

Appendix: On the convergence of augmented Lagrangian method for optimal transport between nonnegative densities

Romain Hug^{1,4,5}, Emmanuel Maitre² & Nicolas Papadakis³

¹Univ. Grenoble Alpes, CNRS, LJK, 38000 Grenoble, France

²Grenoble Institute of Technology, LJK, 38000 Grenoble, France

³Univ. Bordeaux, IMB, CNRS, UMR 5251, F-33400 Talence, France.

⁴Johannes Kepler Univ., RICAM, 4040 Linz, Austria.

⁵Univ. Artois, LML, 62307 Lens, France

A Burgers Equation and regularity results on the velocity field of dynamic optimal transport

The aim of this first section is to give a sketch of proof for Proposition 4.1 (Burgers equation is satisfied by the velocity field $v = v_\phi$) and Proposition 6.2 (uniform control of the gradients of fields $v = v_\phi$ and $v_\gamma = v_{\gamma\phi}$) of our paper *On the convergence of augmented Lagrangian method for optimal transport between nonnegative densities* [6].

With Proposition 6.2, we can state **important regularity results** presented in subsection A.3 below in Corollary 1 and Corollary 2 (Remark 6.2 in the main paper): $\nabla_{t,x}v \in L^p(0,1; L^q_{loc}(\mathbb{R}^d))$ for all $1 < p \leq +\infty$ and $1 \leq q < +\infty$ such that $1/p + 1/q > 1$, and especially $v \in W^{1,p}_{loc}([0,1] \times \mathbb{R}^d)$ (i.e. $\forall \Omega \subset \mathbb{R}^d$, $v \in W^{1,p}((0,1) \times \Omega)$) for all $1 \leq p < 2$.

We first recall Hypothesis Γ_1 .

Property (Γ_1). ϕ and ϕ^* are convex, continuous and achieve a minimum on \mathbb{R}^d .

The statements of the two results we prove in this appendix are the following.

Proposition 1 (Proposition 4.1 in [6]). *With the property (Γ_1), v satisfies the Burger's equation, that is to say, in the distribution sense:*

$$\partial_t v + \frac{1}{2} \nabla_x |v|^2 = 0, \tag{A-1}$$

which is a generalized form of $\partial_t v + v \cdot \nabla_x v = 0$.

Proposition 2 (Proposition 6.2 in [6]). *We assume that ϕ satisfies the property (Γ_1). Let $R' > R > 0$ and $a \in \mathbb{R}^d$ such that $\phi(a) = \inf_{\mathbb{R}^d} \phi$ and let $M = \sup_{x \in B(a, 2(R+|a|))} |\partial \phi(x)|$.*

Then there exists an constant $C > 0$ – independent of ϕ , γ , a , R and R' , such that for all $t_0 \in (0, 1)$ satisfying the condition $t_0 < \min\{1/2, (R' - R)/(M + 2|a|)\}$, and by setting $v_0 = v$, we have the property:

$$\forall \gamma \geq 0, \forall t \in (0, t_0], \int_{B(a, R)} |\nabla_x v_\gamma(t, x)| dx \leq \frac{C}{t_0(1-t_0)} \mathcal{L}^d(B(a, R')). \tag{A-2}$$

Thus $\nabla_x v_\gamma \in L^\infty(0, t_0; L^1(B(a, R)))$, for all $\gamma \geq 0$.

In subsection A.1, we first consider an "ideal" framework, i.e. without breaks. For this purpose, we assume that the potential ϕ is regular and satisfies the following property (Γ_2).

Property (Γ_2). ϕ satisfies (Γ_1), is of class C^1 , and $\nabla_x \phi$ is Lipschitz (i.e. $\phi \in C^{1,1}(\mathbb{R}^d)$).

The propositions are first established under the assumption (Γ_2) and then extended to the general framework for ϕ satisfying the property (Γ_1). To that end, a regularization of the potential ϕ with Moreau envelope is considered in the subsection A.2. The final proofs (by compilation of previous results) of Propositions 1 and 2 are stated in subsection A.3. In subsection A.4, we give additional results concerning the regularity of the velocity field. These complementary results are not directly necessary for the proofs of uniqueness of the main paper.

A.1 Regular case (Γ_2)

We assume that ϕ satisfies the hypothesis (Γ_2). In this "ideal" case (without breaks), we want to show that the velocity field v satisfies the Burgers equation in the sense of distributions (A-1), as well as the control of the gradient of the velocity field: $\nabla_{t,x} v \in L_{loc}^\infty([0,1], L^1(\mathbb{R}^d))$.

The operator \mathbf{p}_ϕ can be interpreted as a spatial "inverse" of the operator $X(t, \cdot) = \nabla_x(\phi_t) = (1-t)\text{id} + t\nabla\phi$. It is indeed invertible when ϕ satisfies the hypothesis (Γ_2). In other words, we have

$$y = \mathbf{p}_\phi(t, x) \Leftrightarrow x = (1-t)y + t\nabla\phi(y). \quad (\text{A-3})$$

As a consequence, the velocity v can be defined from \mathbf{p}_ϕ by :

$$v(t, x) = \nabla\phi(\mathbf{p}_\phi(t, x)) - \mathbf{p}_\phi(t, x) = \frac{x - \mathbf{p}_\phi(t, x)}{t}, \quad (\text{A-4})$$

and can be continuously extended on $[0,1] \times \mathbb{R}^d$ (i.e. in $t=0$): $v(0, \cdot) = \nabla\phi - \text{id}$.

Remark 1. We note that $\phi_t = (1/2)(1-t)|\cdot|^2 + t\phi$ is of class C^1 , strictly convex and superlinear, since ϕ is convex and $|\cdot|^2$ is strictly convex and superlinear. Thus, for all $t \in (0,1)$, $X(t, \cdot) = (1-t)\text{id} + t\nabla\phi = \nabla\phi_t$ is bijective of inverse $\mathbf{p}_\phi(t, \cdot) = \nabla_x(\phi_t)^*$.

Proposition 3. Under the property (Γ_2), v satisfies the Burgers equation (A-1) in the sense of distributions.

Sketch of the proof: If the potential ϕ is of class C^1 , then by differentiating the advection relation $\nabla_x\phi - \text{id} = \partial_t\nabla_x\phi_t = v(t, \nabla_x\phi_t)$, we obtain (see Remark 2):

$$0 = \partial_{tt}\nabla_x\phi_t = (\partial_tv + v \cdot \nabla_x v)(\nabla_x\phi_t).$$

□

In order to prove the second regularity result $\nabla_{t,x} v \in L_{loc}^\infty([0,1], L^1(\mathbb{R}^d))$, we now present intermediate results on the potential ϕ .

Proposition 4. We assume that ϕ satisfies the property (Γ_2). Let $R' > R > 0$ and $a \in \mathbb{R}^d$ such that $\phi(a) = \inf_{\mathbb{R}^d} \phi$. Then, for all for all $t \in [0, t_0]$, there exists $t_0 \in (0,1)$ such that for all $t \in [0, t_0]$, with

$$t_0 < \min \left\{ \frac{1}{2}, \frac{R' - R}{M + 2|a|} \right\} \quad \text{for} \quad M = \sup_{x \in B(a, 2(R+|a|))} |\partial\phi(x)|, \quad (\text{A-5})$$

we have $\mathbf{p}_\phi(t, B(a, R)) \subset \mathbf{p}_\phi(t_0, B(a, R'))$.

Sketch of the proof: Let $x \in B(a, R)$ and $t, t_0 \in [0, 1)$ such that $t_0 > 0$ and $t \in [0, t_0]$. We have $\mathbf{p}_\phi(t, x) \in \mathbf{p}_\phi(t_0, B(a, R'))$ if and only if there exists $y \in B(a, R')$ such that $\mathbf{p}_\phi(t_0, y) = \mathbf{p}_\phi(t, x)$. According to the definition of \mathbf{p}_ϕ (by the equivalence (A-3)), for $y \in \mathbb{R}^d$ we have:

$$\mathbf{p}_\phi(t_0, y) = \mathbf{p}_\phi(t, x) \Leftrightarrow y = (1 - t_0)\mathbf{p}_\phi(t, x) + t_0\nabla\phi(\mathbf{p}_\phi(t, x)). \quad (\text{A-6})$$

Let us take $y = (1 - t_0)\mathbf{p}_\phi(t, x) + t_0\nabla\phi(\mathbf{p}_\phi(t, x))$ and look for a sufficient condition on t_0 for $y \in B(a, R')$. We have $\mathbf{p}_\phi(t, (1 - t)a) = a$ and then $|\mathbf{p}_\phi(t, x) - a| \leq (R + |a|)/(1 - t_0)$. By taking $t_0 \leq 1/2$, we thus have $\mathbf{p}_\phi(t, x) \in B(a, 2(R + |a|))$.

For all $t \in [0, t_0]$, for $x \in B(a, R)$ and $y = (1 - t_0)\mathbf{p}_\phi(t, x) + t_0\nabla\phi(\mathbf{p}_\phi(t, x))$, we get

$$|y - a| \leq (1 - t_0)|\mathbf{p}_\phi(t, x) - a| + t_0|\nabla\phi(\mathbf{p}_\phi(t, x)) - a| \leq R + t_0(M + 2|a|).$$

If we assume $t_0 < \min\{1/2, (R' - R)/(M + 2|a|)\}$, then for all $t \in [0, t_0]$ and $x \in B(a, R)$, $y = (1 - t_0)\mathbf{p}_\phi(t, x) + t_0\nabla\phi(\mathbf{p}_\phi(t, x)) \in B(a, R')$ i.e. $\mathbf{p}_\phi(t_0, y) = \mathbf{p}_\phi(t, x)$, and therefore we have the inclusion $\mathbf{p}_\phi(t, B(a, R)) \subset \mathbf{p}_\phi(t_0, B(a, R'))$. \square

Lemma 1. *We assume that ϕ satisfies the property (Γ_2) . For every $t \in [0, 1)$, $\mathbf{p}_\phi(t, \cdot)$ is differentiable almost everywhere on \mathbb{R}^d . Moreover, for almost every $x \in \mathbb{R}^d$, $\nabla\phi$ is differentiable in $\mathbf{p}_\phi(t, x)$ and $D^2\phi$ is such that:*

$$\nabla_x \mathbf{p}_\phi(t, x) = (tD^2\phi(\mathbf{p}_\phi(t, x)) + (1 - t)I)^{-1} \quad (\text{A-7})$$

where $I \in \mathcal{M}_d(\mathbb{R})$ is the identity matrix.

Proof: Let $t \in [0, 1)$. The operator $\mathbf{p}_\phi(t, \cdot)$ is Lipschitz and bijective, and from (A-3) its inverse is $\mathbf{p}_\phi(t, \cdot)^{-1} = t\nabla\phi + (1 - t)\text{id} = X(t, \cdot)$. Recall that by hypothesis $\nabla\phi$ is assumed to be Lipschitz.

Rademacher's Theorem (see for instance [3] p.81) states that if $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a locally Lipschitz function, then f is \mathcal{L}^n -almost everywhere Fréchet-differentiable (and its differential in the sense of Fréchet coincides with its differential in the sense of distributions).

According to this last Theorem, $\nabla\phi$ and $\mathbf{p}_\phi(t, \cdot)$ are differentiable almost everywhere on \mathbb{R}^d and their gradients coincide with their derivatives in the sense of distributions. Thus the set F of points in \mathbb{R}^d where $\nabla\phi$ is not differentiable is of zero Lebesgue measure. Since $\nabla\phi$ is assumed to be Lipschitz, then so does $X(t, \cdot)$, which gives $\mathcal{L}^d(X(t, F)) = 0$ ([3] p. 75). As $X(t, F)$ is the set of points $x \in \mathbb{R}^d$ for which $\nabla\phi$ is not differentiable in $\mathbf{p}_\phi(t, x)$, this means that $\nabla\phi$ is differentiable in $\mathbf{p}_\phi(t, x)$ for almost all $x \in \mathbb{R}^d$. Hence $\mathbf{p}_\phi(t, \cdot)$ is differentiable at almost every $x \in \mathbb{R}^d$, $\nabla\phi$ is differentiable at $\mathbf{p}_\phi(t, x)$ and $I = (tD^2\phi(\mathbf{p}_\phi(t, x)) + (1 - t)I)\nabla_x \mathbf{p}_\phi(t, x)$. The potential ϕ being convex, $D^2\phi(\mathbf{p}_\phi(t, x))$ is symmetric positive, and hence by coercivity, $tD^2\phi(\mathbf{p}_\phi(t, x)) + (1 - t)I$ is symmetric positive definite and therefore invertible in $\mathcal{M}_2(\mathbb{R})$, which concludes the proof. \square

Remark 2 (An useful example of the application of Rademacher's Theorem). *We know that v is locally Lipschitz on $(0, 1) \times \mathbb{R}^d$. Thus, according to Rademacher's Theorem, v is differentiable almost everywhere on $(0, 1) \times \mathbb{R}^d$, and its differential corresponds to its derivative in the sense of distributions. In particular, we have*

$$v \cdot \nabla_x v = \frac{1}{2} \nabla_x |v|^2, \quad (\text{A-8})$$

in the sense of Frechet (\mathcal{L}^{d+1} -almost everywhere) and in the sense of distributions.

Remark 3 (On the contribution of property (Γ_2) in the previous proof). *Note that for every $t \in [0, 1)$, the operator $\mathbf{p}_\phi(t, \cdot)$ is differentiable almost everywhere on \mathbb{R}^d , and the operator $\nabla_x \mathbf{p}_\phi(t, \cdot)$*

is thus well defined. This property is satisfied regardless of the regularity of ϕ , since the proximal operator is always Lipschitz. The additional property brought here by the regularity C^1 of ϕ is in fact the bijectivity of the operator $\mathbf{p}_\phi(t, \cdot)$.

The fact that $\nabla\phi$ is locally Lipschitz on \mathbb{R}^d is crucial to ensure that $D^2\phi$ is well defined almost everywhere. However, the fact that $\nabla\phi$ is **globally** Lipschitz on \mathbb{R}^d ensures that $\mathbf{p}_\phi(t, \cdot)$ does not send sets of \mathbb{R}^d of positive measure to negligible sets (see [3] p. 75), such as sets where ϕ is not twice differentiable. Such global regularity ensures that the operator $D^2\phi(\mathbf{p}_\phi(t, \cdot))$ is well defined almost everywhere. Then, the assumption of property (Γ_2) allows us to consider $\nabla_x \mathbf{p}_\phi(t, \cdot)$ as a function of $\mathbf{p}_\phi(t, \cdot)$ almost everywhere.

We now have all the elements to state the following proposition, which is one of the main results of this subsection, concerning the control of the gradient of the velocity field v . This result is namely required to control the solutions of the transport problem generated by the field v in the uniqueness results of Section 6 of the main paper [6].

For convenience, we will use the norm $|\cdot|_1$ on $\mathcal{M}_d(\mathbb{R})$, defined by $|A|_1 = \sum_{i,j} |a_{ij}|$, instead of the operator norm associated to the euclidean norm on \mathbb{R}^d .

Proposition 5. *We assume that ϕ satisfies the property (Γ_2) . Let $R' > R > 0$ and $a \in \mathbb{R}^d$ such that $\phi(a) = \inf_{\mathbb{R}^d} \phi$. Then there exist constants C and C' (independent of ϕ , a , R and R') such that for all $t_0 \in (0, 1)$ satisfying the condition (A-5), we have the property:*

$$\forall t \in (0, t_0], \int_{B(a,R)} |\nabla_x v(t, x)| dx \leq C \int_{B(a,R')} |\nabla_x v(t_0, x)| dx \leq \frac{C'}{t_0(1-t_0)} \mathcal{L}^2(B(a, R')) \quad (\text{A-9})$$

so that $\nabla_x v \in L^\infty([0, t_0], L^1(B(a, R)))$.

Sketch of proof: Let $t \in (0, 1)$. Remember that for all $x \in \mathbb{R}^d$, $v(t, x) = \nabla\phi(\mathbf{p}_\phi(t, x)) - \mathbf{p}_\phi(t, x)$. Then, for almost all $x \in \mathbb{R}^d$, $v(t, \cdot)$ is differentiable on x and

$$\nabla_x v(t, x) = \nabla_x \mathbf{p}_\phi(t, x) (D^2\phi(\mathbf{p}_\phi(t, x)) - I) = (t D^2\phi(\mathbf{p}_\phi(t, x)) + (1-t)I)^{-1} (D^2\phi(\mathbf{p}_\phi(t, x)) - I). \quad (\text{A-10})$$

Since, $v(t, \cdot)$ is Lipschitz, $|\nabla_x v(t, \cdot)|_1 \in L^\infty(\mathbb{R}^d) \subset L^1_{loc}(\mathbb{R}^d)$. The function $X(t, \cdot) = t\nabla\phi + (1-t)\text{id}$ is Lipschitz and bijective (and $\mathbf{p}_\phi(t, \cdot)$ is its inverse), we can therefore apply the generalized Change of Variable Theorem ([3] p. 117) and obtain:

$$\int_{B(a,R)} |\nabla_x v(t, x)|_1 dx = \int_{\mathbf{p}_\phi(t, B(a,R))} |\det(t D^2\phi(y) + (1-t)I)| \times |(t D^2\phi(y) + (1-t)I)^{-1} (D^2\phi(y) - I)|_1 dy. \quad (\text{A-11})$$

For every $y \in \mathbb{R}^d$ where $\nabla\phi$ is differentiable, the matrix $D^2\phi$ is symmetric positive and can therefore be diagonalized in an orthonormal basis. We will consider $\lambda_1(y), \dots, \lambda_d(y) \geq 0$ the eigenvalues associated with $D^2\phi$. From the equivalence between the $|\cdot|_1$ norm and the Frobenius norm, we obtain the following relation:

$$C_1 \int_{B(a,R)} |\nabla_x v(t, \cdot)|_1 dx \leq \sum_{j=1}^d \int_{\mathbf{p}_\phi(t, B(a,R))} |\lambda_j(y) - 1| \prod_{\substack{i=1 \\ i \neq j}}^d (t\lambda_i(y) + (1-t)) dy \leq C_2 \int_{B(a,R)} |\nabla_x v(t, \cdot)|_1 dx, \quad (\text{A-12})$$

where the constants C_1 and C_2 only depend on the constants of equivalence between the $|\cdot|_1$ norm and the Frobenius norm, and are independent of t and R . Let us take $t_0 \in (0, 1)$ verifying

the condition (A-5) of Proposition 4. Hence, we have $\mathbf{p}_\phi(t, B(a, R)) \subset \mathbf{p}_\phi(t_0, B(a, R'))$. The condition (A-5) also gives $t_0 \leq 1/2$, and so for any $t \in (0, t_0]$, we have $1-t \leq 1 \leq 2(1-t_0)$. Since $\lambda_i(y) \geq 0$ ($i = 1, \dots, d$) for almost all $y \in \mathbb{R}^d$, we have $0 \leq t\lambda_i(y) + (1-t) \leq 2(t_0\lambda_i(y) + (1-t_0))$. Thus, thanks to the inequality (A-12), we can conclude:

$$\int_{B(a, R)} |\nabla_x v(t, x)|_1 dx \leq \int_{B(a, R')} |\nabla_x v(t_0, x)|_1 dx, \quad (\text{A-13})$$

since $\prod_{\substack{i=1 \\ i \neq j}}^d (t\lambda_i(y) + (1-t)) \leq 2^{d-1} \prod_{\substack{i=1 \\ i \neq j}}^d (t_0\lambda_i(y) + (1-t_0))$. Finally, as already mentioned at the beginning of this proof, there exists a constant $c > 0$ such that $|\nabla_x v(t_0, \cdot)|_1$ is bounded by $c/t_0(1-t_0)$, which completes the proof of the Proposition. \square

This proof is useful for another lemma concerning the control of v . Gathering the different results of this subsection concerning the control of the gradient of the field v , we can control the solutions of the transport problem and thus obtain uniqueness results.

Lemma 2. *We assume that ϕ satisfies the property (Γ_2) . Then there exists a constant $C > 0$ such that for all $t \in (0, 1)$,*

$$\|\nabla_x v(t, \cdot)\|_{L^\infty(\mathbb{R}^d)} \leq \frac{C}{1-t} \left(\|D^2\phi\|_{L^\infty(\mathbb{R}^d)} + d \right)$$

Proof: As in the previous proof, using a diagonalization of $D^2\phi$ and the equivalence between $|\cdot|_1$ and Frobenius norms, as $\lambda_i(y) \geq 0$, we obtain, for every $y \in \mathbb{R}^d$ where $\nabla\phi$ is differentiable:

$$\begin{aligned} C_1 |(tD^2\phi(y) + (1-t)I)^{-1}(D^2\phi(y) - I)|_1 &\leq \sum_{i=1}^d \left| \frac{\lambda_i(y) - 1}{t\lambda_i(y) + (1-t)} \right|_1 \leq \frac{1}{1-t} \left(d + \sum_{i=1}^d \lambda_i(y) \right) \\ &\leq \frac{C_2}{1-t} (|D^2\phi(y)|_1 + d). \end{aligned} \quad (\text{A-14})$$

We can then conclude by injecting the equation (A-10) into this last inequality. \square

A.2 General case (Γ_1)

In the context of our optimal transport problem, we have proved at Proposition 3.2 of the main paper [6] that we can assume that the potential ϕ satisfies the property (Γ_1) .

We are now able to extend the results of the previous subsection to potentials ϕ which satisfy (Γ_1) , so that ϕ may have non-differentiability points **involving breaks in the transport plan**. More precisely we can extend Propositions 3 and 5 to the case where ϕ only satisfies (Γ_1) in order to show Propositions 1 and 2. In particular we show that $\nabla_x v$ is uniformly integrable in the neighborhood of $t = 0$, and in Corollary 1 we argue symmetrically to show that it is the same in the neighborhood of $t = 1$, and then on all $[0, 1]$. To that end we consider the regularization $\gamma\phi$ of ϕ with the Moreau envelope [2] defined for all $x \in \mathbb{R}^d$ and for all $\gamma > 0$ as

$$\gamma\phi(x) = \inf_{y \in \mathbb{R}^d} \frac{1}{2\gamma} |x - y|^2 + \phi(y) = \frac{1}{2\gamma} |x - \text{Prox}_{\gamma\phi}(x)|^2 + \phi(\text{Prox}_{\gamma\phi}(x)). \quad (\text{A-15})$$

For all $\gamma > 0$, we also define the velocity field v_γ for all $t \in (0, 1)$ and $x \in \mathbb{R}^d$ by

$$v_\gamma(t, x) = v_{\gamma\phi}(t, x) = \nabla\phi(\mathbf{p}_{\gamma\phi}(t, x)) - \mathbf{p}_{\gamma\phi}(t, x) = \frac{x - \mathbf{p}_{\gamma\phi}(t, x)}{t}, \quad (\text{A-16})$$

where $\mathbf{p}_{\gamma\phi}(t, x) = \text{Prox}_{\frac{t}{1-t}\gamma\phi}\left(\frac{x}{1-t}\right)$. We also recall that $\gamma\phi$ is of class C^1 and that for all $x \in \mathbb{R}^d$,

$$\nabla(\gamma\phi)(x) = \frac{x - \text{Prox}_{\gamma\phi}(x)}{\gamma} \in \partial\phi(\text{Prox}_{\gamma\phi}(x)). \quad (\text{A-17})$$

The potential $\gamma\phi$ is then γ^{-1} -Lipschitz.

We now show that if ϕ satisfies property (Γ_1) , then $\gamma\phi$ satisfies property (Γ_2) . The results of subsection A.1 are then applied on ϕ_γ , that corresponds to transport plans without breaks.

Lemma 3. *If ϕ satisfies property (Γ_1) , then $\gamma\phi$ satisfies property (Γ_2) for all $\gamma > 0$.*

Sketch of the proof: As $(\gamma f)^* = f^* + \frac{\gamma}{2}|\cdot|^2$, then if f^* is in a Hölder space $C^{1,1}$ and admits a minimum on \mathbb{R}^d , the same holds for $(\gamma f)^*$. Notice on the other hand that the functions ϕ and $\gamma\phi$ have the same minima on \mathbb{R}^d . \square

Lemma 4. *We assume that ϕ satisfies property (Γ_1) , then $(\nabla\gamma\phi)_{\gamma>0}$ is locally bounded, uniformly with respect to $\gamma > 0$.*

Proof: We can show that for every minimum a of ϕ on \mathbb{R}^d and $r > 0$, we have the inclusion $\nabla(\gamma\phi)(B(a, r)) \subset \partial\phi(B(a, r))$. To that end, we use the relation $\nabla(\gamma f) = \gamma^{-1}(\text{id} - \text{Prox}_{\gamma f}) = \text{Prox}_{f^*/\gamma}(\cdot/\gamma)$ [2], and then, with definition of proximal operator ($y = \text{Prox}_f(x) \Leftrightarrow x - y \in \partial f(y)$), we have $\nabla(\gamma f)(x) \in \partial f(\text{Prox}_{\gamma f}(x))$ for all $x \in \mathbb{R}^d$. Moreover, we have the inclusion $\text{Prox}_{\gamma\phi}(B(a, r)) \subset B(a, r)$, due to the non-expansiveness of the operator $\text{Prox}_{\gamma\phi}$ and the fact that a is a fixed point for these operators. The union of the subdifferentials of ϕ on $B(a, r)$, denoted by $\partial\phi(B(a, r))$, is bounded in \mathbb{R}^d , since the potential ϕ is convex and locally Lipschitz. \square

Then, by weak compactness of Sobolev spaces, we can conclude that for all $x \in \mathbb{R}^d$, the set of adherence values of the family $(\nabla\gamma\phi(x))_\gamma$, when $\gamma > 0$ tends to 0 is included in $\partial\phi(x)$, and that if ϕ is differentiable at x , then $\nabla\gamma\phi(x)$ converges to $\nabla\phi(x)$ when $\gamma > 0$ tends to 0. The functions $\nabla\gamma\phi$ converge simply almost everywhere to $\nabla\phi$ when γ tends to 0.

We recall that $\|v_\gamma\|_{L^\infty((0,1)\times\omega)} \leq 5(\max\{M_\gamma, M_\gamma^*\} + \sup(\omega))$, with $M_\gamma = \|\nabla\gamma\phi\|_{L^\infty(\omega)}$ and $M_\gamma^* = \|\partial(\gamma\phi)^*\|_{L^\infty(\omega)}$. According to Lemma 4, the constants M_γ are uniformly bounded with respect to $\gamma > 0$. The same holds for the constants M_γ^* with respect to $\gamma \in (0, \gamma_0]$ by noticing that $\partial(\gamma\phi)^* = \partial\phi^* + \gamma \text{id}$ (property of conjugation of an inf-convolution: $(f \square g)^* = f^* + g^*$ [2]). Therefore, $(v_\gamma)_{\gamma_0 \geq \gamma > 0}$ is uniformly bounded on $(0, 1) \times \omega$ by a constant independent of $\gamma \in (0, \gamma_0]$.

We have the relation $\nabla_x(\phi_t)^* = (\text{id} - \text{Prox}_{(1-t)(t\phi)^*})/(1-t)$ for all $t \in (0, 1)$. In addition,

$$\text{Prox}_{(1-t)(t\gamma\phi)^*} = \text{Prox}_{(1-t)(t\phi)^* + (1-t)\frac{\gamma}{2t}|\cdot|^2} = \text{Prox}_{(1-t)(t\phi)^*}(\text{id} - \gamma(1-t)[\text{id} + (1-t)v_\gamma(t, \cdot)]).$$

As we have $\mathbf{p}_\phi(t, \cdot) = \nabla_x(\phi_t)^*$ and $v_\phi(t, x) = [x - \mathbf{p}_\phi(t, x)]/t$ and using the non-expansiveness of the proximal operator, we get:

$$|v_\gamma(t, \cdot) - v(t, \cdot)| = \frac{1}{t(1-t)} |\text{Prox}_{(1-t)(t\gamma\phi)^*} - \text{Prox}_{(1-t)(t\phi)^*}| \leq \frac{\gamma}{t} |\text{id} + (1-t)v_\gamma(t, \cdot)|.$$

Since the fields v_γ are uniformly bounded on $(0, 1) \times \omega$ independently of $\gamma \in (0, \gamma_0]$, there exists a constant M independent of $\gamma \in (0, \gamma_0]$, such that $\|v(t, \cdot) - v_\gamma(t, \cdot)\|_{L^\infty(\omega)} \leq M(\gamma/t)$ for all $\gamma \in (0, \gamma_0]$ and all $t \in (0, 1)$. As $(v_\gamma)_{\gamma_0 \geq \gamma > 0}$ is uniformly bounded on $(0, 1) \times \omega$, we have:

$$\forall \alpha \in [0, 1], \forall \gamma \in (0, \gamma_0], \forall t \in (0, 1), \|v(t, \cdot) - v_\gamma(t, \cdot)\|_{L^\infty(\omega)} \leq C \left(\frac{\gamma}{t}\right)^\alpha. \quad (\text{A-18})$$

for all $\alpha \in [0, 1]$, and with C a constant independent of γ .

A.3 Finalization of the proof of Propositions 1 and 2

We now have enough elements to prove Propositions 1 and 2.

Sketch of proof of Proposition 1: When ϕ satisfies property (Γ_1) , a transport plan may admit "breaks". With Proposition 3 the Burgers equation is nevertheless satisfied by the v_γ 's. According to inequality (A-18), we can conclude that v_γ converges to the field v in $L^p((0, 1) \times \omega)$ when γ goes to 0, for every $1 < p < +\infty$ and every bounded open set ω in \mathbb{R}^d . From the weak compactness of sobolev spaces, we get that $v_\gamma(t, \cdot)$ converges weakly in $W^{1,p}(\omega)$ to $v(t, \cdot)$, for all $t \in (0, 1)$, since $\|\nabla_x v_\gamma(t, \cdot)\|_{L^\infty(\omega)} \leq c/t(1-t)$. As the fields v_γ are uniformly bounded on $(0, 1) \times \omega$ independently of $\gamma \in (0, \gamma_0]$, we can apply these properties of convergence in the weak formulation of Burgers equation and conclude. \square

Sketch of proof of Proposition 2: As already mentioned in the proof of Lemma 4, we have the inclusion $\nabla(\gamma\phi)(B(a, r)) \subset \partial\phi(B(a, r))$, and thus, by setting $r = 2(R + |a|)$, we have $M_\gamma = \sup_{B(a, r)} |\nabla\gamma\phi| \leq M$. Hence, a time t_0 verifying the hypothesis of the statement (independent of γ), also satisfies the hypothesis of Proposition 5 for all $\gamma > 0$. We can therefore apply this last proposition for such a t_0 for all v_γ to obtain (A-2). For the case $\gamma = 0$ (i.e. $v_0 = v$), the weak convergence of $v_\gamma(t, \cdot)$ to $v(t, \cdot)$ in $W^{1,p}(\omega)$ can be used to conclude, by weak lower semi-continuity of the L^1 norm on $B(a, r)$. \square

A.4 An independent but notable regularity result

We now show that Proposition 2 can be generalized to $\nabla_{t,x}v \in L^\infty(0, 1; L^1(\Omega))$ for every bounded open set Ω of \mathbb{R}^d . This property is not required in the proof of the results presented in the main paper [6]. It nevertheless gives a new insight of the regularity and the control of the velocity field of an isotropic optimal transport for the L^2 distance.

Corollary 1 (Corollary 6.1 in [6]). *Assume that ϕ satisfies the property (Γ_1) . Let Ω be a bounded open set. Then $\nabla_{t,x}v \in L^\infty(0, 1; L^1(\Omega))$ and there exists $K > 0$ such that for all $t \in (0, 1)$,*

$$\int_{\Omega} |\nabla_{t,x}v(t, x)| dx \leq K. \quad (\text{A-19})$$

Sketch of the proof: We symmetrize the result of Proposition 2 (case $\gamma = 0$) respectively in the neighborhood of $t = 0$ and $t = 1$, and apply the upper bound $c/t(1-t)$ in the middle. \square

Corollary 2 (Remark 6.2 in [6]). *We assume that ϕ satisfies the property (Γ_1) . Let Ω be a bounded open set. For all $p, q \geq 1$ such that $1/p + 1/q > 1$, we have $\nabla_{t,x}v \in L^p(0, 1; L^q(\Omega))$. In particular $v \in W^{1,p}((0, 1) \times \Omega)$ for all $1 \leq p < 2$.*

Sketch of the proof: We jointly use the result of Corollary 1 with the estimate $c/t(1-t)$. We thus partially bound from above $|\nabla_x v(t_0, \cdot)|_1$ in order to be able to apply Corollary 1. More precisely, for all $\omega \subset \subset \Omega$:

$$\left(\int_{\omega} |\nabla_x v(t, x)|^p dx \right)^{\frac{q}{p}} \leq \frac{c^a}{t^a(1-t)^a} \left(\int_{\omega} |\nabla_x v(t, x)| dx \right)^{\frac{q}{p}} \leq \frac{C}{t^a(1-t)^a},$$

with $a = q(p-1)/p$, and $t \mapsto \frac{C}{t^a(1-t)^a}$ is integrable on $(0, 1)$. For the particular case $W^{1,p}$, it is sufficient to take $p = q < 2$. \square

Corollary 1 and Corollary 2 give the most consistent regularity that one can have for general velocity field v . Indeed, according to Corollary 2, a velocity field v defined with respect to a potential ϕ (verifying the property (Γ_1)) for all $t \in (0, 1)$ by

$$v(t, \cdot) = v_\phi(t, \cdot) = \frac{\text{id} - \mathbf{p}_\phi(t, \cdot)}{t} = \frac{1}{t} (\text{id} - (t\phi + (1-t)\text{id})^*), \quad (\text{A-20})$$

is an element of $W_{loc}^{1,p}([0, 1] \times \mathbb{R}^d)$. Hence the restriction of v to any bounded open set Ω of \mathbb{R}^d is an element of $W^{1,p}((0, 1) \times \Omega)$, for all $p < 2$. The question that arises naturally is whether $v = v_\phi$ could not be an element of $H_{loc}^1([0, 1] \times \mathbb{R}^d)$. In general, this is not the case (see for instance Caffarelli's counter-example on mass splitting).

B Convergence of Benamou-Brenier's algorithm

This section is dedicated to the proof 8.2 in the main paper. We first recall general results on non-expansive operators that will imply the weak and strong convergence of the algorithm. In what follows, $(H, \langle \cdot, \cdot \rangle)$ is an Hilbert space and \mathbf{M} is an operator from H to H .

Definition 1 (Non-expansive Operator). *The operator \mathbf{M} is called **non-expansive** if and only if it is 1-Lipschitz, and **firmly non-expansive** if and only if we have $\|\mathbf{M}x - \mathbf{M}y\|^2 \leq \langle x - y, \mathbf{M}x - \mathbf{M}y \rangle$, for all $x, y \in H$. The operator \mathbf{M} is called **quasi-firmly non-expansive** on a subset A of H , containing the set of fixed points of \mathbf{M} , if and only if, for any fixed point x^* of \mathbf{M} , we have for all $x \in A$: $\|x - \mathbf{M}x\|^2 + \|x - x^*\|^2 \geq \|\mathbf{M}x - x^*\|^2$.*

Theorem 1. *Let \mathbf{M} be a non-expansive operator on H , and quasi-firmly non-expansive on $\mathbf{M}(H)$ (the image of H by \mathbf{M}). Assume that the set $\text{Fix}(\mathbf{M})$ of the fixed points of \mathbf{M} is non-empty. Let $(x_n)_n$ be a sequence of elements of H satisfying for every $n \in \mathbb{N}$ the estimate: $\|\mathbf{M}(x_n) - x_{n+1}\| \leq \epsilon_n$, where $(\epsilon_n)_n$ is a non-negative real sequence satisfying $\sum_n \epsilon_n < +\infty$. Then $(x_n)_n$ weakly converges in H to a fixed point of \mathbf{M} .*

Theorem 2 (H. Bauschke [1]). *Let \mathbf{M} be a non-expansive operator over H and assume that the set $\text{Fix}(\mathbf{M})$ of the fixed points of \mathbf{M} is non-empty. Let $(\lambda_n)_{n \geq 0}$ be a sequence of parameters of $[0, 1)$, converging to 0, and satisfying: $\sum_n \lambda_n = +\infty$ and $\sum_n |\lambda_{n+1} - \lambda_n| < +\infty$. Given a and x_0 in H , we define the sequence (x_n) by the recurrence $x_{n+1} = \lambda_n a + (1 - \lambda_n) \mathbf{M}_n x_n$ ($\forall n \geq 0$), where we have for all $n \in \mathbb{N}$ the estimate $\|\mathbf{M}_n x_n - \mathbf{M}x_n\| \leq \epsilon_n$, with $\sum_n \epsilon_n < +\infty$. Then the sequence $(x_n)_n$ converges strongly to the L_2 projection of a on the closed convex set $\text{Fix}(\mathbf{M})$.*

The sequence $(\epsilon_n)_n$ represents the numerical errors inherent to the implementation of the algorithm. They are here assumed to be highly controlled, which is not realistic in practice.

Theorem 1 can be shown using classical functional analysis tools (see [7]) such as the Opial's Lemma [8]. Notice that in Theorem 2, it is direct to prove that the set of fixed points of a non-expansive operator is a closed convex set (see Lemma 2.3-7 of [5]). The detailed proofs of these two theorems can be found in [5] (section 2.3 and its appendix B).

We are now able to apply the previous convergence results to the Benamou-Brenier algorithm. To that end, it is sufficient to show that an iteration of the algorithm corresponds to the iteration of a certain non-expansive operator. We consider the space $H = L^2(Q)^{d+1} \times L^2(Q)^{d+1}$, provided with the scalar product

$$\langle (\mu_1, q_1), (\mu_2, q_2) \rangle_H = \langle \mu_1, \mu_2 \rangle_{L^2} + r^2 \langle q_1, q_2 \rangle_{L^2},$$

so that $(H, \langle \cdot, \cdot \rangle)$ is an Hilbert space. Let $\mathbf{B} : H \rightarrow H$ be the operator which associate to (μ, q) the product (μ', q') of the last two steps (B and C) of the algorithm Benamou-Brenier. Here

ψ just acts as an auxiliary variable. Indeed, if (μ^*, q^*, ψ^*) is a saddle point of the Lagrangian $\mathbf{L}(\psi, p, \mu) = \chi_{\mathcal{P}}(q) + G(\psi) + \langle \mu, \nabla_{t,x}\psi - q \rangle_{L^2}$, then $(\mu^*, q^*) = \mathbf{B}(\mu^*, q^*)$. Conversely if (μ^*, q^*) is a fixed point of \mathbf{B} then (μ^*, q^*, ψ^*) is a saddle point of the Lagrangian, where ψ^* is the unique element of S which satisfies $q^* = \nabla\psi^*$. The potential ψ is therefore only required for computational purposes.

Proposition 6. *Operator \mathbf{B} is non-expansive on H and quasi-firmly non-expansive on $\mathbf{B}(H)$.*

Proof: (μ_1, q_1) and (μ_2, q_2) being given, we obtain $(\mu'_1, q'_1) = \mathbf{B}(\mu_1, q_1)$ and $(\mu'_2, q'_2) = \mathbf{B}(\mu_2, q_2)$ with the Benamou-Brenier iterations. For $i = 1, 2$, we look for:

- Step A : Find the unique $\psi'_i \in (H^1/\mathbb{R})(Q)$ such that $G(h) + \langle \mu_i, \nabla h \rangle_{L^2} + r \langle \nabla\psi'_i - q_i, \nabla h \rangle_{L^2} = 0$, $\forall h \in (H^1/\mathbb{R})(Q)$.
- Step B : Find the unique q'_i such that $\langle \mu_i + r(\nabla\psi'_i - q'_i), p - q'_i \rangle_{L^2} \leq 0$, for all $p \in \mathcal{P}$.
- Step C : Define μ'_i by $\mu'_i = \mu_i + r(\nabla_{t,x}\psi'_i - q'_i)$.

Let us start by studying the non-expansiveness of \mathbf{B} . Note that by injecting the equation of step C into step A and step B (for $i = 1$ or $i = 2$), we obtain the two new equations:

$$G(h) + \langle \mu'_i, \nabla h \rangle_{L^2} + r \langle q'_i - q_i, \nabla h \rangle_{L^2} = 0, \quad \forall h \in H^1(Q)/\mathbb{R}, \quad (\text{B-21})$$

$$\langle \mu'_i, p - q'_i \rangle_{L^2} \leq 0, \quad \forall p \in \mathcal{P}. \quad (\text{B-22})$$

We then set $\bar{\mu}^{(\prime)} = \mu_2^{(\prime)} - \mu_1^{(\prime)}$, $\bar{q}^{(\prime)} = q_2^{(\prime)} - q_1^{(\prime)}$ and $\bar{\psi}^{(\prime)} = \psi_2^{(\prime)} - \psi_1^{(\prime)}$. By taking $(\text{B-21})_{i=2} - (\text{B-21})_{i=1}$ with $h = \bar{\psi}^{(\prime)}$, and by summing $(\text{B-22})_{i=2}$ with $p = q'_1$ and $(\text{B-22})_{i=1}$ with $p = q'_2$, we respectively obtain the two following relations:

$$\langle \bar{\mu}^{(\prime)}, \nabla \bar{\psi}^{(\prime)} \rangle + r \langle \bar{q}^{(\prime)} - \bar{q}, \nabla \bar{\psi}^{(\prime)} \rangle = 0, \quad (\text{B-23})$$

$$\langle \bar{\mu}^{(\prime)}, \bar{q}^{(\prime)} \rangle \geq 0. \quad (\text{B-24})$$

Summing these two relations, we have $\langle \bar{\mu}^{(\prime)}, \nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)} \rangle + r \langle \bar{q}^{(\prime)} - \bar{q}, \nabla \bar{\psi}^{(\prime)} \rangle \leq 0$. Observing from Step C that $\bar{\mu}^{(\prime)} = \bar{\mu} + r(\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)})$, we then obtain

$$\begin{aligned} |\bar{\mu}|^2 - |\bar{\mu}^{(\prime)}|^2 &= \langle \bar{\mu} - \bar{\mu}^{(\prime)}, \bar{\mu} + \bar{\mu}^{(\prime)} \rangle = -r \langle \nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}, 2(\bar{\mu} + r(\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)})) - r(\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}) \rangle \\ &= -2r \langle \bar{\mu}^{(\prime)}, \nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)} \rangle + r^2 |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2 \geq 2r^2 \langle \nabla \bar{\psi}^{(\prime)}, \bar{q}^{(\prime)} - \bar{q} \rangle + r^2 |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2. \end{aligned} \quad (\text{B-25})$$

Moreover, we have

$$\begin{aligned} \langle \nabla \bar{\psi}^{(\prime)}, \bar{q}^{(\prime)} - \bar{q} \rangle &= \langle \nabla \bar{\psi}^{(\prime)} - \bar{q} + \bar{q}, \bar{q}^{(\prime)} - \bar{q} \rangle = \langle \nabla \bar{\psi}^{(\prime)} - \bar{q}, \bar{q}^{(\prime)} - \bar{q} \rangle - |\bar{q}^{(\prime)} - \bar{q}|^2 + \frac{1}{2} (|\bar{q}^{(\prime)}|^2 - |\bar{q}|^2 + |\bar{q}^{(\prime)} - \bar{q}|^2) \\ &= \langle \nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}, \bar{q}^{(\prime)} - \bar{q} \rangle + \frac{1}{2} (|\bar{q}^{(\prime)}|^2 - |\bar{q}|^2 + |\bar{q}^{(\prime)} - \bar{q}|^2). \end{aligned} \quad (\text{B-26})$$

By re-injecting (B-26) in (B-25), we get:

$$\begin{aligned} |\bar{\mu}|^2 - |\bar{\mu}^{(\prime)}|^2 &\geq r^2 \left((|\bar{q}^{(\prime)}|^2 - |\bar{q}|^2) + \left(|\bar{q}^{(\prime)} - \bar{q}|^2 + 2 \langle \nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}, \bar{q}^{(\prime)} - \bar{q} \rangle + |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2 \right) \right) \\ &\geq r^2 (|\bar{q}^{(\prime)}|^2 - |\bar{q}|^2) + r^2 |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2, \end{aligned} \quad (\text{B-27})$$

which leads to

$$r^2 |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2 + (|\bar{\mu}^{(\prime)}|^2 + r^2 |\bar{q}^{(\prime)}|^2) \leq (|\bar{\mu}|^2 + r^2 |\bar{q}|^2) \quad (\text{B-28})$$

$$r^2 |\nabla \bar{\psi}^{(\prime)} - \bar{q}^{(\prime)}|^2 + \|\mathbf{B}(\mu_1, q_1) - \mathbf{B}(\mu_2, q_2)\|_H \leq \|(\mu_1, q_1) - (\mu_2, q_2)\|_H, \quad (\text{B-29})$$

so that the operator \mathbf{B} is non-expansive.

We now demonstrate the quasi-firmly non-expansiveness of \mathbf{B} on $\mathbf{B}(H)$. Let (μ^*, q^*) be a fixed point of \mathbf{B} , thus included in $\mathbf{B}(H)$. Using equation (B-28) with $(\bar{\psi}, \bar{q}, \bar{\mu}) = (\psi - \psi^*, q - q^*, \mu - \mu^*)$ and $(\bar{\psi}', \bar{q}', \bar{\mu}') = (\psi' - \psi^*, q' - q^*, \mu' - \mu^*)$, we obtain:

$$r^2 |\nabla(\psi' - \psi^*) - (q - q^*)|^2 + (|\mu' - \mu^*|^2 + r^2 |q' - q^*|^2) \leq (|\mu - \mu^*|^2 + r^2 |q - q^*|^2). \quad (\text{B-30})$$

For the first term, we get

$$\begin{aligned} r^2 |\nabla(\psi' - \psi^*) - (q - q^*)|^2 &= r^2 |\nabla\psi' - q|^2 = |(\mu' - \mu) + r(q' - q)|^2 \\ &= |\mu' - \mu|^2 + 2r \langle \mu' - \mu, q' - q \rangle + r^2 |q' - q|^2, \end{aligned}$$

with $\langle \mu' - \mu, q' - q \rangle \geq 0$, according to the relation (B-24). As in [4], we thus obtain:

$$\begin{aligned} |\mu' - \mu|^2 + r^2 |q' - q|^2 + (|\mu' - \mu^*|^2 + r^2 |q' - q^*|^2) &\leq (|\mu - \mu^*|^2 + r^2 |q - q^*|^2), \\ \|\mathbf{B}(\mu, q) - (\mu, q)\|_H^2 + \|(\mu', q') - (\mu^*, q^*)\|_H^2 &\leq \|(\mu, q) - (\mu^*, q^*)\|_H^2 \end{aligned}$$

so that \mathbf{B} is quasi-firmly non-expansive. \square

References

- [1] H. Bauschke. The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. *Journal of Mathematical Analysis and Applications*, 202(1):150–159, 1996.
- [2] H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.
- [3] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC PRESS, 1992.
- [4] M. Fortin and R. Glowinski. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Studies in Mathematics and its Applications. Elsevier Science, 1983.
- [5] R. Hug. *Mathematical analysis and convergence of an algorithm for dynamic optimal transport: case of non-smooth transport plans*. PhD thesis, Université de Grenoble, 2016.
- [6] R. Hug, E. Maitre, and N. Papadakis. On the convergence of augmented lagrangian method for optimal transport between nonnegative densities. *HAL preprint hal-01128793*, 2017.
- [7] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *Siam J. Numer. Anal.*, 16:964–979, 1979.
- [8] Z. Opial. Weak convergence of the sequence of successive approximations for non- expansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.