Available online at www.sciencedirect.com
ScienceDirect
Ann. I. H. Poincaré – AN 26 (2009) 2539–2580
ANNALES DE L'INSTITUT HENRI POINCARÉ
ANALYSE NON LINÉAIRE
www.elsevier.com/locate/anihpc

Constructing a relativistic heat flow by transport time steps

Robert J. McCann [a], Marjolaine Puel [b,*]

[a] *Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4*
[b] *Laboratoire MIP, Institut de Mathématiques, Université Paul Sabatier, 31062 Toulouse, France*

Received 2 August 2007; accepted 18 June 2009

Available online 5 August 2009

Abstract

An alternative construction to Andreu et al. (2005) [12] is given for $L^1_w([0,T],BV(\Omega))$ solutions to the relativistic heat equation (1) (see Brenier (2003) [14], Mihalas and Mihalas (1984) [37], Rosenau (1992) [40], Chertock et al. (2003) [20], Caselles (2007) [19]) under the assumption of initial data bounded from below and from above. For that purpose, we introduce a time discretized scheme in the style of Jordan et al. (1998) [30], Otto (1996) [38] involving an optimal transportation problem with a discontinuous hemispherical cost function. The limiting process is based on a monotonicity argument and on a bound of the Fisher information by an entropy balance characteristic of the minimization problem.

Résumé

Nous présentons ici une construction alternative à celle d'Andreu et al. (2005) [12] de solution $L^1_w([0,T],BV(\Omega))$ de l'équation de la chaleur relativiste (1) (voir Brenier (2003) [14], Mihalas et Mihalas (1984) [37], Rosenau (1992) [40], Chertock et al. (2003) [20], Caselles (2007) [19]) dans le cas de conditions initiales bornées inférieurement et supérieurement. Pour cela, nous introduisons un schéma discret en temps dans le style de Jordan et al. (1998) [30], Otto (1996) [38] basé sur un problème de transport optimal faisant intervenir une fonction de coût hémisphérique et discontinue. Le passage à la limite lorsque le pas de temps tend vers zéro repose sur un argument de monotonie et une borne de l'information de Fisher par la variation de l'entropie, inégalité caractéristique du problème de transport optimal.

Keywords: Relativistic heat equation; Optimal transportation; Gradient flow; Jordan–Kinderlehrer–Otto scheme

1. Introduction

We consider in this work a relativistic heat equation which has been introduced for example in the paper of Rosenau [40] (see also [20]) or Mihalas and Mihalas [37]; it fills in a gap in the Fokker–Planck theory by imposing an upper bound for the propagation velocity. This equation can be written

$$\partial_t \rho = \operatorname{div}\left(\rho \frac{\nabla \rho}{\sqrt{\rho^2 + |\nabla \rho|^2}}\right) = \operatorname{div}\left(\rho \frac{\nabla \log \rho}{\sqrt{1 + |\nabla \log \rho|^2}}\right). \tag{1}$$

* Corresponding author.
E-mail address: puel@mip.ups-tlse.fr (M. Puel).

0294-1449/$ – see front matter
doi:10.1016/j.anihpc.2009.06.006

Viewing the ordinary heat equation as a transport equation by velocity $\nabla \log \rho$, then Eq. (1) involving the velocity $\frac{\nabla\rho}{\sqrt{\rho^2+|\nabla\rho|^2}}$ motivates the appellation *relativistic*.

The aim of this present work is to construct solutions to Eq. (1) following a strategy introduced by Jordan, Kinderlehrer, Otto [30] and Otto [38], subsequently developed by many authors, M. Agueh [1] and Ambrosio, Gigli, Savaré [3] in particular, and suggested in the relativistic context by Brenier [14]. This strategy applies to finding a solution to general transport equations given by

$$\partial_t \rho = \operatorname{div}\big(\rho \nabla c^*\big(\nabla\big(F'(\rho)\big)\big)\big) \tag{2}$$

where c^* is a convex mobility function on $\mathbb{R}^d$ and F is a convex function on $[0,\infty[$ representing the entropy. It is based on a new point of view on (2) saying that the transport of the density ρ is seen as the gradient flow of the convex function F with respect to a distance induced by the cost function c, the Legendre transform of c^* (see [2,3,5,43] for the notion of gradient flow). The solution is obtained as a limit of a solution of a time discretized scheme and the peculiarity of the method is that this discrete scheme involves an optimal mass transport problem.

This work comes after a series of papers. In [30], the case of the Wasserstein distance $c(z) = \frac{|z|^2}{2}$ and entropy $F(\rho) = \rho\log\rho - V\rho$ were addressed by Jordan, Kinderlehrer and Otto, who obtain the Fokker–Planck equation

$$\partial_t \rho = \operatorname{div}(\rho\nabla V) + \beta^{-1}\Delta\rho \quad \text{where } V \text{ is a given smooth function.}$$

In [38], Otto treated the case of $c(z) = \frac{|z|^q}{q}$ with $F(\rho) = \frac{n\rho^b}{b(b-1)}$ and $b = n + (p-2)/(p-1)$ which leads to the doubly degenerate equation

$$\partial_t \rho = \operatorname{div}\big(\big|\nabla\rho^n\big|^{p-2}\nabla\rho^n\big) \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1,\ p \geqslant 2.$$

And in [1], Agueh considered the case of cost functions satisfying

$$\beta|z|^q \leqslant c(z) \leqslant \alpha\big(|z|^q + 1\big) \quad \text{for all } z \in \mathbb{R}^d, \text{ where } \alpha, \beta > 0 \text{ and } q > 1,$$

which includes a very large class of equations such as the Fokker–Planck equation, the porous medium and fast diffusion equation, the p-Laplacian and the doubly degenerate diffusion equation.

Note that for the relativistic heat equation, the entropy is $F(\rho) = \rho\log\rho - \rho$ and

$$\nabla c^*(z) = \frac{z}{\sqrt{1+|z|^2}},$$

which gives

$$c^*(z) = \sqrt{1+|z|^2} - 1.$$

The corresponding cost function

$$c(z) = \begin{cases} 1 - \sqrt{1-|z|^2} & \text{if } |z| \leqslant 1, \\ +\infty & \text{if } |z| > 1 \end{cases}$$

is hemispherical and discontinuous, hence does not belong to the set of cost function for which this strategy has been successful.

The aim of this paper is to apply this "Optimal Transportation Strategy" to the Cauchy problem for Eq. (1), and more precisely for any cost functions satisfying

$$c(z) = \begin{cases} \tilde{c}(|z|) \geqslant 0 & \text{if } |z| \leqslant 1, \\ +\infty & \text{if } |z| > 1 \end{cases}$$

where $\tilde{c}$ is a continuous strictly convex function on $[0,1]$ belonging to $C^2([0,1[)$ and with $|\nabla c(z)| \to \infty$ when $|z| \to 1$ (and hence $c^* \in C^2(\mathbb{R}^d)$) and, as in the work of Agueh [1], for any strictly convex entropy functions satisfying $F \in C^2(\mathbb{R})$ such that $\frac{F(\lambda)}{\lambda} \to \infty$ when $\lambda \to \infty$ and $\lambda^d F(\lambda^{-d})$ is convex.

The Cauchy problem for such an equation has been recently studied by Andreu, Caselles and Mazón [11,12] and the speed of propagation for the support of its solutions has been studied by Andreu, Caselles, Mazón and Moll in [13]. The convergence of the relativistic equation toward the heat equation as the light speed goes to infinity has been also investigated very recently by Caselles [19]. Those works belong to a large program (see the long series of paper [7,9–12]) where the Cauchy problem is examined for degenerate elliptic and parabolic equations. As in the strategy of Jordan, Kinderlehrer and Otto, the proof of the existence of solution to this equation by Andreu, Caselles and Mazón involves a discrete in time equation. In their work, the time discrete density ρ^h is obtained by solving the elliptic equation on each time interval $[ih; (i+1)h]$

$$\rho_i^h - \rho_{i-1}^h = h \operatorname{div}\big(\rho_i^h \nabla c^*\big(D \log \rho_i^h\big)\big).$$

We will assume in this work that the initial density is bounded from below and from above while in their work, Andreu, Caselles and Mazón assume only that the initial data is non-negative.

Nevertheless, the point of view of this present work is interesting since the method employed — in particular the construction of the discrete scheme — involves the study of a mass transport problem with a discontinuous convex cost function. The limiting process remains a delicate step because of the weak regularity of the discrete sequence constructed via the minimization process, and presents a real difficulty. As far as we know, the construction of optimal transport maps for discontinuous cost functions has not been completed before but is a necessary condition to obtain relativistic phenomena for the corresponding transport equation. However, the Kantorovich duality for discontinuous cost functions has been investigated in the paper of Ambrosio and Pratelli [4], and the use of approximate differentiability as in Ambrosio, Gigli, Savaré [3] also proves a crucial tool.

1.1. Description of the "Optimal transportation strategy"

Fix $\Omega \Subset \mathbb{R}^d$ bounded throughout. The method consists in constructing a time discrete scheme as follows:

Let $P(\Omega)$ be the set of Borel probability measures on Ω, $\rho_0 \in P(\Omega)$ given, find $\rho^h(t,y) \in P([0,T] \times \Omega)$ defined by

$$\begin{cases} \rho^h(0,y) = \rho_0(y), \\ \rho^h(t,y) = \rho_i^h(y) \quad \text{for } t \in \big]ih; (i+1)h\big], \ h \text{ (being the time step)} \end{cases} \tag{3}$$

where $\rho_i^h(y)$ is a solution of the minimization problem $\mathcal{P}_i^h$ defined by

$$I\big(\rho_{i-1}^h, \rho_i^h\big) = \inf_{\rho \in P(\Omega)} I\big(\rho_{i-1}^h, \rho\big)$$

with

$$I\big(\rho_{i-1}^h, \rho\big) = \int_\Omega F\big(\rho(y)\big)\, dy + h \inf_{\gamma \in \Gamma_i^h(\rho_{i-1}^h, \rho)} \int_{\Omega \times \Omega} c\left(\frac{x-y}{h}\right) d\gamma(x,y),$$

Γ_i^h denoting the set of probability measures that have ρ_{i-1}^h and ρ as marginals and c is the convex cost function, the Legendre transform of c^* defined as in [15] by

$$c^*(z) = \sup_{w \in \mathbb{R}^d} \big(w \cdot z - c(w)\big). \tag{4}$$

This process follows the ideas presented in the work of Otto [38] which extends the notion of gradient flow to more general cost functions (cf. also Villani's book [43]). This point of view is particularly simple to explain when the cost function is the quadratic function, minimizing $E(\rho) + h|\frac{\rho_0 - \rho}{h}|^2$ gives formally $\frac{\rho_0 - \rho}{h} = E'(\rho)$ which is a discrete version of $\partial_t \rho = E'(\rho)$, meaning that ρ is a gradient flow of E. In a more general setting, we minimize the entropy among all the densities reachable at time T by moving along geodesics induced by the cost function c.

Let us now describe the different steps of the construction of a solution to (2):

Step 1. Prove that the minimization problem has a unique solution, the optimal transport plan γ.

Step 2. Define an optimal map corresponding to this optimal measure and derive the Euler–Lagrange equation of this minimization problem which gives formally

$$\nabla c\left(\frac{S_i^h(y) - y}{h}\right) = \nabla\big(F'\big(\rho_i^h\big)\big)$$

where S_i^h denotes the optimal map pushing ρ_i^h forward to ρ_{i-1}^h.

Using $\nabla c^*(\nabla c(y)) = y$, this leads to

$$S_i^h(y) - y = h\nabla c^*\big(\nabla\big(F'\big(\rho_i^h\big)\big)\big). \tag{5}$$

Step 3. Obtain an approximate time discrete equation and pass to the limit when the time step goes to zero: By multiplying Eq. (5) by $\rho\nabla\xi$ where ξ is a smooth test function, we obtain in the sense of distributions

$$\rho_i^h - \rho_{i-1}^h = h\,\mathrm{div}\big(\rho_i^h\nabla c^*\big(\nabla\big(F'\big(\rho_i^h\big)\big)\big)\big) + \text{Correction terms.} \tag{6}$$

Solutions to (2) are obtained by passing to the limit in (6) when h goes to zero. The identification of $\overline{A}$, the limit of $\rho_i^h\nabla c^*(\nabla(F'(\rho_i^h)))$ involves a Minty–Browder argument (see for example Evans's book [25]) based on a monotonicity property of the gradient of the mobility function c^*. Indeed, we prove for any test function $\xi(t,x) \geqslant 0$ and $\zeta(t,x) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$

$$\int \xi\big(\overline{A} - \bar\rho\nabla c^*(\zeta)\big)\big(D^{ac}\big(F'(\bar\rho)\big) - \zeta\big) \geqslant 0 \tag{7}$$

which yields $\overline{A} = \bar\rho\nabla c^*(D^{ac}(F'(\bar\rho)))$ by taking $\zeta = D^{ac}(F'(\bar\rho)) + \delta w$, and by passing successively to the limit $\delta \searrow 0$ and $\delta \nearrow 0$.

Formally, to obtain (7), we pass to the $\varepsilon < h \to 0$ limit in

$$\int_0^T\int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\big(\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big) - \nabla c^*\big(\zeta(t,y)\big)\big)\cdot\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big) - \zeta(t,y)\big)\,dy\,dt \geqslant 0.$$

This limiting process is strongly based:

1. On the displacement convexity [35] of the entropy function (ensured by the monotonicity of F and the convexity of $\lambda^d F(\lambda^{-d})$) that yields formally the Fisher information-entropy inequality satisfied by the minimizer ρ^h of problem $\mathcal{P}_i^h$

$$\int_0^T\int_\Omega \rho^h(t,y)\nabla c^*\big(\nabla\big(F'\big(\rho^h(t,y)\big)\big)\big)\cdot\nabla\big(F'\big(\rho^h(t,y)\big)\big)\,dy\,dt \leqslant \int_\Omega \big(F\big(\rho^h(0,y)\big) - F\big(\rho^h(T,y)\big)\big)\,dy \tag{8}$$

or its localized version

$$\int_0^T\int_\Omega \xi(t,y)\rho^h(t,y)\nabla c^*\big(\nabla\big(F'\big(\rho^h(t,y)\big)\big)\big)\cdot\nabla\big(F'\big(\rho^h(t,y)\big)\big)\,dy\,dt$$
$$\leqslant -\int_0^T\int_\Omega \nabla\xi(t,y)\rho^h(t,y)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho^h(t,y)\big)\big)\big)F'\big(\rho^h(t,y)\big)\,dy\,dt - \int_0^T\int_\Omega \xi(t,y)\partial_t F\big(\rho^h(t,y)\big)\,dy\,dt.$$

2. On the corresponding equality satisfied formally by any smooth solution $\tilde\rho$ of (1) and obtained by multiplying the equation by $F'(\tilde\rho(t,y))$ and integrating by parts:

$$\int_\Omega \tilde\rho(t,y)\nabla F'\big(\tilde\rho(t,y)\big)\cdot\nabla c^*\big(\nabla F'\big(\tilde\rho(t,y)\big)\big)\,dy\,dt = \int_\Omega \big[F\big(\rho_0(y)\big) - F\big(\tilde\rho(T,y)\big)\big]\,dy. \tag{9}$$

These relations avoid the problem of nonlinearities in ρ^h since we will prove a strong convergence which allows us to pass to the limit in the right-hand side of (8).

1.2. Assumptions on the cost function and on the entropy function

We will here give some direct consequences of the assumptions on the cost function that justify its relation with relativistic phenomena.

Recall that we deal with cost functions satisfying

$$c(z) = \begin{cases} \tilde{c}(|z|) \geqslant 0 & \text{if } |z| \leqslant 1, \\ +\infty & \text{if } |z| > 1 \end{cases}$$

where $\tilde{c} \in C^2([0,1[)$ is bounded and strictly convex with $\tilde{c}''(\lambda) > 0$ and $\tilde{c}'(\lambda) \to \infty$ as $\lambda \to 1$. Assume also that $\tilde{c}(0) = 0 = \tilde{c}'(0)$; then $c^* \in C^2(\mathbb{R}^d)$.

Recall that we have the two relations $\nabla c(\nabla c^*(z)) = z$ and $z \cdot \nabla c^*(z) = c(\nabla c^*(z)) + c^*(z)$ and note that since c^* is radial, $\nabla c^*(z) = \frac{z}{|z|}\omega(|z|)$ which implies that $z \cdot \nabla c^*(z) = |z||\nabla c^*(z)|$.

Remark then that the discontinuity of c — implying its infinite part — is strongly linked with the bound on $\nabla c^*(z)$. Indeed, since c is defined from c^* by

$$c(z) = \sup_{w \in \mathbb{R}^d} (z \cdot w) - c^*(w),$$

the fact that c is not finite when $|z| > 1$ means that the supremum is not attained and then the relation $z = \nabla c^*(w)$ cannot be matched by any $w \in \mathbb{R}^d$. This means that $|\nabla c^*| \leqslant 1$ (and reciprocally) so we recover that the discontinuity of the cost function is equivalent to the relativistic aspect of the transport. It also implies — since $|\nabla c^*(z)| = \omega(|z|)$ is a non-decreasing function of $|z|$ — that $\lim_{|z|\to\infty} |\nabla c^*(z)| = 1$.

As we said, we consider any strictly convex entropy functions satisfying $F \in C^2(\mathbb{R})$ such that $\frac{F(\lambda)}{\lambda} \to \infty$ when $\lambda \to \infty$ and $\lambda^d F(\lambda^{-d})$ is convex, the last condition being necessary for displacement convexity of the entropy [35].

1.3. Notations and definitions

1.3.1. Optimal transport theory

We now recall two ways to link pairs of probability measures.

Definition 1.1. Let $\rho_1 \in P(\Omega)$ and S be a Borel map $S : \Omega \to \Omega$. We say that ρ_0 is the *push-forward* of ρ_1 through S if for any bounded Borel function φ

$$\int_\Omega \varphi\big(S(x)\big)\, d\rho_1(x) = \int_\Omega \varphi(x)\, d\rho_0(x).$$

Definition 1.2. Given two probability measures ρ_0 and ρ_1, the set of *transport plans* between them refers to joint probability measures on $\Omega \times \Omega$ with ρ_0 and ρ_1 as marginals:

$$\Gamma(\rho_0, \rho_1) = \Bigg\{\gamma(x,y) \in P(\Omega \times \Omega) \text{ s.t. } \int_{\Omega\times\Omega} \varphi(x)\, d\gamma(x,y) = \int_\Omega \varphi(x)\, d\rho_0(x) \text{ and }$$
$$\int_{\Omega\times\Omega} \varphi(y)\, d\gamma(x,y) = \int_\Omega \varphi(y)\, d\rho_1(y)\Bigg\}$$

for all Borel test functions $\varphi : \Omega \to \mathbb{R}$.

Finally, we need to define the c-superdifferential of a function

Definition 1.3. $x \in \partial^c \varphi(y)$ if $|x - y| \leqslant 1$ and for all $v \in \Omega$

$$\varphi(v) \leqslant c(x - v) - c(x - y) + \varphi(y).$$

1.3.2. Definition of BV functions and of functions of BV functions

In this section, we recall definitions that can be found in the series of paper of Andreu, Caselles and Mazón, that we will use throughout the second part of this paper and that are fundamental for the understanding of the notion of solution to Eq. (1) which we construct.

Some functional spaces

Throughout this paper, we will deal with BV functions, the set of $\rho \in L^1(\Omega)$ functions such that the gradient of ρ defined as a distribution is a vector valued Radon measure whose total variation, i.e.

$$\|\rho\|_{BV} = \|D\rho\|_{TV} = \sup\left\{\int_\Omega \rho \operatorname{div}(\xi)\,dx \text{ where } \xi \in C_0^\infty(\Omega;\mathbb{R}^d) \text{ s.t. } |\xi(x)| \leqslant 1\right\}$$

is finite.

A sequence ρ_i of BV functions is said to converge $w^*BV(\Omega)$ toward ρ if $\rho_i \to \rho$ in $L^1_{loc}(\Omega)$ and its gradient $D\rho_i$ converges toward $D\rho$, weak* as measures, i.e. against any compactly supported continuous test function.

We also need to introduce the space $L^1([0,T],BV_2(\Omega))$ where $BV_2(\Omega) = BV(\Omega) \cap L^2(\Omega)$ with the norm $\|\rho\|_{BV_2(\Omega)} = \|\rho\|_{L^2(\Omega)} + \|D\rho\|_{TV}$ contained in the space where the solution will live, $L^1_w([0,T],BV(\Omega))$. The difference between those two spaces comes from the fact that $BV(\Omega)$ is not separable.

Definition 1.4. A function ρ belongs to $L^1([0,T],BV(\Omega))$ if it is a limit almost everywhere in time of a sequence of simple functions, i.e. defined by $\sum \mathbf{1}_{[t_i,t_{i+1}]}\rho_i$ where $\rho_i \in BV(\Omega)$, hence Bochner integrable.

In particular $C([0,T],BV(\Omega)) \subset L^1([0,T],BV(\Omega))$.

Definition 1.5. A function ρ belongs to $L^1([0,T],BV_2(\Omega))$ if it is a limit almost everywhere in time of a sequence of simple functions, i.e. defined by $\sum \mathbf{1}_{[t_i,t_{i+1}]}\rho_i$ where $\rho_i \in BV_2(\Omega)$.

Whereas:

Definition 1.6. $L^1_w([0,T],BV(\Omega))$ is the space of weakly measurable functions $\rho:[0,T] \to BV(\Omega)$ (i.e., $t:[0,T] \to \langle\rho(t);\zeta\rangle$ is measurable for every $\zeta \in BV(\Omega)'$), such that $\int_0^T \|\rho(t)\|_{BV(\Omega)}\,dt < \infty$.

Note that, since $BV(\Omega)$ has a separable predual, it follows for $\rho \in L^1_w([0,T],BV(\Omega))$, that $t:[0,T] \to \|\rho(t)\|_{BV(\Omega)}$ is measurable.

The first space is useful because we know its dual $L^1([0,T],BV_2(\Omega))'$ equals (see Dunford and Schwartz [24])

$$L^\infty([0,T],BV_2(\Omega)') = \left\{\rho \text{ weak}^* \text{ measurable functions, } \rho:[0,T] \to BV(\Omega)'_2 \text{ such that} \right.$$
$$\left. \operatorname*{ess\,sup}_{[0,T]} \sup_{\|\xi\|_{BV_2(\Omega)}\leqslant 1} \left|\int_\Omega \rho\xi\right| < \infty\right\}$$

and we define the duality bracket for $\rho \in L^1([0,T],BV_2(\Omega))$ and $\tilde\rho \in L^\infty([0,T],BV_2(\Omega)')$ by

$$\langle\rho;\tilde\rho\rangle = \int_0^T \langle\rho(t);\tilde\rho(t)\rangle\,dt.$$

As a matter of fact, the solution will live in $L^\infty([0,T]\times\Omega) \cap L^1_w([0,T],BV(\Omega))$ and the equation will hold in the sense of distributions, which leads to one of the principal difficulties of the third step of the proof, namely that the solution cannot be taken as a test function to obtain (9).

Functions defined on BV

As we will see, $D\rho$, the gradient of the BV function ρ can be decomposed as the sum of its absolute continuous part $D^{ac}\rho$ (with respect to the Lebesgue measure), also called the Radon–Nikodym derivative of $D\rho$ (which coincides with the approximate derivative) of ρ, and its singular part $D^s\rho$, divided into a jump part $D^j\rho$ and a Cantor part $D^c\rho$. So we write

$$D\rho = D^{ac}\rho + D^s\rho = D^{ac}\rho + D^j\rho + D^c\rho.$$

Moreover, we will need to define the composition of certain functions with *BV* functions and their derivatives. For example, if $f(x,\lambda,\xi)$ is an integrand depending on the space variable x, on $\lambda\in\mathbb{R}$ and on the vector field ξ, we will define the functional $\mathcal{F}(\rho,D\rho)$ of a *BV* function ρ as in the paper of Dal Maso [22] by

$$\mathcal{F}(\rho,D\rho)=\int_0^T\int_\Omega f\big(x,\rho,D^{ac}\rho\big)\,dt\,dx+\int_\Omega f^0\left(x,\rho,\frac{D\rho}{|D\rho|}\right)\big|D^c\rho\big| + \int_{J_\rho}\left(\int_{\rho_-(x)}^{\rho_+(x)} f^0\big(x,s,\nu_\rho(x)\big)\,ds\right)d\mathcal{H}^{d-1}(x) \tag{10}$$

where f^0 is the recession function equal to $\lim_{t\to 0} tf(x,\lambda,\frac{\xi}{t})$, J_ρ is the set of approximate jump points of ρ, $\nu_\rho = D\rho/|D\rho|$ the Radon–Nikodym derivative of $D\rho$ with respect to its total variation $|D\rho|$. Indeed with those definition, the jump part of the singular part of $D\rho$ can be written

$$D^j\rho=(\rho^+-\rho^-)\nu_\rho\mathcal{H}^{d-1}_{|J_\rho}$$

where $\mathcal{H}^{d-1}$ is the $d-1$ Hausdorff measure on $\mathbb{R}^d$.

For example, in the theorem cited below, there is in fact no ambiguity in the right-hand side term of Eq. (14) since the corresponding recession function is zero.

We can also compute g^0 in the case where $f=g(\lambda,\xi)=\lambda\xi\cdot\nabla c^*(\xi)$. We obtain easily using the properties of the cost function that

$$\lim_{t\to 0} tf\left(x,\lambda,\frac{\xi}{t}\right)=\lim_{t\to 0}\lambda|\xi|\left|\nabla c^*\left(\frac{\xi}{t}\right)\right|=\lambda|\xi|. \tag{11}$$

Finally, we also need the result of De Cicco, Fusco and Verde about the L^1-lower semi-continuity in *BV*.

Theorem 1.7 *(L^1-lower semi-continuity in BV). (See [23].) Let Ω be an open set of $\mathbb{R}^d$ and $h:\Omega\times\mathbb{R}\times\mathbb{R}^d\to[0,\infty)$ a locally bounded Caratheodory function (that is, measurable with respect to x in Ω for every $(\lambda,\xi)\in\mathbb{R}\times\mathbb{R}^d$, and continuous with respect to (λ,ξ) for almost every x in Ω) such that for every $(\lambda,\xi)\in\mathbb{R}\times\mathbb{R}^d$, the function $h(\cdot,\lambda,\xi)$ is of class $C^1(\Omega)$. Let us assume that*:

(i) *$h(x,\lambda,\cdot)$ is convex in $\mathbb{R}^d$ for every $(x,\lambda)\in\Omega\times\mathbb{R}$;*
(ii) *$h(x,\cdot,\xi)$ is continuous in $\mathbb{R}$ for every $(x,\xi)\in\Omega\times\mathbb{R}^d$.*

Then the functional $\mathcal{H}$ defined by

$$\mathcal{H}(\rho,D\rho)=\int_0^T\int_\Omega h\big(x,\rho,D^{ac}\rho\big)\,dt\,dx+\int_\Omega h^0\left(x,\rho,\frac{D\rho}{|D\rho|}\right)\big|D^c\rho\big|+\int_{J_\rho}\left(\int_{\rho_-(x)}^{\rho_+(x)} h^0\big(x,s,\nu_\rho(x)\big)\,ds\right)d\mathcal{H}^{d-1}(x)$$

is lower semi-continuous with respect to $L^1(\Omega)$ convergence.

This extension of the functional $\mathcal{F}$ to *BV* functions allows us to extend the definition of

$$\int_0^T\int_\Omega DF'(\rho)\cdot\nabla c^*\big(D\big(F'(\rho)\big)\big)$$

to $\rho\in L^1([0,T],BV(\Omega))$ by decomposing $\xi\cdot\nabla c^*(\xi)=c(\nabla c^*(\xi))+c^*(\xi)$ and by applying Theorem 1.7 to $c^*(\xi)$.

Definition of the measure $(z,D\rho)$

There is another way of defining $(DF'(\rho);\nabla c^*(D(F'(\rho))))$, for $\rho\in L^\infty([0,T]\times\Omega)$ and $\operatorname{div}(\nabla c^*(D(F'(\rho))))\in L^1([0,T]\times\Omega)$ which is inspired by [6]. But in this present work, this is not relevant since the last property is not satisfied even at the time discrete level. This make the notion of entropic solution more difficult to introduce in this framework and this property and the uniqueness of solution coming from it is not addressed here.

However, we define $(z; D\rho)$ as a distribution when $z \in L^\infty([0,T] \times \Omega; \mathbb{R}^d)$, $\rho \in L^1([0,T], BV_2(\Omega)) \cap L^\infty([0,T] \times \Omega)$, when $\operatorname{div} z = B$ in the sense of distributions, B is the weak* limit of a bounded sequence B^h in the dual of $L^1([0,T], BV_2(\Omega))$ and z is the weak* limit of a bounded sequence z^h. Indeed, for any test function ξ, we write

$$\big\langle \xi; (z; D\rho)\big\rangle = -\int_0^T \langle B; \rho\xi\rangle_{L^1([0,T],BV_2(\Omega))'; L^1([0,T],BV_2(\Omega))}\, dt - \int_0^T \int_\Omega z\rho\nabla\xi\, dt\, dy.$$

But if z is the weak* limit of a bounded sequence z^h, we still have to prove that

$$\big\langle \xi; (z; D\rho)\big\rangle = \lim_{h\to 0}\big\langle \xi; \big(z^h; D\rho\big)\big\rangle = \lim_{h\to 0}\left(-\int_0^T\int_\Omega \operatorname{div} z^h \rho\xi\, dt\, dy - \int_0^T\int_\Omega z^h\rho\nabla\xi\, dt\, dy\right).$$

For that to be true, we need that $\operatorname{div} z^h = B^h$ for h and t fixed in $L^1(\Omega)$. Indeed, in this case, we can write

$$\begin{aligned}
\big\langle \xi; (z; D\rho)\big\rangle &= -\int_0^T \langle B; \rho\xi\rangle_{L^1([0,T],BV_2(\Omega))'; L^1([0,T],BV_2(\Omega))}\, dt\, dy - \int_0^T\int_\Omega z\rho\nabla\xi\, dt\, dy \\
&= -\lim_{h\to 0}\int_0^T\int_\Omega B^h\rho\xi\, dt\, dy - \int_0^T\int_\Omega z\rho\nabla\xi\, dt\, dy \\
&= -\lim_{h\to 0}\int_0^T\int_\Omega \operatorname{div} z^h\rho\xi\, dt\, dy - \int_0^T\int_\Omega z\rho\nabla\xi\, dt\, dy \\
&= -\lim_{h\to 0}\left(\int_0^T\int_\Omega \operatorname{div} z^h\rho\xi\, dt\, dy + \int_0^T\int_\Omega z^h\rho\nabla\xi\, dt\, dy\right).
\end{aligned}$$

Moreover, $(z; D\rho)$ is a Radon measure since by Theorem 1.5 in [6], we have

$$\begin{aligned}
\big\langle \xi; (z; D\rho)\big\rangle &= \lim_{h\to 0}\left[-\int_0^T\int_\Omega \operatorname{div} z^h\rho\xi\, dt\, dy - \int_0^T\int_\Omega z^h\rho\nabla\xi\, dt\, dy\right] \\
&\leqslant \|z\|_{L^\infty([0,T]\times\Omega)}\|\xi\|_{L^\infty([0,T]\times\Omega)}\int_0^T\int_\Omega |D\rho|\, dy\, dt.
\end{aligned}$$

We define the measure $(z; D^s\rho)$ as the subtraction

$$\big(z; D^s\rho\big) = (z; D\rho) - zD^{ac}\rho\mathcal{L} = \lim_{h\to 0}\big(z^h; D^s\rho\big)$$

where $\mathcal{L}$ denotes the Lebesgue measure. It is proven in [31] that $(z^h; D^s\rho)$ is a singular measure and that

$$\big|\big(z^h; D^s\rho\big)\big| \leqslant \big\|z^h\big\|_{L^\infty(\Omega)}\big|D^s\rho\big|.$$

Then we also have when $\xi \geqslant 0$

$$\big\langle \xi; \big(z; D^s\rho\big)\big\rangle \leqslant \|z\|_{L^\infty([0,T]\times\Omega)}\int_0^T\int_\Omega \xi\big|D^s\rho\big|\, dy\, dt. \tag{12}$$

1.4. Statement of the result

Throughout this paper, we assume $\Omega \Subset \mathbb{R}^d$ to be a bounded convex domain, and $0 < m < \rho_0 < M$ which implies $\int_\Omega F(\rho_0) < \infty$. The support $\operatorname{spt}\mu$ of a measure μ on $\mathbb{R}^d$ refers to the smallest closed set of full mass. Let us now state our main result.

Theorem 1.8. *Let $\Omega \Subset \mathbb{R}^d$ be a bounded convex domain, and $\rho_0 \in P(\Omega)$ satisfy $0 < m < \rho_0 < M$. Let $c : \mathbb{R}^d \to [0, \infty]$ be a cost function on $\mathbb{R}^d$ given by $c(z) = \tilde{c}(|z|)$, where $\tilde{c} \in C^0([0,1]) \cap C^2([0,1[)$ with $|\nabla c(z)| \to \infty$ as $|z| \to 1$, $\tilde{c}(0) = 0 = \tilde{c}'(0)$ and $\tilde{c}''(\lambda) > 0$ on $[0,1[$. Let $F \in C^2(\mathbb{R})$ be a convex function such that $\frac{F(\lambda)}{\lambda} \to \infty$ when $\lambda \to \infty$ and $\lambda^d F(\lambda^{-d})$ is convex.*

(i) ***Characterization of the support of the optimal measure: Finite speed of propagation.***
If γ_i^h is the optimal measure for the minimization problem $(\mathcal{P}_i^h)$ with second marginal ρ_i^h, then $\rho^{ih} \in BV(\Omega)$ and

$$\operatorname{spt}\gamma_i^h \subset \left\{(x,y) \mid \frac{|x-y|}{h} < 1\right\} \cup Z_i^h \quad \text{with } \gamma_i^h\big(Z_i^h\big) = 0.$$

(ii) ***Euler–Lagrange equation: A discrete scheme.***
There exists a one-to-one map $S_i^h \in L^\infty(\Omega;\Omega)$ defined Lebesgue-a.e. by

$$S_i^h(y) = y + h\nabla c^*\big(D^{ac}\big(F'\big(\rho_i^h(y)\big)\big)\big) \tag{13}$$

such that $\gamma_i^h = (S_i^h \times id)_\# \rho_i^h$ and $D^{ac}(F'(\rho_i^h(y))) = F''(\rho_i^h(y)) D^{ac}\rho_i^h$.

(iii) ***Convergence of the measure ρ^h.***
Let ρ^h be the piecewise constant function defined from ρ_i^h by (3). As $h \to 0$, a subsequence of ρ^h converges strongly in $L^1([0,T]\times\Omega)$ towards $\bar\rho \in L^\infty([0,T]\times\Omega) \cap L^1_w([0,T], BV(\Omega))$ and $D\rho^h \rightharpoonup D\bar\rho$ in the sense of measures.

(iv) ***Limiting equation.***
Up to a subsequence, the limit of ρ^h, $\bar\rho$ belongs to $L^1_w([0,T], BV(\Omega)) \cap L^\infty([0,T]\times\Omega)$ and satisfies in the sense of distributions

$$\partial_t \bar\rho = \operatorname{div}\big(\bar\rho \nabla c^*\big(D^{ac}\big(F'(\bar\rho)\big)\big)\big). \tag{14}$$

Let us now make some remarks about this theorem.

- The first point of the theorem gives the finite speed of propagation, indeed, since in a time interval of length h, the displacement is bounded by h, the speed of propagation is bounded by 1. This property characterizes a relativistic transport.
- In (13), since $F \in C^2(]0,1[)$ and ρ^h is bounded from below, $F'(\rho^h)$ belongs to $BV(\Omega)$ and the chain rule for BV functions gives $D^{ac}(F'(\rho_i^h(y))) = F''(\rho_i^h(y)) D^{ac}\rho_i^h$.
- The second point of this theorem is the most important one. First, it claims the existence of an optimal map for the minimization problem $\mathcal{P}_i^h$ and it gives the Euler–Lagrange equation. The existence of an optimal map for a discontinuous cost function cannot be obtained by the same argument as for a smooth cost function. Indeed, the proof of the existence of a map for a smooth test function by the Kantorovich duality (see Gangbo and McCann [28], Villani [43]) is based on a uniform Lipschitz bound on the potential function given by the module of continuity of the cost function. More precisely, let γ_{opt} be the optimal measure for the minimization problem, the Kantorovich duality gives

$$\int_{\Omega\times\Omega} hc\left(\frac{x-y}{h}\right) d\gamma_{\text{opt}}(x,y) = \int_\Omega \phi(x)\, d\rho_{i-1}(x) + \int_\Omega \psi(y)\, d\rho_i(y),$$

where the potential functions ψ and ϕ are linked by the following relation

$$hc\left(\frac{x-y}{h}\right) - \phi(x) - \psi(y) \geqslant 0 \quad \forall (x,y) \in \Omega\times\Omega, \tag{15}$$

which becomes

$$hc\left(\frac{x-y}{h}\right)-\phi(x)-\psi(y)=0 \quad \forall(x,y)\in \operatorname{spt}\gamma_{\mathrm{opt}}. \tag{16}$$

When the cost function c is smooth, the relations (15)–(16) imply that the potential function ψ is Lipschitz — and then differentiable — and then it implies also the equality between the gradients

$$\nabla\psi(y)=\nabla c\left(\frac{x-y}{h}\right) \tag{17}$$

that gives directly the shape of the optimal map $x=S(y)=y+h\nabla c^*(\nabla\psi(y))$ since $\nabla c^*=(\nabla c)^{-1}$. In this present work, the potential function will not be Lipschitz anymore and then we have to find another argument to prove its almost everywhere differentiability. Moreover, to write an equality like (17), we need that the support of the optimal measure γ is — up to a negligible set — included in $\{(x,y)$ such that $\frac{|x-y|}{h}<1\}$ to be able to define the gradient of c.

So we introduce a mollified problem and pass to the limit. The sequence ρ^h will not be Lipschitz but will be in $BV(\Omega)$ which gives only the almost everywhere approximate differentiability. As Ambrosio, Gigli and Savaré showed in Theorem 6.4.2 [3], this is sufficient to define an optimal map. The strategy of cost mollification has often been used, for example in proof of the existence of a map for (non-strictly) convex cost via decomposition on one-dimensional rays by Ambrosio and Pratelli [4] (see also [26]).

Note that the regularity results for the optimal map of Ma, Trudinger and Wang [34] and Loeper [33] do not apply for this sign of cost function.

Note that the shape of the Euler–Lagrange equation involves ∇c^* and then we recover the finite speed of propagation (the relativistic effect) since ∇c^* is bounded.

- The proof is strongly based on a Fisher information-entropy inequality which thanks to the lower bound on the density gives the $L^1([0,T],BV(\Omega))$ bound on the solution.
- To pass to the limit when h goes to zero, we want to use a monotonicity argument (see Otto [38], Lions's book [32] or Evan's book [25]) to identify the nonlinear term. The identification of the weak limit of the flux is based on (8) and on the entropy equation (9) where in fact we would like to use the solution as a test function in the limiting equation. Then we have to define all terms of those relations very carefully and find the most precise framework in which the equation holds. For that purpose, we need the generalization of functionals to BV introduced in the previous paragraph and the introduction of the distribution $(z;D\rho)$ because in one sense, it replaces the product $\operatorname{div} z\rho$ when $\rho\in BV(\Omega)$ and $z\in L^\infty(\Omega)$.
- Note that the correction terms in (6) prevent the divergence term from being L^1 which means that the time discrete function is less regular than in Andreu, Caselles and Mazón [12] even if we recover the expected regularity in the continuous time limit. To avoid this difficulty, we introduce a different approximation of the right-hand side which will be in L^1 which is a crucial point in the definition of the distribution $(z;D\rho)$.
- The notion of entropic solution and the uniqueness [12] of solution in the class of $L^1_w([0,T],BV_2(\Omega))$ is not available here because of the approximation in the discrete in time equation and is not addressed in this paper.
- Particular case: Dimension 1.

  In dimension 1, this problem becomes radically simpler (see [18]). Let $\rho\in P(\Omega)$, we can extend it to $\mathbb{R}$ and following Villani's notations in [43], we introduce cumulative distribution functions

  $$l(x)=\int_{-\infty}^{x} d\rho = d\rho\big[(-\infty,x]\big].$$

  The function l is right-continuous, non-decreasing, and has limits $l(-\infty)=0$ and $l(+\infty)=1$. Then we define the generalized inverse of l on $[0,1]$

  $$l^{-1}(t)=\inf\big\{x\in\mathbb{R} \text{ s.t. } l(x)>t\big\}.$$

  In the particular case of dimension 1, the optimal transport plan and the optimal map does not depend on the convex cost function as soon as $c(x,y)=c(|x-y|)$ (cf. [36]). The optimal transport plan is given by

  $$\gamma_i=\big(l_{i-1}^{-1}\times l_i^{-1}\big)_\# d\mathcal{H}^1$$

where l_{i-1} is the cumulative distribution function for ρ_{i-1} and l_i is the cumulative distribution function for ρ_i. The optimal map is given — when ρ_{i-1} does not give mass to points — by

$$\rho_{i-1} = \left(l_{i-1}^{-1} \circ l_i\right)_{\#}\rho_i.$$

Then the Euler–Lagrange equation given by (5) gives directly

$$l_{i-1}^{-1} \circ l_i(y) - y = h\nabla c^*\left(\nabla\left(F'\left(l_i'(y)\right)\right)\right)$$

or

$$\frac{l_{i-1}^{-1}(x) - l_i^{-1}(x)}{h} = \nabla c^*\left(\nabla\left(F'\left(l_i'\left(l_i^{-1}(x)\right)\right)\right)\right)$$
$$= \nabla c^*\left(\nabla\left(F'\left(\frac{1}{(l_i^{-1})'(l_i \circ l_i^{-1}(x))}\right)\right)\right) = \nabla c^*\left(\nabla\left(F'\left(\frac{1}{(l_i^{-1})'(x)}\right)\right)\right)$$

which corresponds to an implicit Euler scheme for l^{-1}.

In the following section, we prove the existence of the infimum for problem $(\mathcal{P}_i(h))$. In Section 3, we prove some optimal transport theory (construction of an optimal map corresponding to the optimal measure and study of its properties) and we derive the Euler–Lagrange equation of the minimization problem (5). Then, in Section 4, we construct the piecewise constant function ρ^h which satisfies a time discrete version of (1). Finally, in the last section, we pass to the limit when the time step goes to zero, we use an argument of monotonicity to identify the limiting equation with Eq. (1).

2. Existence of solution to the minimization problem

Let us first study the minimization problem $(\mathcal{P}) = (\mathcal{P}_1^1)$ for the first step when $h = 1$ (we drop the superscript $h = 1$ and subscript $i = 1$ in this case):

$\rho_0 \in P(\Omega)$ given, find $\rho_1 \in P(\Omega)$ such that

$$I(\rho_0, \rho_1) = \inf_{\rho \in P(\Omega)} I(\rho_0, \rho)$$

with

$$I(\rho_0, \rho) = \int_\Omega F'\left(\rho(y)\right) dy + \inf_{\gamma \in \Gamma(\rho_0, \rho)} \int_{\Omega \times \Omega} c(x - y)\, d\gamma(x, y),$$

$\Gamma(\rho_0, \rho)$ denoting the set of measures that have ρ_0 and ρ as marginals.

Notations. We will use the following notations: given $\gamma \in \Gamma(\rho_0, \rho)$,

$$E(\rho) = \int_\Omega F\left(\rho(y)\right) dy =: \tilde{E}(\gamma), \quad \text{when seen as a function of } \gamma,$$
$$\tilde{W}(\gamma) = \int c(x - y)\, d\gamma(x, y), \quad \text{and}$$
$$W_c(\rho_0, \rho) = \inf_{\gamma \in \Gamma(\rho_0, \rho)} \tilde{W}(\gamma).$$

Following Agueh [1], we can prove that

Proposition 2.1. *Recall that we assume $m < \rho_0 < M$.*

(i) *There exists a unique minimum ρ_1^R satisfying*

$$I\left(\rho_0, \rho_1^R\right) = \inf_{\rho < R} I(\rho_0, \rho).$$

(ii) *There is a maximum principle that insures that $m < \rho_1^R < M$ when R is chosen such that $2M < R$.*

Proof. Let us denote $I_R = \inf_{\rho<R} I(\rho_0, \rho)$.

(i) To prove the existence of a minimizer, we use that $I_R \leqslant I(\rho_0, \rho_0) = E(\rho_0) < \infty$ and thanks to Jensen's inequality, since $\int_\Omega \rho(y)\,dy = 1$, we can write that $I_R \geqslant \inf_\rho E(\rho) \geqslant |\Omega| F(\frac{1}{|\Omega|})$. Thus I_R is finite. Let $\rho^{(n)}$ be a minimizing sequence. Since $\rho^{(n)} < R$, the sequence is bounded in $L^\infty(\Omega)$ and since Ω is bounded, up to a subsequence, $\rho^{(n)}$ converges in $L^1(\Omega)'$ towards $\rho_1^R \in L^\infty(\Omega)$ satisfying $\rho_1^R \leqslant R$. Because of the lower semi-continuity of $I(\rho)$, we obtain

$$I\big(\rho_0, \rho_1^R\big) \leqslant \lim_{n\to\infty} \inf I\big(\rho_0, \rho^{(n)}\big) \leqslant I_R \leqslant I\big(\rho_0, \rho_1^R\big).$$

The uniqueness comes from the strict convexity of I. Note that

$$E\big(\rho_1^R\big) \leqslant I\big(\rho_0, \rho_1^R\big) \leqslant I(\rho_0, \rho_0) = E(\rho_0).$$

(ii) To prove the maximum principle, we argue by contradiction to prove that $m < \rho_1^R < M$. Note that it implies also that ρ_1^R does not depend on R. Let us do the argument for the upper bound, the same proof leads to the lower bound. For that, assume that $H = \{x \mid \text{such that } \rho_1^R > M\}$ has a strictly positive Lebesgue measure; we can then construct a new measure that would be a better measure, which gives us a contradiction. Starting from γ_{opt}, the optimal measure associated to ρ_1^R, let us construct a sequence of measures γ^η with second marginal ρ^η depending on η, that will make the cost decrease proportionately to η while the entropy increases proportionately to η^2. Then for η small enough, the total $I(\rho_0, \rho^\eta)$ will be smaller than $I(\rho_0, \rho_1^R)$.

To be precise, we recall the construction of this sequence which follows the argument made by Martial Agueh in [1].

Note that $\gamma_{\mathrm{opt}}(H^c \times H) > 0$ where $H^c = \mathbb{R}^d \setminus H$; otherwise

$$M|H| < \int_H \rho_1^R(y)\,dy = \gamma_{\mathrm{opt}}\big(\mathbb{R}^d \times H\big) = \gamma_{\mathrm{opt}}(H \times H) \leqslant \gamma_{\mathrm{opt}}\big(H \times \mathbb{R}^d\big) = \int_H \rho_0(x)\,dx \leqslant M|H|.$$

On $H^c \times H$, for a part of the measure depending on η, we will leave x in place instead of sending x to y, i.e. we define the action of the sequence γ^η against a test function ξ by

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \xi(x,y)\,d\gamma^\eta(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \xi(x,y)\,d\gamma_{\mathrm{opt}}(x,y) + \eta \int_{H^c\times H} \big(\xi(x,x) - \xi(x,y)\big)\,d\gamma_{\mathrm{opt}}(x,y).$$

The corresponding marginal ρ^η can also be defined as $\rho_1^R + \eta(v_0 - v_1)$ where v_0 is the first marginal of $\nu = \gamma_{\mathrm{opt}} \mathbf{1}_{H^c\times H}$, the restriction of γ_{opt} to the set $H^c \times H$ (respectively, v_1 is the second marginal of ν). Since $\nu \ll \gamma_{\mathrm{opt}}$, we have

- $0 \leqslant v_0 \leqslant M$ a.e., and $0 \leqslant v_1 \leqslant R$ a.e.,
- $v_0 = 0$ on H and $v_1 = 0$ on H^c. (18)

Then $0 \leqslant \rho^\eta \leqslant M + \eta M \leqslant R$ on H^c and $M - \eta v_1 \leqslant \rho^\eta \leqslant R$ on H, so we obtain $0 \leqslant \rho^\eta \leqslant R$ (η is chosen such that $\eta R < M$ and $2M \leqslant R$). Moreover,

$$\int_\Omega \rho^\eta(y)\,dy = 1 + \eta\big(\gamma_{\mathrm{opt}}\big(H^c \times H\big) - \gamma_{\mathrm{opt}}\big(H^c \times H\big)\big) = 1.$$

So we effectively have constructed a new measure γ^η belonging to $\Gamma(\rho_0, \rho^\eta)$ with $\rho^\eta \leqslant R$. We still have to prove that $I(\rho_0, \rho^\eta) < I(\rho_0, \rho_1^R)$. For that, we will compute

$$\begin{aligned} W_c\big(\rho_0, \rho^\eta\big) - W_c\big(\rho_0, \rho_1^R\big) &\leqslant \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\,d\gamma^\eta(x,y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x-y)\,d\gamma_{\mathrm{opt}}(x,y) \\ &= -\eta \int_{H^c\times H} c(x-y)\,d\gamma_{\mathrm{opt}}(x,y). \end{aligned}$$

This means that with the new measure, we have managed to lower the cost by a quantity of order η. But the entropy has to increase: let us check how much by computing

$$\int_{\Omega} \left(F\left(\rho^{\eta}(y)\right) - F\left(\rho_1^R(y)\right)\right) dy.$$

Using the convex property of F and (18), we decompose

$$\begin{aligned}\int_{\Omega} \left(F\left(\rho^{\eta}(y)\right) - F\left(\rho_1^R(y)\right)\right) dy &= \int_{H^c} \left(F\left(\rho^{\eta}(y)\right) - F\left(\rho_1^R(y)\right)\right) dy + \int_{H} \left(F\left(\rho^{\eta}(y)\right) - F\left(\rho_1^R(y)\right)\right) dy \\ &\leqslant \eta\left[\int_{H^c} F'\left(\rho_1^R(y) + \eta v_0(y)\right) v_0(y)\, dy - \int_{H} F'\left(\rho_1^R(y) - \eta v_1(y)\right) v_1(y)\, dy\right] \\ &\leqslant \eta\left[\int_{H^c} F'\left(M + \eta v_0(y)\right) v_0(y)\, dy - \int_{H} F'\left(M - \eta v_1(y)\right) v_1(y)\, dy\right]\end{aligned}$$

since $\rho_1^R(y) \leqslant M$ if $y \in H^c$ and $\rho_1^R(y) \geqslant M$ if $y \in H$, and F' is non-decreasing. But also

$$\begin{aligned}&\int_{H^c} F'\left(M + \eta v_0(y)\right) v_0(y)\, dy - \int_{H} F'\left(M - \eta v_1(y)\right) v_1(y)\, dy \\ &\quad = \int_{H^c \times H} \left(F'\left(M + \eta v_0(x)\right) - F'\left(M - \eta v_1(y)\right)\right) d\gamma_{\mathrm{opt}}(x, y) \\ &\quad \leqslant \eta \sup|F''| \sup_{\Omega} |v_0 - v_1|.\end{aligned}$$

Since F is $C^2(\mathbb{R}^+)$, the entropy increases by a quantity of order η^2, which means that for η small enough, $I(\rho_0, \rho^{\eta}) < I(\rho_0, \rho_1^R)$. This contradiction establishes Proposition 2.1. □

3. Optimal transport theory: a time discrete equation

In this work we deal with a convex cost function which may be discontinuous and takes infinity as a value. As we said in the Introduction, for this kind of cost function, we cannot apply the classical result of Gangbo and McCann [28,29] or Caffarelli [17] to define an optimal map associated to the optimal measure involved in initial time step $\mathcal{P}_0^1$. To construct a map in the present case, we will use the properties of the optimal map for a mollified case using strongly the double minimization process (combining the cost minimization with respect to γ for a fixed ρ with the entropy minimization with respect to ρ). Indeed, note that for given ρ_0, it is easy to construct ρ_1 such that the value of

$$W_c(\rho_0, \rho_1) = \inf_{\gamma \in \Gamma(\rho_0, \rho_1)} \int c(x - y)\, d\gamma(x, y),$$

the double minimization always produces ρ_1 for which the infimum is finite.

3.1. Previous results

We recall in this section Propositions 2.6 and 2.7 of the paper of M. Agueh [1] (see also [35]) and a result due to D. Cordero-Erausquin [21] and F. Otto [38] also quoted in [1] that we will use for the mollified approximation of $\mathcal{P}$. L. Ambrosio, Gigli, and G. Savaré's Theorem 6.2.7 [3] could also be used.

Proposition 3.1. *(See [21,38].) Let $\rho_0, \rho_1 \in P(\Omega)$ and assume that $c : \mathbb{R}^d \to [0, \infty[$ is strictly convex and satisfies $c, c^* \in C^2(\mathbb{R}^d)$. Denote by S, the c-optimal map that pushes ρ_1 forward to ρ_0, and define the interpolant map S_t, and the interpolant measure μ_{1-t}, by*

$$S_t = (1 - t)Id + tS \quad \textit{and} \quad \mu_{1-t} = (S_t)_{\#}\rho_1,$$

for $t \in [0, 1]$. Then

(i) *S_t is injective for $t<1$, and μ_{1-t} is absolutely continuous with respect to Lebesgue.*

Moreover, there exists a subset B of Ω, of full measure for $\mu_1=\rho_1(y)\,dy$, such that, for $y\in B$ and $t\in[0,1]$,

(ii) *$\nabla S(y)$ is diagonalizable with positive eigenvalues.*
(iii) *The pointwise Jacobian* $\det\nabla S$ *satisfies*

$$0\neq\rho_1(y)=\rho_{1-t}\big(S_t(y)\big)\det\big[(1-t)Id+t\nabla S(y)\big], \tag{19}$$

where ρ_{1-t} is the density function of μ_{1-t}.

In addition, if $\rho_1>0$ a.e., then:

(iv) *The pointwise divergence* $\operatorname{div}S$ *is integrable on Ω, and*

$$\int_\Omega \operatorname{div}\big(S(y)-y\big)\xi(y)\,dy\leqslant-\int_\Omega\big\langle S(y)-y;\nabla\xi\big\rangle\,dy,$$

for $\xi\geqslant 0$ in $C_c^\infty(\mathbb{R}^d)$.

Proposition 3.2 *(Optimal map for smooth cost function). (See [1].) Let $\rho_0\in P(\Omega)$ be such that $m\leqslant\rho_0\leqslant M$ a.e. Assume that $F:[0,\infty[\,\to\mathbb{R}$ is strictly convex, and satisfies $F\in C^2((0,\infty))$, and $c:\mathbb{R}^d\to[0,\infty[$ is strictly convex, of class C^1 such that $0=c(0)<c(|z|)$ for $|z|\neq 0$ and c coercive, i.e. $\lim_{|z|\to\infty}c(z)=\infty$. If ρ_1 denotes the minimizer for $(\mathcal{P}_0^1)$, then the following holds*:

$$\int_{\Omega\times\Omega}\nabla c(x-y)\cdot\xi(y)\,d\gamma(x,y)+\int_\Omega P_F\big(\rho_1(y)\big)\operatorname{div}\xi(y)\,dy=0,$$

for $\xi\in C_c^\infty(\Omega;\mathbb{R}^d)$; here $P_F(\lambda)=\lambda F'(\lambda)-F(\lambda)$ for $\lambda\in(0,\infty)$, and γ is the c-optimal measure in $\Gamma(\rho_0,\rho_1)$. Moreover,

(i) *$P_F(\rho_1)\in W^{1,\infty}(\Omega)$.*
(ii) *If S is the c-optimal map that pushes ρ_1 forward to ρ_0, then*

$$S(y)-y=\nabla c^*\big[\nabla\big(F'\big(\rho_1(y)\big)\big)\big], \tag{20}$$

for a.e. $y\in\Omega$, and for $\xi\in C^2(\overline{\Omega})$,

$$\left|\int_\Omega\big(\rho_1(y)-\rho_0(y)\big)\xi(y)\,dy+\int_\Omega\rho_1(y)\nabla c^*\big[\nabla\big(F'\big(\rho_1(y)\big)\big)\big]\cdot\nabla\xi(y)\,dy\right|$$
$$\leqslant\frac{1}{2}\sup_{x\in\overline{\Omega}}\big|D^2\xi(x)\big|\int_{\Omega\times\Omega}|x-y|^2\,d\gamma(x,y).$$

Proposition 3.3 *(Displacement convexity of energy (above-tangent form)). (See [1].) Let $\rho_0,\rho_1\in P(\Omega)$ be density functions of two Borel probability measures μ_0 and μ_1 on $\mathbb{R}^d$, respectively. Let $c:\mathbb{R}^d\to[0,\infty[$ be strictly convex, such that $c,c^*\in C^2(\mathbb{R}^d)$. Let $F:[0,\infty)\to\mathbb{R}$ be differentiable on $]0,\infty[$, such that $F(0)=0$, and $\lambda\to\lambda^dF(\lambda^{-d})$ be convex non-increasing on $]0,\infty[$. Then, the internal energy inequality holds, i.e.*

$$\int_\Omega F\big(\rho_0(y)\big)\,dy-\int_\Omega F\big(\rho_1(y)\big)\,dy\geqslant-\int_\Omega P_F\big(\rho_1(y)\big)\operatorname{div}\big(S(y)-y\big)\,dy.$$

In addition, if $P_F(\rho_1)\in W^{1,\infty}(\Omega)$ and $\rho_1>0$ a.e., then

$$\int_\Omega F\big(\rho_0(y)\big)\,dy-\int_\Omega F\big(\rho_1(y)\big)\,dy\geqslant\int_\Omega\nabla\big[F'\big(\rho_1(y)\big)\big]\cdot\big(S(y)-y\big)\rho_1(y)\,dy. \tag{21}$$

3.2. *Construction of an optimal map for* $(\mathcal{P}_0^1)$*: a mollification process*

Definition of the mollified problem

Recall that the hemispherical cost

$$c(z) = \begin{cases} 1 - \sqrt{1-|z|^2} & \text{where } |z| \leqslant 1, \\ +\infty & \text{where } |z| > 1, \end{cases}$$

is the motivating example, whose Legendre transform (4) is given by the hyperboloid

$$c^*(z) = \sqrt{1+|z|^2} - 1.$$

In a more general setting, we deal with radial cost functions $c:\mathbb{R}^d \to [0,\infty]$ satisfying $c(z) = \tilde{c}(|z|)$ where $\tilde{c} \in C^2([0,1[) \cap C^0([0,1])$ satisfies $\tilde{c}''(\lambda) > 0$ on $[0,1[$ and $\tilde{c}(0) = 0 = \tilde{c}'(0)$ and $|\nabla c(z)| \to \infty$ as $|z| \to 1$. Let mollify those general cost functions c by the Yosida regularization [16]

$$c^\epsilon(z) = \inf_{w\in\mathbb{R}^d}\left(c(z-w) + \frac{|w|^2}{2\varepsilon}\right).$$

Note that from $w = 0$ we obtain $c^\varepsilon(z) \leqslant c(z)$. In fact, the Legendre transform $c^{\varepsilon*}$ of this mollified cost is a strict convexification of c^*, namely

$$c^{\varepsilon*}(z) = c^*(z) + \frac{\varepsilon}{2}|z|^2.$$

Note that the mollified c^ε is finite and convex, hence continuous.

Here, we recall an argument of [1] to justify the regularity of c^ε. The function $c^{\varepsilon*}(z)$ is a $C^2(\mathbb{R}^d)$ function nonnegative and strictly convex and $\nabla c^\varepsilon = (\nabla c^{\varepsilon*})^{-1}$ and then c^ε belongs to $C^1(\mathbb{R}^d)$ and the function

$$z \to \frac{\operatorname{cof}(D^2 c^{\varepsilon*}[\nabla c^\varepsilon(z)])}{\det(D^2 c^{\varepsilon*}[\nabla c^\varepsilon(z)])} := D^2 c^\varepsilon(z)$$

is well defined and is continuous on $\mathbb{R}^d$. Then $c^\varepsilon \in C^2(\mathbb{R}^d)$.

Consider the sequence of approximate minimization problems $(\mathcal{P}^\varepsilon)$.

Find γ^ε *with first marginal* ρ_0 *and second marginal* ρ^ε *such that*

$$I^\varepsilon(\rho_0, \rho^\varepsilon) = \tilde{I}^\varepsilon(\gamma^\varepsilon) = \min_{\gamma\in P_0(\Omega\times\Omega)}\big(\tilde{E}(\gamma) + \tilde{W}_{c^\varepsilon}(\gamma)\big)$$

where $P_0(\Omega\times\Omega) = \{\gamma \text{ such that } \int_{\Omega\times\Omega}\varphi(x)\,d\gamma(x,y) = \int_\Omega \varphi(x)\,d\rho_0(x)\}$ i.e. the probabilities with first marginal ρ_0.

Here ρ_0 represents any ρ_i and is assumed to satisfy $m \leqslant \rho_0 \leqslant M$.

The mollification allows us to apply the previous results quoted above to this sequence of problems since c^ε has the regularity required.

Lemma 3.1 *(The Kantorovich duality for the mollified problem). For* $\varepsilon > 0$ *fixed, we have*

$$\int_{\Omega\times\Omega} c^\varepsilon(x-y)\,d\gamma^\varepsilon(x,y) = \int_\Omega \phi^\varepsilon(x)\,d\rho_0(x) + \int_\Omega \psi^\varepsilon(y)\,d\rho^\varepsilon(y) \tag{22}$$

where the potential $\psi^\varepsilon = -F'(\rho^\varepsilon)$ *and* ϕ^ε *is the corresponding c-transform given by*

$$\phi^\varepsilon(x) = \inf_{y\in\overline{\Omega}}\big(c^\varepsilon(x-y) - \psi^\varepsilon(y)\big). \tag{23}$$

Proof. First recall the Kantorovich duality theory for smooth cost described in Rachev and Rüschendorf [39] which gives that

$$\begin{aligned}\inf_{\gamma\in\Gamma(\rho_0^\varepsilon,\rho^\varepsilon)}\int_{\Omega\times\Omega} c^\varepsilon(x-y)\,d\gamma(x,y) &= \sup_{\phi,\psi\in T_c}\int_\Omega \phi(x)\,d\rho_0(x) + \int_\Omega \psi(y)\,d\rho^\varepsilon(y)\\ &= \sup_{\psi,\psi^c}\int_\Omega \psi^c(x)\,d\rho_0(x) + \int_\Omega \psi(y)\,d\rho^\varepsilon(y)\end{aligned}$$

where $\phi, \psi \in T_c$ mean that $\psi(y)+\phi(x) \leqslant c^\varepsilon(x-y)$ for all $x, y \in \Omega$, and ψ^c is given by (23). Moreover the supremum is attained. In our study, since ρ^ε is the second marginal corresponding to the optimal measure of the complete minimization problem $\mathcal{P}^\varepsilon$, the Euler–Lagrange equation (20) for this problem is (see [1])

$$S^\varepsilon(y) - y = \nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho^\varepsilon(y)\big)\big)\big)$$

where $S^\varepsilon(y)$ is the optimal map associated to the potential ψ^ε by

$$S^\varepsilon(y) - y = -\nabla c^{\varepsilon *}\big(\nabla \psi^\varepsilon(y)\big).$$

This means that up to a constant (that we can fix to be zero without loss of generality since $\int_\Omega \rho^\varepsilon(y)\,dy = 1$), we have

$$\psi^\varepsilon(y) = -F'\big(\rho^\varepsilon(y)\big). \qquad \square$$

Lemma 3.2 *(Existence of a limiting measure).*

(i) *Up to a subsequence, γ^ε converges, as $\varepsilon \to 0$, towards the probability γ^∞ in $C^0(\Omega)'$.*
(ii) *The support of the limiting measure γ^∞ is included in $\Delta_1 = \{(x, y)$ such that $|x-y| \leqslant 1\}$.*
(iii) *Identification of the limiting measure*:
The limiting measure is the optimal measure for the initial problem, i.e.

$$\gamma^\infty = \gamma_{\text{opt}} \quad or \quad \tilde{I}\big(\gamma^\infty\big) = \min_{\gamma \in P_0(\Omega\times\Omega)} \big(\tilde{E}(\gamma) + \tilde{W}_c(\gamma)\big).$$

Proof. (i) Since γ^ε is a sequence of probability measures, it is relatively compact and converges up to a subsequence in $C^0(\Omega)'$.

(ii) Let us define $N_\delta = \{(x, y) \in \Omega \times \Omega$ such that $|x-y| \geqslant 1+\delta\}$, we claim that $\gamma^\varepsilon(N_\delta) \to 0$ when $\varepsilon \to 0$. Then we obtain that $\gamma^\infty(N_\delta) = 0$ for any $\delta > 0$, hence $\operatorname{spt}\gamma^\infty \subset \Delta_1$.

Note that,

$$C \geqslant E(\rho_0) \geqslant \tilde{I}^\varepsilon\big(\gamma^\varepsilon\big) \geqslant \int_{N_\delta} c^\varepsilon(x-y)\,d\gamma^\varepsilon(x, y) \geqslant \tilde{c}^\varepsilon(1+\delta)\gamma^\varepsilon(N_\delta).$$

Since $\tilde{c}^\varepsilon(1+\delta) \to \infty$ when $\varepsilon \to 0$, the previous inequality implies that $\gamma^\varepsilon(N_\delta) \to 0$ when $\varepsilon \to 0$ and then (ii) holds.

(iii) First of all, we prove that $\tilde{I}^\varepsilon(\gamma^\varepsilon) \to \tilde{I}(\gamma^\infty)$.

Let us first prove that

$$\limsup_{\varepsilon\to 0}\big(\tilde{I}\big(\gamma^\infty\big) - \tilde{I}^\varepsilon\big(\gamma^\varepsilon\big)\big) \leqslant 0. \tag{24}$$

The lower semi-continuity of the entropy leads to the first part of the inequality

$$\tilde{E}\big(\gamma^\infty\big) \leqslant \liminf_{\varepsilon\to 0} \tilde{E}\big(\gamma^\varepsilon\big).$$

For the part involving the cost function, we use the fact that $c^\varepsilon \geqslant 0$ and that the support of the limiting measure γ^∞ is a subset of Δ_1. Indeed, it leads to

$$\begin{aligned}
&\int_{\Omega\times\Omega} c(x-y)\,d\gamma^\infty(x, y) - \int_{\Omega\times\Omega} c^\varepsilon(x-y)\,d\gamma^\varepsilon(x, y) \\
&\quad \leqslant \int_{\Delta_1} \big[c(x-y)\,d\gamma^\infty(x, y) - c^\varepsilon(x-y)\,d\gamma^\varepsilon(x, y)\big] \\
&\quad = \int_{\Delta_1} \big(c(x-y) - c^\varepsilon(x-y)\big)\,d\gamma^\varepsilon(x, y) + \int_{\Delta_1} c(x-y)\big(d\gamma^\infty(x, y) - d\gamma^\varepsilon(x, y)\big).
\end{aligned}$$

Up to the introduction of a $C^0(\Omega)$ extension of the restriction of c to the unit ball, since the sequence $\gamma^\varepsilon \to \gamma^\infty$ in the sense of measures (or in $C^0(\Omega)'$), the second term goes to zero.

To prove that the first term goes to zero, we will prove that

$$\left\| c(z) - c^{\varepsilon}(z) \right\|_{L^{\infty}(\{|z| \leqslant 1\})} \leqslant M(\varepsilon) \tag{25}$$

where $M(\varepsilon) \to 0$ when $\varepsilon \to 0$.

Indeed, for any $t \in [0,1]$ and $t > \theta > 0$

$$\tilde{c}(t) - \tilde{c}(t-\theta) \leqslant \tilde{c}(1) - \tilde{c}(1-\theta)$$

since for θ fixed, $t \to \tilde{c}(t) - \tilde{c}(t-\theta)$ is a non-decreasing function. Note also that

$$c(z) - c^{\varepsilon}(z) = \sup_{w \in B(z,1)} \left(c(z) - \tilde{c}\big(|z-w|\big) - \frac{|w|^2}{2\varepsilon} \right) = \sup_{w \in B(z,1)} \left(\tilde{c}\big(|z|\big) - \tilde{c}\big(|z| - |w|\big) - \frac{|w|^2}{2\varepsilon} \right).$$

Since for any $w \in B(z,1)$,

$$c(z) - \tilde{c}\big(|z| - |w|\big) - \frac{|w|^2}{2\varepsilon} \leqslant \tilde{c}(1) - \tilde{c}\big(1 - |w|\big) - \frac{|w|^2}{2\varepsilon} \leqslant \tilde{c}(1) - \tilde{c}^{\varepsilon}(1),$$

we finally obtain

$$c(z) - c^{\varepsilon}(z) \leqslant \tilde{c}(1) - \tilde{c}^{\varepsilon}(1).$$

By taking $M(\varepsilon) = \tilde{c}(1) - \tilde{c}^{\varepsilon}(1)$, (25) holds.

Since $\gamma^{\varepsilon}(\Delta_1) \leqslant 1$, we have

$$\int_{\Delta_1} \big(c(x-y) - c^{\varepsilon}(x-y)\big)\, d\gamma^{\varepsilon}(x,y) \leqslant M(\varepsilon)$$

and then (24) holds.

Let us now prove the other inequality

$$\tilde{I}\big(\gamma^{\infty}\big) \geqslant \limsup_{\varepsilon \to 0} \tilde{I}\big(\gamma^{\varepsilon}\big).$$

Recall that we have the following inequalities,

$$\tilde{I}^{\varepsilon}\big(\gamma^{\varepsilon}\big) \leqslant \tilde{I}^{\varepsilon}(\gamma) \leqslant \tilde{I}(\gamma)$$

for any γ which has ρ^0 as first marginal, including $\gamma = \gamma^{\infty}$.

Thus, we obtain the expected inequality since

$$\limsup_{\varepsilon \to 0} \tilde{I}^{\varepsilon}\big(\gamma^{\varepsilon}\big) \leqslant \tilde{I}(\gamma)$$

for any measure γ which has ρ_0 as first marginal and in particular for γ^{∞}.

To conclude the proof of (iii), we use the last inequality to obtain that

$$\tilde{E}\big(\gamma^{\infty}\big) + \int_{\Omega \times \Omega} c(x-y)\, d\gamma^{\infty}(x,y) \leqslant \inf_{\gamma} \left[\tilde{E}(\gamma) + \int_{\Omega \times \Omega} c(x-y)\, d\gamma(x,y) \right]$$

and then γ^{∞} is equal to γ_{opt}, the minimizer for the initial problem $(\mathcal{P})$ and $\rho^{\infty} = \rho_1$. □

Lemma 3.3 *(The Kantorovich duality for the limiting problem).*

(i) *There exist $\phi^{\infty} \in L^{\infty}(\Omega)$ and $\psi^{\infty} \in L^{\infty}(\Omega) \cap BV(\Omega)$ such that, up to a subsequence, the potentials $(\psi^{\varepsilon}, \phi^{\varepsilon})$ from* (26)–(27) *satisfy*

$$\psi^{\varepsilon} \rightharpoonup \psi^{\infty} \quad \textit{weak in } BV(\Omega),\ \textit{strong in } L^1(\Omega),$$
$$\rho^{\varepsilon} \rightharpoonup \rho^{\infty} = \rho_1 = (F^*)'\big(-\psi^{\infty}\big) \quad \textit{weak in } BV(\Omega),\ \textit{strong in } L^1(\Omega),$$
$$\phi^{\varepsilon} \rightharpoonup \phi^{\infty} \quad \textit{weak* in } L^{\infty}(\Omega) = L^1(\Omega)'.$$

(ii) *The Kantorovich duality holds for the limiting problem*

$$\int_{\Omega\times\Omega} c(x-y)\,d\gamma_{\mathrm{opt}}(x,y) = \int_{\Omega} \phi^{\infty}(x)\,d\rho_0(x) + \int_{\Omega} \psi^{\infty}(y)\,d\rho_1(y). \tag{26}$$

Moreover,

$$c(x-y) - \phi^{\infty}(x) - \psi^{\infty}(y) \geqslant 0 \tag{27}$$

almost everywhere on $\Omega\times\Omega$ and the equality holds γ_{opt}-almost everywhere.

Proof. (i) Since for all ε, $\psi^{\varepsilon}(y) = -F'(\rho^{\varepsilon}(y))$ with $m < \rho^{\varepsilon} < M$, up to a subsequence, ψ^{ε} converges weak* in $L^{\infty}(\Omega)$ towards $\psi^{\infty} \in L^{\infty}(\Omega)$. Moreover, using (21) and (20), applied for c^{ε}, we obtain

$$\int_{\Omega} \nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\cdot \rho^{\varepsilon}(y)\nabla c^{*}\big(\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\big) + \varepsilon\rho^{\varepsilon}(y)\big|\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\big|^{2}\,dy \leqslant \int_{\Omega}\big(F\big(\rho_0(y)\big) - F\big(\rho^{\varepsilon}(y)\big)\big)\,dy$$

which gives in particular

$$\int_{\Omega} \rho^{\varepsilon}(y)\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\cdot\nabla c^{*}\big(\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\big)\,dy \leqslant \int_{\Omega}\big(F\big(\rho_0(y)\big) - F\big(\rho^{\varepsilon}(y)\big)\big)\,dy \tag{28}$$

and then, since $m < \rho^{\varepsilon} < M$ (see [1]), $\nabla\rho^{\varepsilon}$ is a bounded sequence in $L^{1}(\Omega)$. Indeed,

$$\int_{\Omega}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy = \int_{|\nabla\rho^{\varepsilon}|<1}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy + \int_{|\nabla\rho^{\varepsilon}|>1}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy \leqslant |\Omega| + \int_{|\nabla\rho^{\varepsilon}|>1}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy.$$

Since we can write $|\nabla(F'(\rho^{\varepsilon}))| = |\nabla\rho^{\varepsilon}||F''(\rho^{\varepsilon})|$, we have

$$\begin{aligned}\int_{|\nabla\rho^{\varepsilon}|>1}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy &= \int_{|\nabla\rho^{\varepsilon}|>1}\frac{|\nabla(F'(\rho^{\varepsilon}(y)))|}{|F''(\rho^{\varepsilon}(y))|}\,dy\\ &\leqslant \frac{1}{K}\int_{\Omega}\big|\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\big|\,dy\end{aligned}$$

where $K = \inf_{[m,M]} F''$.

Recall that we have the relation $|z||\nabla c^{*}(z)| = |z\cdot\nabla c^{*}(z)|$ which implies $|z| \leqslant \frac{|z\cdot\nabla c^{*}(z)|}{w(s)}$ when $|z| > s$ since $|\nabla c^{*}(z)| = \omega(|z|)$ is a non-decreasing function of $|z|$.

Moreover, when $|\nabla\rho^{\varepsilon}| > 1$, $|\nabla(F'(\rho^{\varepsilon}))| \geqslant K$ and then

$$\begin{aligned}\int_{|\nabla\rho^{\varepsilon}|>1}\big|\nabla\rho^{\varepsilon}(y)\big|\,dy &\leqslant \frac{1}{Km|\nabla c^{*}(K)|}\int_{\Omega}\rho^{\varepsilon}(y)\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\cdot\nabla c^{*}\big(\nabla\big(F'\big(\rho^{\varepsilon}(y)\big)\big)\big)\\ &\leqslant C\int_{\Omega}\big[F\big(\rho_0(y)\big) - F\big(\rho^{\varepsilon}(y)\big)\big]\,dy\end{aligned}$$

and then

$$\big\|\nabla\rho^{\varepsilon}\big\|_{L^{1}(\Omega)} \leqslant C \tag{29}$$

where the constant C does not depend on ε.

Then, $\rho_1 \in BV(\Omega)$ and consequently, ψ^{∞}, the limit of ψ^{ε} belongs also to $BV(\Omega)$ and the convergence is strong in $L^{1}(\Omega)$ (cf. [27, Theorem 4, Section 5.2.3]).

Finally, since ϕ^{ε} is defined from ψ^{ε} by (23), ϕ^{ε} is a bounded sequence in $L^{\infty}(\Omega)$. Indeed by (23)

$$-\psi^{\varepsilon}(y) \leqslant -\psi^{\varepsilon}(y) + c^{\varepsilon}(x-y) = \phi^{\varepsilon}(x) \leqslant -\psi^{\varepsilon}(x) \quad\Rightarrow\quad \forall x\in\Omega,\quad -\sup_{[m,M]} F' \leqslant \phi^{\varepsilon}(x) \leqslant -\inf_{[m,M]} F'.$$

Then, up to the extraction of a subsequence ϕ^{ε} converges in weak* in $L^{\infty}(\Omega)$.

(ii) In the previous lemma, we proved that

$$\lim_{\varepsilon\to 0} \tilde{E}\big(\gamma^\varepsilon\big) + \int_{\Omega\times\Omega} c^\varepsilon(x-y)\,d\gamma^\varepsilon(x,y) = \tilde{E}(\gamma_{\mathrm{opt}}) + \int_{\Omega\times\Omega} c(x-y)\,d\gamma_{\mathrm{opt}}(x,y). \tag{30}$$

On the other hand, (22) says that

$$\tilde{E}\big(\gamma^\varepsilon\big) + \int_{\Omega\times\Omega} c^\varepsilon(x-y)\,d\gamma^\varepsilon(x,y) = \tilde{E}\big(\gamma^\varepsilon\big) + \int_\Omega \phi^\varepsilon(x)\,d\rho_0 + \int_\Omega \psi^\varepsilon(y)\,d\rho^\varepsilon.$$

Since ϕ^ε converges toward ϕ^∞ in weak* in $L^\infty(\Omega) = L^1(\Omega)'$, we have

$$\lim_{\varepsilon\to 0} \int_\Omega \phi^\varepsilon(x)\,d\rho_0(x) = \int_\Omega \phi^\infty(x)\,d\rho_0(x).$$

Moreover $\rho^\varepsilon \to \rho_1$ strongly in $L^1(\Omega)$ and ρ^ε is bounded in $L^\infty(\Omega)$, so the dominated convergence theorem yields

$$\begin{aligned}
\tilde{E}\big(\gamma^\varepsilon\big) + \int_\Omega \psi^\varepsilon(y)\,d\rho^\varepsilon(y) &= \int_\Omega F\big(\rho^\varepsilon(y)\big)\,dy - \int_\Omega \rho^\varepsilon(y) F'\big(\rho^\varepsilon(y)\big)\,dy \\
&\to \int_\Omega F\big(\rho_1(y)\big)\,dy - \int_\Omega \rho_1(y) F'\big(\rho_1(y)\big)\,dy \\
&= E(\gamma_{\mathrm{opt}}) + \int_\Omega \psi^\infty(y)\,d\rho_1(y).
\end{aligned}$$

Combined with (30), this gives

$$\tilde{E}(\gamma_{\mathrm{opt}}) + \int_{\Omega\times\Omega} c(x-y)\,d\gamma_{\mathrm{opt}}(x,y) = \tilde{E}(\gamma_{\mathrm{opt}}) + \int_\Omega \phi^\infty(x)\,d\rho_0(x) + \int_\Omega \psi^\infty(y)\,d\rho_1(y)$$

which leads to (26).

Finally, since for all ε,

$$c^\varepsilon(x-y) - \phi^\varepsilon(x) - \psi^\varepsilon(y) \geqslant 0$$

using the fact that $c^\varepsilon(x-y) \leqslant c(x-y)$ we have

$$c(x-y) - \phi^\varepsilon(x) - \psi^\varepsilon(y) \geqslant 0$$

which gives (27) at the limit.

Since (26) holds, the equality holds in the support of γ_{opt}. □

Lemma 3.4 *(Existence of an optimal map).*

(i) *The support of γ_{opt} is included, up to a negligible set, in the c-superdifferential of the potential function ψ^∞, i.e. if $(x,y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$, $x \in \partial^c\psi^\infty(y)$.*
(ii) *The set $\{y \in \Omega;\ \exists x_1, \neq x_2$ with $|x_1 - y| < 1$ and $(x_1, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$ and $(x_2, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})\}$ is a negligible set.*
(iii) *Let γ_{opt} be the optimal measure for the minimization problem $(\mathcal{P})$,*

$$\mathrm{spt}\,\gamma_{\mathrm{opt}} \subset \big\{(x,y) \text{ with } |x-y| < 1\big\} \cup Z_0 \quad \text{with } \gamma_{\mathrm{opt}}(Z_0) = 0.$$

More precisely, for each $\delta \in\,]0,1[$, we obtain an estimate of the measure of the set $Z_\delta = \{(x,y) \in \mathrm{spt}\,\gamma_{\mathrm{opt}}$ with $|x-y| > 1-\delta\}$,

$$\gamma_{\mathrm{opt}}(Z_\delta) \leqslant \frac{E(\rho_0) - E(\rho_1)}{(1-\delta)|\nabla c(1-\delta)|}. \tag{31}$$

(iv) *We define Lebesgue almost everywhere a one-to-one map by*

$$S(y) = y - \nabla c^*\big(\nabla\psi^\infty(y)\big) = y + \nabla c^*\big(\nabla\big(F'\big(\rho_1(y)\big)\big)\big) \quad \textit{satisfying } \gamma_{\mathrm{opt}} = (S \times id)_\# \rho_1 \tag{32}$$

where, when $\rho \in BV(\Omega)$, $\nabla\rho \in L^1(\Omega)$ denotes the approximate derivative (or Radon–Nikodym derivative of $D\rho$ with respect to the Lebesgue measure also called the absolute continuous part $D^{ac}\rho$).

Remark 3.1. As we said before in the setting of the result, there is no ambiguity in the formulation of Eq. (32) since the recession function associated to ∇c^* is zero and then, we could also write $\nabla c^*(D\rho_1)$.

Before proving this lemma, let us recall the following result.

Lemma 3.5. *If both $\psi_1(z)$ and $\psi_2(z)$ are approximately differentiable at $z = y$, if $\psi_1(y) = \psi_2(y)$ and $\psi_1(z) \geqslant \psi_2(z)$ in a neighborhood of y, then*

$$\nabla\psi_1(y) = \nabla\psi_2(y).$$

Proof. Step 1. Let us first consider the case where $\psi_1 = 0$. Assume $\tilde\psi(z)$ approximately differentiable at $z = y$, $\tilde\psi(y) = 0$ and $\tilde\psi(z) \leqslant 0$ near $z = y$. Let us prove that $\nabla\tilde\psi(y) = 0$.

For a contradiction, suppose $\nabla\tilde\psi(y) = \lambda e_n$ where e_n is a given direction and $\lambda > 0$.

For any $a > 0$, the definition of approximate differentiability [27] asserts that

$$\mathcal{D} = \left\{ z \in \mathbb{R}^n \text{ such that } \frac{|\tilde\psi(z) - \lambda(z_n - y_n)|}{|z - y|} < a \right\}$$

has full Lebesgue density at $z = y$.

In particular, it must intersect the cone

$$\mathcal{C} = \left\{ z \in \mathbb{R}^n \text{ such that } z_n - y_n > \sqrt{|z_1 - y_1|^2 + \cdots + |z_{n-1} - y_{n-1}|^2} \right\}$$

inside each ball $B(y, \frac{1}{k}) = \{z \in \mathbb{R}^n \text{ such that } |z - y| \leqslant \frac{1}{k}\}$.

Fix $a < \frac{\lambda}{\sqrt{2}}$ and choose a sequence $z^k \in \mathcal{C} \cap B(y, \frac{1}{k}) \cap \mathcal{D}$. Since $z^k \in \mathcal{C}$, $2|z_n^k - y_n|^2 > |z^k - y|^2$ and then we find

$$-\lambda\big|z_n^k - y_n\big| < -a\big|z^k - y\big| < \tilde\psi\big(z^k\big) - \lambda\big(z_n^k - y_n\big)$$

which means that $\tilde\psi(z^k) > 0$ which leads to a contradiction.

Step 2. Apply Step 1 to $\tilde\psi(z) = \psi_2(z) - \psi_1(z)$. □

Proof of Lemma 3.4. (i) On the support of γ_{opt}, we have almost everywhere

$$c(x - y) = \phi^\infty(x) + \psi^\infty(y)$$

and then, since

$$\phi^\infty(x) = \inf_{v \in \overline{\Omega}} \big(c(x - v) - \psi^\infty(v)\big),$$

we obtain for all $(x, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$ and for all $v \in \Omega$

$$\psi^\infty(v) \leqslant c(x - v) - c(x - y) + \psi^\infty(y) \tag{33}$$

which means that the support of γ_{opt} is included in the c-superdifferential of the potential function ψ^∞.

(ii) The proof of this point is based on the approximate differentiability of ψ^∞. Indeed, since $\psi^\infty \in BV(\Omega)$, by Theorem 4 of Section 6.1.3 in [27], ψ^∞ is approximately differentiable almost everywhere. Let us denote $\nabla\psi^\infty$ its approximate derivative which belongs to $L^1(\Omega)$. (As we said before, the absolute continuous part of $D\psi^\infty$ coincides with the approximate derivative of ψ^∞ so in this section, we will keep this notation to insist on the fact that we talk about $L^1(\Omega)$ functions.) We will prove that this avoids the two possible problems,

Case 1. $\exists(x^1, x^2)$ such that $(x^1, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$, $(x^2, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$, $|x^1 - y| < 1$ and $|x^2 - y| < 1$.

Case 2. $\exists(x^1, x^2)$ such that $(x^1, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$, $(x^2, y) \in \mathrm{spt}(\gamma_{\mathrm{opt}})$, $|x^1 - y| < 1$ and $|x^2 - y| = 1$.

To prove a contradiction in the first case, apply Lemma 3.5 to the functions $\psi^1(v) = \psi^\infty(y) - c(x^1 - y) + c(x^1 - v)$ and $\psi^2(v) = \psi^\infty(v)$ from (33) which leads to $\nabla\psi^\infty(y) = -\nabla c(x^1 - y)$, and then apply Lemma 3.5 a second time to $\psi^1(v) = \psi^\infty(y) - c(x^2 - y) + c(x^2 - v)$ and $\psi^2(v) = \psi^\infty(v)$ which leads to $\nabla\psi^\infty(y) = -\nabla c(x^2 - y)$. Which leads to a contradiction since $\nabla c(x^2 - y) = \nabla c(x^1 - y)$ implies that $x^1 = x^2$.

In the second case, we assume that $\exists x^1$ and x^2 such that $|x^1 - y| < 1$ and $|x^2 - y| = 1$ with

$$\psi^\infty(z) - \psi^\infty(y) \leqslant c(x^i - z) - c(x^i - y) \quad \text{for } i = 1, 2.$$

Note that the previous point says that if ψ^∞ is approximately differentiable at y, $\nabla\psi^\infty(y) = -\nabla c(x^1 - y)$. The problem is that $\nabla c(x^2 - y)$ cannot be defined since $|x^2 - y| = 1$.

To get a contradiction, denote $B(y, r) = \{|z - y| \leqslant r\}$, we will prove that exists a set $\mathcal{S}$ such that

$$\frac{|\mathcal{S} \cap B(y, r)|}{|B(y, r)|} = C \quad \text{where } C \text{ is a constant} \tag{34}$$

satisfying for any $z \in \mathcal{S}$, for any $a > 0$,

$$\psi^\infty(z) - \psi^\infty(y) - \nabla\psi^\infty(y) \cdot (z - y) < -a|z - y|$$

which contradicts the fact that ψ^∞ is approximately differentiable at the point y.

Without loss of generality, we assume that $x^2 = y + e_n$ (or $x_i^2 = y_i + \delta_{i,n}$).

Take $\mathcal{S} = \{z \text{ such that } z_n - y_n > \sqrt{|z_1 - y_1|^2 + \cdots + |z_{n-1} - y_{n-1}|^2}\}$. $\mathcal{S} \cap B(y, r)$ satisfies (34).

For any $z \in \mathcal{S}$, we claim that:

(a) Since $z \in \mathcal{S}$, then $|z - y| < \sqrt{2}|z_n - y_n|$.
(b) $|z - x^2| < 1$.

Indeed, we have

$$\begin{aligned} |z - x^2|^2 &= |z_1 - y_1|^2 + \cdots + |z_{n-1} - y_{n-1}|^2 + |z_n - (1 + y_n)|^2 \\ &= |z - y|^2 + 1 - 2(z_n - y_n) \\ &\leqslant 1 + 2|z_n - y_n|^2 - 2(z_n - y_n). \end{aligned}$$

Then, if $0 < z_n - y_n < 1$ — which can be ensured by taking $r < 1$ — we get $|z - x^2| < 1$.

(c) The inequality

$$-a|z - y| - \nabla\psi^\infty(y) \cdot (z - y) \leqslant \psi^\infty(z) - \psi^\infty(y) \leqslant c(x^i - z) - c(x^i - y) \tag{35}$$

is not possible.

Indeed, we can write

$$\begin{aligned} -a|z - y| - \nabla\psi^\infty(y) \cdot (z - y) &= -a|z - y| - \lambda \cdot (z - y) \\ &\geqslant \sqrt{2}\Big[-a - n \max_i(|\lambda_i|)\Big]|z_n - y_n| = C|z_n - y_n| \end{aligned}$$

and

$$\begin{aligned} c(x^2 - z) - c(x^2 - y) &= c(x^2 - z) - \tilde{c}(1) \\ &\leqslant \tilde{c}\big(\sqrt{1 - 2(z_n - y_n) + |z - y|^2}\big) - \tilde{c}(1) \\ &\leqslant \nabla\tilde{c}\big(\sqrt{1 - 2(z_n - y_n) + |z - y|^2}\big)\big(\sqrt{1 + |z - y|^2 - 2(z_n - y_n)} - 1\big) \end{aligned}$$

since $\tilde{c}(1) - \tilde{c}(1 - \theta) \geqslant \nabla\tilde{c}(1 - \theta)\theta$. And then since $|z - y| < \sqrt{2}|z_n - y_n|$, we obtain

$$c(x^2 - z) - c(x^2 - y) \leqslant \nabla\tilde{c}\big(\sqrt{1 - 2(z_n - y_n) + 2|z_n - y_n|^2}\big)\big(|z_n - y_n|^2 - (z_n - y_n) + o(|z_n - y_n|)\big).$$

So if (35) for $i = 2$ holds, we would have for $\delta = (z_n - y_n) > 0$

$$C \frac{1}{\nabla\tilde{c}(\sqrt{1-2\delta+2\delta^2})} \leqslant -1 + \delta + o(1)$$

which is not possible for any C since passing to the limit $\delta \to 0$, it would give $0 \leqslant -1$.

(iii) We will first prove (31) from where it is easy to deduce the first claim of (iii).

Define $\gamma_{1-\eta}(x, y) = (x, \eta x + (1-\eta)y)_{\#}\gamma_{\rm opt}$ and $\rho_{1-\eta}$ its second marginal (the use of this interpolation map has been introduced in McCann's paper [35]).

We will use the fact that

$$I(\rho_0, \rho_1) \leqslant I(\rho_0, \rho_{1-\eta})$$

for any η to obtain an estimate on $\gamma_{\rm opt}(Z_\delta)$.

On the first hand, we use the convexity of the entropy. Indeed, the entropy satisfies

$$E(\rho_{1-\eta}) \leqslant \eta E(\rho_0) + (1-\eta)E(\rho_1).$$

The convexity of $t \to E(\rho_{1-t})$ is classical (see Agueh [1] and McCann [35]) for $t \in [0, 1[$ and $t \in]0, 1]$ then for $t \in [0, 1]$.

On the other hand, we compute the difference of both terms involving the cost.

Let us compute

$$W_c(\rho_0, \rho_1) - W_c(\rho_0, \rho_{1-\eta}) \geqslant \int_{\Omega\times\Omega} c(x-y)\, d\gamma_{\rm opt}(x, y) - \int_{\Omega\times\Omega} c(x-y)\, d\gamma_{1-\eta}(x, y)$$
$$\geqslant \int_{Z_\delta} \big(c(x-y) - c\big((1-\eta)(x-y)\big)\big)\, d\gamma_{\rm opt}(x, y)$$

since $c((1-\eta)(x-y)) \leqslant c(x-y)$ yields

$$\int_{\Omega\times\Omega/Z^\delta} \big(c(x-y) - c\big((1-\eta)(x-y)\big)\big)\, d\gamma_{\rm opt}(x, y) \geqslant 0.$$

And then, using the fundamental theorem of calculus, we obtain

$$W_c(\rho_0, \rho_1) - W_c(\rho_0, \rho_{1-\eta}) \geqslant \int_{Z_\delta} \eta|x-y|\big|\nabla c\big((1-\theta)(x-y)\big)\big|\, d\gamma_{\rm opt}(x, y)$$

where $\theta \in [0, \eta]$. Since $|\nabla c(z)| = \nabla\tilde{c}(|z|)$ is increasing with respect to $|z|$ and

$$|1-\theta||x-y| \geqslant |1-\eta| \inf_{(x,y)\in Z^\delta} |x-y|,$$

it leads to

$$W_c(\rho_0, \rho_1) - W_c(\rho_0, \rho_{1-\eta}) \geqslant \eta\gamma_{\rm opt}(Z_\delta) \inf_{Z_\delta}\big(|x-y|\big)\Big|\nabla\tilde{c}\Big((1-\eta)\inf_{Z^\delta}|x-y|\Big)\Big|$$
$$\geqslant \eta\gamma_{\rm opt}(Z_\delta)(1-\delta)\nabla\tilde{c}\big((1-\eta)(1-\delta)\big).$$

Finally, since

$$W_c(\rho_0, \rho_1) + E(\rho_1) \leqslant W_c(\rho_0, \rho_{1-\eta}) + E(\rho_{1-\eta})$$

we have

$$W_c(\rho_0, \rho_1) - W_c(\rho_0, \rho_{1-\eta}) \leqslant E(\rho_{1-\eta}) - E(\rho_1) \leqslant \eta\big(E(\rho_0) - E(\rho_1)\big)$$

and then, dividing by η and letting $\eta \to 0$, we obtain

$$\gamma_{\rm opt}(Z_\delta) \leqslant \frac{E(\rho_0) - E(\rho_1)}{(1-\delta)|\nabla\tilde{c}(1-\delta)|}$$

which goes to zero when $\delta \to 0$.

(iv) Points (ii) and (iii) imply that almost everywhere in the support of $\gamma_{\rm opt}$, $|x-y|<1$ and $-\nabla c(x-y)=\nabla\psi^\infty(y)$ which means that

$$x = S(y) = y - \nabla c^*\big(\nabla\psi^\infty(y)\big).$$

It means that since $\operatorname{spt}\gamma \subset \Delta_1$, it is in fact supported almost everywhere on the graph of a one-to-one function which is the optimal map. □

Remark 3.2 *(The optimal map in dimension 1).* In dimension 1, the whole argument is much simpler since the map is already known and does not depend on the cost function and then does not depend on h or ε.

3.3. The optimal map for an arbitrary h

From now on, we deal with the complete minimization problem when $h>0$ and for any problem $\mathcal{P}_i^h$. We use the notation $c_h(x-y)=c(\frac{x-y}{h})$.

Proposition 3.4.

(i) *Let γ_i^h be the optimal measure for the minimization problem $(\mathcal{P}_i^h)$,*

$$\operatorname{spt}\gamma_i^h \in \left\{(x,y)\ \Big|\ \frac{|x-y|}{h}<1\right\}\cup Z_i^h \quad \textit{with } \gamma_{\rm opt}^h\big(Z_i^h\big)=0.$$

(ii) *More precisely, we obtain an estimate of the measure of the set $Z_{i\delta}^h=\{(x,y)\in\operatorname{spt}\gamma_i^h \mid \frac{|x-y|}{h}>1-\delta\}$,*

$$\gamma_i^h\big(Z_{i\delta}^h\big)\leqslant \frac{E(\rho_{i-1})-E(\rho_i^h)}{h(1-\delta)|\nabla c(1-\delta)|}.$$

(iii) *We define Lebesgue almost everywhere a one-to-one map $S_i^h(y)\in L^\infty(\Omega,\Omega)$ by*

$$S_i^h(y)=y-h\nabla c^*\big(\nabla\psi_i^h(y)\big)=y+h\nabla c^*\big(\nabla\big(F'\big(\rho_i^h(y)\big)\big)\big) \quad \textit{satisfying } \gamma_i^h=\big(S_i^h\times id\big)_{\#}\rho_i^h \tag{36}$$

where ρ_i^h and ψ_i^h belong to $BV(\Omega)$.

Proof. To prove (i) to (iii), apply the previous result replacing $x-y$ by $\frac{x-y}{h}$ and $E(\rho_0)-E(\rho_1^h)$ by $\frac{E(\rho_{i-1})-E(\rho_i^h)}{h}$.
Note also that since Ω is a bounded domain, $S_i^h(y)\in L^\infty(\Omega,\Omega)$. □

4. From the discrete equation to the continuous equation

In this section, we have to pass to the limit when the time step goes to zero.
For this purpose, we use a monotonicity argument quoted in Evans [25], Lions [32] or in Otto [38] and since we deal with BV functions, we use also very delicate concepts defined in Andreu, Caselles and Mazón [8,12].

4.1. Construction, compactness and convergence of the measure ρ^h

In the sequel, we will assume without loss of generality that h is chosen such that $\frac{T}{h}$ is an integer.

Notation. In the sequel, the gradient of ρ^h will involve both its absolute continuous part and its singular part and then, we will now use the notation $D^{ac}\rho^h$ instead of $\nabla\rho^h$ for the absolute continuous part of $D\rho^h$. We shall also need the following space of distributions $u(t,y)$ on $[0,T]\times\Omega$:

$$L^1\big([0,T];\big(W^{2,\infty}(\Omega)\big)'\big)=\left\{u\ \Big|\ \int_0^T\int_\Omega u\xi\,dt\,dy\leqslant C_u\|\xi\|_{W^{2,\infty}(\Omega)} \text{ for any test function } \xi\in C_c^\infty(\Omega)\right\}.$$

Proposition 4.1 *(Compactness and convergence of ρ^h). Let ρ^h be the piecewise constant in time function defined by* (3). *Then*:

- (i) *The sequence ρ^h is bounded in $L^1_w([0,T]; BV(\Omega))$ and its time derivative $\partial_t \rho^h$ is a bounded sequence in $L^1([0,T]; (W^{2,\infty}(\Omega))')$.*
- (ii) *A subsequence of ρ^h converges strongly in $L^1([0,T]\times\Omega)$ and weakly* in $L^1_w([0,T], BV(\Omega))$ to a limit $\overline{\rho} \in L^\infty([0,T]\times\Omega) \cap L^1_w([0,T], BV_2(\Omega))$ as $h \to 0$.*

Proof. (i) Recall that for a fixed h, $\rho_i^h(y)$ is the $L^1(\Omega)$ strong and $w^*BV(\Omega)$ limit of $\rho_i^{\varepsilon h}$ solution of the mollified problem $\mathcal{P}_i^\varepsilon$. Then we have

$$\|D(\rho^h)\|([0,T]\times\Omega) \leqslant \liminf_{\varepsilon\to 0} \|\nabla \rho^{\varepsilon h}\|_{L^1([0,T]\times\Omega)} = \liminf_{\varepsilon\to 0} \sum_i h\|\nabla \rho_i^{\varepsilon h}\|_{L^1(\Omega)} \leqslant C$$

where the constant C depends neither on h nor on ε (cf. (29)) and then ρ^h is a bounded sequence in $L^1_w([0,T], BV(\Omega))$.

Concerning the time derivative, we have

$$\begin{aligned}
\int_0^T \left|\int_\Omega \partial_t \rho^h \xi(x)\,dx\right| dt| &= \sum_{i=1}^{\frac{T}{h}} \left| h \int_\Omega \frac{(\rho_i^h - \rho_{i-1}^h)}{h} \xi(x)\,dx \right| \\
&= \sum_{i=1}^{\frac{T}{h}} \left| h \int_\Omega \frac{(\xi(y)-\xi(x))}{h}\, d\gamma_i^h(x,y)\right| dt \\
&\leqslant \sum_{i=1}^{\frac{T}{h}} \left| \int_\Omega (x-y)\cdot\nabla\xi(y)\,d\gamma_i^h(x,y)\right| dt + Ch\|D^2\xi\|_{L^\infty(\Omega)} \\
&\leqslant \sum_{i=1}^{\frac{T}{h}} h \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho_i^h(y)\big)\big)\big)\rho_i^h(y)\cdot\nabla\xi(y)\,dy\,dt + Ch\|D^2\xi\|_{L^\infty(\Omega)} \\
&\leqslant C\|\nabla\xi\|_{L^\infty(\Omega)} + Ch\|D^2\xi\|_{L^\infty(\Omega)}.
\end{aligned}$$

(ii) The sequence ρ^h is bounded in $L^1_w([0,T], BV(\Omega))$, and $BV(\Omega)$ is compactly imbedded in $L^1(\Omega)$. On the other hand $\partial_t\rho^h$ is bounded in $L^1([0,T], (W^{2,\infty}(\Omega))')$ and $L^1(\Omega) \subset (W^{2,\infty}(\Omega))'$. The Aubin Lemma (see [32,41,42]) then implies that ρ^h is relatively compact in $L^1([0,T]\times\Omega)$. □

4.2. Properties of the sequence $\rho^{\varepsilon h}$

We first recall a result proved by Agueh in [1] presented here for $C^2(\mathbb{R}^d)$ cost functions but proved in fact for a larger set of cost functions.

Proposition 4.2 *(Displacement convexity of the L^∞ norm). (See [1].) Let $\rho_0, \rho_1 \in P(\Omega)$ be such that $\rho_0, \rho_1 \leqslant M$ a.e., and assume that $0 \leqslant c \in C^2(\mathbb{R}^d)$ strictly convex satisfies $c(0)=0$ with Legendre transform $c^* \in C^2(\mathbb{R}^d)$. Denote by S the c-optimal map that pushes ρ_1 forward ρ_0, and define the interpolant map*

$$S_t = (1-t)id + tS, \quad \textit{for } t\in[0,T].$$

Then $\|(S_t)_\#\rho_1\|_{L^\infty(\Omega)} \leqslant M$, meaning for non-negative functions $\xi \in C_c(\mathbb{R}^d)$ we have

$$\int_\Omega \xi\big(S_t(y)\big)\rho_1(y)\,dy \leqslant M\int_\Omega \xi(x)\,dx. \tag{37}$$

The proof of this proposition for regular cost function consists in introducing $\rho_{1-t} = (S_t)_\#\rho_1$ and prove that $\rho_{1-t} \leqslant M$.

The next proposition establishes local and global inequalities (38) and (40) relating the generalized Fisher information integrated along the curve $t \in [0, T] \to \rho^{\epsilon h}$ to the net change in entropy. We hereafter refer to such bounds as *entropy-information inequalities*.

Proposition 4.3 *(Localized entropy-information inequality).*

(i) *At each instant in time, $\rho^{\varepsilon h}$ belongs to $W^{1,\infty}(\Omega)$ and lies in a ball in $L^\infty(\Omega) \cap W^{1,1}(\Omega)$ whose radius is independent of $t \in [0, T]$ and $\varepsilon > 0$. Moreover the entropy-information inequality*

$$\int_0^T \int_\Omega \rho^{\varepsilon h}(t, y)\nabla F'\big(\rho^{\varepsilon h}(t, y)\big) \cdot \nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t, y)\big)\big)\big)\, dy$$
$$\leqslant \int_\Omega \big(F\big(\rho^h(0, y)\big) - F\big(\rho^{\varepsilon h}(T, y)\big)\big)\, dy + \sum_1^{\frac{T}{h}} o_\varepsilon(1) \tag{38}$$

is satisfied where $o_\varepsilon(1) \to 0$ when $\varepsilon \to 0$. This integration in time yields

$$\big\|\nabla\rho^{\varepsilon h}\big\|_{L^1([0,T]\times\Omega)} + \sqrt{\varepsilon}\big\|\nabla\rho^{\varepsilon h}\big\|_{L^2([0,T]\times\Omega)} \leqslant C + \sum_1^{\frac{T}{h}} o_\varepsilon(1) \tag{39}$$

where the constant C is now independent of $h > 0$ as well as of $\varepsilon > 0$.

(ii) *Space compactness*: $\rho^{\varepsilon h}$ *satisfies*

$$\sum_1^{\frac{T}{h}} h \int_\Omega \big|\rho_{i-1}^{\varepsilon h}(y) - \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big)\big|\, dy = O(h) + \sum_1^{\frac{T}{h}} o_\varepsilon(1).$$

(iii) *Time compactness*: $\rho^{\varepsilon h}$ *satisfies*

$$\sum_1^{\frac{T}{h}} h \int_\Omega \big|\rho_i^{\varepsilon h}(y) - \rho_{i-1}^h(y)\big|\, dy\, dt = \delta(\varepsilon, h)$$

with $0 = \lim_{h\to 0}\lim_{\varepsilon\to 0}\delta(\varepsilon, h)$.

(iv) *The sequence $\rho^{\varepsilon h}(t, y)$ satisfies a localized entropy-information inequality for any test function $\xi(t, y) \geqslant 0$*

$$\int_0^T \int_\Omega \xi(t, y)\rho^{\varepsilon h}(t, y)\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t, y)\big)\big)\big) \cdot \nabla\big(F'\big(\rho^{\varepsilon h}(t, y)\big)\big)\, dy\, dt$$
$$\leqslant -\int_0^T \int_\Omega \nabla\xi(t, y)\rho^{\varepsilon h}(t, y) \cdot \nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t, y)\big)\big)\big)F'\big(\rho^{\varepsilon h}(t, y)\big)\, dt\, dy$$
$$- \int_0^T \int_\Omega \big[\xi(t, y)\partial_t\big(F\big(\rho^{\varepsilon h}(t, y)\big)\big)\big]\, dt\, dy + \tilde{\delta}(\varepsilon, h) \tag{40}$$

where $0 = \lim_{h\to 0}\lim_{\varepsilon\to 0}\tilde{\delta}(\varepsilon, h)$.

Proof. (i) Each term $\rho^{\varepsilon h} \leqslant M$ and Eq. (21) implies that $\|h\nabla\rho_i^{\varepsilon h}\|_{L^1(\Omega)}$ is bounded independently of ε and by summing

$$h\int_\Omega \rho_i^{\varepsilon h}(y)\nabla F'\big(\rho_i^{\varepsilon h}(y)\big)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy \leqslant \int_\Omega \big(F\big(\rho_{i-1}^h(y)\big) - F\big(\rho_i^{\varepsilon h}(y)\big)\big)\,dy$$

on each time interval, we obtain (38) since $\lim_{\varepsilon\to 0}\int_\Omega (F(\rho_i^{\varepsilon h}(y)) - F(\rho_i^h(y)))\,dy = 0$.

Then (39) holds. Moreover, for ε fixed, we can use an argument of [1] to prove that $\nabla\rho^{\varepsilon h} \in L^\infty(\Omega)$. Indeed since

$$\nabla\rho_i^{\varepsilon h}F''\big(\rho_i^{\varepsilon h}\big) = \nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big) = \frac{\rho_i^{\varepsilon h}(y)}{\rho_i^{\varepsilon h}(y)}\nabla c^\varepsilon\left(\frac{S_i^{\varepsilon h}(y)-y}{h}\right)$$

we obtain for any test function ξ

$$\begin{aligned}\left|\int_\Omega \nabla\rho_i^{\varepsilon h}(y)\xi(y)\,dy\right| &\leqslant \frac{1}{m\inf_{[m,M]}F''}\left|\int_\Omega \rho_i^{\varepsilon h}(y)\nabla c^\varepsilon\left(\frac{S_i^{\varepsilon h}(y)-y}{h}\right)\xi(y)\,dy\right| \\ &= \frac{1}{m\inf_{[m,M]}F''}\left|\int_{\Omega\times\Omega} \nabla c^\varepsilon\left(\frac{x-y}{h}\right)\xi(y)\,d\gamma_i^{\varepsilon h}(x,y)\right| \\ &\leqslant \frac{1}{m\inf_{[m,M]}F''}\|\xi\|_{L^1(\Omega)}\sup_{x,y\in\Omega\times\Omega}\left|\nabla c^\varepsilon\left(\frac{x-y}{h}\right)\right|\end{aligned}$$

and since $c^\varepsilon \in C^1(\mathbb{R}^d)$ we deduce that $\rho_i^{\varepsilon h} \in W^{1,\infty}(\Omega)$.

(ii) Since $\rho_i^{\varepsilon h} \in W^{1,\infty}(\Omega)$ (it may be approximated by a sequence of $C^1(\Omega)$ function), the fundamental theorem of calculus (FTOC) yields

$$\begin{aligned}I_a &= \sum_1^{\frac{T}{h}} h\int_\Omega \big|\rho_{i-1}^{\varepsilon h}(y) - \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big)\big|\,dy \\ &= \sum_1^{\frac{T}{h}} h\int_\Omega \left|\int_0^1 \nabla\big(\rho_{i-1}^{\varepsilon h}\big)\big(y + s\big(S_i^{\varepsilon h}(y) - y\big)\big)\cdot\big(S_i^{\varepsilon h}(y) - y\big)\,ds\right|\,dy.\end{aligned}$$

Since $S_i^{\varepsilon h}(y) - y = h\nabla c^*(\nabla(F'(\rho_i^{\varepsilon h}(y)))) + h\varepsilon\nabla(F'(\rho_i^{\varepsilon h}(y)))$, it leads to

$$\begin{aligned}I_a &\leqslant \sum_1^{\frac{T}{h}} h^2\int_\Omega\int_0^1 \big|\nabla\big(\rho_{i-1}^{\varepsilon h}\big)\big(y + s\big(S_i^{\varepsilon h}(y) - y\big)\big)\big|\,ds\,dy \\ &\quad + \sum_1^{\frac{T}{h}} h^2\varepsilon\int_\Omega\int_0^1 \big|\nabla\big(\rho_{i-1}^{\varepsilon h}\big)\big(y + s\big(S_i^{\varepsilon h}(y) - y\big)\big)\big|\big|\nabla\big(F'\big(\rho_{i-1}^{\varepsilon h}\big)\big)(y)\big|\,dy\,ds\end{aligned}$$

and then

$$\begin{aligned}I_a &\leqslant \frac{1}{m}\int_0^1 \sum_1^{\frac{T}{h}} h^2\int_\Omega \big|\nabla\rho_{i-1}^{\varepsilon h}\big(y + s\big(S_i^{\varepsilon h}(y) - y\big)\big)\big|\rho_i^{\varepsilon h}(y)\,dy\,ds \\ &\quad + \frac{h\varepsilon}{\sqrt{m}}\sup_{[m,M]}|F''|\int_0^1 \left(\sum_1^{\frac{T}{h}} h\int_\Omega \big|\nabla\big(\rho_{i-1}^{\varepsilon h}\big)\big(y + s\big(S_i^{\varepsilon h}(y) - y\big)\big)\big|^2\rho_i^{\varepsilon h}(y)\,dy\right)^{\frac{1}{2}} \\ &\quad \times\left(\sum_1^{\frac{T}{h}} h\int_\Omega \big|\nabla\big(\rho_{i-1}^{\varepsilon h}\big)(y)\big|^2\,dy\right)^{\frac{1}{2}}ds\end{aligned}$$

which becomes using (37) (by approaching $|\nabla\rho_{i-1}^{\varepsilon h}|$ by $C^\infty(\Omega)$ non-negative function) in both terms

$$I_a \leqslant \frac{h}{m} M \int_0^1 \sum_1^{\frac{T}{h}} h \int_\Omega \left|\nabla\rho_{i-1}^{\varepsilon h}(y)\right| dy\, ds$$
$$+ \frac{h}{\sqrt{m}}\sqrt{M} \sup_{[m,M]} |F''| \int_0^1 \left(\sum_1^{\frac{T}{h}} h \int_\Omega \varepsilon\left|\nabla\left(\rho_{i-1}^{\varepsilon h}\right)(y)\right|^2 dy\right)^{\frac{1}{2}} \left(\sum_1^{\frac{T}{h}} h \int_\Omega \varepsilon\left|\nabla\left(\rho_{i-1}^{\varepsilon h}\right)(y)\right|^2 dy\right)^{\frac{1}{2}} ds$$
$$\leqslant \left(C + \sum_1^{\frac{T}{h}} o_\varepsilon(1)\right) h$$

by the $L^1([0,T]\times\Omega)$ bound on the sequence $\nabla\rho^{\varepsilon h}$ and the $L^2([0,T]\times\Omega)$ bound on $\sqrt{\varepsilon}\nabla\rho^{\varepsilon h}$ written in (38).

(iii) Once again, we adapt arguments involved in [1]. We will in fact prove that

$$\int_h^T \int_\Omega \left(F'\left(\rho^{\varepsilon h}(t,y)\right) - F'\left(\rho^h(t-h,y)\right)\right)\left(\rho^{\varepsilon h}(t,y) - \rho^h(t-h,y)\right) dy\, dt$$
$$= \sum_1^{\frac{T}{h}} h \int_\Omega \left(F'\left(\rho_i^{\varepsilon h}(y)\right) - F'\left(\rho_{i-1}^h(y)\right)\right)\left(\rho_i^{\varepsilon h}(y) - \rho_{i-1}^h(y)\right) dy \leqslant Ch$$

which means when $\varepsilon < h \to 0$ by denoting ρ^+ the $L^1([0,T]\times\Omega)$ limit of $\rho^{\varepsilon h}(t,y)$ and ρ^- the $L^1([0,T]\times\Omega)$ limit of $\rho^h(t-h,y)$ that $\rho^+ = \rho^-$ since at the limit

$$0 \leqslant \left(F'(\rho^+) - F'(\rho^-)\right)(\rho^+ - \rho^-) \leqslant 0.$$

And then $\rho^+ = \rho^-$ or

$$\lim_{\varepsilon<h\to 0} \int_h^T \int_\Omega \left|\rho^{\varepsilon h}(t,y) - \rho^h(t-h,y)\right| dy\, dt = 0.$$

But by denoting $\Phi_i(y) = F'(\rho_i^{\varepsilon h}(y)) - F'(\rho_{i-1}^{\varepsilon h}(y))$ we can rewrite

$$I_b = \sum_1^{\frac{T}{h}} h \int_\Omega \left(F'\left(\rho_i^{\varepsilon h}(y)\right) - F'\left(\rho_{i-1}^{\varepsilon h}(y)\right)\right)\left(\rho_i^{\varepsilon h}(y) - \rho_{i-1}^h(y)\right) dy$$
$$+ \sum_1^{\frac{T}{h}} h \int_\Omega \left(F'\left(\rho_{i-1}^{\varepsilon h}(y)\right) - F'\left(\rho_{i-1}^h(y)\right)\right)\left(\rho_i^{\varepsilon h}(y) - \rho_{i-1}^h(y)\right) dy$$
$$= \sum_1^{\frac{T}{h}} h \int_\Omega \left(\Phi_i(y) - \Phi_i\left(S_i^{\varepsilon h}(y)\right)\right)\rho_i^{\varepsilon h}(y)\, dy + h \sum_1^{\frac{T}{h}} o_\varepsilon(1).$$

And then by the FTOC, we obtain

$$I_b = \sum_1^{\frac{T}{h}} h \int_\Omega \int_0^1 \left(S_i^{\varepsilon h}(y) - y\right) \cdot \nabla\Phi_i\left(y + s\left(S_i^{\varepsilon h}(y) - y\right)\right)\rho_i^{\varepsilon h}(y)\, dy\, ds + h \sum_1^{\frac{T}{h}} o_\varepsilon(1)$$
$$\leqslant \left(C + \sum_1^{\frac{T}{h}} o_\varepsilon(1)\right) h$$

using the same argument as above by applying (37) for both terms, the uniform $L^1([0,T]\times\Omega)$ bound on the sequence $\nabla\rho^{\varepsilon h}$ and the $L^2([0,T]\times\Omega)$ bound of $\sqrt{\varepsilon}\nabla\rho^{\varepsilon h}$.

(iv) We now want to prove the localized inequality (40) for the mollified problem. For a fixed test function $\xi(t,x)\geqslant 0$, take h small enough such that $\operatorname{spt}\xi\subset[h,T]\times\Omega$.

We want an estimate for the following quantity

$$\int_0^T\int_\Omega \xi(t,y)\nabla F'\big(\rho^{\varepsilon h}(t,y)\big)\rho^{\varepsilon h}(t,y)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big)\,dy\,dt$$
$$=\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla F'\big(\rho_i^{\varepsilon h}(y)\big)\rho_i^{\varepsilon h}(y)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy+O(h)+h\sum_1^{\frac{T}{h}} o_\varepsilon(1).$$

Indeed by inequality (21),

$$\left|\int_0^T\int_\Omega \xi(t,y)\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big)\,dy\,dt\right.$$
$$\left.-\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy\right|$$
$$\leqslant\|\partial_t\xi\|_{L^\infty}\sum_1^{\frac{T}{h}} h\int_\Omega\big(F\big(\rho^{\varepsilon h}(0,y)\big)-F\big(\rho^{\varepsilon h}(T,y)\big)\,dy\big)$$
$$\leqslant\left(C+\sum_1^{\frac{T}{h}} o_\varepsilon(1)\right)h.$$

Then by introducing $P_F(\lambda)=\lambda F'(\lambda)-F(\lambda)$, we can write

$$\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla F'\big(\rho_i^{\varepsilon h}(y)\big)\rho_i^{\varepsilon h}(y)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy$$
$$=\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla\big(P_F\big(\rho_i^{\varepsilon h}(y)\big)\big)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy.$$

Moreover since $\xi\geqslant 0$,

$$\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla\big(P_F\big(\rho_i^{\varepsilon h}(y)\big)\big)\cdot\nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy$$
$$\leqslant\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla\big(P_F\big(\rho_i^{\varepsilon h}(y)\big)\big)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy.$$

So we will consider the quantity

$$I_0=\sum_1^{\frac{T}{h}} h\int_\Omega \xi(t_i,y)\nabla\big(P_F\big(\rho_i^{\varepsilon h}(y)\big)\big)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\,dy$$

$$= -\sum_{1}^{\frac{T}{h}} h \int_{\Omega} \nabla\xi(t_i, y) P_F\big(\rho_i^{\varepsilon h}(y)\big) \cdot \nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\, dy$$

$$- \sum_{1}^{\frac{T}{h}} h \int_{\Omega} \xi(t_i, y) P_F\big(\rho_i^{\varepsilon h}(y)\big) \operatorname{div} \nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\, dy$$

$$= I_1 + I_2.$$

But since $S_i^{\varepsilon h}(y) - y = h \nabla c^{\varepsilon *}(\nabla(F'(\rho_i^{\varepsilon h}(y))))$ we obtain

$$I_2 = -\sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi(t_i, y) P_F\big(\rho_i^{\varepsilon h}(y)\big) \operatorname{div}\big(S_i^{\varepsilon h}(y) - y\big)\, dy.$$

We now use the intermediate result proved by Agueh in [1]

$$-P_F\big(\rho_i^{\varepsilon h}(y)\big) \operatorname{div}\big(S_i^{\varepsilon h}(y) - y\big) = -P_F\big(\rho_i^{\varepsilon h}(y)\big) \operatorname{tr}\big(\nabla S_i^{\varepsilon h}(y) - Id\big)$$

$$= \rho_i^{\varepsilon h}(y)\left[-d\left(\frac{1}{(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{(d-1)} P_F\left(\left(\frac{1}{(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{-d}\right)\left(\frac{\operatorname{tr}(\nabla S_i^{\varepsilon h}(y))}{d(\rho_i^{\varepsilon h}(y))^{1/d}} - \frac{1}{(\rho_i^{\varepsilon h}(y))^{1/d}}\right)\right]$$

$$\leqslant \rho_i^{\varepsilon h}(y)\left[F\left(\left(\frac{\operatorname{tr}\nabla S_i^{\varepsilon h}(y)}{d(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{-d}\right)\left(\frac{\operatorname{tr}\nabla S_i^{\varepsilon h}(y)}{d(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{-d} - F\left(\left(\frac{1}{(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{-d}\right)\left(\frac{1}{(\rho_i^{\varepsilon h}(y))^{1/d}}\right)^{-d}\right]$$

$$\leqslant F\left(\frac{\rho_i^{\varepsilon h}(y)}{\det \nabla S_i^{\varepsilon h}(y)}\right) \det \nabla S_i^{\varepsilon h}(y) - F\big(\rho_i^{\varepsilon h}(y)\big)$$

since $\lambda^d F(\lambda^{-d})$ (which derivative is $-d\lambda^{d-1} P_F(\lambda^{-d})$) is convex and non-decreasing and $\det \nabla S_i^{\varepsilon h}(y) \leqslant (\frac{\operatorname{tr}\nabla S_i^{\varepsilon h}(y)}{d})^d$ (see the displacement convexity in [35]).

Since $\xi(t_i, y) \geqslant 0$, we obtain

$$I_2 \leqslant \sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi(t_i, y)\left(F\left(\frac{\rho_i^{\varepsilon h}(y)}{\det \nabla S_i^{\varepsilon h}(y)}\right) \det \nabla S_i^{\varepsilon h}(y) - F\big(\rho_i^{\varepsilon h}(y)\big)\right) dy$$

and then

$$I_2 \leqslant \sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi\big(t_i, S_i^{\varepsilon h}(y)\big) F\left(\frac{\rho_i^{\varepsilon h}(y)}{\det \nabla S_i^{\varepsilon h}(y)}\right) \det \nabla S_i^{\varepsilon h}(y) - \xi(t_i, y) F\big(\rho_i^{\varepsilon h}(y)\big)\, dy$$

$$+ \sum_{1}^{\frac{T}{h}} \int_{\Omega} \big(\xi(t_i, y) - \xi\big(t_i, S_i^{\varepsilon h}(y)\big)\big) F\left(\frac{\rho_i^{\varepsilon h}(y)}{\det \nabla S_i^{\varepsilon h}(y)}\right) \det \nabla S_i^{\varepsilon h}(y)\, dy.$$

The two right-hand terms $I_{21} + I_{22}$ can be treated as follows using the relation

$$\rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big) \det \nabla S_i^{\varepsilon h}(y) = \rho_i^{\varepsilon h}(y)$$

and we obtain

$$I_{21} = \sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi\big(t_i, S_i^{\varepsilon h}(y)\big) F\big(\rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big)\big) \frac{\rho_i^{\varepsilon h}(y)}{\rho_{i-1}^h(S_i^{\varepsilon h}(y))} - \xi(t_i, y) F\big(\rho_i^{\varepsilon h}(y)\big)\, dy$$

$$= \sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi(t_i, y)\big(F\big(\rho_{i-1}^h(y)\big) - F\big(\rho_i^{\varepsilon h}(y)\big)\big)\, dy$$

$$= \sum_{1}^{\frac{T}{h}} \int_{\Omega} \xi(t_i, y)\big(F\big(\rho_{i-1}^{\varepsilon h}(y)\big) - F\big(\rho_i^{\varepsilon h}(y)\big)\big)\,dy + \sum_{1}^{\frac{T}{h}} o_\varepsilon(1)$$
$$= -\int_0^T \int_{\Omega} \partial_t F\big(\rho^{\varepsilon h}(t,y)\big)\xi(t,y)\,dy\,dt + \sum_{1}^{\frac{T}{h}} o_\varepsilon(1)$$

and

$$I_{22} = \sum_{1}^{\frac{T}{h}} \int_{\Omega} \big(\xi(t_i,y) - \xi\big(t_i, S_i^{\varepsilon h}(y)\big)\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \det \nabla S_i^{\varepsilon h}(y)\,dy$$
$$= -\sum_{1}^{\frac{T}{h}} \int_{\Omega} \nabla\xi(t_i,y)\big(S_i^{\varepsilon h}(y) - y\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \det \nabla S_i^{\varepsilon h}(y)\,dy + O(h)$$
$$= -\sum_{1}^{\frac{T}{h}} \int_{\Omega} h\nabla\xi(t_i,y)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \det \nabla S_i^{\varepsilon h}(y)\,dy + O(h).$$

We will use that fact that

$$\det \nabla S_i^{\varepsilon h}(y) = \frac{\rho_i^{\varepsilon h}(y)}{\rho_{i-1}^{h}(S_i^{\varepsilon h}(y))} = 1 + \frac{\rho_i^{\varepsilon h}(y) - \rho_{i-1}^{h}(S_i^{\varepsilon h}(y))}{\rho_{i-1}^{h}(S_i^{\varepsilon h}(y))}$$

which yields

$$I_{22} = -\sum_{1}^{\frac{T}{h}} \int_{\Omega} h\nabla\xi(t_i,y)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big) F\big(\rho_i^{\varepsilon h}(y)\big)\,dy$$
$$+ \sum_{1}^{\frac{T}{h}} \int_{\Omega} h\nabla\xi(t_i,y)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\big(F\big(\rho_i^{\varepsilon h}(y)\big) - F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big)\big)$$
$$- \sum_{1}^{\frac{T}{h}} \int_{\Omega} h\nabla\xi(t_i,y)\cdot\nabla c^{\varepsilon *}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \frac{\rho_i^{\varepsilon h}(y) - \rho_{i-1}^{h}(S_i^{\varepsilon h}(y))}{\rho_{i-1}^{h}(S_i^{\varepsilon h}(y))} + O(h).$$

Finally, we write that

$$\left| -\sum_{1}^{\frac{T}{h}} \int_{\Omega} h\nabla\xi(t_i,y)\cdot\nabla c^{*}\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \frac{\rho_i^{\varepsilon h}(y) - \rho_{i-1}^{h}(S_i^{\varepsilon h}(y))}{\rho_{i-1}^{h}(S_i^{\varepsilon h}(y))} \right|$$
$$\leqslant C\|\nabla\xi\|_{L^\infty([O,T]\times\Omega)}\|F\|_{L^\infty([m,M])} \sum_{1}^{\frac{T}{h}} h \int_{\Omega} \big|\rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y)\big|\,dy$$

and

$$\left| -\sum_{1}^{\frac{T}{h}} \int_{\Omega} \varepsilon h\nabla\xi(t_i,y)\cdot\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big) F\big(\rho_{i-1}^{h}\big(S_i^{\varepsilon h}(y)\big)\big) \frac{\rho_i^{\varepsilon h}(y) - \rho_{i-1}^{h}(S_i^{\varepsilon h}(y))}{\rho_{i-1}^{h}(S_i^{\varepsilon h}(y))} \right|$$
$$\leqslant C\sqrt{\varepsilon}\|\nabla\xi\|_{L^\infty(\Omega_T)}\|F\|_{L^\infty([m,M])}\|F''\|_{L^\infty([m,M])}\big\|\sqrt{\varepsilon}\nabla\rho^{\varepsilon h}\big\|_{L^2(\Omega_T)}$$

$$\times \left(\sum_1^{\frac{T}{h}} h \int_\Omega \left| \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y) \right| dy \right)^{\frac{1}{2}}.$$

Moreover, since

$$\left| F\big(\rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big)\big) - F\big(\rho_i^{\varepsilon h}(y)\big) \right| \leqslant \|F'\|_{L^\infty([m,M])} \left| \rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y) \right|,$$

we have

$$\sum_1^{\frac{T}{h}} \int_\Omega h \nabla\xi(t_i, y) \cdot \nabla c^*\big(\nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big)\big(F\big(\rho_i^{\varepsilon h}(y)\big) - F\big(\rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big)\big)\big)$$

$$\leqslant C \|\nabla\xi\|_{L^\infty([O,T]\times\Omega)} \|F'\|_{L^\infty([m,M])} \sum_1^{\frac{T}{h}} h \int_\Omega \left| \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y) \right| dy$$

and

$$\sum_1^{\frac{T}{h}} \int_\Omega h\varepsilon \nabla\xi(t_i, y) \cdot \nabla\big(F'\big(\rho_i^{\varepsilon h}(y)\big)\big)\big(F\big(\rho_i^{\varepsilon h}(y)\big) - F\big(\rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big)\big)\big)$$

$$\leqslant C\sqrt{\varepsilon} \|\nabla\xi\|_{L^\infty(\Omega_T)} \|F'\|_{L^\infty([m,M])} \|F''\|_{L^\infty([m,M])} \left\| \sqrt{\varepsilon} \nabla\rho^{\varepsilon h} \right\|_{L^2(\Omega_T)}$$

$$\times \left(\sum_1^{\frac{T}{h}} h \int_\Omega \left| \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y) \right| dy \right)^{\frac{1}{2}}$$

where $\Omega_T = [0, T] \times \Omega$.

And then, since

$$\begin{aligned}
\sum_1^{\frac{T}{h}} h \int_\Omega \left| \rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big) - \rho_i^{\varepsilon h}(y) \right| dy &\leqslant \sum_1^{\frac{T}{h}} h \bigg[\int_\Omega \left| \rho_{i-1}^h\big(S_i^{\varepsilon h}(y)\big) - \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) \right| dy \\
&\quad + \int_\Omega \left| \rho_{i-1}^{\varepsilon h}\big(S_i^{\varepsilon h}(y)\big) - \rho_{i-1}^{\varepsilon h}(y) \right| dy \\
&\quad + \int_\Omega \left| \rho_{i-1}^{\varepsilon h}(y) - \rho_{i-1}^h(y) \right| dy + \int_\Omega \left| \rho_{i-1}^h(y) - \rho_i^{\varepsilon h}(y) \right| dy \bigg] \\
&\leqslant \sum_1^{\frac{T}{h}} o_\varepsilon(1) + O(h) + \delta(\varepsilon, h)
\end{aligned}$$

we obtain (40) using (ii) and (iii). □

Proposition 4.4 *(Approximate equation and entropy-information). The piecewise constant in time function ρ^h satisfies*:

(i) *The approximate heat equation*

$$\int_0^T \int_\Omega \partial_t \rho^h(y) \xi(t, y)\, dy\, dt = O(h) + \int_0^T \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho^h(y)\big)\big)\big)\rho^h(y) \cdot \nabla\xi(t, y)\, dy\, dt \tag{41}$$

for any test function ξ, where $D^{ac}(F'(\rho^h(y))) = D^{ac}\rho^h(y)F''(\rho^h(y))$ and $D^{ac}\rho^h$ denotes the Radon–Nikodym derivative of the measure $D\rho^h$ with respect to the Lebesgue measure.

(ii) *The entropy-information inequality*

$$\mathcal{G}\big(F'\big(\rho^h\big); D\big(F'\big(\rho^h\big)\big)\big) \leqslant \int_\Omega \big[F\big(\rho_0(y)\big) - F\big(\rho^h(T,y)\big)\big]\,dy \tag{42}$$

where the left-hand side represents the integral (46) *of a function generalizing* $g(z,\xi) = (F^*)' \circ (F(z))\xi\nabla c^*(\xi)$.

Remark 4.1. Note however that the same kind of localized inequality as (40) is also valid for ρ^h.

Proof. (i) We will first prove the following discrete in time equation localized in time

$$\int_\Omega \frac{\rho_i^h(y) - \rho_{i-1}^h(y)}{h}\xi(y)\,dy = \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho_i^h(y)\big)\big)\big)\rho_i^h(y)\nabla\xi(y)\,dy + O(h) \tag{43}$$

holds for any test function $\xi \in C^\infty(\Omega)$. Multiplying (36) by $\rho_i^h\nabla\xi$, we obtain for any test function $\xi \in C_c^\infty(\Omega)$

$$\int_\Omega \left(\frac{S_i^h(y) - y}{h}\right)\nabla\xi(y)\rho_i^h(y)\,dy = \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho_i^h(y)\big)\big)\big)\rho_i^h(y)\nabla\xi(y)\,dy.$$

To obtain properly the discrete in time equation, we have to compute for any test function ξ

$$\int_\Omega \frac{\rho_i^h(y) - \rho_{i-1}(y)}{h}\xi(y)\,dy = \frac{1}{h}\int_{\Omega\times\Omega} \big(\xi(y) - \xi(x)\big)\,d\gamma_i^h(x,y).$$

By using the FTOC applied to the test function ξ, we have

$$\left|\int_\Omega \frac{1}{h}\big[\big(\xi(y) - \xi(x)\big) - (x-y)\nabla\xi(y)\big]\,d\gamma_i^h(x,y)\right| \leqslant \frac{1}{2h}\sup_{z\in\Omega}\big|D^2\xi(z)\big|\int_\Omega |x-y|^2\,d\gamma_i^h(x,y) \leqslant C\frac{h}{2}$$

and since

$$\int_\Omega \left(\frac{S_i^h(y) - y}{h}\right)\nabla\xi(y)\rho_i^h(y) = \int_{\Omega\times\Omega} \frac{x-y}{h}\nabla\xi(y)\,d\gamma_i^h(x,y),$$

we obtain the expected discrete in time equation

$$\left|\int_\Omega \frac{\rho_i^h(y) - \rho_{i-1}^h(y)}{h}\xi(y)\,dy - \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho_i^h(y)\big)\big)\big)\rho_i^h(y)\nabla\xi(y)\,dy\right|$$
$$\leqslant \frac{1}{2h}\sup_{z\in\Omega}\big|D^2\xi(z)\big|\int_\Omega |x-y|^2\,d\gamma_i^h(x,y) \tag{44}$$

and then we have (43). Let us write (43) on each interval $]ih,(i+1)h]$ and sum it with respect to i. Since for any test function $\xi \in C^\infty([0,T]\times\Omega)$

$$\left|\sum_1^{\frac{T}{h}} h\int_\Omega \frac{\rho^h((i+1)h,y) - \rho^h(ih,y)}{h}\xi(t_i,y)\,dy\,dt - \int_0^T\int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho^h(y)\big)\big)\big)\rho^h(y)\nabla\xi(t_i,y)\,dy\,dt\right|$$
$$\leqslant \sum_1^{\frac{T}{h}} h\frac{1}{h}\sup_{[0,T]\times\Omega}\big|D_x^2\xi\big|\int_\Omega |x-y|^2\,d\gamma_i^h(x,y)$$

we obtain (41) because $|x-y| < h$ and $\sum_1^{\frac{T}{h}} h = T$ and

$$\left|\sum_1^{\frac{T}{h}} h \int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho^h(y)\big)\big)\big)\rho^h(y)\nabla\xi(t_i,y)\,dy\,dt - \int_0^T\int_\Omega \nabla c^*\big(D^{ac}\big(F'\big(\rho^h(y)\big)\big)\big)\rho^h(y)\nabla\xi(t,y)\,dy\,dt\right|$$
$$\leqslant C\|\partial_t\nabla\xi\|_{L^\infty}h.$$

(ii) We want to prove that

$$\int_0^T\int_\Omega \rho^h\nabla c^*\big(D\big(F'\big(\rho^h\big)\big)\big)\cdot D\big(F'\big(\rho^h\big)\big) \leqslant \liminf_{\varepsilon\to 0}\int_0^T\int_\Omega \rho^{\varepsilon h}\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big)\cdot\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)$$
$$\leqslant \liminf_{\varepsilon\to 0}\int_\Omega \big[F\big(\rho_0(y)\big) - F\big(\rho^{\varepsilon h}(T,y)\big)\big]\,dy \tag{45}$$

but since ρ^h is in BV, we have to give a sense to the left-hand side.

Remark 4.2. In the above expression, we define $\rho^{\varepsilon h}(t,y)$ the same way we defined $\rho^h(t,y)$ from $\rho_i^h(y)$ through (3).

Using the same notation as in Andreu, Caselles and Mazón [12], let us denote

$$g\big(F'(\rho), D\big(F'(\rho)\big)\big) = (F^*)'\circ F'(\rho)D\big(F'(\rho)\big)\cdot\nabla c^*\big(D\big(F'(\rho)\big)\big).$$

We can then define the generalized version of the Fisher information for BV functions by defining

$$\mathcal{G}(\rho, D\rho) = \int_0^T\int_\Omega g\big(\rho, D^{ac}\rho\big)\,dt\,dx + \int_\Omega g^0\bigg(\rho, \frac{D\rho}{|D\rho|}\bigg)\big|D^c\rho\big| + \int_{J_\rho}\bigg(\int_{\rho_-(x)}^{\rho_+(x)} g^0\big(s,\nu_\rho(x)\big)\,ds\bigg)\,d\mathcal{H}^{d-1}(x) \tag{46}$$

where g^0 is the recession function equal to $\lim_{t\to 0} t g(x,z,\frac{\xi}{t}) = |\xi|$ cf. (11).

To get the entropy-information inequality, we want to prove a lower semi-continuity for $\mathcal{G}$:

$$\mathcal{G}\big(F'\big(\rho^h\big), D\big(F'\big(\rho^h\big)\big)\big) \leqslant \liminf_{\varepsilon\to 0}\int_0^T\int_\Omega g\big(F'\big(\rho^{\varepsilon h}\big), \nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big). \tag{47}$$

For this purpose, let us decompose this quantity in two parts, which will be semi-continuity. Indeed, we write when $\rho\in W^{1,1}(\Omega)$

$$g\big(F'(\rho), \nabla\big(F'(\rho)\big)\big) = (F^*)'\circ F'(\rho)\nabla\big(F'(\rho)\big)\nabla c^*\big(\nabla F'(\rho)\big)$$
$$= (F^*)'\circ F'(\rho)\big(c\big(\nabla c^*\big(\nabla\big(F'(\rho)\big)\big)\big) + c^*\big(\nabla\big(F'(\rho)\big)\big)\big).$$

Using the result of De Cicco, Fusco and Verde on lower semi-continuity for BV functions [23], since c^* is convex and $(F^*)'$ is continuous, we obtain

$$\int_0^T\int_\Omega (F^*)'\circ\big(F'\big(\rho^h\big)\big)c^*\big(D\big(F'\big(\rho^h\big)\big)\big) \leqslant \liminf_{\varepsilon\to 0}\int_0^T\int_\Omega (F^*)'\circ\big(F'\big(\rho^{\varepsilon,h}\big)\big)c^*\big(\nabla\big(F'\big(\rho^{\varepsilon,h}\big)\big)\big).$$

In another hand, since $\nabla c^*(\nabla(F'(\rho^{\varepsilon h})))$ is bounded in $L^\infty([0,T]\times\Omega)$, it converges in $w^*L^\infty([0,T]\times\Omega)$. It still remains to identify this limit. For any test function $\xi\in\mathcal{D}(\Omega)$, we have, for a fixed t

$$\int_\Omega \xi(y)\nabla c^*\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\rho^{\varepsilon h}(t,y)\,dy$$
$$= \int_\Omega \xi(y)\nabla c^{\varepsilon *}\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\rho^{\varepsilon h}(t,y)\,dy - \varepsilon\int_\Omega \xi(y)\big|\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big|\rho^{\varepsilon h}(t,y)\,dy$$

$$= \int_{\Omega\times\Omega} \xi(y)(x-y)\,d\gamma_i^{h,\varepsilon}(x,y) + O(\sqrt{\varepsilon}) \to \int_{\Omega\times\Omega} \xi(y)(x-y)\,d\gamma_i^{h}(x,y)$$

$$= \int_{\Omega} \xi(y)\nabla c^*\big(D^{ac}\big(F'\big(\rho^h(t,y)\big)\big)\big)\rho^h(t,y)\,dy$$

and then the uniqueness of the limit says that $\nabla c^*(\nabla(F'(\rho^{\varepsilon h}))) \to \nabla c^*(D^{ac}(F'(\rho^h)))$ $w^*L^\infty([0,T]\times\Omega)$ and then

$$\big|\nabla c^*\big(D^{ac}\big(F'\big(\rho^h(t,y)\big)\big)\big)\big| \leqslant \liminf_{\varepsilon\to 0}\big|\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big)\big|.$$

Since c is non-decreasing on $\mathbb{R}^+$, we have

$$c\big(\nabla c^*\big(\big|D^{ac}\big(F'\big(\rho^h\big)\big)\big|\big)\big) \leqslant c\big(\big|\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big)\big|\big)$$

and then

$$\int_0^T\int_\Omega \rho^h c\big(\nabla c^*\big(D^{ac}\big(F'\big(\rho^h\big)\big)\big)\big) \leqslant \liminf_{\varepsilon\to 0}\int_\Omega \rho^h c\big(\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big)\big).$$

Moreover — using that $\rho^{\varepsilon h}$ converges strongly in $L^1([0,T]\times\Omega)$ towards ρ^h — we have

$$\lim_{\varepsilon\to 0}\int_0^T\int_\Omega \big(\rho^h - \rho^{\varepsilon h}\big)c\big(\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}\big)\big)\big)\big) = 0.$$

Hence we recover the expected lower semi-continuity property (47). □

Remark 4.3. In our case $g^0(x,z,\xi) = |\xi|$ and then

$$\int_\Omega g^0\bigg(\rho, \frac{D\rho}{|D\rho|}\bigg)\big|D^c\rho\big| + \int_{J_\rho}\Bigg(\int_{\rho_-(x)}^{\rho_+(x)} g^0\big(s,\nu_\rho(x)\big)\,ds\Bigg)\,d\mathcal{H}^{d-1}(x)$$

$$= \int_\Omega \big|D^c\rho\big| + \int_{J_\rho}\big(\rho_+(x) - \rho_-(x)\big)\,d\mathcal{H}^{d-1}(x) \geqslant 0$$

so the singular terms are positive and we can write the Fisher information for the absolute continuous part of $D(F'(\rho^h))$, i.e.

$$\int_0^T\int_\Omega \rho^h(t,y)\nabla c^*\big(D^{ac}\big(F'\big(\rho^h(t,y)\big)\big)\big)\cdot D^{ac}\big(F'\big(\rho^h(t,y)\big)\big) \leqslant \int_\Omega \big[F\big(\rho_0(y)\big) - F\big(\rho^h(T,y)\big)\big]\,dy.$$

4.3. Limiting equation

The following proposition represents the keystone which enables us to apply the Minty–Browder technique to complete the construction.

Proposition 4.5 *(The continuous time limit).*

(i) *The time derivative $\partial_t\bar\rho$ belongs to $L^1([0,T],BV(\Omega))'$ and is the time derivative of $\bar\rho$ in $L^1([0,T],BV(\Omega))'$, i.e. for any test function $\rho\in L^1([0,T],BV_2(\Omega))$ admitting a weak derivative Θ in $L^1_w([0,T],BV(\Omega))\cup L^\infty([0,T]\times\Omega)$ as defined in* [10]*, meaning*

$$\rho(t,y) = \int_0^t \Theta(s,y)\,ds$$

we have

$$\int_0^T \int_\Omega \partial_t \overline{\rho}(t,y)\rho(t,y)\,dt\,dy = -\int_0^T \int_\Omega \overline{\rho}(t,y)\Theta(t,y)\,dt\,dy. \tag{48}$$

(ii) *Let $\overline{A} \in L^\infty([0,T]\times\Omega)$ be the weak* limit of $A(\rho^h) = \rho^h \nabla c^*(D^{ac}(F'(\rho^h)))$, for any test function $\xi \geqslant 0$, we have then the corresponding localized inequality*

$$-\int_\Omega \partial_t F\big(\overline{\rho}(t,y)\big)\xi(t,y)\,dt\,dy - \int_\Omega \nabla\xi(t,y)F'\big(\overline{\rho}(t,y)\big)\overline{A}(t,y)\,dt\,dy$$
$$\leqslant \int_0^T \int_\Omega \xi(t,y)\big(\overline{A}(t,y); D^{ac}\big(F'\big(\overline{\rho}(t,y)\big)\big)\big)\,dy\,dt + \big(T^s;\xi(t,y)\big) \tag{49}$$

where T^s is a purely singular measure. We shall use (49) *in lieu of integration by parts.*

Remark 4.4. Note that we would originally expect instead of inequality (49) an equality obtained by integration by part from the equation

$$\partial_t \overline{\rho} = \operatorname{div}(\overline{A})$$

multiplied by $\xi F'(\overline{\rho})$. The problem is that $F'(\overline{\rho})$ is not regular enough — since it belongs to $L^1_w([0,T], BV_2(\Omega))$ — to be used as a test function.

Proof. Let us prove (i) which is fundamental to obtain (ii). We want to prove that $\partial_t\overline{\rho}$ belongs to $L^1([0,T], BV(\Omega))'$. Remark that

$$\int_0^T \int_\Omega \partial_t \overline{\rho}(t,y)\xi(t,y)\,dt\,dy = \lim_{h\to 0}\int_0^T \int_\Omega \partial_t \rho^h(t,y)\xi(t,y)\,dt\,dy$$
$$= \lim_{h\to 0}\sum_1^{\frac{T}{h}} \int_{t_{i-1}}^{t_i}\int_\Omega \frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\xi(t,y)\,dt\,dy.$$

Indeed

$$\left|\sum_1^{\frac{T}{h}} \int_{t_{i-1}}^{t_i}\int_\Omega \frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\big(\xi(t,y)-\xi(t_i,y)\big)\,dt\,dy\right|$$
$$= \left|\sum_1^{\frac{T}{h}} \int_{t_{i-1}}^{t_i}\int_\Omega \frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\partial_t\xi(t,y)(t_i-t)\,dt\,dy\right| + o(h)$$
$$= \|\partial_t\xi\|_{L^\infty([0,T]\times\Omega)} \lim_{\varepsilon\to 0}\delta(\varepsilon,h) + o(h).$$

We therefore consider

$$\sum_1^{\frac{T}{h}} \int_{t_{i-1}}^{t_i}\int_\Omega \frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\xi(t,y)\,dt\,dy$$
$$= \lim_{\varepsilon\to 0}\sum_{i=1}^{\frac{T}{h}} \int_{t_{i-1}}^{t_i}\int_\Omega \xi(t,y)\frac{(\rho_i^{\varepsilon h}(y)-\rho_{i-1}^{\varepsilon h}(y))}{h}\,dy$$

$$= \lim_{\varepsilon\to 0}\sum_{i=1}^{\frac{T}{h}}\int_{t_{i-1}}^{t_i}\int_{\Omega}\xi(t,y)\frac{(\rho_i^{\varepsilon h}(y)-\rho_{i-1}^h(y))}{h}\,dy+\sum_{i=1}^{\frac{T}{h}}o_\varepsilon(1)$$

$$= \lim_{\varepsilon\to 0}\sum_{i=1}^{\frac{T}{h}}\int_{t_{i-1}}^{t_i}\int_{\Omega}\frac{(\xi(t,y)-\xi(t,S_i^{\varepsilon h}(y)))}{h}\rho_i^{\varepsilon h}(y)\,dy$$

$$= \lim_{\varepsilon\to 0}\sum_{i=1}^{\frac{T}{h}}\int_{t_{i-1}}^{t_i}\int_{\Omega}\int_0^1\nabla\xi\big(t,y+s\big(S_i^{\varepsilon h}(y)-y\big)\big)\cdot\frac{(S_i^{\varepsilon h}(y)-y)}{h}\rho_i^{\varepsilon h}(y)\,dy\,ds.$$

Then since $S_i^{\varepsilon h}(y)-y=h\nabla c^*(\nabla(F'(\rho_i^{\varepsilon h}(y))))+\varepsilon\nabla(F'(\rho_i^{\varepsilon h}(y)))$ we can write

$$\begin{aligned}\sum_1^{\frac{T}{h}}\int_{t_{i-1}}^{t_i}\int_{\Omega}\frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\xi(t,y)\,dt\,dy &\leqslant \lim_{\varepsilon\to 0}\int_0^1\sum_{i=1}^{\frac{T}{h}}\int_{t_{i-1}}^{t_i}\int_{\Omega}\big|\nabla\xi\big(t,y+s\big(S_i^{\varepsilon h}(y)-y\big)\big)\big|\rho_i^{\varepsilon h}(y)\,dy\,ds\\ &\quad+\varepsilon\|\nabla\xi\|_{L^\infty([0,T]\times\Omega)}\big\|\nabla\rho^{\varepsilon h}\big\|_{L^1([0,T]\times\Omega)}\\ &\leqslant C\|\nabla\xi\|_{L^1([0,T]\times\Omega)}\end{aligned}$$

where C does not depend on h. Moreover this integral may also be defined for any $\xi\in BV_2(\Omega)$ since for h fixed, the piecewise constant function $\partial_h\rho^h=\frac{\rho_i^h-\rho_{i-1}^h}{h}$ on $[(i-1)h,ih]$ belongs to $L^\infty([0,T]\times\Omega)$. Then,

$$\sup_{t\in[0,T]}\int_{\Omega}\partial_h\rho^h\xi\,dy\leqslant C\quad\forall\xi\in BV_2(\Omega),$$

and the sequence $\partial_h\rho^h$ is bounded in $L^1([0,T],BV_2(\Omega))'$. This implies that $\partial_t\bar\rho$ is the w^* limit of $\partial_h\rho^h$ and then belong to the dual of $L^1([0,T],BV_2(\Omega))$.

In fact, let us define the distribution z^h by

$$\big\langle z^h;\zeta\big\rangle=\int_0^T\int_{\Omega}\int_0^1\zeta\big(t,y+sh\nabla c^*\big(\nabla\big(F'\big(\rho^h(t,y)\big)\big)\big)\big)\nabla c^*\big(\nabla F'\big(\rho^h(t,y)\big)\big)\rho^h(t,y)\,ds\,dy\,dt.$$

The previous computation proves that z^h is a bounded sequence in $L^\infty([0,T]\times\Omega)$ and that, for h fixed, $\operatorname{div}z^h=\partial_h\rho^h$ in $L^1([0,T]\times\Omega)$.

In another hand, we have equality (48). Indeed,

$$\begin{aligned}\int_0^T\int_{\Omega}\partial_t\bar\rho(t,y)\rho(t,y)\,dt\,dy &= \lim_{\delta\to 0}\lim_{h\to 0}\int_\delta^{T-\delta}\int_{\Omega}\partial_h\rho^h(t,y)\rho(t,y)\,dt\,dy\\ &= \lim_{\delta\to 0}\lim_{h\to 0}\Bigg(\sum_2^{\frac{T}{h}-1}\int_{t_{i-1}}^{t_i}\int_{\Omega}\frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\rho(t,y)\,dt\,dy\\ &\quad+\int_\delta^{t_1}\int_{\Omega}\frac{(\rho_1^h(y)-\rho_0^h(y))}{h}\rho(t,y)\,dt\,dy\\ &\quad+\int_{\frac{T}{h}-1}^{t_{\frac{T}{h}}-\delta}\int_{\Omega}\frac{(\rho_i^h(y)-\rho_{i-1}^h(y))}{h}\rho(t,y)\,dt\,dy\Bigg)\end{aligned}$$

$$= -\lim_{\delta\to 0}\lim_{h\to 0}\int_{\delta}^{T-\delta}\int_{\Omega}\frac{(\rho(t-h,y)-\rho(t,y))}{h}\rho^h(t,y)\,dt\,dy$$

$$= -\lim_{\delta\to 0}\lim_{h\to 0}\int_{\delta}^{T-\delta}\int_{\Omega}\frac{1}{h}\int_{t-h}^{t}\Theta(s,y)\,ds\,\rho^h(t,y)\,dt\,dy$$

$$= -\int_0^T\int_{\Omega}\Theta(t,y)\overline{\rho}(t,y)\,dt\,dy.$$

Indeed, ρ^h converges strongly in $L^1([0,T]\times\Omega)$ and $\frac{1}{h}\int_{t-h}^{t}\Theta(s,y)\,ds$ is bounded in $L^\infty([0,T]\times\Omega)$ — since Θ is bounded in $L^\infty([0,T]\times\Omega)$ — and converges toward Θ.

(ii) First of all, z^h is a bounded sequence in $L^\infty([0,T]\times\Omega)$ and then, up to a subsequence, converges in the sense of distributions toward a distribution z which is equal to $\overline{A}$ in $L^\infty([0,T]\times\Omega)$.

Passing to the limit in the discrete equation (41), we obtain the limiting equation in the sense of distributions.

This point is an adaptation of the argument of [10].

We define as in the introduction for any $\rho\in L^1([0,T],BV_2(\Omega))$ and any test function ζ

$$\int_0^T\big\langle(z,D\rho);\zeta\big\rangle\,dt = -\int_0^T\big\langle\partial_t\overline{\rho}(t,y);\rho(t,y)\zeta(y)\big\rangle_{L^1([0,T],BV_2(\Omega))',L^1([0,T],BV_2(\Omega))}\,dt$$
$$-\int_0^T\int_{\Omega}\operatorname{div} z\rho(t,y)\zeta(y)\,dy\,dt.$$

As we noticed, $F'(\overline{\rho})$ does not belong to $L^1([0,T],BV_2(\Omega))$, so we need to introduce a regularization. In this step, we follow the proposition of [10].

Let the test function $\xi(t,y)=\eta(t)\zeta(y)\geqslant 0$ and introduce $F^\tau(t,y)=\frac{1}{\tau}\int_{t-\tau}^{t}\eta(s)F'(\overline{\rho}(s,y))\,ds$. For τ fixed, small enough such that $\operatorname{spt}\eta\subset[\tau,T-\tau]$, since $F'(\overline{\rho})$ is integrated in time, $F^\tau(t,\cdot)$ belongs to $C^0([0,T],BV_2(\Omega))$ (see [7]) and then belong to $L^1([0,T],BV_2(\Omega))$.

Let us now precise the argument of [10] by the following computation. First of all, Eq. (48) implies

$$\int_0^T\int_{\Omega}F\big(\overline{\rho}(t,y)\big)\partial_t\xi(t,y)\,dt\,dy = \lim_{\tau\to 0}\int_0^T\int_{\Omega}\frac{\eta(t-\tau)-\eta(t)}{-\tau}F\big(\overline{\rho}(t,y)\big)\zeta(y)\,dy\,dt$$
$$= \lim_{\tau\to 0}\int_0^T\int_{\Omega}\frac{F(\overline{\rho}(t+\tau))-F(\overline{\rho}(t))}{-\tau}\eta(t)\zeta(y)\,dy\,dt$$

and then since F is a convex function

$$\int_0^T\int_{\Omega}\frac{F(\overline{\rho}(t+\tau))-F(\overline{\rho}(t))}{-\tau}\eta(t)\zeta(y)\,dy\,dt$$
$$\leqslant -\int_0^T\int_{\Omega}F'\big(\overline{\rho}(t,y)\big)\frac{\overline{\rho}(t+\tau,y)-\overline{\rho}(t,y)}{\tau}\eta(t)\zeta(y)\,dt\,dy$$
$$= \int_0^T\int_{\Omega}\overline{\rho}(t,y)\zeta(y)\frac{F'(\overline{\rho}(t-\tau,y))\eta(t-\tau)-F'(\overline{\rho}(t,y))\eta(t)}{-\tau}\,dy\,dt.$$

Since $\frac{F'(\bar\rho(t-\tau,y))\eta(t-\tau)-F'(\bar\rho(t,y))\eta(t)}{-\tau} \in L^1_w([0,T],BV_2(\Omega)) \cap L^\infty([0,T]\times\Omega)$ is the time derivative of F^τ as defined in [10], i.e.

$$F^\tau(t,y) = \int_0^t \frac{F'(\bar\rho(s-\tau,y))\eta(s-\tau) - F'(\bar\rho(s,y))\eta(s)}{-\tau}\,ds$$

and since $F^\tau \in L^1([0,T],BV_2(\Omega))$ we obtain

$$\int_0^T\int_\Omega \bar\rho(t,y)\zeta(y)\partial_t F^\tau(t,y)\,dy\,dt = -\int_0^T \langle \partial_t\bar\rho, \zeta(y)F^\tau(t,y)\rangle\,dt$$
$$= \int_0^T\int_\Omega F^\tau(t,y)z\cdot\nabla\zeta(y)\,dy\,dt + \int_0^T\int_\Omega \zeta(y)\big(z; D\big(F^\tau(t,y)\big)\big)\,dt\,dy$$

by definition of the distribution (z, DF^τ).

And then

$$\int_0^T\int_\Omega F\big(\bar\rho(t,y)\big)\partial_t\xi(t,y)\,dt\,dy \leqslant \lim_{\tau\to0}\int_0^T\int_\Omega F^\tau(t,y)z\cdot\nabla\zeta(y)\,dy\,dt$$
$$+ \lim_{\tau\to0}\frac{1}{\tau}\int_0^T\int_\Omega\int_{t-\tau}^t \zeta(y)\eta(s)z\cdot D^{ac}\big(F'\big(\bar\rho(s,y)\big)\big)\,dt\,dy\,ds$$
$$+ \lim_{\tau\to0}\frac{1}{\tau}\int_0^T\int_\Omega \zeta(y)\left(z; \int_{t-\tau}^t \eta(s)D^s\big(F'\big(\bar\rho(s,y)\big)\big)\right)dt\,dy\,ds.$$

First by (12) we write

$$\frac{1}{\tau}\int_0^T\int_\Omega \zeta(y)\left(z; \int_{t-\tau}^t \eta(s)D^s\big(F'\big(\bar\rho(s,y)\big)\big)\right)dt\,dy\,ds \leqslant M\frac{1}{\tau}\int_{-\tau}^0\int_0^T\int_\Omega \zeta(y)\eta(t+s)\big|D^s\big(F'\big(\bar\rho(t+s,y)\big)\big)\big|\,dt\,dy\,ds$$

and then we use that for any $0 \leqslant p \in L^1([0,T])$, $\lim_{\tau\to0}\frac{1}{\tau}\int_{-\tau}^0 p(t+s)\,ds = p(t)$ in the sense of measure.

This yields the result since $z = \bar A$ which gives

$$\int_0^T\int_\Omega F\big(\bar\rho(t,y)\big)\partial_t\xi(t,y)\,dt\,dy \leqslant \int_0^T\int_\Omega \eta(t)F'\big(\bar\rho(t,y)\big)\bar A\cdot\nabla\zeta(y)\,dy\,dt$$
$$+ \int_0^T\int_\Omega \zeta(y)\eta(t)\bar A\cdot D^{ac}\big(F'\big(\bar\rho(t,y)\big)\big)\,dt\,dy$$
$$+ \|\bar A\|_{L^\infty([0,T]\times\Omega)}\int_0^T\int_\Omega \eta(t)\zeta(y)\big|D^s\big(F'\big(\bar\rho(t,y)\big)\big)\big|\,dt\,dy.$$

We denote $\langle T^s;\xi\rangle$ the singular term $\|\bar A\|_{L^\infty([0,T]\times\Omega)}\int_0^T\int_\Omega \xi(t,y)|D^s(F'(\bar\rho(t,y)))|\,dt\,dy$. $\square$

4.4. Identification of the limiting equation: a monotonicity argument

By a monotonicity argument, we achieve the proof of Theorem 1.8 by proving the proposition

Proposition 4.6. *The limiting measure $\overline{\rho}$ of the discrete in time measure ρ^h satisfying Eq.* (41) *is solution in the sense of distributions to the following equation*

$$\partial_t \overline{\rho} = \operatorname{div}\big(\overline{\rho} \nabla c^*\big(D\big(F'(\overline{\rho})\big)\big)\big). \tag{50}$$

Proof. We want to prove that $\overline{A} = \overline{\rho} \nabla c^*(D(F'(\overline{\rho})))$. For that purpose, we want to use the Minty–Browder's argument which means that we have to prove that for any $z \in \mathbb{R}^d$,

$$\big(\overline{A} - A(\overline{\rho}, z)\big)\big(D^{ac}\big(F'(\overline{\rho})\big) - z\big) \geqslant 0$$

and in fact, we want to prove for any test function $\xi(t,x) \geqslant 0$ and $\zeta(t,x) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$

$$\int \xi\big(\overline{A} - A(\overline{\rho}, \zeta)\big)\big(D^{ac}\big(F'(\overline{\rho})\big) - \zeta\big) \geqslant 0.$$

Indeed, this yields the expected equality since by taking $\zeta = D^{ac}(F'(\overline{\rho})) + \delta w$, we obtain

$$\big(\overline{A} - A\big(\overline{\rho}; D^{ac}\big(F'(\overline{\rho})\big) + \delta w\big)\big)\delta w \geqslant 0$$

and then by passing successively to the limit $0 > \delta \to 0$ and $0 < \delta \to 0$, we obtain the equality $\overline{A} = A(\overline{\rho}, D^{ac}(F'(\overline{\rho})))$.

To prove this inequality, we want to pass to the limit when ε and h go to zero in the following inequality that comes from the convexity of c^*

$$\Lambda^{\varepsilon h} = \int_0^T \int_\Omega \rho^{\varepsilon h}(t,y)\big(\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big) - \nabla c^*\big(\zeta(t,y)\big)\big) \cdot \big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big) - \zeta(t,y)\big)\, dy\, dt \geqslant 0.$$

Let us develop this quantity as follows

$$\begin{aligned}
\Lambda^{\varepsilon h} = {} & \int_0^T \int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big)\, dy\, dt \\
& + \int_0^T \int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\zeta(t,y) \cdot \nabla c^*\big(\zeta(t,y)\big)\, dy\, dt \\
& - \int_0^T \int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\nabla c^*\big(\zeta(t,y)\big) \cdot \nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\, dy\, dt \\
& - \int_0^T \int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big) \cdot \zeta(t,y)\, dy\, dt.
\end{aligned}$$

Term 1. Proposition 4.3 gives us at the limit

$$\begin{aligned}
& \liminf_{\varepsilon h \to 0} \int_0^T \int_\Omega \xi(t,y)\rho^{\varepsilon h}(t,y)\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\nabla c^*\big(\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\big)\, dy\, dt \\
& \quad \leqslant - \int_0^T \int_\Omega \xi(t,y)\partial_t\big(F\big(\overline{\rho}(t,y)\big)\big)\, dt\, dy - \int_0^T \int_\Omega \nabla\xi(t,x)F'\big(\overline{\rho}(t,y)\big)\overline{A}
\end{aligned}$$

$$= \int_0^T \int_\Omega \xi(t,y)\overline{A}(t,y)\cdot\nabla\big(F'\big(\overline{\rho}(t,y)\big)\big)\,dy\,dt + \big(T^s;\xi\big)$$

by (49).

Term 2. Since $\rho^{\varepsilon h}$ converges in $w^*L^\infty(\Omega)$ towards $\overline{\rho}$, we have

$$\int_0^T \int_\Omega \xi(t,x)\rho^{\varepsilon h}(t,y)\zeta(t,y)\cdot\nabla c^*\big(\zeta(t,y)\big)\,dy\,dt \to \int_0^T \int_\Omega \xi(t,x)\overline{\rho}(t,y)\zeta(t,y)\cdot\nabla c^*\big(\zeta(t,y)\big)\,dy\,dt$$

when $\varepsilon < h \to 0$.

Term 3. Using Theorem 1.7 of De Cicco, Fusco and Verde [23] on the L^1 semi-continuity of functionals $J[\rho, D\rho]$ on BV which are convex with respect to the gradient variable and continuous with respect to $\rho = F'(\rho^{\varepsilon h})$, we obtain

$$\liminf_{\varepsilon\to 0} \int_0^T \int_\Omega \xi(t,y)\nabla c^*\big(\zeta(t,y)\big)\cdot\nabla\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)(F^*)'\big(F'\big(\rho^{\varepsilon h}(t,y)\big)\big)\,dy\,dt$$

$$\geqslant \liminf_{h\to 0} \int_0^T \int_\Omega \xi(t,y)\nabla c^*\big(\zeta(t,y)\big)\cdot D\big(F'\big(\rho^h(t,y)\big)\big)\rho^h(t,y)\,dy\,dt$$

$$\geqslant \int_0^T \int_\Omega \xi(t,y)\nabla c^*\big(\zeta(t,y)\big)\cdot D\big(F'\big(\overline{\rho}(t,y)\big)\big)\overline{\rho}(t,y)\,dy\,dt.$$

In fact, since $J[\rho, D\rho]$ depends linearly on $D\rho$ in this case, we have that both lim infs can be replace by limits which converge, and the inequality above becomes an equality.

Term 4.

$$\int_0^T \int_\Omega \xi(t,x)\rho^{\varepsilon h}(t,y)\nabla c^*\big(\rho^{\varepsilon h}(t,y)\big)\cdot\zeta(t,y)\,dy\,dt \to \int_0^T \int_\Omega \xi(t,x)\overline{A}(t,y)\cdot\zeta(t,y)\,dy\,dt$$

when $\varepsilon < h \to 0$.

Then the inequality $\Lambda^{\varepsilon h} \geqslant 0$ becomes

$$\big(\overline{A}\big(\nabla\big(F'(\overline{\rho})\big)-\zeta\big) - \overline{\rho}\nabla c^*(\zeta)\big(D\big(F'(\overline{\rho})\big)-\zeta\big) + T^s;\xi\big) \geqslant 0.$$

By Corollary 1, p. 53 of [27], we know that $\overline{A}(\nabla(F'(\overline{\rho}))-\zeta) - \overline{\rho}\nabla c^*(\zeta)(D^{ac}(F'(\overline{\rho}))-\zeta) + T^s$ is a Radon measure and then by the Lebesgue decomposition of Radon measure, we can conclude that it's absolute continuous part is positive which means that

$$\big(\overline{A} - A(\overline{\rho},\zeta)\big)\big(D^{ac}\big(F'(\overline{\rho})\big)-\zeta\big) \geqslant 0$$

for all test function ζ which concludes the proof. □

Acknowledgements

The first author acknowledges the support of United States National Science Foundation grant DMS 0354729, Natural Sciences and Engineering Research Council of Canada grant 217006-03, and the hospitality of the Centro Recerca Matematica de Catalunya and the Université Paul Sabatier de Toulouse, where parts of this work were preformed. The second author would like to thank the Research Training Network "Fronts singularities" for its support, and the Department of Mathematics of the Rutgers University where this work was completed. Both authors are grateful to Xavier Cabre and Jose-Antonio Carrillo, who originally proposed this project and contributed energetically to its solution, and to Luigi Ambrosio, Wilfrid Gangbo and Felix Otto for critical comments along the way. They are especially grateful to Fuensanta Andreu, Vicent Caselles and José Mazón for several fruitful discussions that have been crucial for the understanding of this problem. The authors would like to thank the referee for his careful reading of the paper and his very helpful remarks.

References

[1] Martial Agueh, Existence of solutions to degenerate parabolic equations via the Monge–Kantorovich theory, Adv. Differential Equations 10 (3) (2005) 309–360.
[2] Luigi Ambrosio, Steepest descent flows and applications to spaces of probability measures, Lectures Notes, Santander, July 2004.
[3] Luigi Ambrosio, Nicola Gigli, Giuseppe Savaré, Gradient Flows in Metric Spaces and in the Spaces of Probability Measures, Lectures Math. ETH Zurich, Birkhäuser Verlag, Basel, 2005.
[4] Luigi Ambrosio, Aldo Pratelli, Existence and Stability Results in the L^1 Theory of Optimal Transportation, Lecture Notes in Math.
[5] Luigi Ambrosio, Paolo Tilli, Topics on Analysis in Metric Spaces, Oxford Lecture Ser. Math. Appl., vol. 25, Oxford University Press, Oxford, 2004.
[6] Gianni Anzellotti, Pairing between measures and bounded functions and compensated compactness, Ann. Mat. Pura Appl. (4) 135 (1983) 293–318.
[7] Fuensanta Andreu, Vicent Caselles, José M. Mazón, Existence and uniqueness of solution for a parabolic quasilinear problem for linear growth functionals with L^1 data, Math. Ann. 322 (2002) 139–206.
[8] Fuensanta Andreu, Vicent Caselles, José M. Mazón, Parabolic Quasilinear Equations Minimizing Linear Growth Functionals, Progr. Math., vol. 223, Birkhäuser Verlag, 2004.
[9] Fuensanta Andreu, Vicent Caselles, José M. Mazón, A strongly degenerate quasilinear equation: The elliptic case, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 3 (3) (2004) 555–587.
[10] Fuensanta Andreu, Vicent Caselles, José M. Mazón, A strong degenerate quasilinear equation: The parabolic case, Arch. Ration. Mech. Anal. (2005).
[11] Fuensanta Andreu, Vicent Caselles, José M. Mazón, A strongly degenerate quasilinear elliptic equation, Nonlinear Anal. 61 (2005) 637–669.
[12] Fuensanta Andreu, Vicent Caselles, José M. Mazón, The Cauchy problem for a strong degenerate quasilinear equation, J. Eur. Math. Soc. (JEMS) 7 (2005) 361–393.
[13] Fuensanta Andreu, Vicent Caselles, José M. Mazón, Salvador Moll, The speed of propagation of the support of solutions of a tempered diffusion equation, preprint.
[14] Yann Brenier, Extended Monge–Kantorovich theory, in: Optimal Transportation and Applications, Martina Franca, 2001, in: Lecture Notes in Math., vol. 1813, Springer, Berlin, 2003, pp. 91–121.
[15] Haim Brezis, Analyse fonctionnelle et ses applications, Masson, 1983.
[16] Haim Brezis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North-Holland Math. Stud., vol. 5, Notas Mat., vol. 50, North-Holland Publishing Co./American Elsevier Publishing Co., Inc., Amsterdam, London/New York, 1973 (in French).
[17] Luis A. Caffarelli, Boundary regularity of maps with convex potentials. II, Ann. of Math. (2) 144 (3) (1996) 453–496.
[18] Jose Antonio Carrillo, Giuseppe Toscani, Wasserstein metric and large-time asymptotics of nonlinear diffusion equations, in: New Trends in Mathematical Physics, World Sci. Publ., Hackensack, NJ, 2005, pp. 234–244 (in honor of the Salvatore Rionero 70th birthday).
[19] Vicent Caselles, Convergence of the "relativistic" heat equation to the heat equation as $c \to \infty$, Publ. Mat. 51 (1) (2007) 121–142.
[20] Alina Chertock, Alexander Kurganov, Philip Rosenau, Formation of discontinuities in flux-saturated degenerate parabolic equations, Nonlinearity 16 (2003) 1875–1898.
[21] Dario Cordero-Erausquin, Non-smooth differential properties of optimal transport, in: Recent Advances in the Theory and Applications of Mass Transport, in: Contemp. Math., vol. 353, Amer. Math. Soc., Providence, RI, 2004, pp. 61–71.
[22] Gianni Dal Maso, Integral representation on $BV(\omega)$ of Γ-limits of variational integrals, Manuscripta Math. 30 (1980) 387–416.
[23] Virginia De Cicco, Nicola Fusco, Anna Verde, On L^1-lower semicontinuity in BV, J. Convex Anal. 12 (1) (2005) 173–185.
[24] Nelson Dunford, Jaco Schwartz, Linear Operators, Interscience Publishers, New York, 1958.
[25] Lawrence C. Evans, Weak Convergence Methods for Nonlinear Partial Differential Equations, published for the Conference Board of the Mathematical Sciences, Washington, DC, CBMS Reg. Conf. Ser. Math., vol. 74, Amer. Math. Soc., Providence, RI, 1990.
[26] Lawrence C. Evans, Wilfrid Gangbo, Differential equation methods for the Monge Kantorovich mass transfer, Mem. Amer. Math. Soc. 653 (1999).
[27] Lawrence C. Evans, Ronald F. Gariepy, Measure Theory and Fine Properties of Functions, Stud. Adv. Math., CRC Press, 1992.
[28] Wilfrid Gangbo, Robert J. McCann, Optimal maps in Monge's mass transport problem, C. R. Acad. Sci. Paris Sér. I Math. 321 (12) (1995) 1653–1658.
[29] Wilfrid Gangbo, Robert J. McCann, The geometry of optimal transportation, Acta Math. 177 (2) (1996) 113–161.
[30] Richard Jordan, David Kinderlehrer, Felix Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[31] Kohn Robert, Temam Roger, Dual spaces of stresses and strains, with applications to Hencky plasticity, Appl. Math. Optim. 10 (1) (1983) 1–35.
[32] Jacques-Louis Lions, Quelques methodes de résolution des problèmes aux limites non linéaires, Dunod, 1969.
[33] Grégoire Loeper, On the regularity of the polar factorization for time dependent maps, Calc. Var. Partial Differential Equations 22 (3) (2005) 343–374.
[34] Xi-Nan Ma, Neil S. Trudinger, Xu-Jia Wang, Regularity of potential functions of the optimal transportation problem, Arch. Ration. Mech. Anal. 177 (2) (2005) 151–183.
[35] Robert J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1997) 153–179.
[36] Robert J. McCann, Exact solutions to the transportation problem on the line, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 455 (1984) (1999) 1341–1380.
[37] D. Mihalas, B. Mihalas, Foundations of Radiation Hydrodynamics, Oxford University Press, 1984.

[38] Felix Otto, Doubly degenerate diffusion equations as steepest descent, preprint, 1996.
[39] Svetlozar T. Rachev, Ludger Rüschendorf, Mass Transportation Problems, vols. I. and II, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998.
[40] Philip Rosenau, Tempered diffusion: A transport process with propagating fronts and initial delay, Phys. Rev. A 46 (1992) 7371–7374.
[41] Jacques Simon, Compact sets in the space $L^p(0, T; B)$ Ann. Mat. Pura Appl. 146 (1987) 65–96.
[42] Roger Temam, Navier–Stokes Equation. Theory and Numerical Analysis, third ed., Stud. Math. Appl., vol. 2, North-Holland Publishing Co., Amsterdam, 1984.
[43] Cedric Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, Amer. Math. Soc., 2003.