, we illustrate the algorithm on two examples: the porous medium equation and a model of crowd motion with diffusion introduced by Santambrogio and Mészáros in [32]. Section 4 addresses the case of interaction terms, as in aggregation or granular media equations by a semi-implicit variant of the ALG2-JKO scheme. Finally, section 5 extends the method to several systems: interacting species (which are not gradient flows) and diffusive models of crowd motions à la Santambrogio and Mészáros with several populations.

2 The ALG2-JKO scheme

Let us consider one step of the JKO scheme (1.4) in the case where the energy $\mathcal{E}$ is of the form

$$\mathcal{E}(\rho) := \int_{\mathbb{R}^d} E(\rho(x))\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\rho(x)\mathrm{d}x$$

with E a convex internal energy (typical cases being the entropy or a convex power), which corresponds to the time discretization of the PDE:

$$\partial_t \rho = \mathrm{div}(\rho \nabla (E'(\rho) + V)),\ \rho|_{t=0} = \rho_0. \tag{2.1}$$

Our goal is to rewrite (1.4) as a tractable convex problem. To do so, we use the Benamou-Brenier dynamic formula [5] to rewrite W_2^2 as

$$W_2^2(\rho,\nu) := \inf\left\{\int_0^1\int_{\mathbb{R}^d}\frac{|m_t(x)|^2}{\mu_t(x)}\mathrm{d}x\mathrm{d}t \ : \ \partial_t\mu+\mathrm{div}(m)=0,\ \mu|_{t=0,1}=\rho,\nu\right\} \tag{2.2}$$

which is a convex variational program (it is implicit that the energy above is set to $+\infty$ whenever μ becomes negative or when $\mu=0$ and $m\neq 0$ so that momentum m can be written as $m=\mu v$ that is m vanishes where μ does and then $|m|^2/\mu=\mu|v|^2$ is the kinetic energy).

Thanks to (2.2) one can rewrite one step of the JKO scheme (1.4) as the convex minimization:

$$\inf_{(\mu_t,m_t,\mu_1=\mu_t(1,.))}\frac{1}{2\tau}\int_0^1\int_{\mathbb{R}^d}\frac{|m_t(x)|^2}{\mu_t(x)}\mathrm{d}x\mathrm{d}t+\mathcal{E}(\mu_1) \tag{2.3}$$

subject to the constraints that $\mu\geq 0$, $m=0$ when $\mu=0$ and the linear constraint

$$\partial_t\mu+\mathrm{div}(m)=0,\ \mu|_{t=0}=\rho^k. \tag{2.4}$$

One then recovers $\rho^{k+1}=\mu_1$ (and actually even an interpolation $(\mu_t)_{t\in[0,1]}$ between ρ^k and ρ^{k+1}).

Of course we can consider variants, for instance the periodic (in space) case or the case of a smooth bounded domain Ω of $\mathbb{R}^d$. In the latter case, we have to supplement the PDE (2.1) with the Neumann boundary condition:

$$\nabla(E'(\rho)+V))\cdot\nu=0,\ \text{on } \partial\Omega \tag{2.5}$$

this amounts to modify (2.3)-(2.4) as

$$\inf_{(\mu_t,m_t)}\frac{1}{2}\int_0^1\int_{\Omega}\frac{|m_t(x)|^2}{\mu_t(x)}\mathrm{d}x\mathrm{d}t+\tau\mathcal{E}(\mu_1) \tag{2.6}$$

subject to the constraints that $\mu\geq 0$, $m=0$ when $\mu=0$ and the linear constraint

$$\partial_t\mu+\mathrm{div}(m)=0,\ \mu|_{t=0}=\rho^k,\ m\cdot\nu=0 \text{ on } \partial\Omega. \tag{2.7}$$

2.1 Augmented Lagrangian formulation

Convex time-dependent problems like (2.6) subject to a divergence constraint (2.7) appear in various contexts, they are actually particular cases of deterministic Mean-Field Games (a class of games with a continuum of players introduced by Lions and Lasry [27], [28]). Such problems can be solved

by Augmented Lagrangian methods, see in particular [7] for applications to Mean-Field Games, Papadakis, Peyré and Oudet [35] for connections with proximal schemes and Buttazzo, Jimenez and Oudet [13] for applications to congested transport. We now recall the principle of the Augmented Lagrangian approach and explain how to use it in the JKO framework.

As was observed by Benamou and Brenier [5] the convex lsc 1-homogeneous function defined for $(\mu, m) \in \mathbb{R} \times \mathbb{R}^d$ by:

$$\Phi(\mu, m) := \begin{cases} \frac{|m|^2}{\mu}, \text{ if } \mu > 0, \\ 0, \text{ if } \mu = 0 \text{ and } m = 0 \\ +\infty, \text{ otherwise.} \end{cases}$$

is the support function of the convex set

$$K := \{(a, b) \in \mathbb{R}^{d+1},\ a + \frac{1}{2}|b|^2 \leq 0\} \tag{2.8}$$

i.e.

$$\Phi(\mu, m) = \sup_{(a,b)\in K} \{a\mu + b \cdot m\}.$$

Our convex problem (2.6)-(2.7) then is dual to:

$$\inf_{\phi=\phi(t,x)} \{\int_\Omega \phi(0, .)\rho^k + \tau\mathcal{E}^*\Big(-\frac{\phi(1, .)}{\tau}\Big)\ :\ \partial_t\phi + \frac{1}{2}|\nabla\phi|^2 \leq 0\} \tag{2.9}$$

where $\mathcal{E}^*$ is the Legendre tranform of $\mathcal{E}$ (extended by $+\infty$ on $(-\infty, 0]$):

$$\mathcal{E}^*(c) := \sup_{\mu\geq 0}\{\int_\Omega ((c(x) - V(x))\mu(x) - E(\mu(x))\mathrm{d}x\} = \int_\Omega \mathcal{E}^*(x, c(x))\mathrm{d}x$$

where, slightly abusing notations, we have set

$$\mathcal{E}^*(x, c) := \sup_{\mu\geq 0}\Big\{(c - V(x))\mu - E(\mu)\Big\}.$$

We then rewrite the dual as

$$\inf_{\phi=\phi(t,x)} J(\phi) := F(\phi) + G(\Lambda\phi) \tag{2.10}$$

where

$$\Lambda\phi := (D\phi, -\phi(1, .)) = ((\partial_t\phi, \nabla\phi), -\phi(1, .)),\ F(\phi) = \int_\Omega \phi(0, .)\rho^k$$

and for $q = (a, b, c)$

$$G(q) = \int_0^1 \int_\Omega \chi_K(a, b)\mathrm{d}x\mathrm{d}t + \tau\mathcal{E}^*\Big(\frac{c}{\tau}\Big)$$

where χ_K denotes the indicator function

$$\chi_K(a, b) = \begin{cases} 0, \text{ if } (a, b) \in K \\ +\infty, \text{otherwise.} \end{cases}$$

Now the variables $\sigma := (\mu, m, \mu_1)$ play the role of Lagrange multipliers associated to the constraint $q = \Lambda\phi$ i.e. $a = \partial_t\phi$, $m = \nabla\phi$ and $c = -\phi(1, .)$, note in particular that μ_1 is a multiplier associated to the constraint $c = -\phi(1, .)$ it coincides with $\mu(1, .)$ for the saddle-point but not necessarily along the iterations of the augmented Lagrangian algorithm below.

The primal-dual extremality relations are formally equivalent to finding a saddle-point of the Lagrangian

$$L(\phi, q, \sigma) := F(\phi) + G(q) + \sigma \cdot (\Lambda\phi - q), \tag{2.11}$$

in the sense that (ϕ, σ) satisfies the optimality conditions of (2.10) and (2.6)-(2.7) respectively if and only if

$$(\phi, q, \sigma) = (\phi, \Lambda\phi, \sigma)$$

is a saddle-point of L. Now for $r > 0$, we consider the augmented Lagrangian function

$$L_r(\phi, q, \sigma) := F(\phi) + G(q) + \sigma \cdot (\Lambda\phi - q) + \frac{r}{2}|\Lambda\phi - q|^2 \tag{2.12}$$

where $q = (a, b, c)$, $\sigma = (\mu, m, \mu_1)$,

$$\sigma \cdot (\Lambda\phi - q) = \int_0^1 \int_\Omega \Big(\mu(t, x)(\partial_t\phi(t, x) - a(t, x)) + m(t, x) \cdot (\nabla\phi(t, x) - b(t, x))\Big)\mathrm{d}x\mathrm{d}t + \int_\Omega \mu_1(x)(-\phi(1, x) - c(x))\mathrm{d}x$$

and

$$|\Lambda\phi - q|^2 = \int_0^1 \int_\Omega (|\partial_t\phi(t, x) - a(t, x)|^2 + |\nabla\phi(t, x) - b(t, x)|^2)\mathrm{d}x\mathrm{d}t + \int_\Omega (\phi(1, x) + c(x))^2\mathrm{d}x$$

and recall (see for instance [22], [23]) that being a saddle-point of L is equivalent to being a saddle-point of L_r.

The augmented Lagrangian algorithm ALG2 consists, starting from (ϕ^0, q^0, σ^0) to generate inductively a sequence (ϕ^n, q^n, σ^n) as follows:

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

$$\phi^{n+1} := \mathrm{argmin}_\phi \left\{ F(\phi) + \sigma^n \cdot \Lambda\phi + \frac{r}{2}|\Lambda\phi - q^n|^2 \right\}, \tag{2.13}$$

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

$$q^{n+1} := \mathrm{argmin}_q \left\{ G(q) - \sigma^n \cdot q + \frac{r}{2}|\Lambda\phi^{n+1} - q|^2 \right\}, \tag{2.14}$$

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

$$\sigma^{n+1} = \sigma^n + r(\Lambda\phi^{n+1} - q^{n+1}). \tag{2.15}$$

The convergence of ALG2 to a saddle-point is well documented see in particular the general results of Bertsekas and Eckstein [21] in finite dimensions. We therefore have to understand that in the problems above, we have already projected the potentials in (2.10) on a finite-dimensional space of finite-elements and therefore deal with a finite-dimensional problem for which existence of a saddle-point is rather standard and convergence follows from [21].

2.2 Details for the three steps

Step 1 corresponds to a linear elliptic problem in t and x,

$$-r\Delta_{t,x}\phi^{n+1} = \mathrm{div}_{t,x}((\mu^n, m^n) - r(a^n, b^n)), \text{ in } (0,1) \times \Omega, \tag{2.16}$$

together with the boundary conditions

$$r\partial_t\phi^{n+1}(0,.) = \rho^k - \mu^n(0,.) + ra^n(0,.), \tag{2.17}$$

$$r(\partial_t\phi^{n+1}(1,.) + \phi^{n+1}(1,.)) = \mu_1^n - \mu^n(1,.) + r(a^n(1,.) - c^n(.)), \tag{2.18}$$

and

$$r\frac{\partial\phi^{n+1}}{\partial\nu} + (m^n - rb^n) \cdot \nu = 0 \text{ on } \partial\Omega. \tag{2.19}$$

(or periodic boundary conditions if Ω is replaced by the flat torus).

Step 2 splits into two convex pointwise (i.e. for every t and x) minimization subproblems, the first one (minimization with respect to (a,b)) is a projection problem onto the parabola K:

$$\inf_{(a,b)\in K} |D\phi^{n+1}(t,x) + \frac{1}{r}(\mu^n(t,x), m^n(t,x)) - (a,b)|^2 \tag{2.20}$$

i.e.

$$(a^{n+1}(t,x), b^{n+1}(t,x)) = P_K\Big(D\phi^{n+1}(t,x) + \frac{1}{r}(\mu^n(t,x), m^n(t,x))\Big)$$

where the projection P_K onto K is explicit (see [5] or [35]):

$$P_K(\alpha,\beta) = \begin{cases} (\alpha,\beta), \text{ if } (\alpha,\beta)\in K, \\ (\alpha-\lambda, \frac{\beta}{1+\lambda}), \text{ with } \lambda>0 \text{ root of (2.21) otherwise} \end{cases}$$

where (2.21) is the cubic equation (with a single positive root if $(\alpha,\beta)\notin K$):

$$\alpha(1+\lambda)^2 - \lambda(1+\lambda)^2 + \frac{1}{2}|\beta|^2 = 0. \tag{2.21}$$

The second subproblem gives the update for c which is obtained by solving for each $x\in\Omega$

$$c^{n+1}(x) = \text{argmin}_{c\in\mathbb{R}}\left\{\frac{r}{2}|\phi^{n+1}(1,x) - \frac{1}{r}\mu_1^n(x) + c|^2 + \tau\mathcal{E}^*\Big(x, \frac{c}{\tau}\Big)\right\}. \tag{2.22}$$

Remark 2.1. Given a convex l.s.c. function f: $\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, we recall that the proximal operator of f, prox_f is defined by

$$\text{prox}_f(y_0) := \text{argmin}_{y\in\mathbb{R}^n}\left\{\frac{1}{2}|y-y_0|^2 + f(y)\right\},\ \forall y_0\in\mathbb{R}^n$$

so that (2.22) can be rewritten as

$$c^{n+1}(x) = \text{prox}_{\frac{\tau}{r}\mathcal{E}^*(x,\frac{\cdot}{\tau})}\Big(-\phi^{n+1}(1,x) + \frac{1}{r}\mu_1^n(x)\Big).$$

Thanks to the well-known (and actually easy to check) Moreau's identity

$$\text{prox}_f(y) = y - \text{prox}_{f^*}(y), \tag{2.23}$$

we see that it is not necessary to compute E^* to solve (2.22) if the computation of prox_E turns out to be easier.

Step 3 is an explicit update which may be detailed as

$$\mu^{n+1} = \mu^n + r(\partial_t\phi^{n+1} - a^{n+1}), \tag{2.24}$$

$$m^{n+1} = m^n + r(\nabla\phi^{n+1} - b^{n+1}), \tag{2.25}$$

$$\mu_1^{n+1} = \mu_1^n - r(\phi^{n+1}(1,.) + c^{n+1}(.)). \tag{2.26}$$

Note that only the minimization with respect to c (2.22) in step 2 depends on the form of the energy $\mathcal{E}$, we shall give details for this step in each application given in the sequel.

In the Finite Element discretisation of this algorithm, we use $P2$ FE (in time and space) for ϕ and $P1$ FE for σ so that in (2.20), in fact, one has to understand $D\phi^{n+1}$ as its projection onto $P1$ FE. It was implemented in FreeFeem++[1]. In practice, a discretization with 32×32 triangles in space and 4 inner timesteps needs a few hundreds iterations of ALG2 for each JKO time step. This is a few minutes on a standard laptop. Larger discretizations can be done using FreeFem mpi version which uses for instance MUMPS parallel linear solver[2].

3 Applications

We now present two appplications of the ALG2-JKO scheme: the first one deals with the porous medium equation and the second one with a diffusive model of crowd motion recently introduced in [32].

3.1 Application to the porous medium equation

The porous medium equation

$$\partial_t \rho = \mathrm{div}(\rho \nabla (\frac{m}{m-1}\rho^{m-1} + V)) = \Delta \rho^m + \mathrm{div}(\rho \nabla V) \tag{3.1}$$

corresponds to

$$E(\rho) := \frac{1}{m-1}\rho^m,\ \mathcal{E}(\rho) = \int_\Omega E(\rho) + \int_\Omega \rho V.$$

In this case

$$\mathcal{E}^*(c) = \Big(\frac{m-1}{m}\Big)^{\frac{m}{m-1}} \int_\Omega \Big((c(x) - V(x))_+\Big)^{\frac{m}{m-1}} \mathrm{d}x$$

and then (2.22) consists in a pointwise minimization problem: given $x \in \Omega$, setting $\overline{c} = -\phi^{n+1}(1,x) + \frac{\mu_1^n(x)}{r}$ and $V = V(x)$, we have to solve

$$\inf_{c\in\mathbb{R}} \Big\{\frac{1}{2}|c-\overline{c}|^2 + \theta\Big((c-\tau V)_+\Big)^{\frac{m}{m-1}}\Big\} \text{ with } \theta := \frac{1}{r\tau^{\frac{1}{m-1}}}\Big(\frac{m-1}{m}\Big)^{\frac{m}{m-1}} \tag{3.2}$$

[1] http://www.freefem.org
[2] http://mumps.enseeiht.fr/

whose solution is

$$c = \begin{cases} \bar{c}, \text{ if } \bar{c} \leq \tau V, \\ \text{the root in } (\tau V, +\infty) \text{ of (3.3) otherwise} \end{cases}$$

where (3.3) is the equation

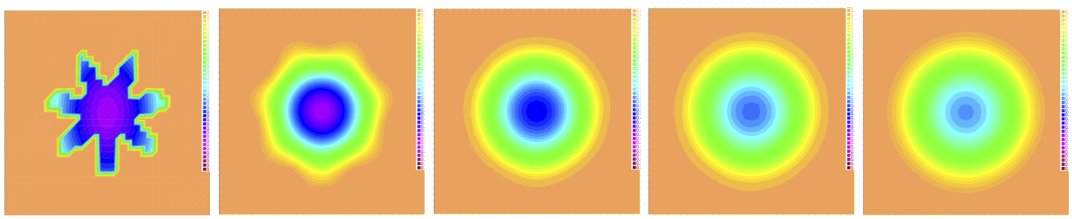
$$\bar{c} = c + \frac{\theta m}{m-1}(c - \tau V)^{\frac{1}{m-1}}. \tag{3.3}$$

The case of a linear diffusion (Fokker-Planck) corresponds to an entropic internal energy $E(\rho) = \rho \log(\rho)$, in which case by similar computations, one finds $c = c(x)$ by solving

$$\bar{c} = c + \frac{1}{r} e^{\frac{c}{\tau} - V - 1}. \tag{3.4}$$

Figure 1 shows the evolution of the density for $m = 3$ and $V = \frac{|x|^2}{2}$. As expected, we converge towards the stationnary Barenblatt profile $BB(x) = (\frac{m-1}{2m} \max(1 - |x|^2, 0))^{\frac{1}{m-1}}$. Discretization is 16×16 in space. We use 8 internal time steps in each JKO step and make sure the ALG2 iterations converge "reasonably" (a 10^{-6} tolerance is prescribed in the optimality system). There are 200 JKO steps (timestep = 0.01).

The plot in figure 2 gives quantitative information on the convergence towards the stationnary profile, decrease of energy and conservation of mass.



$t = 0$ $t = 0.15$ $t = 0.6$ $t = 0.9$ $t = 1.5$

Figure 1: *Density at different time steps for the porous medium equation with a confining potential*

3.2 Application to crowd motions

In [32], Santambrogio and Mészáros considered a model of congested crowd motion with diffusion which leads to

$$\partial_t \rho - \Delta \rho = \operatorname{div}(\rho(\nabla V + \nabla p)), \; p \geq 0, \; \rho \leq 1, \; p(1 - \rho) = 0, \tag{3.5}$$

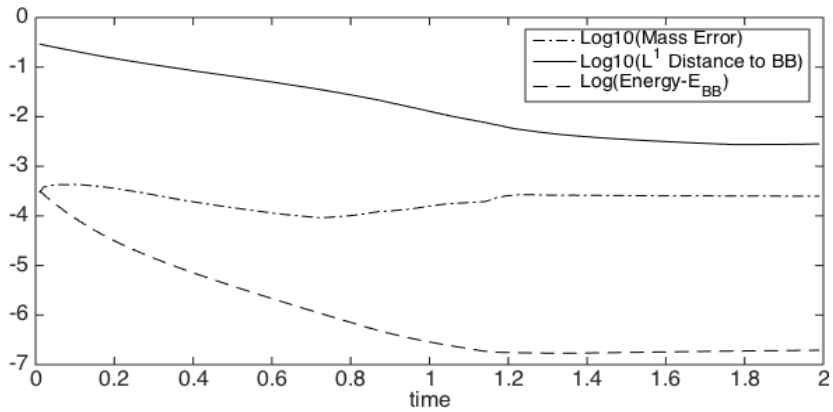


Figure 2: *This simulation shows the time evolution towards the stationary Barenblatt profile for $m = 3$. The first curve shows the mass of the density remains within the discretization error to the constant value (in log scale). The second curve is the L^1 distance of the density to the Barenblatt exact profile $BB(x) = (\frac{m-1}{2m}\max(1-|x|^2, 0))^{\frac{1}{m-1}}$, reached up to discretization error again (in log scale). The last curve shows the decrease of the difference between the density energy (potential + entropy) $\mathcal{E}(\rho)$ and Barenblatt energy $\mathcal{E}(BB)$ in log scale. As expected the decrease behaves like $-2t$ down to numerical discretization error.*

with no flux boundary condition. In the nondiffusive case (no Laplacian in the left hand side), this model is due to Maury, Roudneff-Chupin and Santambrogio [31] who made it clear that it has a gradient flow structure. The diffusive case (3.5), of course also has a gradient flow structure for the following energy

$$\mathcal{E}(\rho) = \int_\Omega E(\rho(x))\mathrm{d}x + \int_\Omega V(x)\rho(x)\mathrm{d}x$$

with

$$E(\rho) := \rho\log(\rho) + \chi_{[0,1]}(\rho) = \begin{cases} \rho\log(\rho), \text{ if } \rho \in [0,1] \\ +\infty, \text{otherwise.} \end{cases}$$

A direct computation gives

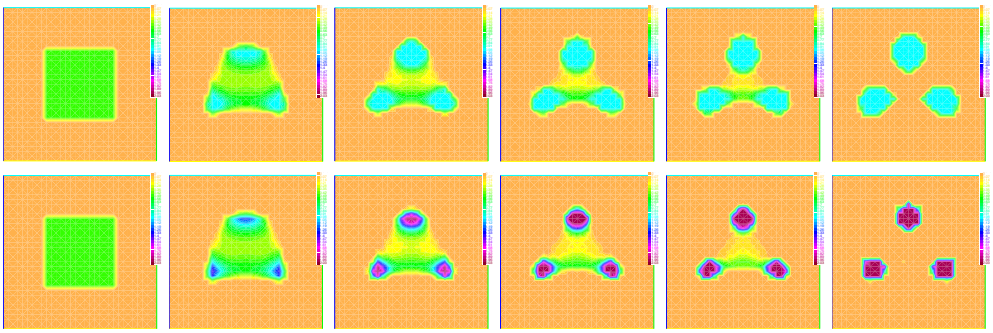
$$\mathcal{E}^*(x,c) = \max\left(\frac{c}{\tau} - V(x), e^{\frac{c}{\tau}-V(x)-1}\right)$$

and then, again setting, $\overline{c} = -\phi^{n+1}(1,x) + \frac{\mu_1^n(x)}{r}$ and $V = V(x)$, (2.22) becomes

$$\min_c \left\{\frac{1}{2}|c-\overline{c}|^2 + \frac{\tau}{r}\max\left(\frac{c}{\tau} - V, e^{\frac{c}{\tau}-V-1}\right)\right\}$$

whose solution is explicit:

$$c = \begin{cases} \overline{c} - \frac{1}{r}, \text{ if } \overline{c} \geq \tau(1+V) + \frac{1}{r}, \\ \text{the root of (3.4), otherwise.} \end{cases}$$



$t = 0$ $t = 0.02$ $t = 0.04$ $t = 0.06$ $t = 0.08$ $t = T_f$

Figure 3: *Crowd motion evolution of one species with potential* $V((x_1, x_2)) = 30(((x_1^2) - 1/4)^2 + (x_2 + 1/4)^2)((x_2 - 1/2)^2 + x_1^2)$. *The first row represents the evolution under the constraint* $\rho \leqslant 1$ *and the second the evolution under the constraint* $\rho \leqslant 2$.

In figure 3, we represent the evolution of one species, the potential has three minima (hot spots where the crowd wants to go) but with two different density constraints. When the density threshold is higher (second row $\rho \leqslant 2$) then, at the end, the density is more concentrated around the three minima of the potential.

4 Semi-implicit variant and nonlocal interactions

If we consider a nonlocal interaction term $\int_{\Omega\times\Omega} W(x,y)\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y$ (with W symmetric and smooth) in the general form of the equation (1.3), the final term $\mathcal{E}$ becomes nonconvex. Therefore, in order to be able to use our Augmented Lagrangian strategy, we have to modify the JKO scheme in a semi-implicit way by replacing the nonconvex bilinear term $\frac{1}{2}\int_{\Omega\times\Omega} W(x,y)\rho(x)\rho(y)\mathrm{d}x\mathrm{d}y$ by the linear one $\int_{\Omega} W(x,y)\rho(x)\rho^k(y)\mathrm{d}x\mathrm{d}y$. This leads to the semi-implicit scheme

$$\rho^{k+1} \in \mathrm{argmin}_{\rho\in\mathcal{P}_2}\left\{\frac{1}{2\tau}W_2^2(\rho,\rho^k) + \mathcal{E}(\rho|\rho^k)\right\} \tag{4.1}$$

where

$$\mathcal{E}(\rho|\mu) := \int_{\Omega} E(\rho(x))\mathrm{d}x + \int_{\Omega} V(x)\rho(x)\mathrm{d}x + \int_{\Omega\times\Omega} W(x,y)\mu(x)\rho(y)\mathrm{d}x\mathrm{d}y. \tag{4.2}$$

We refer to [15] for the convergence of this scheme to the solution of equation (1.3). For each time step, we then have to solve exactly the same type of problems as in section 3.1 except that the potential has to be updated at each step.

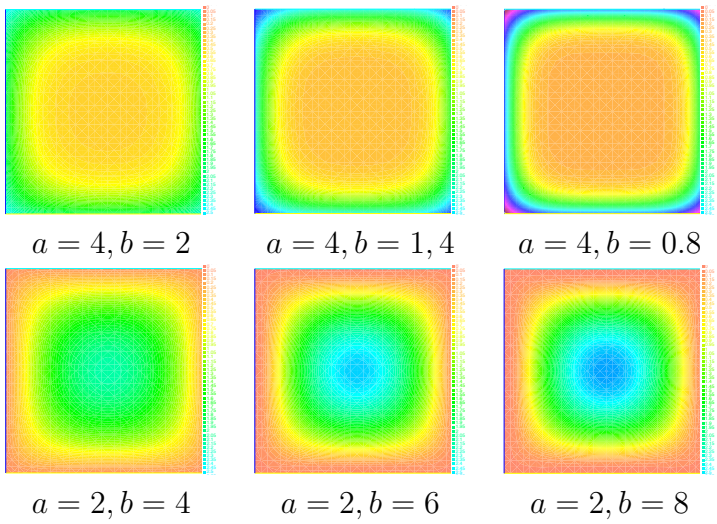


Figure 4: $\Omega = (-0.5, 0.5)^2$, *final state, starting from a uniform measure, with potential* $V = 0$, $E(\rho) = 0.01 * (\rho\log(\rho))$ *and interaction potential* $W(x) = |x|^a - |x|^b$.

5 Systems

Our aim now is to show that the ALG2-JKO scheme can also be used for systems. We first show how the semi-implicit ALG2-JKO scheme can easily be extended to systems which are coupled through interaction drift terms (such systems are not gradient flows in general). We then address the generalization of the Mészáros and Santambrogio model to the case of several populations, in this case, the coupling is through a total density constraint (and then a common pressure field) and the corresponding system actually has a gradient flow structure.

5.1 Interacting species

Let us take two species for the sake of simplicity and consider the evolution of the densities of these two species, coupled only through interaction terms:

$$\partial_t \rho_1 = \mathrm{div}(\rho_1(\nabla E_1'(\rho_1) + \nabla U_1[\rho_1, \rho_2])),\ \partial_t \rho_2 = \mathrm{div}(\rho_2(\nabla E_2'(\rho_2) + \nabla U_2[\rho_1, \rho_2])) \tag{5.1}$$

with energies E_1 and E_2 corresponding to independent linear or non linear diffusion terms and the coupling drift terms given (as in Di Francesco and Fagioli [20]) by convolutions with smooth kernels and individual potentials V_1 and V_2:

$$U_1[\rho_1, \rho_2] = V_1 + W_{11} \star \rho_1 + W_{21} \star \rho_2,\ U_2[\rho_1, \rho_2] = V_2 + W_{12} \star \rho_1 + W_{22} \star \rho_2.$$

The nondiffusive case where $E_1 = E_2 = 0$ was studied by Di Francesco and Fagioli [20]. Note that (5.1) is not a gradient flow except in the particular case where interactions are symmetric i.e. $W_{12} = W_{21}$. The semi-implicit JKO scheme first proposed by Di Francesco and Fagioli in the nondiffusive case consists in defining inductively ρ_1^k and ρ_2^k by

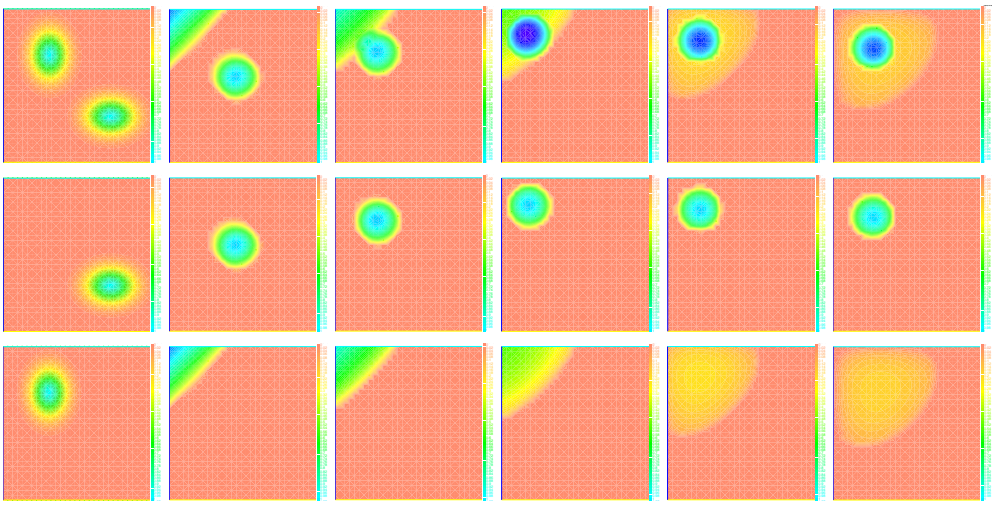
$$\rho_1^{k+1} = \mathrm{argmin}_\rho \frac{1}{2\tau} W_2^2(\rho, \rho_1^k) + \int_\Omega E_1(\rho(x))dx + \int_\Omega U_1[\rho_1^k, \rho_2^k](x)\rho(x)dx \tag{5.2}$$

and

$$\rho_2^{k+1} = \mathrm{argmin}_\rho \frac{1}{2\tau} W_2^2(\rho, \rho_2^k) + \int_\Omega E_2(\rho(x))dx + \int_\Omega U_2[\rho_1^k, \rho_2^k](x)\rho(x)dx. \tag{5.3}$$

The convergence of this scheme is established in Di Francesco and Fagioli [20] in the nondiffusive case and by Carlier and Laborde [15] in the diffusive case. Clearly, each independent subproblem (5.2) and (5.3) can be solved by the ALG2-JKO scheme described in section 2 and the extension to more than two species is direct as shown in the simulations below.

In figure 5, we see the evolution of two species which solve (5.1) with $E_1(\rho) = E_2(\rho) = \frac{1}{2}\rho^3$, $V_1 = V_2 = 0$, $W_{11}(x) = W_{21}(x) = W_{22}(x) = \frac{|x|^2}{2}$ and $W_{12}(x) = -\frac{|x|^2}{2}$. In other words, the first species is attracted by the second one but the second species is repelled by the first one. Since we have attractive self-interactions, the two species do not spread too much.



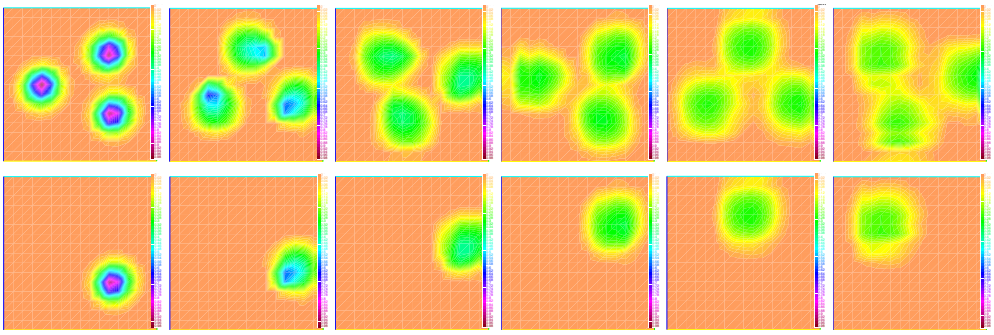
$t = 0$ $t = 0.1$ $t = 0.2$ $t = 0.4$ $t = 0.8$ $t = 1$

Figure 5: *Evolution of two species where the first one is attracted by the other and the second one is repelled by the first one. Top row: display of* $\rho_1 + \rho_2$. *Middle row: display of* ρ_1. *Bottom row: display of* ρ_2.

This scheme can treat systems with more than two species. In figure 6, we represent evolution of three species which run after each other. More precisely, the interaction potentials we use are of the form

$$U_i[\rho_1, \rho_2, \rho_3] = |x|^2 * \rho_{i+1} - |x|^2 * \rho_{i-1},$$

where $\rho_4 := \rho_1$ and $\rho_0 := \rho_3$.



$t = 0$ $t = 0.1$ $t = 0.2$ $t = 0.3$ $t = 0.4$ $t = 0.5$

Figure 6: *Evolution of three species running after each other with linear diffusion. Top row: display of* $\rho_1 + \rho_2 + \rho_3$*. Bottom row: display of* ρ_1*.*

5.2 Crowd motion with diffusion and several species

A natural variant of the model of [32], consists in considering two (or more) populations, each of whom having its own potential but coupled through the constraint that the total density cannot exceed 1 and then subject to a common pressure field. Note that a variant, without diffusion, with fixed initial and final densities and total density *equal* to 1 was treated in [6]. For a linear diffusion (corresponding to a Brownian noise on each species), the two-species crowd dynamics is expressed by the PDEs

$$\begin{cases} \partial_t \rho_1 - \Delta \rho_1 - \mathrm{div}(\rho_1(\nabla V_1 + \nabla p)) = 0, \\ \partial_t \rho_2 - \Delta \rho_2 - \mathrm{div}(\rho_2(\nabla V_2 + \nabla p)) = 0, \\ p \geq 0, \rho_1 + \rho_2 \leq 1,\ p(1 - \rho_1 - \rho_2) = 0. \end{cases}$$

Which is the gradient flow (for the product Wasserstein distance) of the energy

$$\mathcal{E}(\rho_1, \rho_2) := \sum_{i=1}^{2} \int_\Omega (\rho_i \log(\rho_i) + V_i \rho_i) + \int_\Omega \chi_{[0,1]}(\rho_1(x) + \rho_2(x)) \mathrm{d}x.$$

For a more general energy of the form

$$\mathcal{E}(\rho_1, \rho_2) := \mathcal{E}_1(\rho_1) + \mathcal{E}_2(\rho_2) + \mathcal{E}_{\mathrm{mix}}(\rho_1 + \rho_2)$$

with

$$\mathcal{E}_i(\rho_i) := \int_\Omega (E_i(\rho_i) + V_i \rho_i),\ \mathcal{E}_{\mathrm{mix}}(\rho) := \int_\Omega E_{\mathrm{mix}}(\rho(x)) \mathrm{d}x$$

the gradient flow of $\mathcal{E}$ is the following system with a coupling in the diffusion:

$$\partial_t \rho_i = \mathrm{div}(\rho_i \nabla (V_i + E_i'(\rho_1) + E_{\mathrm{mix}}'(\rho_1 + \rho_2))),\ i = 1,\ 2. \tag{5.4}$$

The JKO scheme for this energy then reads

$$(\rho_1^{k+1}, \rho_2^{k+1}) = \text{argmin}_{(\rho_1,\rho_2)} \Big\{ \sum_{i=1}^{2} \frac{1}{2\tau} W_2^2(\rho_i, \rho_i^k) + \mathcal{E}(\rho_1, \rho_2) \Big\} \tag{5.5}$$

which, in the particular case of the linear diffusion crowd motion problem with two species, takes the form

$$(\rho_1^{k+1}, \rho_2^{k+1}) = \text{argmin}_{\ \rho_1+\rho_2\leq 1} \Big\{ \sum_{i=1}^{2} \Big(\frac{1}{2\tau} W_2^2(\rho_i, \rho_i^k) + \int_\Omega (\rho_i \log(\rho_i) + V_i \rho_i) \Big) \Big\}.$$

Arguing as in section 2, setting $\phi = (\phi_1, \phi_2)$, $(D\phi_1, D\phi_2) := (\partial_t\phi_1, \nabla\phi_1, \partial_t\phi_2, \nabla\phi_2)$, $q = (q_1, q_2) = (a_1, b_1, c_1, a_2, b_2, c_2)$, $\sigma = (\sigma_1, \sigma_2) = ((\mu_1, m_1, \widetilde{\mu}_1), (\mu_2, m_2, \widetilde{\mu}_2))$ and defining the convex set K by (2.8), one can rewrite the discretization of (5.5) as a saddle-point problem for the augmented Lagrangian

$$\begin{aligned} L_r(\phi, q, \sigma) = &\sum_{i=1}^{2} \int_\Omega \phi_i(0,x)\rho_i^k(x)\mathrm{d}x + \sum_{i=1}^{2} \int_\Omega \chi_K(a_i(t,x), b_i(t,x))\mathrm{d}x\mathrm{d}t \\ &+ \sum_{i=1}^{2} \int_\Omega \Big((\mu_i, m_i) \cdot (D\phi_i - (a_i, b_i)) + \frac{r}{2} |D\phi_i - (a_i, b_i)|^2 \Big) \mathrm{d}x\mathrm{d}t \\ &+ \sum_{i=1}^{2} \int_\Omega \Big(\frac{r}{2} |\phi_i(1,x) + c_i(x)|^2 \mathrm{d}x - (\phi_i(1,x) + c_i(x))\widetilde{\mu}_i(x) \Big) \mathrm{d}x \\ &+ \tau \mathcal{E}^* \Big(\frac{c_1}{\tau}, \frac{c_2}{\tau} \Big). \end{aligned}$$

One can then again use ALG2 for this Lagrangian, note that since the coupling between the two species only appears in term $\tau\mathcal{E}^*\Big(\frac{c_1}{\tau}, \frac{c_2}{\tau}\Big)$, the only significant difference is in the proximal problem in the variables $c = (c_1, c_2)$ of the second step of the algorithm. Taking $r = 1$ for simplicity and using remark 2.1, we see that the two-populations analogue of (2.22) simply takes the form

$$\begin{aligned} (c_1^{n+1}(x), c_2^{n+1}(x)) &= \text{prox}_{\tau\mathcal{E}^*(x,\frac{\cdot}{\tau})} \Big(-\phi_1^{n+1}(1,x) + \widetilde{\mu}_1^n(x), -\phi_2^{n+1}(1,x) + \widetilde{\mu}_2^n(x) \Big) \\ &= \Big(-\phi_1^{n+1}(1,x) + \widetilde{\mu}_1^n(x), -\phi_2^{n+1}(1,x) + \widetilde{\mu}_2^n(x) \Big) \\ &\quad - \text{prox}_{\tau\mathcal{E}(x,.)} \Big(-\phi_1^{n+1}(1,x) + \widetilde{\mu}_1^n(x), -\phi_2^{n+1}(1,x) + \widetilde{\mu}_2^n(x) \Big) \end{aligned}$$

In the two populations crowd motion model with linear diffusion, setting $V_i = V_i(x)$ and $\overline{c}_i = -\phi_i^{n+1}(1,x) + \widetilde{\mu}_i^n(x)$, $\text{prox}_{\tau\mathcal{E}(x,.)}(\overline{c}_1, \overline{c}_2)$ is obtained by

solving

$$\inf \Big\{ \sum_{i=1}^{2} \frac{1}{2}(c_i - \bar{c}_i)^2 + \tau(c_i \log(c_i) + V_i c_i) \ : \ c_i \geq 0, c_1 + c_2 \leq 1 \Big\}$$

whose solution is again quasi explicit. More precisely

$$\mathrm{prox}_{\tau\mathcal{E}(x,.)}(\bar{c}_1, \bar{c}_2) = (\phi_\tau(\bar{c}_1 - \tau V_1), \phi_\tau(\bar{c}_2 - \tau V_2))$$

if $\phi_\tau(\bar{c}_1 - \tau V_1) + \phi_\tau(\bar{c}_2 - \tau V_2) \leq 1$ where $\phi_\tau(\alpha)$ is the positive root of

$$c + \tau(\log c + 1) = \alpha$$

and

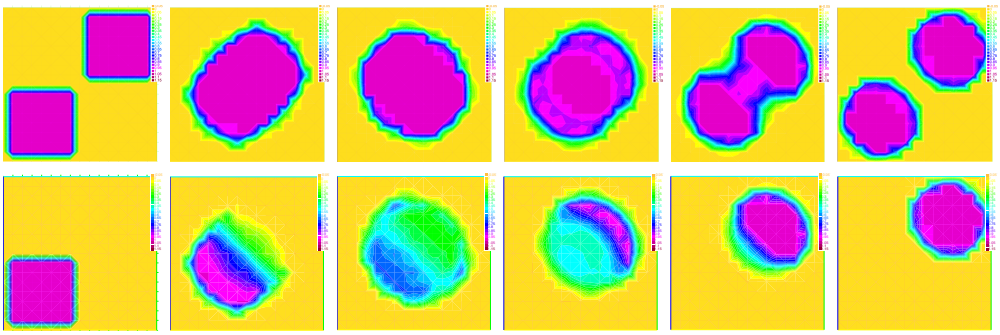
$$\mathrm{prox}_{\tau\mathcal{E}(x,.)}(\bar{c}_1, \bar{c}_2) = (\psi_\tau(\bar{c}_1 - \bar{c}_2 + \tau(V_2 - V_1)), 1 - \psi_\tau(\bar{c}_1 - \bar{c}_2 + \tau(V_2 - V_1)))$$

where $\psi_\tau(\alpha)$ is the root in $(0, 1)$ of

$$c + \tau \log\left(\frac{c}{1-c}\right) = \alpha + 1,$$

otherwise. This proximal computation therefore only involves scalar monotone equations and is therefore not more complicated than what we saw in the case of a single equation.

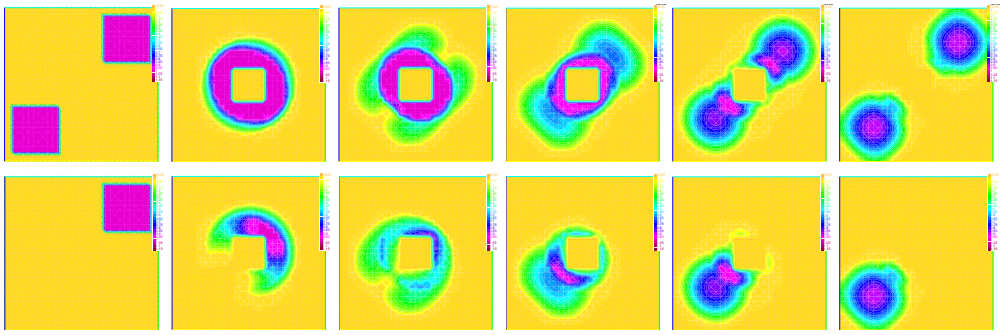
In figure 7, we see two populations which cross each other. When they start to cross each other at time $t = 0.05$, we remark that the density of ρ_1 and ρ_2 decrease and the sum is saturated.



$t = 0$ $t = 0.05$ $t = 0.1$ $t = 0.15$ $t = 0.2$ $t = 0.3$

Figure 7: *Evolution of two species crossing each other with density constraint. Top row: display of $\rho_1 + \rho_2$. Bottom row: display of ρ_1.*

In figure 8, we add a obstacle in the middle. This can be done using a potential with very high value in this area.



$t = 0$ $t = 0.1$ $t = 0.2$ $t = 0.3$ $t = 0.4$ $t = 0.5$

Figure 8: *Evolution of two species crossing each other with density constraint and an obstacle. Top row: display of* $\rho_1 + \rho_2$*. Bottom row: display of* ρ_1*.*

Acknowledgements: G.C. gratefully acknowledges the hospitality of the Mathematics and Statistics Department at UVIC (Victoria, Canada), and the support from the CNRS, from the ANR, through the project ISOTACE (ANR-12- MONU-0013) and from INRIA through the *action exploratoire* MOKAPLAN.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] 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.

[3] J.-D. Benamou, G. Carlier, Q. Mérigot, and É. Oudet. Discretization of functionals involving the monge-ampère operator, 2014.

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

[5] 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.

[6] Jean-David Benamou, Yann Brenier, and Kevin Guittet. Numerical analysis of a multi-phasic mass transport problem. *Contemporary Mathematics*, 353:1–18, 2004.

[7] Jean-David Benamou and Guillaume Carlier. Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations. *J. Optim. Theory Appl.*, 167(1):1–26, 2015.

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

[9] Jean-David Benamou, Francis Collino, and Jean-Marie Mirebeau. Monotone and consistent discretization of the monge-ampere operator. *arXiv preprint arXiv:1409.6694*, 2014.

[10] Jean-David Benamou, Brittany D. Froese, and Adam M. Oberman. Numerical solution of the optimal transportation problem using the Monge-Ampère equation. *J. Comput. Phys.*, 260:107–126, 2014.

[11] Martin Burger, José A. Carrillo, and Marie-Therese Wolfram. A mixed finite element method for nonlinear diffusion equations. *Kinet. Relat. Models*, 3(1):59–83, 2010.

[12] Martin Burger, Marzena Franek, and Carola-Bibiane Schönlieb. Regularized regression and density estimation based on optimal transport. *Appl. Math. Res. Express. AMRX*, (2):209–253, 2012.

[13] 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.

[14] Clément Cancès and Cindy Guichard. Entropy-diminishing CVFE scheme for solving anisotropic degenerate diffusion equations. In *Finite volumes for complex applications. VII. Methods and theoretical aspects*, volume 77 of *Springer Proc. Math. Stat.*, pages 187–196. Springer, Cham, 2014.

[15] Guillaume Carlier and Maxime Laborde. On systems of continuity equations with nonlinear diffusion and nonlocal drifts. Technical report, 2015.

[16] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.

[17] J. A. Carrillo and J. S. Moll. Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. *SIAM J. Sci. Comput.*, 31(6):4305–4329, 2009/10.

[18] José A. Carrillo, Robert J. McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[19] José A. Carrillo, Robert J. McCann, and Cédric Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[20] Marco Di Francesco and Simone Fagioli. Measure solutions for nonlocal interaction PDEs with two species. *Nonlinearity*, 26(10):2777–2808, 2013.

[21] 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.

[22] 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.

[23] 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.

[24] F. S. Patacchini J.A. Carrillo, Y. Huang and G. Wolansky. Numerical study of a particle method for gradient flows. *Preprint*, 2015.

[25] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[26] David Kinderlehrer and Noel J. Walkington. Approximation of parabolic equations using the Wasserstein metric. *M2AN Math. Model. Numer. Anal.*, 33(4):837–852, 1999.

[27] 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.

[28] 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.

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

[30] Daniel Matthes and Horst Osberger. Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation. *ESAIM Math. Model. Numer. Anal.*, 48(3):697–726, 2014.

[31] Bertrand Maury, Aude Roudneff-Chupin, and Filippo Santambrogio. A macroscopic crowd motion model of gradient flow type. *Math. Models Methods Appl. Sci.*, 20(10):1787–1821, 2010.

[32] R.-A. Meszaros and F. Santambrogio. A diffusive model for macroscopic crowd motion with density constraints, 2015.

[33] Horst Osberger and Daniel Matthes. Convergence of a fully discrete variational scheme for a thin-film equation, 2015.

[34] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

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

[36] Gabriel Peyré. Entropic wasserstein gradient flows, 2015.

[37] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser, 2015.

[38] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.

[39] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.