A variational approach to regularity theory in optimal transportation

M. Goldman[*]

June 24, 2019

Abstract

This paper describes recent results obtained in collaboration with M. Huesmann and F. Otto on the regularity of optimal transport maps. The main result is a quantitative version of the well-known fact that the linearization of the Monge-Ampère equation around the identity is the Poisson equation. We present two applications of this result. The first one is a variational proof of the partial regularity theorem of Figalli and Kim and the second is the rigorous validation of some predictions made by Carraciolo and al. on the structure of the optimal transport maps in matching problems.

1   Introduction

Following Caffarelli's groundbreaking papers [9, 8], the classical approach to regularity theory for solutions of the optimal transport problem goes through maximum principle arguments and the construction of barriers (see the review paper [13]). The aim of this note is to describe a recent alternative approach, more variational in nature and based on the fact that the linearization of the Monge-Ampère equation around the identity is the Poisson equation (see [27]). Our main achievement in this direction is an harmonic approximation result which says that if at a given scale the transport plan is close to the identity and if at the same scale both the starting and target measures are close (in the Wasserstein metric) to be constant, then on a slightly smaller scale, the transport plan is actually extremely close to an harmonic gradient field. As in De Giorgi's approach to the regularity theory for minimal surfaces (see [24]) this allows to transfer the good regularity properties of harmonic functions to the transport plan and obtain an "excess improvement by tilting" estimate. This may be used to propagate information from the macroscopic scale down to the microscopic scale through a Campanato iteration.

---

[*]Université de Paris, CNRS, Sorbonne-Université, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France, `goldman@math.univ-paris-diderot.fr`

We give two applications of this result. The first one is a new proof of the partial regularity result of Figalli and Kim [15] (see also [14]). The second one is a validation up to the microscopic scale of the prediction by Caracciolo and al. [11] that for the optimal matching problem between a Poisson point process and the Lebesgue measure, the optimal transport plan is well approximated by the gradient of the solution to the corresponding Poisson equation with very high probability.

The plan of this note is the following. In Section 2 we recall some standard results on optimal transportation. The harmonic approximation theorem is stated together with a sketch of proof in Section 3. We then describe the application to the partial regularity result in Section 4 and to the optimal matching problem in Section 5.

2 The optimal transport problem

Optimal transportation is nowadays a very broad and active field. We give here only a very basic and short introduction to the topic and refer the reader to the monographs [25, 27] for much more details. For μ and λ two positive measures on $\mathbb{R}^d$ with $\mu(\mathbb{R}^d) = \lambda(\mathbb{R}^d)$ the optimal transport problem (in its Lagrangian formulation) is

$$W^2(\mu, \lambda) = \inf_{\pi_1 = \mu, \pi_2 = \lambda} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi, \tag{2.1}$$

where for a coupling π on $\mathbb{R}^d \times \mathbb{R}^d$, π_1 (respectively π_2) denotes the first (respectively the second) marginal of π. Under very mild assumptions on μ and λ (for instance compact supports), an optimal transference plan π exists (see [27]). The optimality conditions are as follows:

Theorem 2.1. *Let π be a coupling between μ and λ.*

(i) *(Knott-Smith) It is optimal if and only if there exists a convex and lower-semicontinuous function ψ (also called the Kantorovich potential) such that* $\mathrm{spt}\, \pi \subseteq Graph(\partial\psi)$.

(ii) *(Brenier) Moreover, if μ does not give mass to Lebesgue negligible sets, then there exists a unique $\nabla\psi$, gradient of a convex function, with $\nabla\psi\#\mu = \lambda$ and $\pi = (id \times \nabla\psi)\#\mu$. In this case we let $T = \nabla\psi$ be the optimal transport map.*

Let us point out that assuming that ψ is regular and that both μ and λ are smooth densities, the condition $T\#\mu = \lambda$ is nothing else than the Monge-Ampère equation

$$\det \nabla^2 \psi = \frac{\mu}{\lambda \circ \nabla\psi}.$$

In particular, if both μ and λ are close to (the same) constant density, then the Monge-Ampère equation linearizes to the Poisson equation (see [27, Ex. 4.1])

$$\Delta\psi = \mu - \lambda. \tag{2.2}$$

We will also use the Eulerian formulation of the optimal transport problem.

Theorem 2.2 (Benamou-Brenier). *There holds*

$$W_2^2(\mu, \lambda) = \inf_{(\rho, j)} \left\{ \int_{\mathbb{R}^d} \int_0^1 \frac{1}{\rho} |j|^2 \; : \partial_t \rho + \nabla \cdot j = 0, \; \rho_0 = \mu, \; \rho_1 = \lambda \right\}. \tag{2.3}$$

Moreover, if π is an optimal transport plan for (2.1), then the density-flux pair (ρ_t, j_t) defined for $t \in [0,1]$ by its action on test functions $(\zeta, \xi) \in C^0(\mathbb{R}^d) \times (C^0(\mathbb{R}^d))^d$ as

$$\int_{\mathbb{R}^d} \zeta d\rho_t = \int_{\mathbb{R}^d \times \mathbb{R}^d} \zeta((1-t)x + ty) d\pi \quad and \quad \int_{\mathbb{R}^d} \xi \cdot dj_t = \int_{\mathbb{R}^d \times \mathbb{R}^d} \xi((1-t)x + ty) \cdot (y - x) d\pi, \tag{2.4}$$

is a minimizer of (2.3).

Let us introduce some further notation. If (ρ, j) is a minimizer of (2.3), we define $(\bar{\rho}, \bar{j})$ the density-flux pair obtained by integrating in time (for instance $\bar{\rho} = \int_0^1 \rho_t$). For $R > 0$ and μ a positive measure on $\mathbb{R}^d$, we denote by $W_{B_R}(\mu, \kappa) = W(\mu \llcorner B_R, \kappa \chi_{B_R} dx)$, the Wasserstein distance between the restriction of μ to the ball B_R and the corresponding constant density $\kappa = \frac{\mu(B_R)}{|B_R|}$.

In order to obtain a local version of the equivalence between (2.1) and (2.3), we will need an L^∞ bound on the displacement (see [20, 19]).

Lemma 2.3. *Let π be a coupling between two measures μ and λ. Assume that $\operatorname{spt} \pi$ is monotone[i] and that for some[ii] $R > 0$, $E + D \ll 1$ where*

$$E = \frac{1}{R^{d+2}} \int_{(B_{6R} \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_{6R})} |x - y|^2 d\pi \tag{2.5}$$

and

$$D = \frac{1}{R^{d+2}} W_{B_{6R}}^2(\mu, \kappa_\mu) + \frac{1}{\kappa_\mu} (\kappa_\mu - 1)^2 + \frac{1}{R^{d+2}} W_{B_{6R}}^2(\lambda, \kappa_\lambda) + \frac{1}{\kappa_\lambda} (\kappa_\lambda - 1)^2. \tag{2.6}$$

Then, for every $(x, y) \in \operatorname{spt} \pi \cap \left((B_{5R} \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_{5R}) \right)$

$$|x - y| \lesssim R (E + D)^{\frac{1}{d+2}}. \tag{2.7}$$

3 The harmonic approximation theorem

We now state the harmonic approximation theorem. By scaling invariance, it is enough to state it at the unit scale $R = 1$. For μ, λ two positive measures and π an optimal coupling between them, we define the "excess" energy E as in (2.5) and the distance to the data D as in (2.6).

---

[i]Meaning that for every (x_1, y_1) and (x_2, y_2) in $\operatorname{spt} \pi$, $(x_1 - x_2) \cdot (y_1 - y_2) \geq 0$.

[ii]We use the short-hand notation $A \ll 1$ to indicate that there exists $\varepsilon > 0$ depending only on the dimension such that $A \leq \varepsilon$. Similarly, $A \lesssim B$ means that there exists a constant $C > 0$ depending on the dimension such that $A \leq CB$.

Theorem 3.1. *([19, Th. 1.4]) For every $0 < \tau \ll 1$, there exist $\varepsilon(\tau) > 0$ and $C(\tau) > 0$ such that provided $E + D \leq \varepsilon$, there exists a radius $R \in (3,4)$ such that if Φ is a solution of (ν denotes here the external normal to ∂B_R)*

$$\Delta\Phi = c \;\; in \; B_R \qquad and \qquad \nu \cdot \nabla\Phi = \nu \cdot \bar{j} \;\; on \; \partial B_R, \tag{3.1}$$

where c is the generic constant for which this equation is solvable, then

$$\int_{(B_1 \times \mathbb{R}^d) \cup (\mathbb{R}^d \times B_1)} |x - y + \nabla\Phi(x)|^2 d\pi \leq \tau E + CD. \tag{3.2}$$

The proof of Theorem 3.1 is actually performed at the Eulerian level. Thanks to Lemma 2.3, it is indeed enough to prove:

Theorem 3.2. *For every $0 < \tau \ll 1$, there exist $\varepsilon(\tau) > 0$ and $C(\tau) > 0$ such that provided $E + D \leq \varepsilon$, there exists a radius $R \in (3,4)$ such that if Φ solves (3.1), then*

$$\int_{B_2} \int_0^1 \frac{1}{\rho} |j - \rho\nabla\Phi|^2 \leq \tau E + CD. \tag{3.3}$$

To simplify a bit the discussion, we will assume from now on that $\lambda = \kappa_\mu = 1$, so that $D = W_{B_6}^2(\mu, 1)$.

The proof of Theorem 3.2 is based on three ingredients. The first of them is the choice of a 'good' radius R. Indeed, as will become apparent in the discussion below, we need a control on various quantities and this seems to be possible only for generic radii. The second ingredient is an almost orthogonality property. The last one is the construction of a competitor for (2.3).

We define the measure $f = \nu \cdot j$ on $\partial B_R \times (0,1)$ and then let $\bar{f} = \int_0^1 f = \nu \cdot \bar{j}$. Before discussing the almost orthogonality property and the construction, let us point out that for our estimates we would need to control the Dirichlet energy $\int_{B_R} |\nabla\Phi|^2$ by $E + D$. Since by elliptic regularity,

$$\int_{B_R} |\nabla\Phi|^2 \lesssim \int_{\partial B_R} \bar{f}^2, \tag{3.4}$$

this is only possible if $\bar{f}$ is controlled in L^2 (or at least in $H^{-1/2}$). In order to solve this issue, (3.3) is first proven with ϕ instead of Φ where ϕ solves

$$\Delta\phi = c \;\; in \; B_R \qquad and \qquad \nu \cdot \nabla\phi = \hat{g} \;\; on \; \partial B_R,$$

where $\hat{g}$ is a regularized version of $\bar{f}$ in the sense that[iii]

$$\int_{\partial B_R} \hat{g}^2 \lesssim E + D \qquad and \qquad W^2(\bar{f}_\pm, \hat{g}_\pm) \lesssim (E + D)^{\frac{d+3}{d+2}}. \tag{3.5}$$

The density $\hat{g}$ is obtained by projection on ∂B_R, using the fact that for 'good' radii, thanks to (2.7), the number of particles crossing ∂B_R is controlled by $E + D$. We will however

---

[iii]For a measure μ, we note $\mu_\pm$ its positive/negative part.

forget here about this difficulty and assume that we may choose $\hat{g} = \bar{f}$ (and thus $\phi = \Phi$). In particular, in view of (3.5), we will assume that we have the bound

$$\int_{\partial B_R} \int_0^1 f^2 \lesssim E + D. \tag{3.6}$$

We may now state the almost-orthogonality property:

Lemma 3.3. *(Orthogonality) For every $0 < \tau \ll 1$, there exist $\varepsilon(\tau) > 0$ and $C(\tau) > 0$ such that if $E + D \leq \varepsilon$,*

$$\int_{B_2} \int_0^1 \frac{1}{\rho} |j - \rho \nabla \Phi|^2 \leq \left(\int_{B_R} \int_0^1 \frac{1}{\rho} |j|^2 - \int_{B_R} |\nabla \Phi|^2 \right) + \tau E + CD. \tag{3.7}$$

Sketch of proof. Expanding the squares we have

$$\int_{B_R} \int_0^1 \frac{1}{\rho} |j - \rho \nabla \Phi|^2 = \left(\int_{B_R} \int_0^1 \frac{1}{\rho} |j|^2 - \int_{B_R} |\nabla \Phi|^2 \right)$$

$$+ 2 \int_{B_R} \int_0^1 (\nabla \Phi - j) \cdot \nabla \Phi + \int_{B_R} (\bar{\rho} - 1) |\nabla \Phi|^2.$$

Let us estimate the two error terms. Using integration by parts we have (assuming without loss of generality that $\int_{B_R} \Phi = 0$)

$$\int_{B_R} \int_0^1 (\nabla \Phi - j) \cdot \nabla \Phi = - \int_{B_R} \int_0^1 (\Delta \Phi - \nabla \cdot j) \Phi + \int_{\partial B_R} \int_0^1 (\bar{f} - f) \Phi$$

$$= - \int_{B_R} \int_0^1 \partial_t \rho \Phi$$

$$= \int_{B_R} \Phi d(\mu - 1)$$

Forgetting higher order terms (and assuming that $\frac{\mu(B_R)}{|B_R|} = 1$), we have (recall that the Wasserstein distance is homogeneous to the H^{-1} norm)

$$\left| \int_{B_R} \Phi d(\mu - 1) \right| \lesssim \left(\int_{B_R} |\nabla \Phi|^2 \right)^{1/2} W_{B_R}(\mu, 1) \overset{(3.4)\&(3.6)}{\lesssim} (E + D)^{1/2} D^{1/2} \overset{\text{Young}}{\leq} \tau E + \frac{C}{\tau} D. \tag{3.8}$$

Regarding the second term, in the case when $\mu = \chi_\Omega$ for some set $B_6 \subseteq \Omega$, we may argue as in [20, Lem. 3.2] and obtain that by McCann's displacement convexity, $\bar{\rho} \leq 1$ and thus $\int_{B_R} (\bar{\rho} - 1) |\nabla \Phi|^2 \leq 0$. For generic measures μ the argument is more subtle and requires a combination of elliptic estimates for (a regularized version of) Φ together with the bound

$$W_{B_R}^2(\bar{\rho}, 1) \lesssim E + D,$$

which holds for 'good' radii.

$\square$

As explained above, the last ingredient is the construction of a competitor:

Lemma 3.4. *For every $0 < \tau \ll 1$, there exist $\varepsilon(\tau) > 0$ and $C(\tau) > 0$ such that if $E + D \leq \varepsilon$, there exists a density-flux pair $(\widetilde{\rho}, \widetilde{j})$ such that*

$$\begin{cases} \partial_t \widetilde{\rho} + \nabla \cdot \widetilde{j} = 0 & in \ B_R \times (0,1) \\ \widetilde{\rho}_0 = \mu, \ \widetilde{\rho}_1 = 1 & in \ B_R \\ \nu \cdot \widetilde{j} = f & on \ \partial B_R \times (0,1) \end{cases} \tag{3.9}$$

and

$$\int_{B_R} \int_0^1 \frac{1}{\widetilde{\rho}} |\widetilde{j}|^2 - \int_{B_R} |\nabla \Phi|^2 \leq \tau E + CD. \tag{3.10}$$

Sketch of proof. We may assume for simplicity that also $\mu = 1$ in B_R. Indeed, otherwise we can connect in the time interval $(0, \tau)$, the measure μ (in B_R) to the constant density 1 at a cost of order $\frac{1}{\tau} W_{B_R}^2(\mu, 1) = \frac{1}{\tau} D$.

Let $0 < r \ll 1$ be a small parameter to be chosen later on. We make the construction separately in the bulk $B_{R-r} \times (0,1)$ and in the boundary layer $B_R \backslash B_{R-r} \times (0,1)$ and set

$$\widetilde{\rho} = \begin{cases} 1 & in \ B_{R-r} \times (0,1) \\ 1 + s & in \ B_R \backslash B_{R-r} \times (0,1), \end{cases} \qquad \widetilde{j} = \begin{cases} \nabla \Phi & in \ B_{R-r} \times (0,1) \\ \nabla \Phi + q & in \ B_R \backslash B_{R-r} \times (0,1), \end{cases}$$

and require that $|s| \leq 1/2$, $\partial_t s + \nabla \cdot q = 0$ in $B_R \backslash B_{R-r} \times (0,1)$, $s_0 = s_1 = 0$ in B_R and $\nu \cdot q = f - \bar{f}$ on $\partial B_R \times (0,1)$, so that (3.9) is satisfied. The existence of an admissible pair (s, q) satisfying the energy bound

$$\int_{B_R \backslash B_{R-r}} \int_0^1 |q|^2 \lesssim r \int_{\partial B_R} \int_0^1 (f - \bar{f})^2 \tag{3.11}$$

as long as $r \gg \left(\int_{\partial B_R} \int_0^1 (f - \bar{f})^2 \right)^{1/(d+1)}$ is obtained arguing by duality, in the same spirit as [2, Lem.3.3] (see [20, Lem. 2.4]).

We may now estimate

$$\int_{B_R} \int_0^1 \frac{1}{\widetilde{\rho}} |\widetilde{j}|^2 - \int_{B_R} |\nabla \Phi|^2 \leq \int_{B_R \backslash B_{R-r}} \int_0^1 \frac{1}{1+s} |\nabla \Phi + q|^2$$

$$\lesssim \int_{B_R \backslash B_{R-r}} |\nabla \Phi|^2 + \int_{B_R \backslash B_{R-r}} \int_0^1 |q|^2$$

$$\overset{(3.11)}{\lesssim} r \int_{\partial B_R} \bar{f}^2 + r \int_{\partial B_R} \int_0^1 (f - \bar{f})^2 \lesssim r \int_{\partial B_R} f^2,$$

where we used that by elliptic regularity, $\int_{B_R \backslash B_{R-r}} |\nabla \Phi|^2 \lesssim r \int_{\partial B_R} \bar{f}^2$. Choosing r to be a large multiple of $\left(\int_{\partial B_R} \int_0^1 (f - \bar{f})^2 \right)^{1/(d+1)}$ yields

$$\int_{B_R} \int_0^1 \frac{1}{\widetilde{\rho}} |\widetilde{j}|^2 - \int_{B_R} |\nabla \Phi|^2 \lesssim \left(\int_{\partial B_R} \int_0^1 f^2 \right)^{\frac{d+2}{d+1}} \overset{(3.6)}{\lesssim} (E + D)^{\frac{d+2}{d+1}},$$

which concludes the proof of (3.10) since $\frac{d+2}{d+1} > 1$ and $E + D \ll 1$. $\qquad\square$

Proof of Theorem 3.2. By (local) minimality of (ρ, j), we have $\int_{B_R} \int_0^1 \frac{1}{\rho} |j|^2 \leq \int_{B_R} \int_0^1 \frac{1}{\tilde{\rho}} |\tilde{j}|^2$ so that combining (3.7) and (3.10) together gives the desired estimate (3.3). $\qquad\square$

4 Application to partial regularity

We now turn to applications of Theorem 3.1 and start with a partial regularity result. Here we are interested in the behavior at small scales.

Let us first recall the main regularity result for optimal transport maps due to Caffarelli [8, 9].

Theorem 4.1. *If μ and λ have compact supports, are absolutely continuous with respect to the Lebesgue measure with densities bounded from above and below on their support and if $\mathrm{spt}\,\lambda$ is convex, then the optimal transport map T from μ to λ is $C^{0,\alpha}$.*

The hypothesis that $\mathrm{spt}\,\lambda$ is convex is not merely technical. Indeed, considering for instance the optimal transport map between one ball and two disjoint balls, it is easy to construct examples where the optimal transport map is discontinuous. However, building on the ideas of Caffarelli to prove Theorem 4.1, Figalli and Kim proved in [15] that even without the convexity assumption on $\mathrm{spt}\,\lambda$, the singular set of T cannot be too big (see also [14] for a generalization to arbitrary non-degenerate cost functions).

Theorem 4.2. *Let μ and λ be probability measures with compact supports, both absolutely continuous with respect to the Lebesgue measure with densities bounded from above and below on their support. Then, there exist open sets $\Omega \subseteq \mathrm{spt}\,\mu$ and $\Omega' \subseteq \mathrm{spt}\,\lambda$ with $|\mathrm{spt}\,\mu \backslash \Omega| = |\mathrm{spt}\,\lambda \backslash \Omega'| = 0$ and such that the optimal transport map T from μ to λ is a $C^{0,\alpha}$ homeomorphism between Ω and Ω'.*

Let us point out that it is actually conjectured that the singular set is much smaller and has the same structure as the singular set of gradients of convex functions i.e. that it is $n - 1$-rectifiable (see [22] for a result in this direction).

A first application of Theorem 3.1 is a new proof of Theorem 4.2 (under the additional hypothesis that μ and λ are Hölder continuous). For the sake of simplicity, we will assume from now on that $\mu = \chi_{\Omega_1}$ and $\lambda = \chi_{\Omega_2}$ for some bounded open sets Ω_i (so that in particular with the notation of Section 3, $D = 0$). As in [14], we derive Theorem 4.2 combining Alexandrov's Theorem (see [27]), which state that T is differentiable a.e., with an $\varepsilon-$regularity theorem.

Theorem 4.3. *([20, Th.1.2]) Let T be the optimal transport map from Ω_1 to Ω_2. For every $\alpha \in (0, 1)$, there exists $\varepsilon(\alpha)$ such that if $R > 0$ is such that $B_{6R} \subseteq \Omega_1 \cap \Omega_2$ and*

$$\frac{1}{R^{d+2}} \int_{B_{6R}} |T - x|^2 \leq \varepsilon,$$

then $T \in C^{1,\alpha}(B_R)$.

By scaling invariance, we may assume that $R = 1$. As already alluded to the proof goes through a Campanato iteration. Indeed, by Campanato's characterization of $C^{1,\alpha}$ spaces (see [10]), it is enough to prove that for every $0 < r \le \frac{1}{6}$,

$$\min_{A,b} \frac{1}{r^{d+2}} \int_{B_r} |T - (Ax + b)|^2 \lesssim r^{2\alpha} \int_{B_1} |T - x|^2.$$

Defining

$$E(T, R) = \frac{1}{R^{d+2}} \int_{B_{6R}} |T - x|^2,$$

this is in turn obtained by using iteratively the following proposition.

Proposition 4.4. *For every $\alpha \in (0, 1)$, there exist $\theta(\alpha) \in (0, 1)$ and $\varepsilon(\alpha)$ such that if $B_{6R} \subseteq \Omega_1 \cap \Omega_2$ and $E(T, R) \le \varepsilon$, there exist a symmetric matrix B with $\det B = 1$ and a vector b such that letting $\hat{T}(x) = B(T(Bx) - b)$,*

$$E(\hat{T}, \theta R) \le \theta^{2\alpha} E(T, R)$$

and $\hat{T}$ is the optimal transport map between $\hat{\Omega}_1 = B^{-1}\Omega_1$ and $\hat{\Omega}_2 = B(\Omega_2 - b)$.

Sketch of proof. By scaling we may assume that $R = 1$.
Let $\tau \ll \frac{\theta^{2\alpha}}{\theta^{d+2}}$ be fixed. Applying Theorem 3.1, we find the existence of a function Φ which is harmonic in B_2 (under our assumptions $c = 0$ in (3.1)) and such that (since $D = 0$)

$$\int_{B_1} |T - (x + \nabla\Phi)|^2 \le \tau E(T, 1) \tag{4.1}$$

and (recall (3.4) and (3.6))

$$\int_{B_2} |\nabla\Phi|^2 \lesssim E(T, 1). \tag{4.2}$$

We then define $b = \nabla\Phi(0)$ and $B = \exp(-A/2)$ where $A = \nabla^2\Phi(0)$. Since Φ is harmonic $\text{Tr} A = 0$ and thus $\det B = 1$. Notice that if $T = \nabla\psi$ for some convex function ψ (by Theorem 2.1), then $\hat{T} = \nabla\hat\psi$ with $\hat\psi(x) = \psi(Bx) - b \cdot x$, which is also a convex function. Therefore $\hat{T}$ is the optimal transport map between $\hat\Omega_1$ and $\hat\Omega_2$. We may now estimate

$$E(\hat{T}, \theta) = \frac{1}{\theta^{d+2}} \int_{B_{6\theta}} |\hat{T} - x|^2$$

$$= \frac{1}{\theta^{d+2}} \int_{B^{-1}(B_{6\theta})} |B(T - b) - B^{-1}x|^2$$

$$\lesssim \frac{1}{\theta^{d+2}} \int_{B_{7\theta}} |T - b - B^{-2}x|^2$$

$$\lesssim \frac{1}{\theta^{d+2}} \int_{B_1} |T - (x + \nabla\Phi)|^2 + \frac{1}{\theta^{d+2}} \int_{B_{7\theta}} |\nabla\Phi - (Ax + b)|^2$$

$$\quad + \frac{1}{\theta^{d+2}} \int_{B_{7\theta}} |(\exp A - \text{id} - A)x|^2$$

$$\overset{(4.1)}{\lesssim} \frac{\tau}{\theta^{d+2}} E(T, 1) + \theta^2 \sup_{B_{7\theta}} |\nabla^3\Phi|^2 + |A|^4 \overset{(4.2)}{\lesssim} \frac{\tau}{\theta^{d+2}} E(T, 1) + \theta^2 E(T, 1) + E(T, 1)^2.$$

8

This concludes the proof since we chose $\tau \ll \frac{\theta^{2\alpha}}{\theta^{d+2}}$ and since $E(T,1) \ll 1$. $\qquad\square$

5 Application to the optimal matching problem

We now present an application to the optimal matching problem. As opposed to the previous section, we are interested here at large scales.

Over the last thirty years, optimal matching problems have been the subject of intensive work. We refer for instance to the monograph [26]. One of the simplest example is the problem of matching the empirical measure of a Poisson point process to the corresponding Lebesgue measure. More specifically, we consider for $L \gg 1$ a Poisson point process μ on the the torus $Q_L = [-L/2, L/2)^d \simeq (\mathbb{R}/L\mathbb{Z})^d$ i.e.

$$\mu = \sum_{i=1}^{n} \delta_{X_i}$$

with X_i iid random variables uniformly distributed in Q_L and n a random variable with Poisson distribution with parameter L^d. The problem is to estimate the random variable

$$\frac{1}{L^d} W_{\text{per}}^2(\mu, \kappa),$$

where W_{per} indicates the Wasserstein distance on the torus Q_L, and to understand the structure of the corresponding optimal transport plans. It is well-known since [1] that[iv]

$$\mathbb{E}\left[\frac{1}{L^d} W_{\text{per}}^2(\mu, \kappa)\right] \sim \begin{cases} \log L & \text{if } d = 2 \\ 1 & \text{for } d \geq 3 \end{cases} \tag{5.1}$$

and thus $d = 2$ is a critical dimension. Recently, Caracciolo and al. used the ansatz that the optimal displacement should be well approximated by $\nabla\varphi_L$, where φ_L solves the Poisson equation (recall (2.2))

$$\Delta\varphi_L = \mu - \kappa \qquad \text{in } Q_L \tag{5.2}$$

to make numerous predictions about the optimal prefactor in (5.1) as well as the correlations (see [11, 12]). At the macroscopic scale, this ansatz has been partially rigorously justified by Ambrosio and al. (see [4, 3] and also [23] for a result about the fluctuations) in dimension 2. To state their result[v], let us introduce some notation. For $t > 0$, denote the heat kernel on Q_L by P_t and let $\varphi_{L,t} = P_t * \varphi_L$, so that $\varphi_{L,t}$ solves

$$\Delta\varphi_{L,t} = P_t * \mu - \kappa \qquad \text{in } Q_L.$$

[iv]We use the notation log for the natural logarithm.

[v]The results of [4, 3] are stated on the unit cube with a (deterministic) number of points $n \to \infty$. However, their results may be easily transposed into our setting by scaling.

Theorem 5.1. *Let $d = 2$, then*

$$\lim_{L \to \infty} \frac{1}{\log L} \mathbb{E}\left[\frac{1}{L^2} W_{per}^2(\mu, \kappa)\right] = \frac{1}{2\pi}. \tag{5.3}$$

Moreover, if π_L is the optimal transport plan between μ and κ, then setting $t_L = \log^4 L$, for $L \gg 1$ there holds

$$\frac{1}{\log L} \mathbb{E}\left[\frac{1}{L^2} \int_{Q_L \times Q_L} |y - x - \nabla\varphi_{L,t_L}|^2 d\pi_L\right] \ll \left(\frac{\log \log L}{\log L}\right)^{1/2}. \tag{5.4}$$

Since by (5.3), the displacement $y - x$ is on average of the order of $\log^{\frac{1}{2}} L$, (5.4) shows that $\nabla\varphi_{L,t_L}$ indeed coincides with the displacement to leading order. This leaves open the description of the optimal transport plan π_L at the microscopic scale. To state our main result, fix a smooth cut-off function (which plays a similar role as the heat kernel in Theorem 5.1)

$$\eta \in C_c^\infty(B_1) \text{ with } \int_{\mathbb{R}^2} \eta = 1, \qquad \text{and set} \qquad \eta_R = \frac{1}{R^2}\eta\left(\frac{\cdot}{R}\right).$$

In [17], we prove the following result (see also [19, Th. 1.2] and [18, Th. 1.1]):

Theorem 5.2. *There exists a stationary random variable $r_* \geq 1$ on Q_L with exponential moments such that if $\bar{x} \in Q_L$ is such that $r_*(\bar{x}) \ll L$, then*

$$\left|\frac{\int_{Q_L \times Q_L} \eta_R(x - \bar{x})(y - x)d\pi}{\int_{Q_L \times Q_L} \eta_R(x - \bar{x})d\pi} - \eta_R * \nabla\varphi_L(\bar{x})\right| \lesssim \frac{\log R}{R} \qquad \forall L \gtrsim R \gtrsim r_*(\bar{x}). \tag{5.5}$$

Moreover, there exists $R = R(\bar{x}) \sim r_*(\bar{x})$ such that defining the shift h by $h(\bar{x}) = \frac{1}{|B_R|}\int_{\partial B_R(\bar{x})}(x - \bar{x})\nu \cdot \nabla\varphi_L$, we have

$$|h(\bar{x})|^2 \lesssim \log L \tag{5.6}$$

and

$$\sup\{|y - x - h(\bar{x})| : (x, y) \in \text{spt } \pi \cap (B_{r_*}(\bar{x}) \times \mathbb{R}^2)\} \lesssim r_*(\bar{x})\left(\frac{\log r_*(\bar{x})}{r_*^2(\bar{x})}\right)^{1/4}. \tag{5.7}$$

With respect to (5.4), (5.7) proves that (circular) averages of $\nabla\varphi_L$ coincide with the displacement $y - x$ up to an error which is of order one. Moreover, (5.5) shows that after averaging, the displacement is actually extremely close to averages of $\nabla\varphi_L$ (notice that the error term $\log R/R$ improves as R increases).

By stationarity, it is enough to prove Theorem 5.2 for $\bar{x} = 0$. The proof is based on the following deterministic result (which is a small post-processing of [19, Th. 1.2]):

Theorem 5.3. *Let μ be a measure on Q_L. If for some $L \gg r \gg 1$,*

$$\frac{1}{R^2} W_{B_R}^2(\mu, \kappa) \lesssim \log R \qquad \text{for all dyadic } L \gtrsim R \gtrsim r, \tag{5.8}$$

then

$$\left| \frac{\int_{Q_L \times Q_L} \eta_R(x)(y-x) d\pi}{\int_{Q_L \times Q_L} \eta_R(x) d\pi} - \int_{Q_L} \eta_R \nabla \varphi_L \right| \lesssim \frac{\log R}{R} \qquad \forall L \gtrsim R \gtrsim r. \tag{5.9}$$

Moreover, there exists $R \sim r$ such that letting $h = \frac{1}{|B_R|} \int_{\partial B_R} x \nu \cdot \nabla \varphi_L$, we have

$$\frac{1}{r^2} \int_{(B_r \times \mathbb{R}^2) \cup (\mathbb{R}^2 \times B_r(h))} |y - x - h|^2 d\pi \lesssim \log r. \tag{5.10}$$

Notice that (5.7) follows from (5.10) and the L^∞ bound (2.7) of Lemma 2.3. In order to obtain Theorem 5.2, Theorem 5.3 is combined with a stochastic argument based on (5.3) and a concentration-of-measure argument which ensures that (5.8) is satisfied for the Poisson point process μ.

The main ingredient for the proof of Theorem 5.3 is a Campanato iteration scheme similar to the one leading to Theorem 4.3 (and mainly based on Theorem 3.1) which allows to transfer the information that (5.10) holds at scale L by (5.8) down to the microscopic scale r. This is inspired by the approach developed by Armstrong and Smart in [6] (and further refined in [16], see also [5]) for quantitative stochastic homogenization. The main ideas of [6] take roots themselves in previous works of Avellaneda and Lin (see [7]) on periodic homogenization. The outcome of the Campanato scheme may be stated as follows (see [19, Prop. 1.9])

Proposition 5.4. *There exists a sequence of approximately geometric radii R_k i.e. $L \geq R_0 \geq \cdots \geq R_K \gtrsim 1$ with $R_{k-1} \geq 2R_k \gtrsim R_{k-1}$, $R_0 \sim L$ and $R_K \sim r$ such that defining recursively the couplings π_k by $\pi_0 = \pi$ and*

$$\pi_k = (id, id - \nabla \Phi_{k-1}(0)) \# \pi_{k-1}$$

where Φ_k solves

$$\Delta \Phi_k = c \ \ in \ B_{R_k} \qquad and \qquad \nu \cdot \nabla \Phi_k = \nu \cdot \bar{j}_k \ \ on \ \partial B_{R_k}$$

with j_k defined as in (2.4) with π_k playing the role of π, we have for $k \in [0, K]$,

$$\frac{1}{R_k^2} \int_{(B_{6R_k} \times \mathbb{R}^2) \cup (\mathbb{R}^2 \times B_{6R_k})} |x - y|^2 d\pi_k \lesssim \log R_k \tag{5.11}$$

and

$$|\nabla \Phi_k(0)|^2 \lesssim \log R_k. \tag{5.12}$$

Let us point out that by invariance of the Lebesgue measure under translations, π_k is the optimal transport plan between μ and the Lebesgue measure for every k (this is the reason why we make the translation in the target space).

Letting $\widetilde{h} = \sum_{k=0}^{K-1} \nabla \Phi_k(0)$ and undoing the iterative definition of π_k, we see that (5.11) directly leads to (5.10) with h replaced by $\widetilde{h}$. The proof of (5.10) is concluded by the estimate (see [19, Prop. 1.10])

$$|h - \widetilde{h}| \lesssim \frac{\log r}{r}.$$

This estimate is also crucial for the proof of (5.9). Let us point out that a naive estimate using (5.12) leads to

$$|\widetilde{h}|^2 \lesssim \log^3 L$$

which is suboptimal. In order to obtain a shift with the optimal estimate (5.6) it is therefore important to take into account cancellations and replace $\widetilde{h}$ by h.

Let us close this note by pointing out that in dimension $d \geq 3$, the optimal transport plans corresponding to a very closely related optimal matching problem, have been used in [21] to construct in the limit $L \to \infty$, a stationary and locally optimal coupling between the Poisson point process on $\mathbb{R}^d$ and the Lebesgue measure. For $d = 2$, such a coupling is expected not to exist. However, using (5.7) and passing to the limit $L \to \infty$, it is possible to construct (at least in the sense of Young measures) a coupling between the Poisson point process on $\mathbb{R}^2$ and the Lebesgue measure, which is locally optimal and has stationary *increments* (see [18, Th.1.2]).

Acknowledgements

References

[1] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings.*, Combinatorica **4** (1984), 259–264.

[2] G. Alberti, R. Choksi, and F. Otto, *Uniform energy distribution for an isoperimetric problem with long-range interactions*, J. Amer. Math. Soc. **22** (2009), no. 2, 569–605.

[3] L. Ambrosio, F. Glaudo, and D. Trevisan, *On the optimal map in the 2-dimensional random matching problem*, 2019, p. 20.

[4] L. Ambrosio, F. Stra, and D. Trevisan, *A PDE approach to a 2-dimensional matching problem*, Probab. Theory Related Fields **173** (2019), no. 1-2, 433–477.

[5] S. Armstrong, T. Kuusi, and J.-C. Mourrat, *Quantitative stochastic homogenization and large-scale regularity*, ArXiv e-prints (2017).

[6] S. N. Armstrong and C. K. Smart, *Quantitative stochastic homogenization of convex integral functionals*, Ann. Sci. Éc. Norm. Supér. (4) **49** (2016), no. 2, 423–481.

[7] M. Avellaneda and F.-H. Lin, *Compactness methods in the theory of homogenization*, Comm. Pure Appl. Math. **40** (1987), no. 6, 803–847.

[8] L. A. Caffarelli, *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity*, Ann. of Math. (2) **131** (1990), no. 1, 129–134.

[9] ———, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc. **5** (1992), no. 1, 99–104.

[10] S. Campanato, *Proprietà di una famiglia di spazi funzionali*, Ann. Scuola Norm. Sup. Pisa (3) **18** (1964), 137–160.

[11] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem*, Physical Review E **90** (2014), no. 1.

[12] S. Caracciolo and G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem. II. Correlation functions*, Physical Review E **91** (2015), no. 6.

[13] G. De Philippis and A. Figalli, *The Monge-Ampère equation and its link to optimal transportation*, Bull. Amer. Math. Soc. (N.S.) **51** (2014), no. 4, 527–580.

[14] ———, *Partial regularity for optimal transport maps*, Publ. Math. Inst. Hautes Études Sci. **121** (2015), 81–112.

[15] A. Figalli and Y.-H. Kim, *Partial regularity of Brenier solutions of the Monge-Ampère equation*, Discrete Contin. Dyn. Syst. **28** (2010), no. 2, 559–565.

[16] A. Gloria, S. Neukamm, and F. Otto, *A regularity theory for random elliptic operators*, ArXiv e-prints (2014).

[17] M. Goldman, M. Huesmann, and F. Otto, *In preparation*.

[18] ———, *A large-scale regularity theory for the Monge-Ampere equation with rough data and application to the optimal matching problem*, arXiv e-prints (2018), arXiv:1808.09250.

[19] ———, *Quantitative linearization results for the Monge-Ampère equation*, arXiv e-prints (2019).

[20] M. Goldman and F. Otto, *A variational proof of partial regularity for optimal transportation maps*, arXiv:1704.05339 (2017).

[21] M. Huesmann and K.-T. Sturm, *Optimal transport from Lebesgue to Poisson*, Ann. Probab. **41** (2013), no. 4, 2426–2478.

[22] J. Kitagawa and R. McCann, *Free discontinuities in optimal transport*, Arch. Ration. Mech. Anal. **232** (2019), no. 3, 1505–1541.

[23] M. Ledoux, *A fluctuation result in dual Sobolev norm for the optimal matching problem*, 2019.

[24] F. Maggi, *Sets of finite perimeter and geometric variational problems*, Cambridge Studies in Advanced Mathematics, vol. 135, Cambridge University Press, Cambridge, 2012, An introduction to geometric measure theory.

[25] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015, Calculus of variations, PDEs, and modeling.

[26] M. Talagrand, *Upper and lower bounds for stochastic processes: modern methods and classical problems*, vol. 60, Springer Science & Business Media, 2014.

[27] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.