Entropic Burgers' equation via a minimizing movement scheme based on the Wasserstein metric

Nicola Gigli[*], Felix Otto[†]

February 27, 2012

Abstract

As noted by the second author in the context of unstable two-phase porous medium flow, entropy solutions of Burgers' equation can be recovered from a minimizing movement scheme involving the Wasserstein metric in the limit of vanishing time step size [4]. In this paper, we give a simpler proof by verifying that the anti-derivative is a viscosity solution of the associated Hamilton Jacobi equation.

AMS Subject classification: 76S05 (35Q35 49Q20 76T99)

Introduction

The aim of this paper is twofold. On one side we give a simpler proof of a result found by the second author ([4]). This amounts in proving that the minimizing movements scheme for the Energy $E(\theta) = \int x\theta dx$ on a two-phase Wasserstein space produces the entropy solution of the Burgers' equation. The difference with respect to the approach of [4] consists in the fact that we pass to the anti-derivative of the limit function and prove that it is the viscosity solution of the associated Hamilton Jacobi equation. This provides some technical simplifications.

On the other hand we discuss, mostly at an informal level, the gradient flow structure that this result suggests. Indeed, on an heuristic level the fact that the Burgers' equation is the equation of the gradient flow of $E(\theta) = \int x\theta dx$ on a two-phase Wasserstein space is obvious (see Section 2.3). It is less obvious why the minimizing movement scheme converges to the *entropy* solution.

[*]University of Nice

[†]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Shortly said, the situation is the following: E is not semi convex on the *two-phase* Wasserstein space and this leads to multiple solutions of the initial value problem (in correspondence with the lack of uniqueness for weak solutions of Burgers' equation). Now, given the structure of the minimizing movements technique, it is natural to argue that the limit curve obtained is the one which locally decreases the energy E fastest or equivalently, which locally has the largest energy dissipation rate $|\nabla E|^2$. This is just a reformulation of Dafermos' observation that the entropy solution is the weak solution which decreases entropy $-\int \theta(1-\theta)dx$ fastest among all weak solutions. This is our explanation why the minimizing movement scheme picks the entropy solution (see also Section 1.2).
Still, it should be noted that this explanation is purely formal (and actually not needed for the rigorous proof). As we discuss in the final sections, there are explicit examples in $\mathbb{R}^2$ of C^1 functions for which there is not a gradient flow trajectory that decreases the energy fastest, and thus in particular the minimizing movements scheme cannot produce such a trajectory (see Section 3.1). Also, it is very possible that a gradient flow trajectory which decreases the energy fastest exists, but is *not* obtained by the minimizing movements scheme (see Section 3.2).

We wish to thank the referee for valuable comments on the preliminary version of the paper.

1 Two words on Wasserstein distance between measures with infinite mass

In this paper we are going to deal with the Wasserstein distance between measures on $\mathbb{R}$ with infinite mass. As this is somehow unusual, we present here the main features of the resulting metric space. Yet, we won't prove the results collected here, because anyway their proofs are just minor variant of those available in the standard setting of probability measures.
Let $H:\mathbb{R}\to\mathbb{R}$ be the Heaviside function

$$H(x) := \left\{ \begin{array}{ll} 1, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{array} \right\}$$

and let $\mathcal{M}^+$ be the set of functions $\theta:\mathbb{R}\to[0,1]$ such that

$$\left.\begin{array}{l} L(\theta) := \sup\{x : \int_{-\infty}^{x} \theta = 0\} > -\infty, \\ R(\theta) := \inf\{x : \int_{x}^{+\infty} 1-\theta = 0\} < +\infty, \end{array}\right\} \qquad \text{bounded “mixing zone”,}$$

$$\int_{\mathbb{R}} \theta - H = 0, \qquad \text{volume constraint.} \tag{1.1}$$

Notice that $H \in \mathcal{M}^+$. Define

$$\mathcal{M}^- := \left\{ \theta : \mathbb{R} \to [0,1] \middle| 1 - \theta \in \mathcal{M}^+ \right\}.$$

Given two functions $\theta_0, \theta_1 \in \mathcal{M}^\pm$ we define their Wasserstein distance W_2 (here and in the following we are identifying a measure with its density) as

$$W_2^2(\theta_0, \theta_1) := \inf_{T : T_\sharp \theta_0 = \theta_1} \int |T(x) - x|^2 \theta_0(x)\, dx.$$

As for the standard transport problem, it is easy to check that there exists a unique non decreasing map T_{opt} such that $(T_{opt})_\sharp \theta_0 = \theta_1$, and that this map is the unique minimizer of the problem above. Furthermore, T_{opt} minimizes also the transport problem written in terms of plans.
Thanks to the volume constraint and the bounded mixing zone, it is immediate to verify that for $L \le \min\{L(\theta_0), L(\theta_1)\}$ and $R \ge \max\{R(\theta_0), R(\theta_1)\}$ it holds $T_{opt}(x) = x$ for any $x \notin [L, R]$. In other words, only a finite amount of mass is moved in the optimal transportation, so that it also holds

$$W_2(\theta_0, \theta_1) = W_2(\theta_0|_{[L,R]}, \theta_1|_{[L,R]}).$$

The fact that the transport problem in $\mathcal{M}^\pm$ is equivalent to a transport problem where the masses involved are finite, yields that several properties of the standard transport problem are true also in the current setting. For instance, the dual formulation holds:

$$\frac{1}{2} W_2^2(\theta_0, \theta_1) = \sup_{(\phi, \tilde{\phi})} \int_L^R \left(\frac{x^2}{2} - \phi_0(x) \right) \theta_0(x) dx + \int_L^R \left(\frac{x^2}{2} - \phi_1(x) \right) \theta_1(x) dx,$$

where the sup is taken among all couples (ϕ_0, ϕ_1) of convex conjugate functions. The sup is always realized and any optimal couple $(\phi^0_{opt}, \phi^1_{opt})$ satisfies $\phi^0_{opt}(x) + \phi^1_{opt}(T_{opt}(x)) = x T_{opt}(x)$ for any $x \in \mathbb{R}$.
Notice that the metric space $(\mathcal{M}^\pm, W_2)$ is *not* complete, indeed: measures with densities θ with either $L(\theta) = -\infty$ or $R(\theta) = +\infty$ may arise as limits. Yet, this will create no troubles to our discussion, because on the one hand we will work with densities ≤ 1 (which rules out singular parts at the limit), and on the other hand solutions of the Burgers equation have finite speed of propagation, so that if we start from a density with bounded mixing zone, also its evolution will have bounded mixing zone (we will recover this property also at the level of time discretized solutions, see (3.10) and (3.13)).
Finally, we remark that as for the standard Wasserstein space, the space $(\mathcal{M}^\pm, W_2)$ can be seen as a sort of infinite dimensional manifold (according to the interpretation provided in [5]), where the scalar product g_θ at some $\theta \in \mathcal{M}^\pm$ is given by

$$g_\theta(\delta\theta, \delta\theta) := \int |v|^2 \theta,$$

for any admissible perturbation $\delta\theta$ - which must satisfy $\int \delta\theta = 0$ - where v is defined from $\delta\theta$ by

$$-\partial_x(v\theta) = \delta\theta, \qquad v \in L^2(\mathbb{R};\theta\mathcal{L}^1).$$

This statement should be understood in the following sense: a curve $(\theta_t) \subset \mathcal{M}^{\pm}$ is absolutely continuous w.r.t. W_2 if and only if it solves the continuity equation

$$\partial_t\theta_t + \partial_x(v_t\theta_t) = 0,$$

for some family of vector fields v_t such that $\int v_t = 0$ for a.e. t. In this case the metric speed $|\dot{\theta}_t|$ of (θ_t) is given by $\sqrt{\int |v_t|^2\theta_t}$ for a.e. t.
Notice that in particular the Benamou-Brenier formula holds:

$$W_2^2(\theta^0, \theta^1) = \inf \int_0^1 \int |v_t|^2\theta_t \, dt,$$

where the infimum is taken among all absolutely continuous curves (θ_t) such that $\theta_0 = \theta^0$ and $\theta_1 = \theta^1$. These latter statements can be made rigorous following the arguments presented in [5] and [1].

2 Heuristics

2.1 Burgers' equation

Let $\lambda \in (0,\infty)$ be a fixed parameter. From the method of characteristics we learn that the initial value problem for

$$\partial_t\theta + \partial_x\left(\frac{\theta(\theta-1)}{\theta+\lambda^{-1}(1-\theta)}\right) = 0 \tag{2.1}$$

generically is not solvable in the class of continuous θ. On the other hand, it is well-known that for given initial data, (2.1) allows several weak solutions θ. A measurable function $\theta\colon (0,\infty)\times\mathbb{R} \to [0,1]$ is called weak solution of (2.1) with initial data given by a measurable function $\theta_0\colon \mathbb{R} \to [0,1]$ if

$$\int_{\mathbb{R}}\int_0^\infty \partial_t\varphi(x,t)\theta(x,t) + \partial_x\varphi(x,t)\frac{\theta(x,t)(\theta(x,t)-1)}{\theta(x,t)+\lambda^{-1}(1-\theta(x,t))}dxdt = \int_{\mathbb{R}} \varphi(x,0)\theta_0(x)dx$$

for all $\varphi \in C_c^\infty([0,\infty)\times\mathbb{R})$.
The most striking example of non-uniqueness occurs when

$$\theta_0(x) = H(x) := \left\{ \begin{array}{ll} 1, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{array} \right\} \tag{2.2}$$

Among the infinitely many weak solutions which start from (2.2), let us mention three:

- The stationary solution:

$$\theta(t,x) = H(x).$$

- The “rarefaction wave”:

$$\theta(t,x) \;=\; \left\{ \begin{array}{ccc} 0 & \text{for} & x \le -\lambda t \\ \frac{1+\lambda^{-1}r}{1+\lambda^{-1}r-r+\sqrt{\lambda(1+\lambda^{-1}r-r)}} & \text{for} & r := \frac{x}{t},\ -\lambda t \le x \le\ t \\ 1 & \text{for} & x \ge t \end{array} \right\}.$$

- An intermediate solution

$$\theta(t,x) \;=\; \left\{ \begin{array}{ccc} 0 & \text{for} & x \le -\frac{1}{1+\lambda^{-1}}t \\ \frac{1}{2} & \text{for} & -\frac{1}{1+\lambda^{-1}}t \le x \le\ \frac{1}{1+\lambda^{-1}}t \\ 1 & \text{for} & x \ge \frac{1}{1+\lambda^{-1}}t \end{array} \right\}.$$

As is well known, the notion of entropy solution restores uniqueness [3]. A function is called entropy solution of (2.1) provided it is a weak solution and satisfies

$$\int_{\mathbb{R}} \int_0^\infty \eta(\theta)\partial_t\varphi + q(\theta)\partial_x\varphi dx\, dt \ge 0,$$

for all $\varphi \in C_c^\infty([0,+\infty) \times \mathbb{R})$, $\varphi \ge 0$, all convex functions $\eta : \mathbb{R} \to \mathbb{R}$ (called entropies), and $q : \mathbb{R} \to \mathbb{R}$ defined by

$$q(x) := \int_0^x \eta'(y) f'(y) dy,$$

where $f(z) := z(z-1)/(z+\lambda^{-1}(1-z))$.
Dafermos [2] has observed that for a scalar conservation law as above, Kruzkov’s (or Lax’) notion of entropy solution is formally equivalent to the following criterium. Select a *strictly* convex entropy η. Then the entropy solution θ^* is characterized by the following property: For every weak solution θ that agrees with θ^* up to time T one has

$$\lim_{t\downarrow T} \frac{\int_{\mathbb{R}} \eta(\theta^*(t,x))dx - \int_{\mathbb{R}} \eta(\theta^*(T,x))dx}{t-T} \le \lim_{t\downarrow T} \frac{\int_{\mathbb{R}} \eta(\theta(t,x))dx - \int_{\mathbb{R}} \eta(\theta(T,x))dx}{t-T}.$$

In words: Among all weak solutions, it is the entropy solution that decreases the total entropy fastest instantaneously. Dafermos proved the equivalence for the class of piecewise smooth functions with smooth discontinuity lines.

2.2 Abstract gradient flow

We are going to discuss in the Section 2.3 the gradient flow structure of the Burger's equation. But before doing so, we collect here some heuristic ideas which drove our intuition in attacking the problem.
Recall that the gradient flow equation for an energy $\mathcal{E}$ defined on, say, $\mathbb{R}^d$ is

$$u'(t) = -\nabla\mathcal{E}(u(t)), \tag{2.3}$$

where the unknown is the curve $u : [0, +\infty) \to \mathbb{R}^d$ and it is specified the initial value $u(0) = u_0 \in \mathbb{R}^d$. It is well known that if the energy $\mathcal{E}$ is C^2 or semi-convex, then there is both existence and uniqueness for equation (2.3). When $\mathcal{E}$ is only C^1, existence is still ensured by Peano's theorem, but uniqueness may fail. A typical example being $\mathcal{E}(x) := -x^{\frac{4}{3}}$ on $\mathbb{R}$: with the initial condition $u_0 = 0$ we have uncountably many solutions, among which the constant $u_1(t) \equiv 0$ and the curve $u_2(t) = \left(\frac{8}{9}\right)^{\frac{3}{2}} t^{\frac{3}{2}}$. Clearly, these two solutions have a very different behavior: the first is stationary, while the second shows the maximal decrease of the energy.
To better formalize this concept, let us observe that the infinitesimal rate of dissipation of the energy is prescribed by the gradient flow equation, as it holds

$$\frac{d}{dt}\mathcal{E}(u(t)) = \nabla\mathcal{E}(u(t)) \cdot u'(t) = -|\nabla\mathcal{E}|^2(u(t)).$$

If $\mathcal{E}$ is C^2, also the second derivative of the energy is given, being equal to

$$\frac{d^2}{dt^2}\mathcal{E}(u(t)) = -\frac{d}{dt}|\nabla\mathcal{E}|^2(u(t)) = 2\nabla\mathcal{E}(u(t)) \cdot \big(\nabla^2\mathcal{E}(u(t)) \cdot \nabla\mathcal{E}(u(t))\big).$$

If $\mathcal{E}$ is only C^1, different solutions may very well have different second order variations of the energy. For instance, with the example $\mathcal{E}(x) := -x^{\frac{4}{3}}$ above we have

$$\frac{d^2}{dt^2}\mathcal{E}(u_1(t)) = 0,$$
$$\frac{d^2}{dt^2}\mathcal{E}(u_2(t)) = -\left(\frac{8}{9}\right)^2 \frac{d^2}{dt^2}t^2 = -2\left(\frac{8}{9}\right)^2.$$

Thus we see that among the two gradient flow trajectories, the one which decreases the energy fastest is characterized by a smaller second order derivative of the energy, or, which is the same, by a larger derivative of $|\nabla\mathcal{E}|^2(u(t))$. Thus if one looks for a concepts that isolates the curve $u_2(t)$ among all the gradient flows trajectories of $\mathcal{E}(x) = -x^{\frac{4}{3}}$ as the one with the highest dissipation of the energy, it seems reasonable to look for the curve with the largest derivative of $|\nabla\mathcal{E}|^2(u(t))$. In practice, in order to avoid situations where the derivative does not exist or is not sufficient to isolate the curve with highest energy dissipation (this is the case, for instance, of the functional $\mathcal{E}(x) := -x^{\frac{3}{2}}$), it is better to substitute the infinitesimal notion of derivative of $|\nabla\mathcal{E}|^2(u(t))$, with a local one, as in the following definition.

Definition 1 (Local maximal increase of the slope and local maximal decrease of the functional)**.** *We say that a solution $u(t)$ of* (2.3) *has the local maximal increase of the slope if for any $t_0 \geq 0$ and any solution $\tilde{u}(t)$ of*

$$\begin{cases} \tilde{u}(0) = u(t_0), \\ \tilde{u}'(t) = -\nabla\mathcal{E}(\tilde{u}(t)), \end{cases} \tag{2.4}$$

there exists $\delta > 0$ such that it holds $|\nabla\mathcal{E}(\tilde{u}(t))| \leq |\nabla\mathcal{E}(u(t+t_0))|$ for any $t \leq \delta$.
Similarly, we say that a solution $u(t)$ of (2.3) *has the local maximal decrease of the energy if for any $t_0 \geq 0$ and any solution $\tilde{u}(t)$ of* (2.4) *there exists $\delta > 0$ such that $\mathcal{E}(\tilde{u}(t)) \geq \mathcal{E}(u(t+t_0))$ for any $t \leq \delta$.*

The two notions of maximal decrease are linked. Indeed, if a solution $u(t)$ has the local maximal increase of the slope, then it has the local maximal decrease of the energy. To see this, observe that, since we are assuming the energy to be C^1, it holds

$$\begin{aligned} \mathcal{E}(u(t_0+t)) - \mathcal{E}(u(t_0)) &= \int_{t_0}^{t_0+t} \frac{d}{ds}\mathcal{E}(u(s))ds \\ &= \int_{t_0}^{t_0+t} \nabla\mathcal{E}(u(s)) \cdot u'(s)ds = -\int_{t_0}^{t_0+t} |\nabla\mathcal{E}(u(s))|^2 ds. \end{aligned} \tag{2.5}$$

Therefore, if we have $\delta > 0$ and a curve $\tilde{u}(t)$ which solves (2.4) and satisfies $|\nabla\mathcal{E}(\tilde{u}(t))| \leq |\nabla\mathcal{E}(u(t+t_0))|$ for any $t \leq \delta$, from the equality $\mathcal{E}(\tilde{u}(0)) = \mathcal{E}(u(t_0))$ and (2.5) for u and $\tilde{u}$ we can conclude that $\mathcal{E}(\tilde{u}(t)) \geq \mathcal{E}(u(t+t_0))$ for any $t \leq \delta$.
It is unclear to us whether the opposite implication is true or not; let us just point out once again that what we are going to say below is just *heuristic* and serves as a motivation for the rigorous result of Section 3 (see also the last section).

There is a well-known approximation scheme which produces solutions of the initial value problem of (2.3): the method of minimizing movements, which consists in the following. We choose a parameter $\tau > 0$ and define a sequence u_n^τ by imposing $u_0^\tau := u_0$ and then recursively choosing u_{n+1}^τ among the minimizers of

$$u \qquad \mapsto \qquad \mathcal{E}(u) + \frac{|u - u_n^\tau|^2}{2\tau}.$$

Then we define the piecewise constant curves $t \mapsto u^\tau(t) := u^\tau_{[t/\tau]}$, where by $[x]$ we denote the integer part of the real number x.
Under only mild assumptions on $\mathcal{E}$, it can be proved that as τ goes to 0, the family of curves $u^\tau(t)$ converges (possibly up to considering some particular sequence $\tau_n \downarrow 0$) locally uniformly to a solution of (2.3).
What we want to point out here, is that given the structure of the minimizing movement technique, when uniqueness of solutions of (2.3) fails, the solution found by this method should be one which satisfies the local maximal decrease of the energy, and the local maximal increase of the slope, if it exists[1].

2.3 Bringing together both concepts

We want to show that the equation

$$\begin{cases} \partial_t \theta + \partial_x \left(\dfrac{\theta(\theta-1)}{\theta + \lambda^{-1}(1-\theta)} \right) = 0 \quad \text{in } \mathbb{R}^+ \times \mathbb{R}, \\ \\ \theta(0,x) = \theta^0(x), \end{cases} \tag{2.6}$$

where the initial value $\theta^0 : \mathbb{R} \mapsto \mathbb{R}$ satisfies $\theta^0 \in \mathcal{M}^+$, $0 \le \theta^0 \le 1$ can be seen as a gradient flow equation. The calculations we run here are purely heuristical, but they serve as motivation for the rigorous result given in the next section.
To justify that equation (2.6) is the equation of a gradient flow, we need to specify a manifold and an energy. Let us recall that a curve $t \mapsto u(t)$ is a gradient flow on a manifold $\mathcal{N}$ for the energy $\mathcal{E}$ if it satisfies

$$\langle \dot{u}(t), w \rangle = - \langle d\mathcal{E}(u(t)), w \rangle \,, \tag{2.7}$$

for any $t > 0$ and any w in the tangent space of $\mathcal{N}$ at $u(t)$.
Our manifold $\mathcal{N}$ is a subspace of the 'manifold' $\mathcal{M}^+ \times \mathcal{M}^-$, where $\mathcal{M}^\pm$ has been introduced in the Section 1:

$$\mathcal{N} := \Big\{ (\theta, \tilde{\theta}) \ : \ \theta \in \ M^+, \ \tilde{\theta} \in \mathcal{M}^-, \ \theta + \tilde{\theta} \equiv 1 \Big\}. \tag{2.8}$$

In the following we will often write $\theta \in \mathcal{N}$ meaning $(\theta, 1-\theta) \in \mathcal{N}$. Also, for $\theta \in \mathcal{N}$, we will often identify θ and $1-\theta$ with the measures $\theta\mathcal{L}^1$, $(1-\theta)\mathcal{L}^1$.
On $\mathcal{N}$ we put two distances. The first, d, is given by

$$d^2\big((\theta^0, \tilde{\theta}^0), (\theta^1, \tilde{\theta}^1)\big) := \frac{1}{\lambda} W_2^2(\theta^0, \theta^1) + W_2^2(\tilde{\theta}^0, \tilde{\theta}^1).$$

[1]let us underline that actually there may be no solution which locally decreases the energy fastest. See the last section for an example.

The second, D, is defined as

$$D^2\big((\theta^0,\tilde\theta^0),(\theta^1,\tilde\theta^1)\big) := \inf \int_0^1 \left(\frac{1}{\lambda}\int |v_t|^2\theta_t + \int |\tilde v_t|^2\tilde\theta_t\right) dt,$$

where the infimum is taken among all curves $t \mapsto (\theta_t, \tilde\theta_t) \in \mathcal{N}$ such that both $t \mapsto \theta_t$ and $t \mapsto \tilde\theta_t$ are absolutely continuous w.r.t. W_2 and $v_t, \tilde v_t$ are the associated velocity vector fields, i.e. it holds $v_t \in L^2(\mathbb{R};\theta_t\mathcal{L}^1)$, $\tilde v_t \in L^2(\mathbb{R};\tilde\theta_t\mathcal{L}^1)$, for a.e. t and

$$\partial_t\theta_t + \partial_x(v_t\theta_t) = 0, \qquad \partial_t\tilde\theta_t + \partial_x(\tilde v_t\tilde\theta_t) = 0.$$

We remark that D is the 'Riemannian distance' associated to the scalar product $\overline{g}_{(\theta,\tilde\theta)}$ defined by

$$\overline{g}_{(\theta,\tilde\theta)}\big((\delta\theta,\delta\tilde\theta),(\delta\theta,\delta\tilde\theta)\big) := \frac{1}{\lambda}\int |v|^2\theta + \int |\tilde v|^2\tilde\theta, \tag{2.9}$$

where $(\delta\theta, \delta\tilde\theta)$ is an admissible variation of $(\theta,\tilde\theta) \in \mathcal{N}$, which means $\int \delta\theta = \int \delta\tilde\theta = 0$ and $\delta\theta = -\delta\tilde\theta$, and $v \in L^2(\mathbb{R},\theta)$, $\tilde v \in L^2(\mathbb{R},\tilde\theta)$ are derived from $\delta\theta, \delta\tilde\theta$ by

$$-\partial_x(v\theta) = \delta\theta, \qquad -\partial_x(\tilde v\tilde\theta) = \delta\tilde\theta. \tag{2.10}$$

Notice that in particular this implies

$$v\theta + \tilde v\tilde\theta \equiv 0. \tag{2.11}$$

In other words, the distance D is the arc distance associated to the embedding of $\mathcal{N}$ into the weighted product $\mathcal{M}^+ \times \mathcal{M}^-$, while d is the corresponding chord distance. In the next section, where we will derive our rigorous result, we will use the distance d, which is more manageable, but for the moment we use D, for which it is easier to derive the gradient flow equation.

The other ingredient needed to have a gradient flow is the energy functional. Our is $\mathcal{E} : \mathcal{N} \to \mathbb{R}$ defined by:

$$\mathcal{E}(\theta,\tilde\theta) := E(\theta) = \int x(\theta(x) - H(x))\,dx. \tag{2.12}$$

Formally, the differential $\operatorname{diff}\mathcal{E}(\theta,\tilde\theta)$ of $\mathcal{E}$ at $(\theta,\tilde\theta)$ computed along the direction $(\delta\theta_0,\delta\tilde\theta_0)$ is easy to compute, and it is given by

$$\operatorname{diff}\mathcal{E}(\theta,\tilde\theta)(\delta\theta_0,\delta\tilde\theta_0) = \int x\delta\theta_0(x)\,dx.$$

The gradient $\operatorname{grad}\mathcal{E}(\theta,\tilde\theta)$ is then, by definition, the tangent vector $(\delta\theta,\delta\tilde\theta)$ defined by the formula

$$\overline{g}_{(\theta,\tilde\theta)}\big((\delta\theta_0,\delta\tilde\theta_0),(\delta\theta,\delta\tilde\theta)\big) = \operatorname{diff}\mathcal{E}(\theta,\tilde\theta)(\delta\theta_0,\delta\tilde\theta_0), \qquad \forall(\delta\theta_0,\delta\tilde\theta_0)$$

Associating to $(\delta\theta_0, \delta\tilde{\theta}_0)$ and $(\delta\theta, \delta\tilde{\theta})$ the vector fields $(v_0, \tilde{v}_0)$ and $(v, \tilde{v})$ respectively via (2.10) and then using the definition (2.9) of the scalar product $\overline{g}$ we get

$$\frac{1}{\lambda}\int v_0 v\theta + \int \tilde{v}_0\tilde{v}\tilde{\theta} = \int v_0\theta, \qquad \text{for any } (v_0, \tilde{v}_0) \text{ satisfying (2.11).}$$

Using twice (2.11) we then get

$$\int v_0\theta\left(v\left(\lambda^{-1} + \frac{\theta}{1-\theta}\right) - 1\right) = 0, \qquad \forall v_0,$$

which identifies $\operatorname{grad}\mathcal{E}(\theta, \tilde{\theta}) = (\delta\theta, -\delta\theta)$ from

$$v = \frac{1-\theta}{\theta + \lambda^{-1}(1-\theta)} \quad \text{and thus} \quad \delta\theta = \partial_x\frac{\theta(\theta-1)}{\theta + \lambda^{-1}(1-\theta)}. \tag{2.13}$$

Hence, we just 'proved' that $t \mapsto (\theta_t, \tilde{\theta}_t)$ is a gradient flow of $\mathcal{E}$ on $\mathcal{N}$ if and only if

$$\partial_t\theta_t = -\partial_x\left(\frac{\theta_t(\theta_t - 1)}{\theta_t + \lambda^{-1}(1-\theta_t)}\right),$$

which is (2.6).

We conclude this introduction showing that in our situation the energy functional is not semi convex and that the lower semicontinuous envelope of $|\operatorname{grad}\mathcal{E}|^2$ vanishes.
We start from the lower semicontinuous envelope of $|\operatorname{grad}\mathcal{E}|^2$. Formula (2.13) yields that $|\operatorname{grad}\mathcal{E}|^2(\theta, 1-\theta)$ is given by

$$\frac{1}{\lambda}\int\left|\frac{1-\theta}{\theta + \lambda^{-1}(1-\theta)}\right|^2\theta + \int\left|\frac{\theta}{\theta + \lambda^{-1}(1-\theta)}\right|^2(1-\theta) = \int\frac{\theta(1-\theta)}{\theta + \lambda^{-1}(1-\theta)}. \tag{2.14}$$

This formula shows that if θ attains only the values $0, 1$, the norm of the gradient is equal to 0. Since the set of such $(\theta, 1-\theta)$'s is dense in $\mathcal{N}$ we get the claim.
The same formula shows that $-|\operatorname{grad}\mathcal{E}|^2 = \int_{-1}^{1}\eta(\theta)$, with η strictly convex in $[-1, 1]$. This shows that $-|\operatorname{grad}\mathcal{E}|^2$ is a suitable entropy for equation (2.1). By the heuristic argument presented in the Section 2.2, we can therefore hope that using the minimizing movement technique, we gain a solution of the Burger's equation which locally decreases $-|\operatorname{grad}\mathcal{E}|^2$ fastest, which in turn implies that this solution is the *entropy solution.*

To prove the lack of semiconvexity we argue by contradiction. Consider the functions

$$\theta_\varepsilon(x) := \left\{\begin{array}{ll} 0 & \text{if } x < -\varepsilon, \\ \dfrac{1}{2} + \dfrac{x}{2\varepsilon} & \text{if } x \in [-\varepsilon, \varepsilon], \\ 1 & \text{if } x > \varepsilon. \end{array}\right\}$$

By formula (2.14) we have that $|\mathrm{grad}\,\mathcal{E}|(\theta_0, 1-\theta_0) = 0$, hence if $\mathcal{E}$ was K-convex for some $K \in \mathbb{R}$ we would have

$$E(\theta_\varepsilon) - E(\theta_0) \geq \frac{K}{2} D^2\big((\theta_\varepsilon, 1-\theta_\varepsilon), (\theta_0, 1-\theta_0)\big), \qquad \forall \varepsilon > 0. \tag{2.15}$$

A simple computation shows that

$$E(\theta_\varepsilon) - E(\theta_0) = -\frac{\varepsilon^2}{6}, \tag{2.16}$$

and that $W_2^2(\theta_\varepsilon, \theta_0) = O(\varepsilon^3)$ (because, roughly said, a mass of order ε is moved of a distance of order ε) and similarly $W_2^2(1-\theta_\varepsilon, 1-\theta_0) = O(\varepsilon^3)$, hence

$$d^2\big((\theta_\varepsilon, 1-\theta_\varepsilon), (\theta_0, 1-\theta_0)\big) = O(\varepsilon^3). \tag{2.17}$$

Finally, since D is the arc distance and d the chord distance of the embedding of $\mathcal{N}$ in $\mathcal{M}^+ \times \mathcal{M}^-$ it is natural to expect (we won't prove this - here we are just at an heuristic level) that it holds

$$\lim_{(\theta_n, \tilde\theta_n) \to (\theta, \tilde\theta)} \frac{D^2\Big((\theta_n, \tilde\theta_n), (\theta, \tilde\theta)\Big)}{d^2\Big((\theta_n, \tilde\theta_n), (\theta, \tilde\theta)\Big)} = 1. \tag{2.18}$$

Equations (2.16), (2.17) and (2.18) contradicts (2.15), so our argument is complete.

3 Rigorous result

In this section we formulate and prove the main result of this paper.
To state the result rigorously, we need to fix some notation. Fix $\theta_0 \in \mathcal{N}$ and $\tau > 0$. Define the discrete solution $\theta^\tau : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ by putting $\theta_0^\tau(x) := \theta_0(x)$, then defining $\theta_{(n+1)\tau}^\tau \in \mathcal{N}$ as the unique minimizer of

$$\int x(\theta(x) - H(x)) + \frac{1}{2\tau\lambda} W_2^2(\theta, \theta_{n\tau}^\tau) + \frac{1}{2\tau} W_2^2(1-\theta, 1-\theta_{n\tau}^\tau) \tag{3.1}$$

among all $\theta \in \mathcal{N}$, and finally interpolating piecewise constantly:

$$\theta_t^\tau(x) := \theta_{n\tau}^\tau(x), \qquad \forall x \in \mathbb{R},\ t \in (n\tau, (n+1)\tau).$$

Also, we define $V^\tau : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ by

$$V_t^\tau(x) := \int_{-\infty}^x \theta_t^\tau(x')dx'.$$

The main results of this paper are the following theorems:

Theorem 2 (Entropy solution)**.** *With the notation just described, as $\tau \downarrow 0$, the functions θ^τ converge weakly (in duality with continuous functions with compact support) to a limit function θ which is the entropy solution of* (2.1) *with initial condition θ_0.*

Theorem 3 (Viscosity solution)**.** *As $\tau \downarrow 0$, the functions V^τ converge locally uniformly to a limit function V which is the viscosity solution of*

$$\begin{cases} \partial_t V + \dfrac{\partial_x V(\partial_x V - 1)}{\partial_x V + \lambda^{-1}(1 - \partial_x V)} = 0 \quad \textit{in } \mathbb{R}^+ \times \mathbb{R}, \\ \\ V(0, x) = \displaystyle\int_{-\infty}^{x} \theta_0(z) dz. \end{cases} \tag{3.2}$$

Furthermore, the function V is 1-Lipschitz in x and $\frac{1}{4}$-Hölder continuous in t.

It is clear that the two theorems are equivalent, because the entropy solution of (2.1) corresponds to the viscosity solution of (3.2) (for instance, because both are obtained as viscous limit as $\varepsilon \downarrow 0$). Therefore we will focus on the proof of Theorem 3.
The hard part of this approach is in deriving appropriate informations for the single time step approximation. We do this work in the next section, after that, we will see that the passage to the limit (done in Section 3.2) will be straightforward.
We underline that in the time discretized problem (3.1) we used the 'chord distance' d and not the 'arc distance' D.

3.1 One step estimates

In this section we study the one-step minimization problem and derive all the properties which will lead to the proof of our main result. Recall that we write $\theta \in \mathcal{N}$ meaning $(\theta, 1-\theta) \in \mathcal{N}$.

Fix $\theta_0 \in \mathcal{N}$, $\tau > 0$ and consider the variational problem :

$$\text{minimize over } \theta \in \mathcal{N} \quad \int x(\theta(x) - H(x))\, dx + \frac{1}{2\tau\lambda} W_2^2(\theta, \theta_0) + \frac{1}{2\tau} W_2^2(1-\theta, 1-\theta_0). \tag{3.3}$$

In order to identify the Euler-Lagrange equation for the minimizer θ_1 of (3.3), we start from some heuristic arguments. Let V_0, W_0 be the cumulated volume functions of θ_0 and $1-\theta_0$ defined by

$$V_0(x) := \int_{-\infty}^{x} \theta_0(y) dy, \qquad W_0(x) := -\int_{x}^{\infty} 1 - \theta_0(y) dy,$$

and define similarly V_1, W_1 from θ_1.
Recalling the dual formulation of the transport problem, we know that (3.3) can be written as

$$\begin{aligned} \min_{\theta} \max_{(\phi,\tilde{\phi}),(\psi,\tilde{\psi})} & \int x(\theta(x) - H(x)) \\ & + \frac{1}{\tau\lambda}\left(\int_L^R (\frac{x^2}{2} - \phi(x))\theta_0(x) + \int_L^R (\frac{x^2}{2} - \tilde{\phi}(x))\theta(x)\right) \\ & + \frac{1}{\tau}\left(\int_L^R (\frac{x^2}{2} - \psi(x))(1-\theta_0(x)) + \int_L^R (\frac{x^2}{2} - \tilde{\psi}(x))(1-\theta(x))\right), \end{aligned} \tag{3.4}$$

for appropriate $L \ll 0$ and $R \gg 0$, where $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ are convex conjugate functions. Thus the minimum θ_1 is part of the solution $(\theta_1, (\phi_0, \phi_1), (\psi_0, \psi_1))$ of this latter saddle point problem. By the properties of the dual formulation of the optimal transport problem we know that $\Phi_0 := \partial_x \phi_0$ is the optimal transport map (because it is increasing) from θ_0 to θ_1. Similarly, putting $\Psi_0 := \partial_x \psi_0$, $\Phi_1 := \partial_x \phi_1$ and $\Psi_1 := \partial_x \psi_1$, we have that Ψ_0, Φ_1, Ψ_1 are the optimal transport maps from $1-\theta_0$ to $1-\theta_1$, from θ_1 to θ_0 and from $1-\theta_1$ to $1-\theta_0$ respectively. In particular, $V_1(x) = V_0(\Phi_1(x))$, $W_1(x) = W_0(\Psi_1(x))$ and it holds

$$V_0(\Phi_1(x)) + W_0(\Psi_1(x)) = x. \tag{3.5}$$

Also, the minimality in θ_1 gives (ignoring the constraint $\theta_1(x) \in [0,1]$, but respecting the volume constraint $\int(\theta_1 - H) = 0$):

$$\tau\lambda x + \frac{x^2}{2} - \phi_1(x) - \lambda\left(\frac{x^2}{2} - \psi_1(x)\right) = \text{const.}. \tag{3.6}$$

Differentiating gives

$$\begin{aligned} &\Phi_1(x) = (1-\lambda)x + \lambda\Psi_1(x) + \tau\lambda, \qquad \text{which is equivalent to} \\ &\Psi_1(x) = (1-\lambda^{-1})x + \lambda^{-1}\Phi_1(x) - \tau. \end{aligned} \tag{3.7}$$

Equations (3.5) and (3.7) are the Euler-Lagrange equations of θ_1. However, the process we used to find them is not rigorous, as the constraint $\theta_1(x) \in [0,1]$ does not allow to deduce (3.6), as in points x where $\theta_1(x) \in \{0,1\}$ only an inequality can be obtained.
To derive rigorously the Euler-Lagrange equations, it is better to proceed the other way around by constructing two functions satisfying (3.5) and (3.7), then to associate to them the corresponding density θ_1 and finally proving that θ_1 is indeed the solution of the variational problem (3.3).
We start constructing Φ_0 and Ψ_0. Recall the definition of $L(\theta), R(\theta)$ in (1.1).

Proposition 4. *For any $x_0 \in \mathbb{R}$ there exists a unique $\Phi_0(x_0)$ and a unique $\Psi_0(x_0)$ such that it holds*

$$\Phi_0(x_0) = V_0(x_0) + W_0((1-\lambda^{-1})\Phi_0(x_0) + \lambda^{-1}x_0 - \tau), \tag{3.8}$$
$$\Psi_0(x_0) = W_0(x_0) + V_0((1-\lambda)\Psi_0(x_0) + \lambda x_0 + \tau\lambda). \tag{3.9}$$

The functions Φ_0, Ψ_0 are both continuous, non decreasing and satisfy

$$\Phi_0(x_0) \in [x_0 - \tau\lambda, x_0], \qquad \Psi_0(x_0) \in [x_0, x_0+\tau], \qquad \forall x_0 \in \mathbb{R}, \tag{3.10}$$

and

$$\begin{aligned} \Phi_0(x_0) &= \begin{cases} x_0 - \tau\lambda, & \text{if } x_0 \le L(\theta_0), \\ x_0, & \text{if } x_0 \ge R(\theta_0) + \tau, \end{cases} \\ \Psi_0(x_0) &= \begin{cases} x_0, & \text{if } x_0 \le L(\theta_0) - \tau\lambda, \\ x_0 + \tau, & \text{if } x_0 \ge R(\theta_0). \end{cases} \end{aligned} \tag{3.11}$$

Also, it holds

$$\begin{aligned} \Phi_0(x_0) &= \Psi_0((1-\lambda^{-1})\Phi_0(x_0) + \lambda^{-1}x_0 - \tau), \\ \Psi_0(x_0) &= \Phi_0((1-\lambda)\Psi_0(x_0) + \lambda x_0 + \tau\lambda), \end{aligned} \tag{3.12}$$

Proof. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$F(x,y) := y - V_0(x) - W_0((1-\lambda^{-1})y + \lambda^{-1}x - \tau).$$

Then F is Lipschitz and from the fact that $0 \le \partial_x W_0 \le 1$ and $\lambda^{-1} > 0$ we get $\partial_y F(x,y) \ge \min\{1, \lambda^{-1}\} > 0$, which grants, for any $x_0 \in \mathbb{R}$, existence and uniqueness of $\Phi_0(x_0)$ such that $F(x_0, \Phi_0(x_0)) = 0$, which is (3.8). The continuity of Φ_0 follows from the one of F. Using the fact that V_0, W_0 are non decreasing, we deduce that $F(x,y)$ is non increasing in x, and therefore that Φ_0 is non decreasing. A direct computation based on the identity $V_0(x) + W_0(x) = x$ for any $x \in \mathbb{R}$ shows that $F(x_0, x_0) \ge 0$ and $F(x_0, x_0 - \tau\lambda) \le 0$ from which the first line in (3.10) follows. To prove the first part of (3.11) observe that $V_0(x) = 0, W_0(x) = x$ for $x \le L(\theta_0)$ and $V_0(x) = x, W_0(x) = 0$ for $x \ge R(\theta_0)$, then check directly that the stated values of Φ_0 fulfill the condition $F(x_0, \Phi_0(x_0)) = 0$ in the appropriate ranges for x_0.
The corresponding properties of Ψ_0 are proven analogously.
Thus it remains to prove (3.12). Consider the first of the two. By definition of Ψ_0, that is true if and only if it holds

$$\begin{aligned} \Phi_0(x_0) =& W_0((1-\lambda^{-1})\Phi_0(x_0) + \lambda^{-1}x_0 - \tau) \\ &+ V_0\big((1-\lambda)\Phi_0(x_0) + \lambda\big((1-\lambda^{-1})\Phi_0(x_0) + \lambda^{-1}x_0 - \tau\big) + \lambda\tau\big). \end{aligned}$$

Since the argument of V_0 in this expression is equal to x_0, the conclusion comes from (3.8). The second identity in (3.12) is proved analogously. □

Define the functions $\Phi_1, \Psi_1 : \mathbb{R} \to \mathbb{R}$ as the - potentially multivalued - inverses of Φ_0, Ψ_0 respectively:

$$\Phi_1(x_1) := \Big\{ x_0 \ : \ \Phi_0(x_0) = x_1 \Big\}, \qquad \Psi_1(x_1) := \Big\{ x_0 \ : \ \Psi_0(x_0) = x_1 \Big\}.$$

Proposition 5. *The functions Φ_1, Ψ_1 are maximally monotone functions, strictly increasing and satisfy*

$$\Phi_1(x_1) \subset [x_1, x_1 + \lambda\tau], \qquad \Psi_1(x_1) \subset [x_1 - \tau, x_1], \tag{3.13}$$

and

$$\begin{aligned} \Phi_1(x_1) &= \begin{cases} x_1 + \tau\lambda, & \text{if } x_1 \leq L(\theta_0) - \tau\lambda, \\ x_1, & \text{if } x_1 \geq R(\theta_0) + \tau, \end{cases} \\ \Psi_1(x_1) &= \begin{cases} x_1, & \text{if } x_1 \leq L(\theta_0) - \tau\lambda, \\ x_1 - \tau, & \text{if } x_1 \geq R(\theta_0) + \tau. \end{cases} \end{aligned} \tag{3.14}$$

Furthermore, it holds

$$\begin{aligned} \Psi_1(x_1) &= (1 - \lambda^{-1})x_1 + \lambda^{-1}\Phi_1(x_1) - \tau, \\ \Phi_1(x_1) &= (1 - \lambda)x_1 + \lambda\Psi_1(x_1) + \tau\lambda \end{aligned} \tag{3.15}$$

and

$$V_0(\Phi_1(z_1)) + W_0(\Psi_1(z_1)) = \{z_1\}. \tag{3.16}$$

Proof. All the properties are trivial consequences of the analogous ones proved in Proposition 4 for Φ_0, Ψ_0. □

¿From (3.16) we deduce that the equations

$$\{V_1(x_1)\} := V_0(\Phi_1(x_1)), \qquad \{W_1(x_1)\} := W_0(\Psi_1(x_1)), \tag{3.17}$$

define the pair of functions $V_1, W_1 : \mathbb{R} \to \mathbb{R}$ for which it holds

$$V_1(x_1) + W_1(x_1) = x_1, \qquad \forall x_1 \in \mathbb{R}. \tag{3.18}$$

By construction, V_1, W_1 are non decreasing, and thus (3.18) forces that both are Lipschitz with $\partial_x V_1, \partial_x W_1 \in [0, 1]$.
Let $\theta_1(x) := \partial_x V_1(x) \in [0, 1]$, notice that from (3.18) we get that $1 - \theta_1(x) = \partial_x W_1(x)$ and that from (3.17) we have

$$\theta_1 = (\Phi_0)_\sharp \theta_0, \qquad 1 - \theta_1 = (\Psi_0)_\sharp (1 - \theta_0). \tag{3.19}$$

Also, from (3.14) and the fact that $V_0(x) = 0$ for $x \ll 0$ and $W_0(x) = 0$ for $x \gg 0$ we deduce that $V_1(x) = 0$ for $x \ll 0$ and $W_1(x) = 0$ for $x \gg 0$ and thus that it holds

$$V_1(x) = \int_{-\infty}^{x} \theta_1(y)dy, \qquad W_1(x) = -\int_{x}^{+\infty} 1 - \theta_1(y)dy.$$

Notice also that by (3.14) and (3.17) we obtain

$$L(\theta_1) \geq L(\theta_0) - \lambda\tau, \qquad R(\theta_1) \leq R(\theta_0) + \tau. \tag{3.20}$$

Proposition 6. *θ_1 is the unique solution of the variational problem* (3.3).

Proof. It is sufficient to prove that for any $L \leq \min\{L(\theta_0), L(\theta_1)\}$, $R \geq \max\{R(\theta_0), R(\theta_1)\}$ and for any $\tilde{\theta} \in \mathcal{N}$ such that $L \leq L(\tilde{\theta})$, $R \geq R(\tilde{\theta})$ it holds

$$\int x(\theta_1(x) - H(x)) + \frac{1}{2\tau\lambda} W_2^2(\theta_1, \theta_0) + \frac{1}{2\tau} W_2^2((1-\theta_1), (1-\theta_0))$$
$$\leq \int x(\tilde{\theta}(x) - H(x)) + \frac{1}{2\tau\lambda} W_2^2(\tilde{\theta}, \theta_0) + \frac{1}{2\tau} W_2^2((1-\tilde{\theta}), (1-\theta_0)),$$

with equality if and only if $\tilde{\theta} = \theta_1$.
Define the functions ϕ_1, ψ_1 by

$$\phi_1(x) := \int_0^x \Phi_1(x')dx', \quad \psi_1(x) := \int_0^x \Psi_1(x')dx',$$

and notice that since Φ_1, Ψ_1 are increasing, ϕ_1, ψ_1 are convex. Let ϕ_0, ψ_0 be the respective Legendre transform. By construction, the couple of functions $\frac{1}{2}x^2 - \phi_0(x)$, $\frac{1}{2}x^2 - \phi_1(x)$ is admissible for the dual formulation of the transport problem, hence it holds

$$\frac{1}{2} W_2^2(\theta_0, \tilde{\theta}) \geq \int_L^R (\frac{1}{2}x^2 - \phi_0(x))\theta_0(x)dx + \int_L^R (\frac{1}{2}x^2 - \phi_1(x))\tilde{\theta}(x)dx. \tag{3.21}$$

For the same reason, it holds

$$\frac{1}{2} W_2^2(1-\theta_0, 1-\tilde{\theta}) \geq \int_L^R (\frac{1}{2}x^2 - \psi_0(x))(1-\theta_0)(x)dx + \int_L^R (\frac{1}{2}x^2 - \psi_1(x))(1-\tilde{\theta})(x)dx. \tag{3.22}$$

And therefore (noticing that $\int x(\theta(x) - H(x)) = \int_L^R x(\theta(x) - H(x))$):

$$\int x(\tilde{\theta}(x) - H(x)) + \frac{1}{2\tau\lambda} W_2^2(\tilde{\theta}, \theta_0) + \frac{1}{2\tau} W_2^2((1-\tilde{\theta}), (1-\theta_0))$$
$$\geq \int_L^R x(\tilde{\theta}(x) - H(x))dx$$
$$+ \frac{1}{\tau\lambda}\left(\int_L^R (\frac{1}{2}x^2 - \phi_0(x))\theta_0(x)dx + \int_L^R (\frac{1}{2}x^2 - \phi_1(x))\tilde{\theta}(x)dx\right)$$
$$+ \frac{1}{\tau}\left(\int_L^R (\frac{1}{2}x^2 - \psi_0(x))(1-\theta_0)(x)dx + \int_L^R (\frac{1}{2}x^2 - \psi_1(x))(1-\tilde{\theta})(x)dx\right).$$

Now we claim that the right hand side of this expression is independent on $\tilde{\theta}$. This is so because the part depending of $\tilde{\theta}$ is given by

$$\frac{1}{\tau\lambda} \int_L^R \tilde{\theta}\Big(\tau\lambda x + \frac{1}{2}x^2 - \phi_1(x) - \lambda\big(\frac{1}{2}x^2 - \psi_1(x)\big)\Big)dx,$$

and the quantity multiplying $\tilde\theta$ is, by (3.15), constant, so that the volume constraint $\int_L^R \tilde\theta - H = 0$ gives the claim.
Now notice that equality holds in (3.21) if and only if $(\Phi_0)_\sharp \theta_0 = \tilde\theta$ and that equality holds in (3.22) if and only if $(\Psi_0)_\sharp (1-\theta_0) = (1-\tilde\theta)$. Thus the conclusion comes from (3.19). □

After having identified the minimizer of the variational problem, our aim is to show that the function $V^\tau : [0,\tau] \times \mathbb{R}$ defined by

$$V^\tau(t,x) := \begin{cases} V_0(x), & \text{if } t \in [0,\tau), \\ V_1(x) & \text{if } t = \tau, \end{cases}$$

is a 'discrete viscosity solution' of (3.2). The rigorous statement is the following.

Proposition 7 (Discrete viscosity solution)**.** *Let $\varphi : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be a smooth function and define its time discretized $\varphi^\tau : [0,\tau] \times \mathbb{R} \to \mathbb{R}$ by*

$$\varphi^\tau(t,x) := \begin{cases} \varphi(0,x), & \text{if } t \in [0,\tau), \\ \varphi(\tau,x), & \text{if } t = \tau. \end{cases}$$

We claim that for any $x_1 \in \mathbb{R}$ it holds

$$\begin{aligned} V^\tau(t,x) - \varphi^\tau(t,x) \geq V^\tau(\tau,x_1) - \varphi^\tau(\tau,x_1), \qquad \forall t,x \in [0,\tau] \times \mathbb{R} \\ \Rightarrow \qquad (\partial_t \varphi + f(\partial_x \varphi))(\tau,x_1) \geq -\mathrm{Rem}(\tau,x_1,\varphi), \end{aligned} \tag{3.23}$$

and that

$$\begin{aligned} V^\tau(t,x) - \varphi^\tau(t,x) \leq V^\tau(\tau,x_1) - \varphi^\tau(\tau,x_1), \qquad \forall t,x \in [0,\tau] \times \mathbb{R} \\ \Rightarrow \qquad (\partial_t \varphi + f(\partial_x \varphi))(\tau,x_1) \leq \mathrm{Rem}(\tau,x_1,\varphi), \end{aligned} \tag{3.24}$$

where $f : \mathbb{R} \to \mathbb{R}$ is given by

$$f(z) := \frac{z(z-1)}{z + \lambda^{-1}(1-z)},$$

and the remainder term $\mathrm{Rem}(\tau,x_1,\varphi)$ is bounded by

$$\left|\mathrm{Rem}(\tau,x_1,\varphi)\right| \leq C\tau \sup\left\{ |\partial_{tt}\varphi(t,x)| + |\partial_{xx}\varphi(t,x)| + |\partial_t\partial_x\varphi(t,x)| \right\}, \tag{3.25}$$

for some universal constant C, the supremum being taken among all $t \in [0,\tau]$ and $x \in [x_1, x_1 + \tau\max\{\lambda,1\}]$.

Proof. Let $x_1 \in \mathbb{R}$ and $x_0 \in \Phi_1(x_1)$. Then by (3.17) we know that it holds

$$V_1(x_1) = V_0(x_0), \tag{3.26}$$

while the identity $V_0(x) + W_0(x) = x$ for any $x \in \mathbb{R}$ and (3.8) yields

$$\lambda^{-1}(x_1 - x_0) + \tau = V_0(x_0) - V_0((1-\lambda^{-1})x_1 + \lambda^{-1}x_0 - \tau). \tag{3.27}$$

With these two we prove (3.23). Notice that from (3.26) and the touching property we get

$$\begin{aligned} \partial_t\varphi(\tau, x_1) &= \frac{\varphi(\tau,x_1) - \varphi(0,x_1)}{\tau} + \mathrm{Rem}_1 \\ &= \frac{\varphi^\tau(\tau,x_1) - \varphi^\tau(0,x_0)}{\tau} + \frac{\varphi(0,x_0) - \varphi(0,x_1)}{\tau} + \mathrm{Rem}_1 \\ \text{(touching)} \quad &\geq \frac{V^\tau(\tau,x_1) - V^\tau(0,x_0)}{\tau} + \partial_x\varphi(\tau,x_1)\frac{x_0 - x_1}{\tau} + \mathrm{Rem}_1 + \mathrm{Rem}_2 \\ (3.26) \quad &= \partial_x\varphi(\tau,x_1)\frac{x_0-x_1}{\tau} + \mathrm{Rem}_1 + \mathrm{Rem}_2, \end{aligned} \tag{3.28}$$

where both Rem_1 and Rem_2 satisfy the bound (3.25).
Using (3.27) we also get

$$\begin{aligned} \partial_t\varphi(\tau, x_1) &= \frac{\varphi(\tau,x_1) - \varphi(0,x_1)}{\tau} + \mathrm{Rem}_1 \\ &= \frac{\varphi^\tau(\tau,x_1) - \varphi^\tau(0,(1-\lambda^{-1})x_1 + \lambda^{-1}x_0 - \tau)}{\tau} \\ &\quad + \frac{\varphi(0,(1-\lambda^{-1})x_1 + \lambda^{-1}x_0 - \tau) - \varphi(0,x_1)}{\tau} + \mathrm{Rem}_1 \\ \text{(touching)} \quad &\geq \frac{V^\tau(\tau,x_1) - V^\tau(0,(1-\lambda^{-1})x_1 + \lambda^{-1}x_0 - \tau)}{\tau} \\ &\quad + \partial_x\varphi(\tau,x_1)\left(\frac{x_0-x_1}{\tau\lambda} - 1\right) + \mathrm{Rem}_1 + \mathrm{Rem}_3 \\ (3.26) \quad &= \left(-1 + \frac{x_0 - x_1}{\tau\lambda}\right)(\partial_x\varphi(\tau,x_1) - 1) + \mathrm{Rem}_1 + \mathrm{Rem}_3, \end{aligned} \tag{3.29}$$

where Rem_3 also satisfies (3.25).
Now observe that the touching property and the fact that $\partial_x V \in [0,1]$ implies $\partial_x\varphi \in [0,1]$, and that (3.13) gives $\frac{x_0-x_1}{\tau} \in [0,\lambda]$. Hence (3.23) follows from (3.28) and (3.29) noticing that

$$\min_{a\in[0,\lambda]} \max\Big\{ab, \Big(1 - \frac{a}{\lambda}\Big)(1-b)\Big\} = -f(b), \qquad \forall b \in [0,1].$$

The implication (3.24) is proved following exactly the same lines. □

3.2 Passage to the limit

Here we conclude the proof of Theorem 3. It will be split in two propositions: in Proposition 10 we prove the compactness of $\{V^\tau\}$ w.r.t. local uniform convergence and the regularity of any limit function V, and in Proposition 11 that any limit V is indeed a viscosity solution of (3.2). This latter fact will also implies, thanks to uniqueness of viscosity solutions, that the limit function V is unique.

In order to get the time regularity of V^τ we start analyzing the one of θ^τ. In what follows we will write, for simplicity, $d(\theta,\tilde\theta)$ in place of $d((\theta,1-\theta),(\tilde\theta,1-\tilde\theta))$ for $\theta,\tilde\theta\in\mathcal{N}$, i.e.

$$d^2(\theta,\tilde\theta) := \frac{1}{\lambda}W_2^2(\theta,\tilde\theta)+W_2^2(1-\theta,1-\tilde\theta).$$

Proposition 8 (Compactness and time regularity at the level of θ's)**.** *The curves $t\mapsto\theta_t^\tau\in\mathcal{N}$ are relatively compact w.r.t. d and any limit curve $t\mapsto\theta_t$ is $\frac{1}{2}$-Hölder continuous.*

Proof. Let $L<R\in\mathbb{R}$. The subspace $\mathcal{N}_L^R$ of $\mathcal{N}$ of θ's such that $L(\theta)\geq L$ and $R(\theta)\leq R$ is compact w.r.t. weak convergence. Given that we are restricting the attention to the measures on the interval $[L,R]$, weak topology is equivalent to W_2 topology and thus also equivalent to d-topology.

Now fix $T>0$ and notice that (3.20) ensures that $L(\theta_t^\tau)\geq L(\theta_0)-\lambda T$ and $R(\theta_t^\tau)\leq R(\theta_0)+T$ for any $\tau>0$ and $t\in[0,T]$. Hence the set $\{\theta_t^\tau\}_{\tau>0,t\in[0,T]}$ is relatively compact in $\mathcal{N}$. Thus a standard diagonalization argument shows that for any sequence $\tau_k\to0$ there exists a subsequence, not relabeled, such that for every $t\in\mathbb{Q}\cap[0,T]$ the functions $\theta_t^{\tau_k}$ converge to some $\theta_t\in\mathcal{N}$ when $k\to\infty$. To conclude the proof it is sufficient to show that actually there is convergence for any t and that the limit curve is $\frac{1}{2}$-Hölder continuous. To this aim, notice that from the minimality of $\theta^\tau_{(n+1)\tau}$ we get

$$\frac{1}{2\tau}d^2(\theta^\tau_{(n+1)\tau},\theta^\tau_{n\tau})\leq E(\theta^\tau_{n\tau})-E(\theta^\tau_{(n+1)\tau}).$$

Therefore, fixing $t\geq s\geq0$ and adding up these inequalities for n which varies from $[s/\tau]$ to $[t/\tau]-1$ we obtain

$$\frac{1}{2\tau}\sum_{n=[s/\tau]}^{[t/\tau]-1}d^2(\theta^\tau_{(n+1)\tau},\theta^\tau_{n\tau})\leq E(\theta^\tau_{[s/\tau]\tau})-E(\theta^\tau_{[t/\tau]\tau})\leq C, \tag{3.30}$$

where $C:=\sup_{\mathcal{N}_L^R}E-\inf_{\mathcal{N}_L^R}E$, $L:=L(\theta_0)-\lambda T$, $R:=R(\theta_0)+T$, and $C<\infty$ because of continuity and compactness.

To conclude, fix $0\leq s\leq t\in\mathbb{Q}$ and let τ go to 0 in the following chain of inequalities:

$$
\begin{aligned}
d(\theta^\tau_{[t/\tau]\tau},\theta^\tau_{[s/\tau]\tau}) &\le \sum_{n=[s/\tau]}^{[t/\tau]-1} d(\theta^\tau_{(n+1)\tau},\theta^\tau_{n\tau})\\
&\le \left(\sum_{n=[s/\tau]}^{[t/\tau]-1} d^2(\theta^\tau_{(n+1)\tau},\theta^\tau_{n\tau})\right)^{1/2}\left(\left[\frac{t}{\tau}\right]-\left[\frac{s}{\tau}\right]\right)^{1/2}\\
&\overset{(3.30)}{\le} \sqrt{C\tau}\sqrt{\frac{t-s}{\tau}+1}\le\sqrt{C(t-s+\tau)}.
\end{aligned}
$$

□

Thus we have some limit function $\theta_t(x)$. Let $V_t(x) := \int_{-\infty}^x \theta_t(y)dy$. Observe that the W_2-convergence of the $\theta_{\tau_k}(t,\cdot)$ to $\theta(t,\cdot)$ implies the local uniform convergence of V_t^τ to V_t. In order to pass from the modulus of continuity of $t\mapsto\theta_t$ to the one of $t\mapsto V_t$ we will use the following simple lemma.

Lemma 9. *Let $V_1, V_2 : \mathbb{R}\to[0,\infty)$ be two 1-Lipschitz and non decreasing functions such that $V_1(x) = V_2(x) = x$ for $x \gg 0$. Then it holds:*

$$
|V_1(x)-V_2(x)| \le \left(\int_{\mathbb{R}} |V_1(z)-V_2(z)|dz\right)^{1/2} \qquad \forall x\in\mathbb{R} \tag{3.31}
$$

$$
\int_{\mathbb{R}} |V_1(x)-V_2(x)|dx = \int_{\{y>0\}} |V_1^{-1}(y)-V_2^{-1}(y)|dy, \tag{3.32}
$$

where V_i^{-1} is defined, as usual, as $V_i^{-1}(y) := \inf\{x : V(x)\ge y\}$, $i = 1,2$.

Proof. To prove (3.31) fix x, suppose that $V_1(x) \ge V_2(x)$ and note that since both functions are 1-Lipschitz and non decreasing the parallelogram whose vertices are $(x, V_2(x))$, $(x + V_1(x) - V_2(x), V_1(x))$, $(x, V_1(x))$, $(x + V_2(x) - V_1(x), V_2(x))$ is contained in the set of those couples (x', y) such that $V_2(x') \le y \le V_1(x')$.
The identity (3.32) is obvious. □

Proposition 10 (Compactness and time regularity at the level of V's)**.** *The set of curves $t\mapsto V_t^\tau$ is relatively compact w.r.t. uniform convergence and any limit function V_t is 1-Lipschitz w.r.t. the x variable and $\frac{1}{4}$-Hölder continuous w.r.t. the time variable t. Moreover if $\tau_n \downarrow 0$ is such that $\theta_t^{\tau_n}\to\theta_t$ weakly for any $t\ge0$ and $V_t^\tau\to V_t$ uniformly, then $V_t(x) = \int_{-\infty}^x \theta_t(y)dy$ for any $t\ge0$.*

Proof. It is well known - and trivial - that weak convergence of measures implies pointwise convergence of the cumulated distribution functions out of a countable set. In our case, the measures have density ≤ 1, hence the distribution functions are Lipschitz and thus weak convergence of the

densities implies local uniform convergence of the distribution functions. To pass from local uniform convergence to uniform convergence recall that for $\tilde{\theta} \in \mathcal{N}$ it holds $\tilde{V}(x) = 0$ for $x \leq L(\tilde{\theta})$ and $\tilde{V}(x) = x$ for $x \geq R(\tilde{\theta})$, where $\tilde{V}(x) := \int_{-\infty}^{x} \tilde{\theta}$ and that in our setting there is finite speed of propagation (see (3.10)). Hence the last statement is proved and relative compactness follows from the compactness claim in Proposition 8.
The fact that any limit function $V_t(x)$ is 1-Lipschitz in x follows from the fact that $\theta_t(x) \leq 1$ for any t, x, so we turn to the time regularity. In the following we will indicate by V_t^{-1} the inverse of the function $x \to V_t(x)$ and by $\|V_t - V_s\|_p$ the L^p norm (depending on t, s) of the function $x \to V_t(x) - V_s(x)$. Fix $\tau > 0$ and recall that for any $n, m \in \mathbb{N}$ it holds

$$W_2(\theta^\tau_{n\tau}, \theta^\tau_{m\tau}) = \|(V^\tau_{n\tau})^{-1} - (V^\tau_{m\tau})^{-1}\|_2. \tag{3.33}$$

Fix $T > 0$ and notice that arguing as in the proof of Proposition 8 we get that for any $t \in [0, T]$ and $\tau > 0$ it holds $V_t^\tau(x) = 0$ for $x \leq L(\theta_0) - \lambda T$ and $V_t^\tau(x) = x$ for $x \geq R(\theta_0) + T$. Hence the Hölder inequality gives

$$\|(V_t^\tau)^{-1} - (V_s^\tau)^{-1}\|_1 \leq \sqrt{R(\theta_0) - L(\theta_0) + (\lambda + 1)T}\, \|(V_t^\tau)^{-1} - (V_s^\tau)^{-1}\|_2, \tag{3.34}$$

for any $t \in [0, T]$ and $\tau > 0$.
Hence for $t \in [0, T]$, $\tau > 0$, $x \in \mathbb{R}$ and some generic $C \geq \sqrt{R(\theta_0) - L(\theta_0) + (\lambda + 1)T}$ it holds:

$$\begin{aligned}
|V_t^\tau(x) - V_s^\tau(x)| &\overset{(3.31)}{\leq} \left(\|V_t^\tau - V_s^\tau\|_1\right)^{1/2} \overset{(3.32)}{=} \left(\|(V_t^\tau)^{-1} - (V_s^\tau)^{-1}\|_1\right)^{1/2} \\
&\overset{(3.34)}{\leq} C\left(\|(V_t^\tau)^{-1} - (V_s^\tau)^{-1}\|_2\right)^{1/2} \leq C \left(\sum_{n=[s/\tau]}^{[t/\tau]-1} \left\|(V^\tau_{(n+1)\tau})^{-1} - (V^\tau_{n\tau})^{-1}\right\|_2 \right)^{1/2} \\
&\leq C \left(\left[\frac{t}{\tau}\right] - \left[\frac{s}{\tau}\right] \right)^{1/4} \left(\sum_{n=[s/\tau]}^{[t/\tau]-1} \left\|(V^\tau_{(n+1)\tau})^{-1} - (V^\tau_{n\tau})^{-1}\right\|_2^2 \right)^{1/4} \\
&\overset{(3.33)}{\leq} C \left(\frac{t - s + \tau}{\tau} \right)^{1/4} \left(\sum_{n=[s/\tau]}^{[t/\tau]-1} W_2^2(\theta^\tau_{(n+1)\tau}, \theta^\tau_{n\tau}) \right)^{1/4} \\
&\leq C \left(\frac{t - s + \tau}{\tau} \right)^{1/4} \left(\sum_{n=[s/\tau]}^{[t/\tau]-1} \lambda d^2(\theta^\tau_{(n+1)\tau}, \theta^\tau_{n\tau}) \right)^{1/4} \overset{(3.30)}{\leq} C\big(\lambda(t - s + \tau)\big)^{1/4},
\end{aligned}$$

from which the stated Hölder continuity of the limit function follows. □

Proposition 11 (Any limit V is a viscosity solution)**.** *Any limit function* $(t, x) \mapsto V_t(x)$ *produced by Proposition 10 is a viscosity solution of* (3.2)*. In particular, the limit* V *is unique.*

Proof. We already know by Proposition 10 that any limit V is continuous. Now let $(t,x) \mapsto \varphi_t(x)$ be a smooth function, and $t_0 > 0$, $x_0 \in \mathbb{R}$ arbitrary. We need to prove that

$$\begin{aligned} V_t(x) - \varphi_t(x) \geq V_{t_0}(x_0) - \varphi_{t_0}(x_0), \qquad \forall t \geq 0, x \in \mathbb{R} \\ \Rightarrow \qquad \partial_t \varphi_{t_0}(x_0) + f(\partial_x \varphi_{t_0}(x_0)) \geq 0, \end{aligned} \tag{3.35}$$

with $f(z) := \frac{z(z-1)}{z+\lambda^{-1}(1-z)}$. This will show that V is a viscosity supersolution. The proof of the subsolution property will follow along the same lines.
Up to modify a bit φ, it is not restrictive to assume that $V - \varphi$ has a strict minimum in (t_0, x_0). For $\tau > 0$ let $\varphi^\tau : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be the time discretization of φ defined by $\varphi^\tau_t(x) := \varphi(\tau[t/\tau], x)$ and notice that φ^τ converges to φ locally uniformly as $\tau \downarrow 0$. Now let $\tau_n \downarrow 0$ be such that V^{τ_n} converges uniformly to a limit V and observe that the local uniform convergence of $V^{\tau_n} - \varphi^{\tau_n}$ implies that for $n \gg 1$ the function $V^{\tau_n} - \varphi^{\tau_n}$ has a minimum (possibly not unique) in some point (t_n, x_n) such that $(t_n, x_n) \to (t_0, x_0)$ as $n \to \infty$. Therefore applying Proposition 7 (in particular the implication in (3.23)) we deduce

$$\partial_t \varphi_{t_n}(x_n) + f(\partial_x \varphi_{t_n}(x_n)) \geq -C\tau_n,$$

where C depends only on the behavior of φ in a neighborhood of (t_0, x_0). Letting $n \to \infty$ (3.35) is proved. □

4 Two examples on the maximal decrease of energy

4.1 Lack of maximal decrease

Let us underline once again that the proof of the existence of an entropy solution of the Burger's equation via the minimizing movement scheme, was motivated by the discussion made in section 1.2, but this motivation was just heuristic, and the rigorous result is actually independent on that.
In this appendix, we want to produce an example of a C^1 function on $\mathbb{R}^2$ for which it does *not* exist a gradient flow trajectory that ' locally decreases the energy fastest'. This means, in particular, that the minimizing movement approach cannot produce such a curve.
For $n \in \mathbb{N}$ consider the function $g_n : [0, +\infty) \to \mathbb{R}$ defined by

$$g_n(x) := \begin{cases} -nx^{\frac{3}{2}} & \text{if } x \leq n^{-7}, \\ -n^{-\frac{19}{2}} & \text{if } x \geq n^{-7}. \end{cases}$$

Now let, for any $n \in \mathbb{N}$, $\tilde{g}_n : [0, +\infty) \to \mathbb{R}$ be a regularization of g_n near the point $x = n^{-7}$, so that $\tilde{g}_n \in C^\infty(0, +\infty)$, g_n is convex and decreasing around

n^{-7} and $\tilde{g}_n = g_n$ far from n^{-7}. Let $\chi : [0,1] \to [0,1]$ be a C^∞ function such that $\chi(0) = 0$, $\chi(1) = 1$ and such that its derivatives of any order are 0 in 0 and 1.

We will define our C^1 function $f : \mathbb{R}^2 \to \mathbb{R}$ by using polar coordinates. For better clarity in the formula, we put $a_n := \frac{1}{n}$, $b_n := \frac{a_n + a_{n+1}}{2}$. Let

$$f(r,\theta) := \begin{cases} 0 & \text{if } \theta \in [\pi/2, 2\pi], \\ \left(1 - \chi\left(\dfrac{\theta - 1}{\pi/2 - 1}\right)\right) \tilde{g}_1(r) & \text{if } \theta \in [1, \pi/2], \\ \chi\left(\dfrac{\theta - b_n}{a_n - b_n}\right) \tilde{g}_n(r) & \text{if } \theta \in [b_n, a_n], \\ \left(1 - \chi\left(\dfrac{\theta - a_{n+1}}{b_n - a_{n+1}}\right)\right) \tilde{g}_{n+1}(r) & \text{if } \theta \in [a_{n+1}, b_n], \end{cases}$$

so that on the half-lines $\{(r, a_n)\}_{r \geq 0}$ the function f is given by

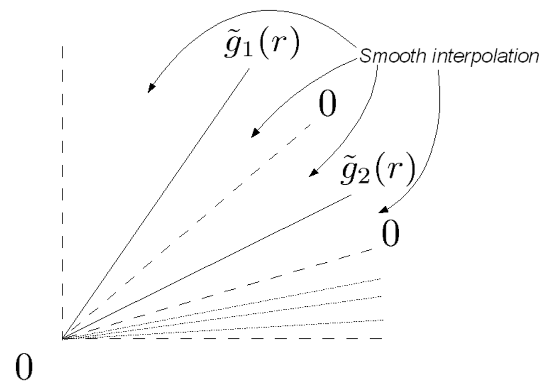

Figure 1: Schematic description of the behavior of f

$$f(r, a_n) = \tilde{g}_n(r),$$

on the 'bisector' half-lines $\{(r, b_n)\}_{r \geq 0}$ its value is 0, and between these lines we have a smooth approximation.

By definition, it is clear that f is continuous on the whole $\mathbb{R}^2$, and that is C^∞ on $\mathbb{R}^2 \setminus \{(r,0)\}_{r\geq 0}$, so that we only have to check that f is continuously differentiable on the positive axis $\{(r,0)\}_{r\geq 0}$. To this aim observe that

$$\begin{aligned} \left| \frac{d}{dr} f(r,\theta) \right| &\leq |\tilde{g}_n'(r)|, \\ \left| \frac{1}{r} \frac{d}{d\theta} f(r,\theta) \right| &\leq \max_{x \in [0,1]} \{|\chi'(x)|\} \frac{|\tilde{g}_n(r)|}{r|a_n - b_n|}, \end{aligned}$$

for $\theta \in [b_{n-1}, b_n]$. For the derivative w.r.t. r we have

$$|\tilde{g}_n'(r)| \leq |g_n'^{-}(n^{-7})| = \tfrac{3}{2} n^{-\frac{5}{2}} \to 0,$$

as $n \to \infty$ uniformly on r, where $g_n'^{-}$ is the left derivative of g_n. To analyze the derivative w.r.t. θ we distinguish the case r far from 0 and r close to 0. In the first case, say $r > \delta > 0$, from the fact that $|a_n - b_n| = O(n^{-2})$ we have, for n sufficiently large, the bound

$$\frac{|\tilde{g}_n(r)|}{r|a_n - b_n|} \leq \frac{n^{-\frac{19}{2}}}{\delta O(n^{-2})} = O(n^{-\frac{15}{2}}) \to 0.$$

For the case r close to 0 from the concavity of g_n on $[0, n^{-7}]$ we have

$$\frac{|\tilde{g}_n(r)|}{r} \leq \frac{|g_n(n^{-7})|}{n^{-7}}$$

and therefore

$$\frac{|\tilde{g}_n(r)|}{r|a_n - b_n|} \leq \frac{|g_n(n^{-7})|}{n^{-7} O(n^{-2})} = n^{-\frac{1}{2}} \to 0.$$

Thus $f \in C^1(\mathbb{R}^2)$.
Now consider the gradient flow equation

$$\begin{cases} u(0) = 0, & \\ u'(t) = -\nabla f(u(t)), & t \geq 0. \end{cases} \tag{4.1}$$

We claim that there is no solution $\overline{u}(t)$ of this equation such that for any other solution $u(t)$ it holds $f(\overline{u}(t)) \leq f(u(t))$ on some interval of the kind $[0, \delta]$, for some $\delta > 0$. That is, there is no solution that locally decreases the energy fastest near 0.
In order to prove this, consider a generic solution $u(t) = (r(t), \theta(t))$ of (4.1). We may assume without loss of generality that $r(t) > 0$ for $t > 0$, so that $\theta(t)$ is well defined and continuous for positive times. We claim that it holds $\theta(t) \neq b_n$ for any $t > 0$, $n \in \mathbb{N}$. Indeed, if equality holds for some t_0, n_0, then, since f is C^∞ around $u(t_0) = (r(t_0), \theta(t_0))$ and $\nabla f(u(t_0)) = 0$, we must have $u(t) = u(t_0)$ for any $t \in [0, +\infty)$, which contradicts the initial condition $u(0) = 0$. Therefore, by continuity, we know that it must hold

$$\theta(t) \in (b_{n-1}, b_n), \qquad \forall t \in (0, +\infty), \tag{4.2}$$

for some fixed $n \in \mathbb{N}$. Now, it is not hard to see (we omit the details) that among all solutions of (4.1) satisfying (4.2), there is one which decreases the energy fastest: the curve $u_n(t) = (r_n(t), a_n)$, where $r_n(t)$ is the unique solution of

$$\begin{cases} r_n(0) & = 0, & \\ r_n'(t) & = -\tilde{g}_n'(u(t)), & t \geq 0, \\ r_n(t) & > 0 & t > 0. \end{cases} \tag{4.3}$$

Therefore, if a solution of (4.1) which decreases the energy fastest exists, it must be equal to $u_n(t)$ for some $n \in \mathbb{N}$. However, it is not hard to check that for $n < m$ it holds $f(u_m(t)) < f(u_n(t))$ in some interval of the kind $(0, \delta_m)$ (see the figures describing the graph of g_n - notice that $\delta_m \downarrow 0$ as $m \to \infty$). Therefore a solution which locally decreases the energy fastest does not exist.

4.2 Convergence of minimizing movements to a limit curve which does not decrease locally the energy fastest

In this final section we give an explicit example of C^1 function on $\mathbb{R}^2$ such that: for some point (x_0, y_0) there exists more than one gradient flow trajectory of f starting from x_0, one - and only one - of these trajectories decreases the energy fastest, and the minimizing movement scheme converges to a unique limit curve which is *not* the one decreasing the energy fastest.

Let $\gamma_1, \gamma_2 : [0, +\infty) \to \mathbb{R}^2$ be the curves defined by

$$\gamma_1(t) := (e^{-t+1}, 2-t), \qquad \forall t \geq 0,$$

$$\gamma_2(t) := \begin{cases} (e^{-t+1}, 2-t) & \text{if } t \leq 1 \\ (2 - t - \frac{(t-1)^2}{2}, 2 - t + \frac{(t-1)^2}{2}) & \text{if } t \geq 1, \end{cases}$$

(see also the Figure 2)

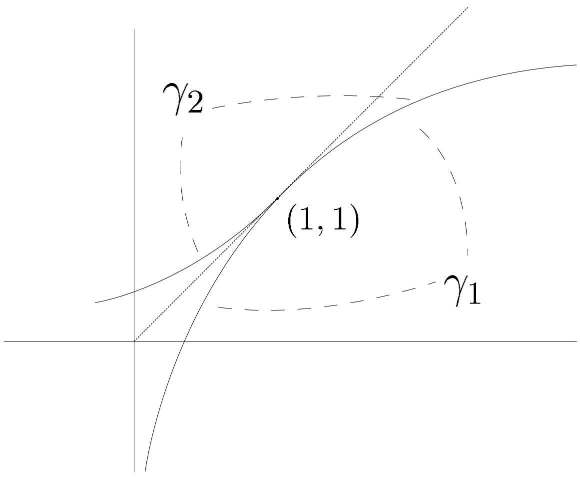

Figure 2: Trajectories of γ_1 and γ_2.

and $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$ the functions

$$f_1(x, y) := \frac{x^2}{2} + y,$$

$$f_2(x, y) := x + y - \frac{1}{2} - \frac{2}{3}|y - x|^{\frac{3}{2}}.$$

Finally, let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by:

$$f(x,y) := \begin{cases} f_1(x,y) & \text{if } x \geq y, \\ f_2(x,y) & \text{if } (x,y) = \gamma_2(t), \text{ for some } t > 1, \\ f_3(x,y) & \text{otherwise,} \end{cases}$$

as in the Figure 3, where f_3 is chosen so that $f_3 \geq f_2$ on $\mathbb{R}^2$, f is C^1 in the whole $\mathbb{R}^2$ and C^2 in $\mathbb{R}^2 \setminus \{(1,1)\}$. Observe that such a choice of f_3 is possible because

$$f_1(1,1) = f_2(1,1) \qquad\qquad \nabla f_1(1,1) = \nabla f_2(1,1),$$

and $f_1 \geq f_2$ on the diagonal $\{x = y\}$.

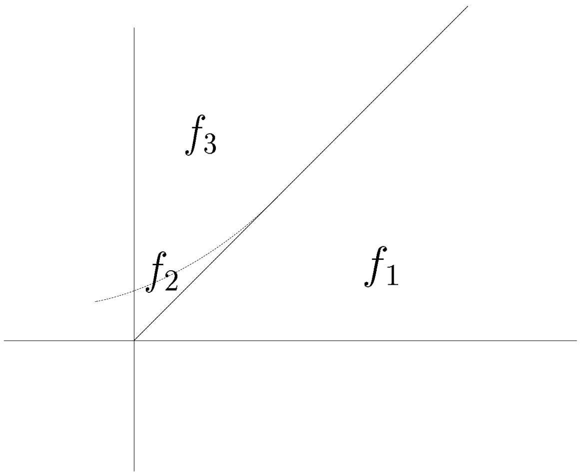

Figure 3: Schematic behavior of f.

We claim that:

A) Both γ_1 and γ_2 are gradient flow trajectories of f starting from $(e,2)$,

B) Among all gradient flows trajectories of f starting from $(e,2)$, γ_2 is the only one which locally decreases the energy fastest,

C) The minimizing movements scheme starting from $(e,2)$ converges to γ_1.

If we prove these, our example is concluded.
(A) is trivial. For **(B)** start observing that all the gradient flow trajectories (g.t.f. in the following) starting from $(e,2)$ coincide with γ_1 and γ_2 for $t \leq 1$, because f is C^2 in $\mathbb{R}^2 \setminus \{(1,1)\}$. For the same reason, it is enough to check that γ_2 is the g.f.t. which decreases the energy fastest at time $t = 1$. Observe that $\nabla f(x,x) = \nabla f_1(x,x)$ and that for $x < 1$ we have

$(\partial_x f_1)(x,x) < (\partial_y f_1)(x,x)$, therefore for every g.f.t. γ of f starting from $(1,1)$ there exists $T>0$ such that either $\{\gamma(t)\}_{0<t<T}$ stays below the diagonal or it stays above (because the diagonal can be crossed only from above to below, thus there exists at most one time $T>0$ such that $\gamma(T)\in\{(x,x)\}$). The class of g.f.t. starting from $(1,1)$ which initially stay below the diagonal contains only one element: the curve $t\mapsto\gamma_1(t+1)$. The class of g.f.t. which initially stay above the diagonal contains infinitely many: we claim that the curve $t\mapsto\gamma_2(t+1)$ is the one that decreases the energy fastest within this class. Consider the gradient flows of f_2 starting from $(1,1)$: it is easy to check that the curve $t\mapsto\gamma_2(t+1)$ is one of them, and that it is actually the one which decreases f_2 the fastest. Since $f_3\geq f_2$ in the set $\{y\geq x\}$, $t\mapsto\gamma_2(t+1)$ decreases the energy f fastest among all the gradient flows starting from $(1,1)$. Thus, to conclude the proof of **(B)** it is enough to compare the energies along γ_1 and γ_2.
Simple calculations show that:

$$\begin{aligned} f_1\big(\gamma_1(1+t)\big) &= \frac{e^{-2t}}{2}+1-t=\frac{3}{2}-2t+2t^2+O(t^3),\\ f_2\big(\gamma_2(1+t)\big) &= \frac{3}{2}-2t-\frac{2}{3}t^3, \end{aligned}$$

so that $f\big(\gamma_2(1+t)\big)=f_2\big(\gamma_2(1+t)\big)<f_1\big(\gamma_1(1+t)\big)=f\big(\gamma_1(1+t)\big)$ for small $t>0$.
Now we pass to **(C)**. If we apply the minimizing movements approach to the function f_1 starting from $(e,2)$, we get the sequence of discrete solutions (x_τ^n,y_τ^n) given by

$$\begin{aligned} x_\tau^n &:= \frac{e}{(1+\tau)^n},\\ y_\tau^n &:= 2-n\tau, \end{aligned}$$

and it is obvious that as $\tau\downarrow 0$ the points $(x_\tau^{[t/\tau]},y_\tau^{[t/\tau]})$ converge to $\gamma_1(t)$ (here $[\cdot]$ stands for the integer part). Therefore it is enough to prove that the minimizing movements scheme for f starting from $(e,2)$ never selects a point lying above the diagonal. To prove this, we use the following lemma:

Lemma 12. *There exists constants $a,c>0$ such that for every $0<\tau<a$ and $n\in\mathbb{N}$ it holds*

$$\frac{e}{(1+\tau)^n}-\Big(2-n\tau\Big)>c\tau.$$

The next proposition contains the proof of **(C)**:

Proposition 13. *The minimizing movements scheme for f starting from $(e,2)$ converges to γ_1.*

Proof. We know that up to passing to a sequence $\tau_k \downarrow 0$ the minimizing movements scheme converges to a gradient flow for f starting from $(e, 2)$. Observe that it is sufficient to show the convergence to γ_1 in the region $\{x > 0, y > 1/\sqrt{2}\}$, because outside such region f is C^2 and thus any gradient flow coinciding with γ_1 inside the region must actually coincide with γ everywhere.
Let a, c be as in the previous lemma and notice that thanks to such lemma it is sufficient to show that for $(x_0, y_0) \in \mathbb{R}^2$ such that

$$\begin{aligned} x_0 - y_0 &> c\tau, \\ x_0 &> 0, \\ y_0 &> \frac{1}{\sqrt{2}} \\ 0 < \tau &< \min\left\{a, c, \frac{1}{2\sqrt{2}}\right\} \end{aligned} \tag{4.4}$$

there is no minizer of

$$(x, y) \mapsto f(x, y) + \frac{|x - x_0|^2 + |y - y_0|^2}{2\tau}, \tag{4.5}$$

such that $x \leq y$. Indeed, once this is proved, an easy induction argument based on Lemma 12 shows that the minimizing movements scheme starting from $(e, 2)$ and with τ sufficiently small stays below the diagonal at least until it goes out the region $\{x > 0, y > 1/\sqrt{2}\}$, which, as said, is sufficient to conclude.
Thus assume that x_0, y_0, τ satisfy (4.4). In the set $\{x \leq y\}$ we have $f_2 \leq f_3 = f$, and therefore

$$\inf_{x \leq y} f(x, y) + \frac{|x - x_0|^2 + |y - y_0|^2}{2\tau} \geq \inf_{x \leq y} f_2(x, y) + \frac{|x - x_0|^2 + |y - y_0|^2}{2\tau}.$$

Thus to prove that the minimizers of (4.5) stay below the diagonal is sufficient to prove that

$$\inf_{x \leq y} f_2(x, y) + \frac{|x - x_0|^2 + |y - y_0|^2}{2\tau} > \inf_{x \geq y} f_1(x, y) + \frac{|x - x_0|^2 + |y - y_0|^2}{2\tau}. \tag{4.6}$$

Let us compute the value of the infimum in the left hand side of the above expression. It is immediate to verify that a minimizing point (x', y') for the expression exists: either it lies in the set $\{x < y\}$ or it lies along the diagonal. Assume it is in $\{x < y\}$, then by explicit computation of ∇f_2 we would have

$$\begin{aligned} \frac{x' - x_0}{\tau} &= -1 - \sqrt{y' - x'} \\ \frac{y' - y_0}{\tau} &= -1 + \sqrt{y' - x'}, \end{aligned}$$

putting $z_0 := x_0 - y_0$ and $z' := x' - y'$, subtracting the two equalities above and taking the squares we get

$$z'^2 + z'(4\tau^2 - 2z_0) + z_0^2 = 0.$$

In order for this equation to admit a solution z' it must hold

$$\tau^4 - z_0\tau^2 > 0,$$

which is impossible by the assumptions $\tau < c$ (because $z_0 = x_0 - y_0 > c\tau > \tau^2$).

Thus the minimum of the LHS of (4.6) lies on the diagonal. It is then obvious that it is given by $(x', y') = (\frac{x_0+y_0}{2}, \frac{x_0+y_0}{2})$. Simple calculations show that (4.6) is then equivalent to

$$x_0 + y_0 - \frac{1}{2} + \frac{1}{\tau}\left(\frac{x_0 - y_0}{2}\right)^2 > \frac{x_0^2}{2(1+\tau)} + y_0 - \frac{\tau}{2},$$

which can be rewritten as

$$x_0^2\left(\frac{1}{4} - \frac{\tau}{2(1+\tau)}\right) + x_0\left(\frac{y_0}{2} - \tau\right) + \frac{y_0^2}{4} + \frac{\tau^2}{2} - \frac{\tau}{2} > 0,$$

which is always true for x_0, y_0, τ as in the hypothesis. □

We conclude with the proof of Lemma 12.

Proof of Lemma 12 We will prove that there exists constants $a, c > 0$ such that for every $\tau, x > 0$ with $\tau < a$ it holds

$$\frac{e}{(1+\tau)^x} - \left(2 - x\tau\right) > c\tau. \tag{4.7}$$

We distinguish two casee: either $x \leq \frac{1}{\tau}$ or $x \geq \frac{1}{\tau}$. Assume that $x \leq \frac{1}{\tau}$ and observe that for $g(\tau) := (1+\tau)^{1/\tau}$ it holds

$$g(\tau) = e - e\tau + o(\tau),$$

therefore for $a > 0$ sufficiently small and some $c_0 > 0$ it holds

$$g(\tau) < e - c_0\tau, \qquad \forall \tau \leq a,$$

which gives

$$\frac{e}{(1+\tau)^{\frac{1}{\tau}}} > 1 + c_0\frac{\tau}{(1+\tau)^{1/\tau}} > 1 + \tau\frac{c_0}{e} = 1 + \tau c_1.$$

Now write $x = \frac{1}{\tau} - y$ and observe that

$$\begin{aligned}\frac{e}{(1+\tau)^x} &= \frac{e}{(1+\tau)^{1/\tau}}(1+\tau)^y > (1+\tau c_1)(1+\tau)^y \\ &> (1+\tau c_1)(1+y\tau) > 1 + \tau y + \tau c_1 = (2 - x\tau) + c_1\tau.\end{aligned}$$

Now assume that $x \le \frac{1}{\tau}$. Clearly it is sufficient to prove that

$$\frac{e}{(1+\tau)^x} - \Big(2 - x\tau\Big) > \frac{e}{(1+\tau)^{1/\tau}} - 1.$$

Put $x = \frac{1}{\tau} + y$ to reduce the previous inequality to

$$\frac{e}{(1+\tau)^{1/\tau}} \left(1 - \frac{1}{(1+\tau)^y}\right) < \tau y.$$

Recall that $e < (1+\tau)^{1+1/\tau}$ for every $\tau > 0$ so that $\frac{e}{(1+\tau)^{1/\tau}} < 1+\tau$. Plugging this bound in the inequality above and making simple manipulations we reduce our claim to

$$(1+\tau)^{1-y} > 1 + \tau(1-y),$$

which is true for any $\tau, y > 0$. □

We thank E. Esselborn for a careful reading of the manuscript.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, Gradient flows in metric spaces and in spaces of probability measures, Birkhäuser, 2005.

[2] C. Dafermos, The entropy rate admissibility criterion for solutions of hyperbolic conservation laws, *J. Differential Equations* **14** (1973), 202-212.

[3] S. N. Kruzkov, First order quasilinear equations in several independent variables, *Math. Sb.* **123** (1970), 228-255; English translation in *Math. USSR Sbornik* **10** (1970), 217-243

[4] F. Otto, Evolution of microstructure in unstable porous media flow: a relaxational approach, *Comm. Pure Appl. Math.* **52** (1999), no. 7, 873–915.

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