Durham University

MSc Mathematics Thesis 2023-24
Department of Mathematical Sciences

Partial Regularity for Optimal Transport Maps Between Uniform Measures

Author:
William Ford

Supervisor:
Dr Alpár Mészáros

September 2024

Acknowledgements

This piece of work is a result of my own work except where it forms an assessment based on group project work. In the case of a group project, the work has been prepared in collaboration with other members of the group. Material from the work of others not involved in the project has been acknowledged and quotations and paraphrases suitably indicated.

I would like to thank my supervisor Dr Alpár Mészáros for his advice and support throughout this year and in preparing this report. His help is greatly appreciated.

Contents

Introduction

History of optimal transportation theory

The story of optimal transport begins in France with mathematician Gaspard Monge, a few years before the revolution of 1789. Monge was a prominent figure in French society: He served on the committee involved in establishing the metric system, was involved in the founding of the École Polytechnique, and was a friend and aide of Napoleon[8]. Our interest with Monge, however, lies in his 1781 paper [21] submitted to the *Académie des sciences*, titled "*Mémoire sur la théorie des déblais et des remblais*". The paper discusses the problem of moving piles of earth to another location/arrangement in the most economical way.

Specifically, the Monge problem (in its modern form in the language of measure theory, which was yet to be developed) is as follows. Given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, representing a distribution of mass, one considers maps $T : \mathbb{R}^d \to \mathbb{R}^d$ that "push" the mass distribution of μ onto that of ν meaning

$$\mu(T^{-1}A) = \nu(A) \quad \text{for each measurable } A \subset \mathbb{R}^d,$$

written $T_\# \mu = \nu$. This is saying that the mass arriving at each A is precisely the amount that was sent there under T. We wish to choose such a T minimising the transport cost

$$\int_{\mathbb{R}^d} |T(x) - x| \mathrm{d}\mu(x),$$

which represents an average of the Euclidean distances particles have moved. In the case that μ and ν are absolutely continuous, one can imagine these mass distributions as piles of sand given by the graphs of the densities, and we wish to move one pile to the other in the most cost-efficient way. If μ and ν have densities f and g and T is differentiable, then this condition is equivalent to the Jacobian equation

$$g(T(x)) \det DT(x) = f(x). \tag{1}$$

This problem has some clear generalisations - we need not be in a Euclidean space, and we might also hope to understand a more general cost function. For spaces X and Y between which we are transporting $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, and a cost density $c(x, y) : X \times Y \to \mathbb{R}$, Monge's problem generalises to

$$\text{(MP)} \quad \min \left\{ \int_X c(x, T(x)) \mathrm{d}\mu(x) \quad : \quad T_\# \mu = \nu \right\}.$$

In general, this problem can be ill-posed in multiple ways: uniqueness can fail, sometimes a minimiser does not exist, and sometimes there is no transport map at all, as we demonstrate in the following canonical example.

Example 0.1 (No transport exists)**.** Let $\mu = \delta_0 \in \mathcal{P}(\mathbb{R})$ be a Dirac mass, and consider the target measure $\nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$; two Dirac masses. Then no transport map can exist as the image measure $T_{\#}\mu$ under any map T will be a single Dirac $\delta_{T(0)}$. The philosophical limitation here is that the Monge problem does not allow "splitting" of mass - we cannot break up the Dirac and send half of the mass to each of the atoms of ν.

Fast forward to 1942 in the Soviet Union. Mathematician and economist Leonid Kantorovich proposes a similar problem in [18], discussing applications "location of consumption stations with respect to production station" for the optimal planning of railways, and "levelling a land area" in which Kantorovich states a similar motivation to that of Monge.

Kantorovich's idea is that instead of maps, one considers "transport plans" represented by measures $\gamma \in \mathcal{P}(X \times Y)$. Here $\gamma(A \times B)$ represents the amount of mass transported from the set $A \subseteq X$ to $B \subseteq Y$. One views $B \mapsto \gamma(\{x\} \times B)$ as describing the image distribution of the mass starting at x (the crucial point being that this disintegration measure need not be a Dirac mass, so we can "split" the mass at x across the posterior space). Then the restrictions on the prior and posterior mass distributions become conditions on the marginals of γ: we want that for each measurable $A \subseteq X$ and $B \subseteq Y$

$$\mu(A) = \gamma(A \times Y) \quad \text{and} \quad \nu(B) = \gamma(X \times B);$$

so that the total mass leaving X is distributed according to μ and the total mass which arrives in Y is distributed according to ν. Simple calculations show this is equivalent[1] to asking $\pi_{x\#}\gamma = \mu$ and$\pi_{y\#}\gamma = \nu$. We write $\Pi(\mu, \nu)$ for all such measures. We observe that this is never empty, as one has at least the product measure $\mu \times \nu$; so unlike in Example 0.1, there is always a transport plan (in this case the product measure is the only transport plan, and crucially this splits the mass at 0 in μ, which (MP) does not allow). This leads to Kantorovich's problem

$$\text{(KP)} \quad \min \left\{ \int_{X \times Y} c(x, y) \mathrm{d}\gamma(x, y) \quad : \quad \gamma \in \Pi(\mu, \nu) \right\}.$$

There are multiple reasons why (KP) is more tractable - first the linearity of the problem, and second that the constraint is closed under weak convergence. We will explore these in detail in Chapter 1.

[1] Here, π_x, π_y are the projections from $X \times Y$ onto the respective coordinates.

Structure of the report

- In Chapter 1, we discuss the core theory regarding the Monge and Kantorovich problems. Focusing on the more tractable (KP) first, we prove a general existence result, characterise optimal plans for an arbitrary continuous cost function, and formulate and prove the equivalence of a dual problem. We then consider (MP) and its relations to (KP), proving a general equivalence between the minimisation costs for suitably regular source measure μ, existence and uniqueness of an optimal map for the case $c(x, y) = h(y - x)$ for h strictly convex, and discuss the quadratic case $h(z) = |z|^2$ and some of its niceties. Finally we look at the Benamou-Brenier formulation of optimal transport, which recasts the problem in the language of continuum mechanics.

- In Chapter 2, we consider some function spaces due to S. Campanato and C. Morrey, which we show, for certain parameters, are isomorphic to Hölder spaces. Campanato spaces consist of functions that can locally be approximated by degree k polynomials suitably strongly in L^q norm at each point. We show that for certain parameters this guarantees k derivatives, with the kth order ones Hölder continuous. These results will aid us in establishing regularity for optimal transport maps and elliptic PDE in the following chapters.

- In Chapter 3, we step away from optimal transport to discuss the regularity of weak solutions to a certain class of linear elliptic PDE. We discuss when solutions lie in both Sobolev and Hölder spaces, depending on the coefficients of the equation. The former is established using a difference quotient which we uniformly bound in L^q norm, from which functional analysis tools give a convergent subsequence whose limit we show is a weak derivative. The Hölder regularity is established obliquely, showing that the solutions lie in Campanato and Morrey spaces.

- In Chapter 4, we present the recent work of Goldman and Otto [16, 15], a partial regularity $C^{1,\alpha}$ result for optimal transport maps between uniform measures. This is established using the Benamou-Brenier formulation to construct a harmonic function who's gradient approximates the transport displacement of the optimal map T in the Lagrangian setting. We then prove a so called "epsilon regularity" result which allows us to iterate approximating T better with harmonic gradients. This is sufficiently fast that the first order Taylor series of the harmonic gradient is a sufficient competitor to show the map lies in a certain Campanato space, which gives the Hölder regularity.

1 Lagrangian and Eulerian formulations of optimal transport

This chapter covers the core results regarding the Monge and Kantorovich problems introduced in the introduction, as well as considering a third framework from the Eulerian viewpoint due to Benamou and Brenier.

1.1 The Kantorovich problem

1.1.1 Existence

We begin by stating a general existence theorem for (KP) defined in the introduction, following the references [25, Ch. 1] and [2, Ch. 1]. We start by stating explicitly what we mean by weak convergence.

Definition 1.1. *We say a sequence $\mu_n \in \mathcal{P}(X)$ converges narrowly to μ if $\mu_n \rightharpoonup \mu$ weakly in duality with $C_b(X)$, so that $\int_X \phi \mathrm{d}\mu_n \to \int_X \phi \mathrm{d}\mu$ for all $\phi \in C_b(X)$. Narrow convergence is, in general, non-metrisable, but it is a relatively powerful notion: if $\mu_n \rightharpoonup \mu$ then $\mu_n(A) \to \mu(A)$ for any Borel set $A \subset X$.*

To show the existence of a minimiser we use the Prokhorov theorem which gives the equivalence of tight and compact sets of measures w.r.t the narrow topology (see appendix Theorem B.2). Once we have the compactness, the proof follows by standard variational arguments.

Theorem 1.2. *Let X and Y be Polish spaces, with $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ be l.s.c. and bounded from below. Then (KP) admits a solution.*

Proof. [Step 1: $\Pi(\mu, \nu)$ is compact w.r.t the narrow topology .]

We first show $\Pi(\mu, \nu)$ is tight. Fix $\varepsilon > 0$ and note that since singletons are compact and hence tight, the converse of Prokhorov gives the existence of $K_X \Subset X$ and $K_Y \Subset Y$ with $\mu(X \setminus K_X) < \varepsilon/2$ and $\nu(Y \setminus K_Y) < \varepsilon/2$. Then for any $\gamma \in \Pi(\mu, \nu)$ we have

$$\gamma((X \times Y) \setminus (K_X \times K_Y)) \leq \gamma((X \setminus K_X) \times Y) + \gamma(X \times (Y \setminus K_Y)) = \mu(X \setminus K_X) + \nu(Y \setminus K_Y) < \varepsilon.$$

Thus we have that $\Pi(\mu, \nu)$ is tight with $K_X \times K_Y$ compact sufficing for the ε challenge. Prokhorov then gives relative compactness. For the full compactness, consider consider a convergent sequence $\gamma_n \in \Pi(\mu,\nu)$ with $\gamma_n \rightharpoonup \gamma$. Then taking $\phi \in C_b(X \times Y)$ as $\phi = \chi_{A\times Y}$ for any measurable A gives

$$\mu(A) = \int_{X\times Y} \chi_{A\times Y} \mathrm{d}\gamma_n \to \int_{X\times Y} \chi_{A\times Y} \mathrm{d}\gamma = \pi_{x\#}\gamma(A),$$

and since the LHS is constant γ has X marginal μ, and a symmetrical argument works on Y so that $\gamma \in \Pi(\mu,\nu)$ and $\Pi(\mu,\nu)$ is narrowly compact.

[Step 2: $K(\gamma) = \int c\mathrm{d}\gamma$ is lower semi-continuous w.r.t the narrow topology.]

Take $\gamma_n \rightharpoonup \gamma$, then as c is l.s.c. and bounded below, there exists a sequence of continuous functions c_k monotone increasing from below to c[1]. By definition of narrow convergence $\int c_k \mathrm{d}\gamma_n \to \int c_k \mathrm{d}\gamma$ for each k, so that

$$\liminf_n \int c(x,y)\mathrm{d}\gamma_n \geq \liminf_n \int c_k(x,y)\mathrm{d}\gamma_n = \int c_k(x,y)\mathrm{d}\gamma,$$

by monotonicity. Then monotone convergence in k gives the result.

[Step 3: A minimiser for (KP) exists.] This follows from standard arguments from the calculus of variations. Take a minimising sequence, pass to a subsequence by the compactness of $\Pi(\mu,\nu)$, then the limit measure is a minimiser by lower semi-continuity. □

1.1.2 Characterisation of optimal plans for a continuous cost

In this section, we seek to understand (KP) in the case of continuous cost, mostly following references [2, 25].

Definition 1.3. *For a function $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$, we define the c-transform of φ by*

$$\varphi^c : Y \to \mathbb{R} \quad \varphi^c(y) = \inf_{x\in X} c(x,y) - \varphi(x).$$

and similarly the $\bar{c}$-transform of a function $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ by

$$\psi^{\bar{c}} : X \to \mathbb{R} \quad \psi^{\bar{c}}(x) = \inf_{y\in Y} c(x,y) - \psi(y).$$

We say a function is c-concave if it is the c or $\bar{c}$ transform of some function.

Definition 1.4. *For a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$, we define the c-superdifferential of φ by pairs x, y satisfying.*

$$\partial^c\varphi := \{(x,y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x,y)\}$$

[1] For $f : X \to \mathbb{R}$ l.s.c. one can take $f_k(x) = \inf_y f(y) + kd(x,y)$.

Remark 1.5. We define c-concavity in terms of an inf, and this means one can easily verify that for any c-concave function, $\varphi^{c\bar{c}} = \varphi$. For the case $c(x, y) = -\langle x, y\rangle$, the c-transform becomes the (negated) Legendre transform for concave functions. We usually define concave functions by some interpolation inequality rather than as an inf of affine functions, so establishing that the double Legendre transform of any concave function returns to itself is not so trivial.

Definition 1.6. *For a cost $c : X \times Y \to \mathbb{R}$, a set $\Gamma \subset X \times Y$ is said to be c-cyclically monotone (c-CM) if for any collection of points $(x_1, y_1), \dots (x_n, y_n) \subset \Gamma$, and any permutation $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$, we have*

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}).$$

By relabeling points, one can replace the need for $\sigma(i)$ with simply $i + 1 \in \mathbb{Z}/n$.

As in [2]; consider $\mu = \sum \frac{1}{n}\delta_{x_i}$ and $\nu = \sum \frac{1}{n}\delta_{y_i}$ as unweighted sums of Dirac masses in a Euclidean space, with the same number of atoms in each. It is clear that having c-CM support is a necessary and sufficient condition for an optimal plan; this simply says that there is no way we can swap where mass is sent to reduce the overall cost.

If the cost is continuous, this can be seen to also be necessary for general non-atomic measures, despite the above only being a condition on countable collections of points. Any neighbourhood of any point in the support of some plan γ has some mass transported through it, and by continuity, if $\operatorname{Spt}\gamma$ was not c-CM we would find some regions of positive mass on which for some permutation, the sum of costs was strictly larger. We could then construct a competitor with lower transport cost by permuting where these regions of mass are sent. We formalise this in the below theorems as found in [2, 25], showing that this criterion completely characterises optimisers of (KP) when c is continuous.

Theorem 1.7. *Let $\gamma \in \Pi(\mu, \nu)$ be optimal for (KP) with cost c be continuous. Then* $\operatorname{Spt}\gamma$ *is c-cyclically monotone.*

Proof. Assume γ has support not c-cyclically monotone, call this $\Gamma := \operatorname{Spt}\gamma$. Then there exist pairs of points $(x_1, y_1), \dots (x_n, y_n) \subset \Gamma$ such that

$$\sum_{i=1}^{n} c(x_i, y_{i+1}) < \sum_{i=1}^{n} c(x_i, y_i).$$

By continuity, we can find neighbourhoods $x_i \in U_i$, $y_i \in V_i$ such that

$$\sum_{i=1}^{n} c(u_i, v_{i+1}) - \sum_{i=1}^{n} c(u_i, v_i) < 0 \quad \text{for all } u_i, v_i \in U_i, V_i.$$

We now modify γ to create a new plan $\tilde{\gamma} = \gamma + m$ with lower transport cost. To maintain $\tilde{\gamma} \in \Pi(\mu, \nu)$, we require that m is a signed measure with zero X and Y marginals and $\gamma \geq m^-$ so that $\tilde{\gamma}$ is a positive measure. Since $x_i, y_i \in \Gamma$ we have $\gamma(U_i \times V_i) > 0$ for each i, so that we can define the measures $\gamma_i = \frac{1}{\gamma(U_i \times V_i)} \gamma|_{U_i \times V_i}$, with marginals μ_i and ν_i. Then for each i, we can build a measure $\tilde{\gamma_i} \in \Pi(\mu_i, \nu_{i+1})$, and define m by

$$m = \frac{\min_i \gamma(U_i \times V_i)}{n} \sum_{i=1}^{k} (\tilde{\gamma_i} - \gamma_i).$$

Then $\tilde{\gamma} = \gamma + m$ is admissible and has a lower transport cost than γ, so γ is not optimal. □

The following is another characterisation of optimal plans, found in [2, 25].

Theorem 1.8. *Suppose that c is continuous, bounded below, assume there exists $f \in L^1(\mu)$ and $g \in L^1(\nu)$ such that*

$$c(x, y) \leq f(x) + g(y). \tag{1.1}$$

Suppose that $\Gamma \subset X \times Y$ is a c-cyclically monotone set, then there exists a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ (which is not constantly $-\infty$) such that $\max(\varphi, 0) \in L^1(\mu)$ *and $\Gamma \subset \partial^c \varphi$.*

Proof. Omitted. The details can be found in the above references. □

1.1.3 The dual formulation

Since (KP) is a linear minimisation with linear constraints, it admits a natural dual problem, given by maximising some functional with linear constraints. We give two proofs: the first is more derivative à la [25], using Theorem B.3 for an infinite-dimensional inf-sup exchange. The idea is to turn the constraint on the marginals into a sup, then attempt to justify exchanging inf and sup due to convexity and concavity in the different variables (here it is just linearity). The second instead uses the cyclical-monotonicity as in [2].

Theorem 1.9 (Duality)**.** *Let $c : X \times Y \to R$ be continuous and bounded below, and satisfying (1.1). Consider the problems*

$$\textit{(KP)} \quad \min \left\{ \int_{X \times Y} c(x, y) \mathrm{d}\gamma \quad : \quad \gamma \in \Pi(\mu, \nu) \right\}$$

and

$$\textit{(DP)} \quad \sup_{\varphi, \psi} \left\{ \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu \ : \ \varphi(x) + \psi(y) \leq c(x, y) \right\}, \tag{1.2}$$

where the sup is over $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. We have sup *(DP)* = min *(KP), and the maximum of (DP) is attained.*

Proof using saddle point theory. With this in mind relaxing to allow any $\gamma \in \mathcal{P}(X \times Y)$, the constraint can be expressed as

$$\sup_{\varphi,\psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \int_{X\times Y} (\varphi(x) + \psi(y))\mathrm{d}\gamma = \begin{cases} 0 & \gamma \in \Pi(\mu,\nu) \\ +\infty \text{ if not,} \end{cases}$$

where here we sup over $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$. By adding the constraint to (KP), the problem becomes

$$\inf_{\gamma} \sup_{\varphi,\psi} \int_{X\times Y} c\mathrm{d}\gamma + \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \int_{X\times Y} (\varphi(x) + \psi(y))\mathrm{d}\gamma.$$

Since the expression is linear in all of γ, φ, ψ, we want to interchange inf and sup, and we justify this by Theorem B.3, as taking $\varphi, \psi \equiv 0$, we know $\Pi(\mu, \nu)$ is narrowly compact and values of the transport cost are bounded over this set by (1.1). Thus the problem becomes

$$\sup_{\varphi,\psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \inf_{\gamma} \int_{X\times Y} c(x,y) - \varphi(x) + \psi(y)\mathrm{d}\gamma.$$

Here the inner inf has value 0 if $\varphi(x) + \psi(y) \leq c(x,y)$ and $-\infty$ otherwise, as we inf over all positive measures. Thus we can write this inf instead as a constraint, giving the dual formulation

$$\text{(DP)} \quad \sup_{\varphi,\psi} \left\{ \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu \ : \ \varphi(x) + \psi(y) \leq c(x,y) \right\}.$$

□

Now as in [2] we show duality another way, also establishing that the sup is a max.

Proof using c-cyclical monotonicity. For $\sup(\text{DP}) \leq \inf(\text{KP})$, take any admissible φ, ψ, γ. We have

$$\int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu = \int_{X\times Y} \varphi + \psi \mathrm{d}\gamma \leq \int_{X\times Y} c\mathrm{d}\gamma.$$

Passing a sup and inf on either side gives the required inequality.

For $\sup(\text{DP}) \geq \inf(\text{KP})$, take an optimal γ, then by Theorem 1.8, we have a pair of potentials with $\max(\varphi, 0) \in L^1(\mu)$ and $\max(\varphi^c, 0) \in L^1(\nu)$ for which $\operatorname{Spt}\gamma$ lies in the c-supergradient. Thus

$$\int_{X\times Y} c\mathrm{d}\gamma = \int_{X\times Y} \varphi + \varphi^c \mathrm{d}\gamma = \int \varphi \mathrm{d}\mu + \varphi^c \mathrm{d}\gamma,$$

so that $\varphi \in L^1(\mu)$ and $\varphi^c \in L^1(\nu)$ and the duality is proven. □

Remark 1.10. Solutions to the dual problem are referred to as Kantorovich potentials.

1.2 The Monge problem

We now turn towards the historical predecessor, Monge's formulation. Some natural questions which we will answer are:

- What is the relation between transport maps and transport plans?
- Are the values of the minimisation problems the same?
- When does an optimal map exist? When is it unique?

1.2.1 Equivalence of the minimisation cost for μ atomless

We first want to understand the relationship to (KP). Let $T : X \to Y$ be a transport map between μ and ν. Then we can define a measure $\gamma_T \in \mathcal{P}(X \times Y)$ by

$$\gamma_T := (\mathrm{Id} \times T)_{\#}\mu.$$

It is easily verified that for any $A \subset X$ and $B \subset Y$

$$\gamma_T(A \times Y) = \mu(A); \qquad \gamma_T(X \times B) = \nu(B)$$

so that $\gamma_T \in \Pi(\mu, \nu)$. We also note that $\operatorname{Spt}\gamma \subseteq \{(x, T(x)) : x \in X\}$. Turning to the respective costs of (MP) and (KP) we have

$$\int_{X\times Y} c(x,y)\mathrm{d}\gamma_T = \int_{X\times Y} c(x,T(x))\mathrm{d}\gamma_T = \int_X c(x,T(x))\mathrm{d}\mu,$$

where we first used the support concentration to replace y with $T(x)$, and then that the integrand only depended on x to replace the integration with the X marginal. Thus the transport costs are the same, so we see that this plan encapsulates the behaviour of map T.

An immediate corollary of this is also that $\min(\mathrm{KP}) \leq \inf\ (\mathrm{MP})$, as the plan-induced maps are a subset of $\Pi(\mu, \nu)$.

Proposition 1.11. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and assume that μ is atomless. Then there exists at least one transport map $T_{\#}\mu = \nu$.*

Proof. Omitted, a simple construction can be found as Lemma 1.28/Corollary 1.29 in [25]. □

Our example from the introduction demonstrated that in some cases, no optimal maps exist. One might ask that if at least one map exists, can we find a minimiser? The answer, in general, is no. However, we do have a strong result on the approximation of plans by maps, as found in [25, Thrm. 1.32]. We first need the following lemma.

Lemma 1.12. *Let X be a compact metric space and $\rho \in \mathcal{P}(X)$. Let $E_n = \{E_{i,n}\}_{i \in I_n}$ be a sequence of finite partitions of X with the diameter of the largest element*

$$|E_n| := \max_i \operatorname{Diam}(E_{i,n}) \to 0 \text{ as } n \to \infty.$$

Let $\rho_n \in \mathcal{P}(X)$ be such that they give the same mass to each element of each partition E_n, that is, for each n and i, $\rho_n(E_{i,n}) = \rho(E_{i,n})$. Then $\rho_n \rightharpoonup \rho$ narrowly.

Proof. We set $m_{i,n} := \rho_n(E_{i,n}) = \rho(E_{i,n})$. For any $\phi \in C(X)$ we have

$$\begin{aligned} \left| \int_X \phi \mathrm{d}\rho_n - \int_X \phi \mathrm{d}\rho \right| &\leq \sum_{i \in I_n} \left| \int_{E_{i,n}} \phi \mathrm{d}\rho_n - \int_{E_{i,n}} \phi \mathrm{d}\rho \right| \\ &\leq \sum_{i \in I_n} \omega(\operatorname{Diam}(E_{i,n})) m_{i,n} \leq |E_n| \to 0, \end{aligned}$$

where ω is the modulus of continuity of ϕ. The control on the 2rd line is simply that the quantity above could be largest when ρ and ρ_n both assign all their mass on $E_{i,n}$ to points whose values are as far apart as the modulus of continuity allows. □

Theorem 1.13. *Let $\Omega \subset \mathbb{R}^d$ be compact and $\mu, \nu \in \mathcal{P}(\Omega)$ with μ atomless. The set of map-induced plans γ_T is dense in $\Pi(\mu, \nu)$ w.r.t the narrow topology.*

Proof. Take an arbitrary $\gamma \in \Pi(\mu, \nu)$, we will construct a sequence of maps T_n for which $\gamma_{T_n} \rightharpoonup \gamma$.

Fix n, we partition Ω into a finite number of sets $K_{i,n}$ of a diameter smaller than $\frac{1}{2n}$, for example suitably small cubes. We set $E_{i,j,n} := K_{i,n} \times K_{j,n}$, which form a partition of $\Omega \times \Omega$, with max diameter less than $\frac{1}{n}$. On account of Lemma 1.12, we must construct a transport map T_n for which γ_{T_n} gives the same mass to each $E_{i,j,n}$. We set $Col_{i,n} := K_{i,n} \times \Omega$ and let $\gamma_{i,n}$ be the restriction of γ to this, this is the measure which describes where the mass starting in $K_{i,n}$ ends up in Ω. We denote the marginals of this measure as $\mu_{i,n}$ and $\nu_{i,n}$, and by Proposition 1.11, define $T_{i,n}$ as a transport map between these marginals. Now each $\mu_{i,n}$ are concentrated on different sets of a partition, they each describe where a certain part of the original mass goes. Thus $\sum_i \mu_{i,n} = \mu$ and $\sum_i \nu_{i,n} = \nu$, and we can define a map $T_n = \sum_i T_{i,n}$ which transports μ to ν.

By construction, γ_{T_n} gives the same mass as γ to each $E_{i,j,n}$, so $\gamma_{T_n} \rightharpoonup \gamma$ by Lemma 1.12. □

Corollary 1.13.1. *For μ atomless and c continuous,* inf *(MP) =* min *(KP).*

1.2.2 Existence of optimal maps for strictly convex cost

The results in the previous section do not tell us when there exists an optimiser to (MP), or equivalently when the optimal plan of (KP) is induced by a map. In this section, we answer this question in the case $X = Y = \Omega \subset \mathbb{R}^d$ for Ω compact, and cost function $c(x, y) = h(x - y)$ with h strictly convex, as found in [25, 2, Sect. 1.3]. We first need the following, motivated by [27].

Proposition 1.14. *For $c(x, y) = h(x - y)$ with h strictly convex, c-concave functions are locally Lipschitz and hence differentiable a.e.*

Proof. We sketch the main ideas, details are found in [27, Lem. 2]. Since convex functions on $\mathbb{R}^d$ are locally Lipschitz, then on a compact domain they are Lipschitz. In light of [25, Box.1.8] the c-transform of a function inherits the modulus of continuity of the cost, so any c-concave functions are Lipschitz. The differentiability follows from Rademacher's Theorem. □

The following is presented as in [25, Thrm. 1.17].

Theorem 1.15. *Let $\mu, \nu \in \mathcal{P}(\Omega)$ for some compact $\Omega \subset \mathbb{R}^d$, with μ absolutely continuous and $\mu(\partial\Omega) = 0$. Let $c(x, y) = h(x-y)$ for h strictly convex, bounded below and satisfying (1.1). Then the optimal plan for (KP) is unique and induced by a map, and taking a Kantorovich potential φ, we have an explicit form for the map*

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x)).$$

Proof. We prove this using duality. Let γ be an optimal plan for (KP), we aim to show that for γ-a.e $x_0, y_0 \in \operatorname{Spt}\gamma$, x_0 uniquely specifies y_0 so that we can define a map μ-a.e. By Theorem 1.8, there exist Kantorovich potentials φ, φ^c such that

$$\varphi(x) + \varphi^c(y) \leq h(x - y) \text{ on } \Omega \times \Omega \text{ and } \varphi(x) + \varphi^c(y) = h(x - y) \text{ on } \operatorname{Spt}\gamma.$$

Since μ is absolutely continuous, then φ is differentiable μ-a.e. and consequently for γ-a.e. $x_0, y_0 \in \operatorname{Spt}\gamma$ we have by first-order optimisation of the c-transform inf that

$$x \mapsto h(x - y_0) - \varphi(x) \text{ is minimised at } x = x_0,$$

so by differentiability given by the previous proposition, for γ-a.e x_0, y_0

$$\nabla_x h(x_0 - y_0) = \nabla \varphi(x_0). \tag{1.3}$$

Our assumption of strict convexity on h means that ∇h is strictly monotone in each coordinate and hence invertible[2], so that (noting also that the boundary points are negligible by assumption)

$$y_0 = x_0 - (\nabla h)^{-1}(\nabla \varphi(x_0)) \text{ for } \gamma\text{-a.e. } x_0, y_0 \in X \times Y.$$

[2]Potentially with the need to view it as a multi-valued subgradient if h is not C^1, but in any case this multi-valued map is still invertible.

Since the boundary and non-differentiability points are μ negligible, we can define μ-a.e. a map $T : X \to Y$ by

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x)).$$

Observe that (1.3) uniquely specifies the gradient of any Kantorovich potential, and hence the map T is uniquely specified up to negligible sets. Consequently, we also have uniqueness of the optimal γ. □

Remark 1.16. We need the strict convexity here, as [25, Ex. 2.16] demonstrates. For example with the classic Monge cost $c(x,y) = |x-y|$, considering one dimensional transport on $\mathbb{R}$ between uniform densities $\mu = \chi_{[0,2]}$ and $\nu = \chi_{[1,3]}$, both the maps

$$T(x) = x+1; \qquad S(x) = \begin{cases} x+2 & x \in [0,1] \\ x & x \in (1,2]. \end{cases}$$

are optimal as they have the same transport cost and T is the monotone map which is always optimal in 1D for convex (not needing strictly) costs, see [25, Thrm. 2.9].

Remark 1.17. We note that if both μ and ν possess the regularity stated in the theorem then there exists a unique map optimising the transport from ν to μ also, and this must also be induced by the same transport plan, so that the two maps must be a.e. inverses of one another.

There is an extremely powerful characterisation of this in the quadratic case due to Y. Brenier [6]. We show this as a result of the following Lemma from [25, Prop. 1.21].

Lemma 1.18. *Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$, and define $u_\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ by $u_\varphi(x) = \frac{|x|^2}{2} - \varphi(x)$. Then $u_{\varphi^c} = (u_\varphi)^*$ where $*$ denotes the Legendre transform. Thus a function φ is c-concave if and only if $\frac{|x|^2}{2} - \varphi(x)$ is convex and l.s.c.*

Proof. We simply compute

$$u_{\varphi^c}(x) = \frac{|x|^2}{2} - \varphi^c(x) = \sup_y \frac{|x|^2}{2} - \frac{1}{2}|x-y|^2 + \varphi(y) = \sup_y \left(\langle x, y \rangle - \left(\frac{|y|^2}{2} - \varphi(y) \right) \right).$$

Since the double c-transform of a c-concave function is itself, we thus have that the double Legendre transform of u_φ is itself, which characterises convexity. □

Theorem 1.19 (Brenier's Theorem)**.** *Let $c(x,y) = \frac{1}{2}|x-y|^2$, $\mu \in \mathcal{P}(\mathbb{R}^d)$, and $u : \mathbb{R}^d \to \mathbb{R}$ convex (hence a.e. differentiable by Alexandrov theorem, Theorem 4.2). Then the map $T = \nabla u : \mathbb{R}^d \to \mathbb{R}^d$ is optimal for transport between μ and $T_\# \mu$.*

Proof. We prove only the case for μ absolutely continuous and only prove that the optimal map is the gradient of a convex function (not that any convex gradient is necessarily optimal) the general case can be seen as a consequence of c-cyclical monotonicity for the cost $c(x, y) = -\langle x, y \rangle$ and its relation to the subgradient of a convex function due to Rockafellar[23].

For μ a.c. the above theorem applies since the cost is strictly convex, and hence there exists an optimal map between μ and $T_{\#}\mu$, which is of the form

$$T(x) = x - \nabla \varphi(x) = \nabla \left(\frac{|x|^2}{2} - \varphi(x) \right),$$

where φ is a Kantorovich potential. The above lemma applies so that T is a.e. the gradient of a convex function. □

Remark 1.20. The above say that one can talk of optimal maps without reference to a target measure for quadratic cost. Gradients of convex functions are monotone in every variable, so most efficiently move mass around as no mass crosses over each other.

1.3 The Eulerian formulation: Benamou-Brenier

The Monge and Kantorovich frameworks are somewhat limited in describing continuous transportation of mass, in the sense that they only consider the start and end point $t = 0$ and $t = 1$, and not the evolution of the mass distribution for all $t \in [0, 1]$, specified by some family of measures ρ_t with $\rho_0 = \mu$ and $\rho_1 = \nu$. In this section, we consider a formulation due to JD. Benamou and Y. Brenier, originally presented in [3].

An important consequence of our existence theorem for (KP) (Theorem 1.2) is that we can always assign a transport cost between two measures. We focus only on the case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2$, but it is easy to see that the following can be generalised to $c(x, y) = |x - y|^p$ for arbitrary $p \geq 1$. For any $\mu, \nu \in \mathcal{P}(\Omega)$, we define the Wasserstein distance

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

To guarantee that this is finite, we actually restrict to the subspace $\mathcal{P}_2(\mathbb{R}^d)$, which consists of all those $\mu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moments $\int |x|^2 \mathrm{d}\mu < \infty$.

This satisfies the symmetry and positivity properties to justify calling this a distance (if $\mu = \nu$ then there is a coupling concentrated along the diagonal which thus has zero transport cost). The triangle inequality is a little more involved, but can be seen to follow from a disintegration argument found in [25, Lem. 5.5]. We note that this distance behaves extremely differently to, for example, taking the L^2 difference of two probability densities. The W_2 distance captures a horizontal transport movement, while the aforementioned measures a vertical overlap.

We aim to understand the most efficient way to continuously deform one mass distribution into another, and this amounts to finding a geodesic in $\mathcal{P}_2(\mathbb{R}^d)$ (w.r.t the Wasserstein distance) between our measures μ and ν. Continuous deformations are naturally described by continuity equations. We will say a family of measures $\rho_t \in \mathcal{P}(\mathbb{R}^d)$ and $\mathbb{R}^d$ valued vector measure fluxes $j_t \in \mathcal{M}^d(\mathbb{R}^d)$ for $t \in [0,1]$ solve the continuity equation

$$\partial_t \rho + \nabla \cdot j = 0, \tag{1.4}$$

if they are a distributional solution in the sense that for all $\phi \in C_c^1(\mathbb{R}^d \times (0,1))$

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \phi \mathrm{d}\rho_t \mathrm{d}t + \int_0^1 \int_{\mathbb{R}^d} \nabla \phi \cdot \mathrm{d}j_t \mathrm{d}t = 0.$$

The following theorem we present similarly to the original paper [3, Prop. 1.1], noting that the result can be significantly generalised to a very broad statement equating absolutely continuous curves in $\mathcal{P}_2$ (in the W_2 metric derivative sense) with solutions to the continuity equation, as found in [25, Thrm. 5.14].

Theorem 1.21 (Benamou-Brenier). *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, be compactly supported measures. We have*

$$W_2(\mu,\nu)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{1}{\rho} |j|^2 : \rho_0 = \mu,\ \rho_1 = \nu \right\}, \tag{1.5}$$

where the inf is taken over all pairs of time-indexed measures ρ_t and fluxes j_t satisfying (1.4) with distributional boundary conditions $\rho_0 = \mu$ and $\rho_1 = \nu$, and we define the quantity

$$\frac{1}{\rho}|j|^2 := \begin{cases} |\frac{\mathrm{d}j}{\mathrm{d}\rho}|^2 \mathrm{d}\rho & \text{if } j \ll \rho \\ +\infty & \text{otherwise.} \end{cases} \tag{1.6}$$

Proof. We give a formal proof à la [3], making some additional regularity assumptions along the way. A rigorous approach by convolution deals with the general case in [25, Ch.5]. Take any pair of time-indexed measures ρ_t and fluxes j_t that satisfy (1.4) with boundary conditions $\rho_0 = \mu$ and $\rho_1 = \nu$, and for which the RHS of (1.5) is finite. The finiteness demands that for a.e. t we can define $v = \frac{\mathrm{d}j}{\mathrm{d}\rho}$. We write the flow instead in Lagrangian coordinates[3], setting

$$X(x,0) = x, \quad \partial_t X(x,t) := v(X(x,t),t).$$

Then for all test functions $\phi \in C_c^1(\mathbb{R}^d \times (0,1))$ we have

$$\int_{\mathbb{R}^d} \int_0^1 \phi(x,t) \mathrm{d}\rho_t \mathrm{d}t = \int_{\mathbb{R}^d} \int_0^1 \phi(X(x,t),t) \mathrm{d}\rho_0 \mathrm{d}t,$$

and specifically $X(\,\cdot\,,1) : \mathbb{R}^d \to \mathbb{R}^d$ is a transport map between μ and ν as a consequence of the boundary conditions. Consequently,

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^d} |v(x,t)|^2 \mathrm{d}\rho_t \mathrm{d}t &= \int_0^1 \int_{\mathbb{R}^d} |v(X(x,t),t)|^2 \mathrm{d}\rho_0 \mathrm{d}t \\ &= \int_{\mathbb{R}^d} \int_0^1 |\partial_t X(x,t)|^2 \mathrm{d}t \mathrm{d}\rho_0 \geq \int_{\mathbb{R}^d} |X(x,1) - X(x,0)|^2 \mathrm{d}\rho_0 \end{aligned} \tag{1.7}$$

[3]Here we assume that v is suitably regular to justify this calculation.

where the second line follows from Jensen's inequality. The RHS is the Monge cost for $X(\cdot, 1)$ and hence is greater than or equal to $W_2(\mu,\nu)^2$. To show the infimum is attained, let T minimise the Monge formulation and set $T_t(x) = tT(x) + (1-t)x$ as a constant speed interpolation of the mass between x and $T(x)$ for each point. Then $\rho_t = (T_t)_{\#}\mu$ and $j_t = T_t\#[(T-\mathrm{Id})\mu]$[4] solve the Benamou-Brenier formulation, with velocity $v_t = \frac{\mathrm{d}j}{\mathrm{d}\rho} = T(x) - x$ for each $t \in [0,1]$. Thus returning to (1.7), the Benamou-Brenier cost of this interpolation is

$$\int_0^1 \int_{\mathbb{R}^d} |v_t|^2 \mathrm{d}\rho_t \mathrm{d}t = \int_0^1 \int_{\mathbb{R}^d} |T(x) - x|^2 \mathrm{d}\rho_0 \mathrm{d}t = W_2(\mu,\nu)^2$$

as T minimises the Monge cost, which gives the result. □

Remark 1.22. These interpolations with T_t only are admissible as $\mathbb{R}^d$ is convex, on a more general space this may fail. One can easily construct some counter-examples for which no geodesic exists when the domain is an open non-convex set.

In applications, one considers functionals on $\mathcal{P}_2(\Omega)$, and instead of desiring convexity in the sense of linear interpolations, they wish for convexity along geodesics[5] - which in this case are continuity equation mass flows minimising (1.5). This theory was initially developed by McCann in his thesis [20], here we present a simple version between uniform measures as found in [16], the paper which will be the subject of Chapter 4.

Proposition 1.23 (McCann's displacement convexity)**.** *Let $E, F \subset \mathbb{R}^d$ be two bounded sets of equal measure and T minimise the Monge problem between $\chi_E \mathrm{d}x$ and $\chi_F \mathrm{d}x$. Set $T_t(x) = (1-t)T(x) + tx$, and set $\rho_t = (T_t)_{\#}\mu$ and $j_t = T_t\#[(T-\mathrm{Id})\mu]$ as a minimising pair of the Benamou-Brenier formulation(1.5). Then $\rho_t \leq 1$ for all $t \in [0,1]$.*

Proof. We assume that ρ and j are absolutely continuous for all t, proof of this can be found in [29, Prop. 5.9]. Thus the Jacobian equation (1) for each t reads

$$\rho(t, T_t(x)) \det DT_t(x) = \chi_E(x),$$

which holds for a.e. x. For $t = 1$ in particular, $\det DT(x) = \chi_E$ so that

$$(\det DT_t(x))^{\frac{1}{d}} = (\det(1-t)\,\mathrm{Id} + tT(x))^{\frac{1}{d}} \leq (1-t)(\det \mathrm{Id})^{\frac{1}{d}} + t(\det DT(x))^{\frac{1}{d}} = 1$$

by concavity of $\det(\cdot)^{1/d}$ on non negative symmetric matricies, see Appendix. Thus $\rho_t \leq 1$ for all t. □

[4] Here this is the push forward of the vector measure $(T-\mathrm{Id})\mu$, given simply by the push forward in each component.

[5] In Euclidean spaces there is no distinction here, geodesics are precisely interpolations. But here interpolations are geodesics for the L^2 distance between densities rather than the W_2 distance.

2 Morrey, Campanato and Hölder spaces

2.1 Definitions, Morrey $\cong$ Campanato

In this chapter, we study some function spaces for which certain energy-type semi-norms are finite. We then establish that these spaces are embedded inside the Hölder spaces for certain parameters. These results will be extremely useful for establishing the regularity of solutions to Elliptic PDE and optimal transport maps in the following chapters.

We set $\Omega(x_0, r) := \Omega \cap B_r(x_0)$ as the r "ball" inside Ω. Throughout the rest of this Chapter, we assume $\Omega \subset \mathbb{R}^n$ to be a connected bounded domain with the following regularity property: There exists a constant $C_\Omega > 0$ such that for all $x_0 \in \Omega$ and $r < \operatorname{Diam}\Omega$ we have

$$|\Omega(x_0, r)| \geq C_\Omega r^d. \tag{2.1}$$

Equivalently, one could say

$$\frac{|\Omega(x_0, r)|}{|B_r(x_0)|} \geq C(\Omega) > 0,$$

so the proportion of any ball that lies in the domain is uniformly bounded away from zero. Intuitively what this says it that each $\Omega(x_0, r)$ must uniformly have enough of its area inside Ω, so that we cannot have any sharp "cusps". For example this fails if the boundary of some $\Omega \subset \mathbb{R}^2$ contains

$$f(x) = \begin{cases} \sqrt{x} & x \geq 0 \\ \sqrt{-x} & x \leq 0. \end{cases}$$

with the domain lying above this curve - then taking a sequence of points inside the domain approaching $x_1, x_2 = 0$ and some fixed radius the proportion of a ball inside vanishes. This condition will be important as the seminorms defined below relate an energy on $\Omega(x_0, r)$ to the growth of r^σ, so making the domain have too sharp a cusp somewhere could cause the seminorm to blow up not because the energy is too small, but the region we are integrating over is. One can show (laboriously) that every Lipschitz or C^1 domain has this property.

We denote by $\mathscr{P}_k$ the space of polynomials in d variables of degree k or less, that is, functions of the form

$$P(x) = \sum_{|\beta| \leq k} c_\beta x^\beta; \qquad x^\beta := x_1^{\beta_1} x_2^{\beta_2} \dots x_d^{\beta_d}$$

for multi-indices β. We will denote by $\beta!$ the multi-factorial $\beta_1!\beta_2!\dots\beta_d!$.

Definition 2.1. *The Hölder space $C^{k,\alpha}(\Omega)$ for $0 < \alpha \leq 1$ consists of all $u \in C^k(\Omega)$ with finite Hölder seminorm*

$$[u]_{k,\alpha} := \sup_{\substack{|\beta|=k \\ x \neq y}} \frac{|D^\beta u(x) - D^\beta u(y)|}{|x-y|^\alpha}.$$

Remark 2.2. We have $C^{k,\alpha} = \mathscr{P}_k$ if $\alpha > 1$. Indeed let $\alpha = 1 + \varepsilon$ with $\varepsilon > 0$. Then for any multi-index β with $|\beta| = k$,

$$\frac{|D^\beta u(x) - D^\beta u(y)|}{|x-y|} \leq |x-y|^\varepsilon.$$

The Hölder condition directly gives continuity and the above equation gives any directional derivative of $D^\beta u$ vanishing everywhere for each $|\beta| = k$ by taking the limit $x \to y$. Thus $u \in \mathscr{P}_k$.

Definition 2.3. *For $1 \leq q \leq \infty$ and $\sigma > 0$, we define the Morrey space $L^{q,\sigma}(\Omega)$ as those $u \in L^q(\Omega)$ for which have finite Morrey seminorm:*

$$\langle u \rangle_{q,\sigma} = \sup_{\substack{x_0 \in \Omega \\ r>0}} \left(\frac{1}{r^\sigma} \int_{\Omega(x_0,r)} |u(x)|^q \mathrm{d}x \right)^{\frac{1}{q}} < \infty.$$

Before defining the Campanato space, we need the following projection result of L^q onto $\mathscr{P}_k$, as proved in [11].

Lemma 2.4. *Fix $u \in L^q(\Omega(x_0, r))$ for $1 \leq q < \infty$, $x_0 \in \Omega$ and $r > 0$. The minimisation problem*

$$\min \left\{ \int_{\Omega(x_0,r)} |u(x) - P(x)|^q \mathrm{d}x; \quad P \in \mathscr{P}_k \right\} \tag{2.2}$$

admits a solution, and it is unique.

Proof. We write a general element of $\mathscr{P}_k$ as $P(x) = \sum_{|\beta| \leq k} \frac{a_\beta}{\beta!}(x - x_0)^\beta$. Thus we can view the coefficients $\{a_\beta\}$ as lying together in some Euclidian space $\mathbb{R}^m$, and the quantity

$$f : \mathbb{R}^m \to \mathbb{R}; \qquad f(\{a_\beta\}) := \|u - P\|^q_{L^q(\Omega(x_0,r))}$$

depends continuously on these parameters. As a consequence of Lemma 2.13 proved in the next section, if we take a sequence of polynomials $P_i \in \mathscr{P}_k$ such that $\left(\sum_{|\beta| \leq k} a_{i,\beta}^2\right)^{\frac{1}{2}} \to \infty$, then $\|P_i\|^q_{L^q(\Omega(x_0,r))} \to \infty$, which in turn gives $f(\{a_{i,\beta}\}) \to \infty$ by the reverse triangle inequality. Thus we have the existence of at least one minimiser since the problem is a coercive, continuous minimisation problem in a Euclidean space. Uniqueness follows from the uniform convexity of L^q spaces. ∎

Definition 2.5. *For $1 \le q < \infty, \sigma > 0$ and $k \in \mathbb{N}$, we define the Campanato space $\mathcal{L}_k^{q,\sigma}(\Omega)$ by all those $u \in L^q(\Omega)$ which have finite Campanato seminorm:*

$$\|u\|_{k,q,\sigma} := \sup_{\substack{x_0\in\Omega\\ r>0}} \left(\frac{1}{r^\sigma} \min_{P\in\mathscr{P}_k} \int_{\Omega(x_0,r)} |u(x) - P(x)|^q \mathrm{d}x \right)^{\frac{1}{q}} < \infty.$$

Remark 2.6. In the case $k = 0$ the degree k polynomials are just constants, and for $q = 2$ the minimiser of the infimum is given by $u_{x_0,r} := \fint_{\Omega(x_0,r)} u$, so that the Campanato seminorm becomes

$$\|u\|_{0,2,\lambda} = \sup_{\substack{x_0\in\Omega\\ r>0}} \left(\frac{1}{r^\lambda} \int_{\Omega(x_0,r)} |u - u_{x_0,r}|^2 \mathrm{d}x \right)^{\frac{1}{2}}.$$

Remark 2.7. Since we are always assuming our domain is bounded, for either space $L^{q,\sigma}$ or $\mathcal{L}_k^{q,\sigma}$ then we can fix any $\rho > 0$ and consider the sup over only $\rho > r > 0$ by an open covering argument, due to the relative compactness of the domain. So really, these conditions are local properties. As a consequence of this, fixing $\rho = 1$, for two parameters $\sigma < \varsigma$, we see that $\mathcal{L}_k^{q,\varsigma} \subset \mathcal{L}_k^{q,\sigma}$, $L^{q,\varsigma} \subset L^{q,\sigma}$ and $C^{k,\varsigma} \subset C^{k,\sigma}$.

Remark 2.8. Membership of $L^{q,\sigma}$ is saying that the L^q energy $\phi_{x_0}(r) = \int_{B_r} |u|^q \mathrm{d}x$ grows in a "σ-Hölder" sense at each point with $\phi_{x_0}(r) \le Cr^\sigma$, and this growth is uniform in space so C can be taken independent of x_0. Membership of $\mathcal{L}_k^{q,\sigma}$ means that the qth power of the L^q distance of u from locally being a degree k polynomial satisfies this Hölder type growth condition instead. Unlike in Remark 2.2, we do not have the restriction $\sigma < 1$ for an interesting space, as the same argument is only enough to show that $\phi'_{x_0}(0) = 0$ (the point being that ϕ denotes a different function at each point unlike before).

Theorem 2.9. *If $0 \le \sigma < d$ and Ω satisfies (2.1) then $L^{2,\sigma}(\Omega) \cong \mathcal{L}_0^{2,\sigma}(\Omega)$.*

Proof. For reasons of space, we only prove the direction $L^{2,\sigma}(\Omega) \subset \mathcal{L}_0^{2,\sigma}(\Omega)$. For the converse direction, see [14, Prop. 5.4]. We have

$$\int_{\Omega(x_0,r)} |u - u_{x_0,r}|^2 \mathrm{d}x \le 2^2 \left(\int_{\Omega(x_0,r)} |u|^2 \mathrm{d}x + |\Omega(x_0,r)||u_{x_0,r}|^2 \right),$$

and Jensen's inequality gives

$$|u_{x_0,r}|^2 \le \frac{1}{|\Omega(x_0,r)|} \int_{\Omega(x_0,r)} |u|^2 \mathrm{d}x.$$

This gives $\|u\|_{0,2,\sigma} \le 2^3 \langle u \rangle_{2,\sigma}$ as required. □

2.2 Campanato's theorem

This section is devoted to proving the following regularity theorem due to Sergio Campanato, originally shown in [11, Th. 5.III]. Although we only need the case $k = 0$ for elliptic PDE as we can differentiate the equation, the general case will be the crux of the proof of partial regularity for optimal transport maps.

Theorem 2.10 (Campanato's Theorem)**.** *Let Ω satisfy (2.1) with $\sigma > d + kq$. Then $C^{k,\alpha}(\Omega) \cong \mathcal{L}_k^{q,\sigma}(\Omega)$ for $\alpha = \frac{\sigma-d-kq}{q}$ and there exist global constants depending only on k, q, σ, d, Ω giving the equivalence of the respective seminorms.*

Remark 2.11. As a consequence of Remark 2.2, if further we have $\sigma > d + (k+1)q$, then $\mathcal{L}_k^{q,\sigma}(\Omega) \cong \mathscr{P}_k$. If a function can be approximated by a degree k polynomial this closely, it must in fact be one.

Remark 2.12. The domain regularity property (2.1) is crucial for a uniform Hölder seminorm in x. As previously mentioned, the Campanato seminorm condition says that the distance from a degree k polynomial on $\Omega(x_0, r)$ vanishes faster than r^σ, but this could be for two reasons: how well behaved u is, or if the Lebesgue measure of $\Omega(x_0, r)$ vanishes very quickly. Property (2.1) gives a uniform control on this, saying a certain proportion of any such set must be in the domain. With Ω open not satisfying this, we still get $u \in C_{\text{loc}}^{k,\alpha}(\Omega)$ however, by applying the result on an open ball around each point.

The direction $C^{k,\alpha}(\Omega) \subseteq \mathcal{L}_k^{q,\sigma}(\Omega)$ is relatively intuitive. We might hope that k times differentiable functions are closely approximated by degree k polynomials, and the natural tool to do this - the degree k Taylor series at each x_0 - turns out to be a good enough candidate to establish this (although it is not necessarily the minimiser of (2.2) on each $\Omega(x_0, r)$).

The converse is considerably more delicate, and we will need several preparatory results to establish this. On account of Remark 2.8, the statement $\mathcal{L}_k^{q,\sigma}(\Omega) \subseteq C^{k,\alpha}(\Omega)$ tells us that if σ is suitably large then the distance locally of u from being a degree k polynomial at each x_0 grows so slowly with r that u must have k continuous derivatives - that if there was any discontinuity of a $D^\beta u$ at some $x_0 \in \Omega$, then here no element of $\mathscr{P}_k$ is locally close enough. To give some intuition, we first sketch the structure of this direction of the proof below.

[Sketch of the converse direction.] We follow Campanato's original proof as presented in [11]. From now on, fix $1 \leq q < \infty$, $k \in \mathbb{Z}_{\geq 0}$ and $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$. For each $x_0 \in \Omega$ and $r > 0$ denote the unique minimiser (c.f. Lemma 2.2)

$$P(x, x_0, r) := \operatorname{argmin} \left\{ \int_{\Omega(x_0,r)} |u(x) - P(x)|^q \mathrm{d}x; \quad P \in \mathscr{P}_k \right\}. \tag{2.3}$$

One should view this as a polynomial in x, with x_0 and r fixed.

[Step 1.] We begin by proving that at each $x_0 \in \Omega$, the value at x_0 of the minimiser on $\Omega(x_0, r)$ converges as $r \to 0$, as do all of its derivatives of degree k or less. That is, for each multi-index $|\beta| \leq k$,

$$a_\beta(x_0, r) := D^\beta P(x, x_0, r)|_{x=x_0} \tag{2.4}$$

has a limit as $r \to 0$. To establish this we first prove some general estimates on the value of derivatives $|D^\beta P(x_0)|$ of degree k polynomials in terms of their L^q norms on $\Omega(x_0, r)$. We apply these estimates to the sequence of polynomials

$$P_h(x) = D^\beta \big[P(x, x_0, r2^{-h}) - P(x, x_0, r2^{-h-1})\big]$$

to show the sequence $D^\beta P(x, x_0, r2^{-h})|_{x=x_0}$ is Cauchy, further deducing its limit is independent of the initial choice of r. For each β, we call this limit

$$v_\beta(x_0) := \lim_{r \to 0} a_\beta(x_0, r). \tag{2.5}$$

[Step 2.] Next, we establish the Hölder continuity of each v_β when $|\beta| = k$, and their differentiability relationship between their derivatives when $|\beta| \leq k - 1$. Specifically, we show that for each orthonormal basis vector e_i,

$$\frac{\partial v_\beta}{\partial x_i}(x) = v_{\beta + e_i}(x); \qquad \text{for } i = 1, 2, \ldots, d.$$

This is established via an inductive argument backwards from the case $|\beta| = k$. The crucial point is establishing also the convergence as $r \to 0$ of $a_\beta(x_0 + e_i r, 2|r|)$ to $v_\beta(x_0)$, the same limit as that of $a(x_0, r)$, as we approach along offset balls.

[Step 3.] Finally, we prove that $v_{(0)} = u$ a.e. (at each Lebesgue point of u) which completes the proof. Here $(0) \in \mathbb{R}^d$ is the empty multi-index.

2.2.1 Limits for the derivatives of the minimising polynomials

We first establish the convergence of the coefficients of minimising polynomials at each point. These results correspond roughly to [11, Sect. 2&3]

Lemma 2.13. *Let $P \in \mathscr{P}_k$, and $\lambda > 0$. There exists $C(k, q, d, \lambda) > 0$ s.t. for each $E \subseteq B_r(x_0)$ with $|E| \geq \lambda r^d$, and any multi-index $|\beta| \leq k$,*

$$|D^\beta P(x_0)|^q \leq \frac{C}{r^{d+q|\beta|}} \int_E |P(x)|^q \mathrm{d}x.$$

Proof. Denote by $\tau_k \subset \mathscr{P}_k$ those polynomials who's coefficients satisfy

$$P(x) = \sum_{|\beta| \leq k} a_\beta x^\beta; \qquad \sum_{|\beta| \leq k} a_\beta^2 = 1.$$

Let $\mathcal{F}$ denote all measurable functions supported inside $B_1(0)$ with $0 \leq f \leq 1$ and $\int_{B_1} f\mathrm{d}x \geq \lambda$. We then define the quantity

$$\gamma(\lambda) = \inf_{P\in\tau_k\ f\in\mathcal{F}} \int_{B_1} |P(x)|^q f(x)\mathrm{d}x.$$

We show this is a minimum. Take sequences P_i, f_i approaching the infimum, then since the coefficients of P_i are uniformly bounded and there are finitely many, we can apply Bolzano-Weierstrass to each coefficient sequence to obtain a convergent subsequence of degree k polynomials, and the limit clearly lies in τ_k also.

Similarly, we can pass to a weak L^2 limit for f since they have $f_i \leq 1$ so are uniformly bounded in L^2 norm by $|B_1|$. It is clear the class $\mathcal{F}$ is stable under weak convergence so the limit is also a member.

Consequently, $\gamma(\lambda) > 0$, since the P-lim must be $\neq 0$ a.e as polynomial, and the f-lim must be positive on a non-negligible set by the integral condition. Thus by taking $f = \chi_E$ we have that for all $P \in \tau_k$,

$$\int_E |P(x)|^q \mathrm{d}x \geq \gamma(\lambda) \quad \text{if } |E| \geq \lambda.$$

Now for a general $P \in \mathscr{P}_k$, $P(x)\left(\sum_{|\beta|\leq k} a_\beta^2\right)^{-\frac{1}{2}} \in \tau_k$, so for each $E \subset B_1$ with $|E| \geq \lambda$,

$$\left(\sum_{|\beta|\leq k} a_\beta^2\right)^{\frac{q}{2}} \leq \frac{1}{\gamma(\lambda)} \int_E |P(x)|^q \mathrm{d}x$$

and specifically

$$|a_\beta|^q \leq \frac{1}{\gamma(\lambda)} \int_E |P(x)|^q \mathrm{d}x. \tag{2.6}$$

Now take $P \in \mathscr{P}_k$ and $E \subset B_r(x_0)$ measurable satisfying the regularity property. The change of variables $T(x) = \frac{x-x_0}{r}$ gives

$$\int_E |P(x)|^q \mathrm{d}x = r^d \int_{T(E)} |P(x_0 + ry)|^q \mathrm{d}y,$$

with $T(E) \subset B_1$ and $|T(E)| = \frac{1}{r^d} \int_E \mathrm{d}x \geq \lambda$. Taylor series gives

$$P(x_0 + ry) = \sum_{|\beta|\leq k} \frac{r^{|\beta|}|D^\beta P(x_0)|}{\beta!} y^\beta,$$

and so (2.6) applied to the above polynomial gives the result with $C = \frac{(\beta!)^q}{\gamma(\lambda)}$. □

Lemma 2.14. *Let $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$, and $P(x, x_0, r)$ denote the minimiser of (2.3). There exists $C(q, \sigma) > 0$ s.t. for all x_0, r and $h \in \mathbb{Z}_{\geq 0}$,*

$$\int_{\Omega(x_0, r2^{-h-1})} \left|P(x, x_0, r2^{-h}) - P(x, x_0, r2^{-h-1})\right|^q \mathrm{d}x \leq \left(C\|u\|_{k,q,\sigma}^q\right) 2^{-h\sigma} r^\sigma.$$

Proof. We have

$$\begin{aligned}|P(x,x_0,r2^{-h})-P(x,x_0,r2^{-h-1})|^q \leq & 2^q|P(x,x_0,r2^{-h})-u(x)|^q \\ & + 2^q|P(x,x_0,r2^{-h-1})-u(x)|^q,\end{aligned}$$

So that by definition of P as the minimiser of (2.2) and that $\Omega(x_0,r2^{-1-h})\subset\Omega(x_0,r2^{-h})$,

$$\begin{aligned}&\int_{\Omega(x_0,r2^{-1-h})}\big|P(x,x_0,r2^{-h})-P(x,x_0,r2^{-h-1})\big|^q\mathrm{d}x \\ &\qquad\leq 2^q\big\|P(\,\cdot\,,x_0,r2^{-h})-u\big\|_{L^q(\Omega(x_0,r2^{-h}))} \\ &\qquad\quad + 2^q\big\|P(\,\cdot\,,x_0,r2^{-1-h})-u\big\|_{L^q(\Omega(x_0,r2^{-1-h}))} \\ &\qquad\leq 2^q(1+2^{-\sigma})\|u\|_{k,q,\sigma}^q 2^{-h\sigma}r^\sigma,\end{aligned}$$

since $u\in\mathcal{L}_k^{q,\sigma}$, which gives the estimate with $C=2^q(1+2^{-\sigma})$. □

Lemma 2.15. *Let $u\in\mathcal{L}_k^{q,\sigma}(\Omega)$, Ω satisfy (2.1) and $P(x,x_0,r)$ be the minimiser of (2.3) and $a_\beta(x,r)$ as (2.4). There exists $C(k,q,\sigma,d,\Omega)>0$ s.t. for all x_0, r, $i\in\mathbb{Z}_{\geq 0}$ and $|\beta|\leq k$,*

$$|a_\beta(x_0,r)-a_\beta(x_0,r2^{-i})|\leq (C\|u\|_{k,q,\sigma})\left(\sum_{h=0}^{i-1}2^{h\left(\frac{d+|\beta|-\sigma}{q}\right)}\right)r^{\frac{\sigma-d-q|\beta|}{q}}.$$

Proof. Fix x_0,r,i,β, then the triangle inequality gives

$$\begin{aligned}\big|a_\beta(x_0,r)-a_\beta(x_0,r2^{-i})\big| &\leq \sum_{h=0}^{i-1}\big|a_\beta(x_0,r2^{-h})-a_\beta(x_0,r2^{-h-1})\big| \\ &= \sum_{h=0}^{i-1}\big|D^\beta\big[P(x_0,x_0,r2^{-h})-D^\beta P(x_0,x_0,r2^{-h-1})\big]\big|.\end{aligned}$$

Applying Lemma 2.13[1] to the polynomials $P_h:=D^\beta\big[P(x,x_0,r2^{-h})-P(x,x_0,r2^{-h-1})\big]$ in the sum gives

$$|a_\beta(x_0,r)-a_\beta(x_0,r2^{-i})|\leq C_1^{\frac{1}{q}}r^{-\frac{d}{q}-|\beta|}\sum_{h=0}^{i-1}2^{(h+1)\left(\frac{d}{q}+|\beta|\right)}\left(\int_{\Omega(x_0,r2^{-1-h})}|P_h|^q\mathrm{d}x\right)^{\frac{1}{q}},$$

then applying Lemma 2.14 to each integral gives the required estimate. □

The consequences of the above are that for suitable parameters of our space, a limit for $a_\beta(x_0,r)$ exists at each x_0 as $r\to 0$, presented as in [11, Lem. 3.4].

[1] Here, we pickup the domain constant dependence.

Proposition 2.16. *Let $u \in \mathcal{L}^{q,\sigma}_k(\Omega)$, Ω satisfy (2.1) and $\sigma > d + sq$ for some $\mathbb{Z}_{\geq 0} \ni s \leq k$. For all multi-indices $|\beta| \leq s$, there exist functions $v_\beta : \Omega \to \mathbb{R}$ s.t. for all x_0, r,*

$$|a_\beta(x_0, r) - v_\beta(x_0)| \leq (C\|u\|_{k,q,\sigma})\, r^{\frac{\sigma - d - q|\beta|}{q}}.$$

where $C(\sigma, q, k, d, \Omega) > 0$. Consequently, $a_\beta(x_0, r) \to v_\beta(x_0)$ uniformly in x_0 as $r \to 0$.

Proof. Fix β, r, x_0 with $|\beta| \leq s$. We prove that the sequence $a_\beta(x_0, r2^{-i})$ converges as $i \to \infty$. For i, j with $j > i$, Lemma 2.15 gives

$$|a_\beta(x_0, r2^{-j}) - a_\beta(x_0, r2^{-i})| \leq C_3\|u\|_{k,q,\sigma} \left(\sum_{h=i}^{j-1} 2^{h\left(\frac{d+|\beta|-\sigma}{q}\right)}\right) r^{\left(\frac{\sigma-d-q|\beta|}{q}\right)}.$$

Since $|\beta| \leq s$ and $\sigma > d + sq$ the exponent of 2 is negative, so the tail sum

$$\sum_{h=i}^{\infty} 2^{h\left(\frac{d+|\beta|-\sigma}{q}\right)} \to 0$$

for $i \to \infty$, as the tail of a convergent geometric series. Thus

$$|a_\beta(x_0, r2^{-j}) - a_\beta(x_0, r2^{-i})| \leq Cr^{\left(\frac{\sigma-d-q|\beta|}{q}\right)},$$

so that $a_\beta(x_0, r2^{-i})$ is a Cauchy sequence in i, as the r exponent is positive. We now show the limit does not depend on r. Let $r_1 \leq r_2$, then Lemma 2.13 gives

$$\begin{aligned}
&|a_\beta(x_0, r_1 2^{-i}) - a_\beta(x_0, r_2 2^{-i})| \\
&\quad\leq C_1 \frac{2^{i(n+|\beta|q)}}{r_1^{d+|\beta|q}} \int_{\Omega(x_0, r_1 2^{-i})} \left|P(x, x_0, r_1 2^{-i}) - P(x, x_0, r_2 2^{-i})\right|^q \mathrm{d}x \\
&\quad\leq C_1 2^q \frac{2^{i(d+|\beta|q)}}{r_1^{d+|\beta|q}} \left(\int_{\Omega(x_0, r_1 2^{-i})} \left|P(x, x_0, r_1 2^{-i}) - u(x)\right|^q \mathrm{d}x \right. \\
&\qquad\qquad \left. + \int_{\Omega(x_0, r_2 2^{-i})} \left|P(x, x_0, r_2 2^{-i}) - u(x)\right|^q \mathrm{d}x \right) \\
&\quad\leq C_1 2^q \|u\|^q_{k,q,\sigma} \left(\frac{r_1^\sigma + r_2^\sigma}{r_1^{d+|\beta|q}} \right) 2^{-i(\sigma - d - |\beta|q)} \to 0 \text{ as } i \to \infty.
\end{aligned}$$

Thus the limit is independent of r, so the functions v_β are well defined. For uniform convergence, again from Lemma 2.15 and that the sum is convergent so uniformly bounded in i, we have

$$|a_\beta(x_0, r) - a_\beta(x_0, r2^{-i})| \leq (C\|u\|_{k,q,\sigma})\, r^{\frac{\sigma-d-q|\beta|}{q}}.$$

Passing to the $i \to 0$ limit we have now the convergence uniformly in r. $\square$

2.2.2 Continuity of the limit functions and differentiability relations

We will see that $v_\beta = D^\beta u$ (derivatives which a priori u need not have). First we show $v_{(0)} \in C^{k,\alpha}$ and some differentiability relations, as in [11, Lem. 3.2, Prop. 4.1].

Lemma 2.17. *Let $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$ and Ω satisfy (2.1). There exists $C(k,q,d,\sigma,\Omega) > 0$ such that for each pair of points $x_0, y_0 \in \overline{\Omega}$ and any multi-index $|\beta| = k$,*

$$|a_\beta(x_0, 2|x_0 - y_0|) - a_\beta(y_0, 2|x_0 - y_0|)|^q \le \left(C\|u\|_{k,q,\sigma}^q\right)|x_0 - y_0|^{\sigma - d - kq}.$$

Proof. Set $r = |x_0 - y_0|$, then

$$\begin{aligned}|P(x, x_0, 2r) - P(x, y_0, 2r)|^q \le & 2^q|P(x, x_0, 2r) - u(x)|^q \\ & + 2^q|P(x, y_0, 2r) - u(x)|^q.\end{aligned}$$

Integrating, noting $\Omega(x_0, r) \subset \Omega(y_0, 2r)$ and using $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$ gives

$$\begin{aligned}&\int_{\Omega(x_0,r)} |P(x, x_0, 2r) - P(x, y_0, 2r)|^q \mathrm{d}x \\ &\quad \le 2^q \int_{\Omega(x_0,2r)} |P(x, x_0, 2r) - u(x)|^q \mathrm{d}x + 2^q \int_{\Omega(y_0,2r)} |P(x, y_0, 2r) - u(x)|^q \mathrm{d}x \\ &\quad \le 2^{q+\sigma+1}\|u\|_{k,q,\sigma}^q r^\sigma.\end{aligned}$$

Finally, Lemma 2.13[2] applied to $P(x) = P(x, x_0, 2r) - P(x, y_0, 2r)$, combined with that the kth order derivatives are constant so we can be evaluated anywhere, we find

$$|a_\beta(x_0, 2r) - a_\beta(y_0, 2r)|^q \le \left(C\|u\|_{k,q,\sigma}^q\right) r^{\sigma - d - kq}.$$

□

Proposition 2.18. *Let $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$, Ω satisfy (2.1) and $\sigma > d + kq$. Then for $|\beta| = k$ we have $v_\beta \in C^{0,\alpha}(\Omega)$ with $\alpha = \frac{\sigma - d - kq}{q}$, where v_β are the limits of the derivatives of the minimisers from Proposition 2.16.*

Proof. Fix β with $|\beta| = k$ and $x, y \in \Omega$ with $r := |x - y| \le \frac{\operatorname{Diam}\Omega}{2}$. The triangle inequality gives

$$|v_\beta(x) - v_\beta(y)| \le |v_\beta(x) - a_\beta(x, 2r)| + |v_\beta(y) - a_\beta(y, 2r)| + |a_\beta(x, 2r) - a_\beta(y, 2r)|.$$

Proposition 2.16 controls the first two terms uniformly in the required fashion and Lemma 2.17 controls the third, giving the relation for $|x - y| \le \frac{\operatorname{Diam}\Omega}{2}$. For general $|x - y|$, simply cover $\overline{\Omega}$ with balls of radius $< \frac{\operatorname{Diam}\Omega}{8}$, and since the Ω is connected and bounded, a finite subcover gives us an upper bound on the number of steps a piecewise affine polygonal path from x to y with vertices inside Ω should take, so by repeatedly applying the triangle inequality along the vertices of any such path we get the Hölder condition for any $x, y \in \Omega$. □

[2]Here, again we pick up the dependency on the domain regularity by applying this lemma.

We now show the differentability relations, corresponding to [11, Thrm. 4.II]. Here we present the proof with a different induction assumption than in the original paper, but with the same result.

Proposition 2.19. *Let $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$, Ω satisfy (2.1) and $\sigma > d + kq$ with $k \geq 1$. For each multi index with $|\beta| \leq k-1$, the functions v_β defined in Proposition 2.16 have first order partial derivatives in Ω, and they satisfy*

$$\frac{\partial v_\beta}{\partial x_i} = v_{\beta+e_i} \quad \text{for } i = 1, 2, \ldots, d.$$

Proof. Fix β with $|\beta| \leq k-1$. The previous proposition gives the continuity of all v_γ with $|\gamma| = k$, so we proceed by induction backwards from this case, assuming the hypothesis that v_γ are continuous for all $\gamma \geq \beta$ (for the partial ordering of multi-indices, where all entries $\gamma_i \geq \beta_i$). Let $x_0 \in \Omega$ and r be suitably small so that $x_0 + e_i r \in \Omega$ also. Consider

$$\begin{aligned}\frac{a_\beta(x_0+e_ir, 2|r|) - a_\beta(x_0, 2|r|)}{r} =& \frac{D^\beta\big[P(x_0, x_0+re_i, 2|r|) - P(x_0, x_0, 2|r|)\big]}{r} \\ &+ \frac{D^\beta\big[P(x_0+re_i, x_0+re_i, 2|r|) - P(x_0, x_0+re_i, 2|r|)\big]}{r} \\ :=&(I) + (II).\end{aligned}$$

We claim that $(I) \to 0$ and $(II) \to v_{\beta+e_i}(x_0)$ as $r \to 0$. For (I), as in the proof of Lemma 2.17[3] and using Lemma 2.13 we deduce

$$|(I)|^q \leq \frac{1}{|r|^q}\frac{C}{|r|^{n+|\beta|q}} 2^{q+\sigma+1}\|u\|^q_{k,q,\sigma}|r|^\sigma = C(k, q, \sigma, \beta, \Omega) r^{\sigma-d-(|\beta|+1)q}$$

which vanishes for $r \to 0$ as the exponent of $|r|$ is positive[4].

For (II), set $P_r(x) := P(x, x_0+re_i, 2|r|)$. We claim the induction assumption of continuity gives all coefficients of the polynomial $D^\beta P_r(x)$ converging (a priori they could have diverged). Indeed continuity gives that for each $\gamma \geq \beta$, we have $D^\gamma P_r(x_0) \to v_\gamma(x_0)$. Working backwards from the highest order derivatives, with $|\gamma| = k$ $D^\gamma P_r(x)$ is constant, and is equal (up to a factor of $\gamma!$) to the x^γ coefficient of P_r. Then stepping back one order at a time, all of the coefficients of $D^\gamma P_r(x)$ must have limits, and these limits necessarily must be $v_\gamma(x)$ times the corresponding factorials. That is, setting

$$D^\beta P_r(x) := \sum_{\substack{\beta\leq\gamma \\ |\gamma|\leq k}} c_{r,\gamma} x^{(\gamma-\beta)}; \qquad P(x) := \sum_{\substack{\beta\leq\gamma \\ |\gamma|\leq k}} \frac{v_\gamma(x_0)}{(\gamma-\beta)!}(x-x_0)^{(\gamma-\beta)},$$

[3] Here we can apply this since $|r| = |(x_0 + re_i) - x_0|$, Lemma 2.17 certainly does not hold for any r.
[4] We have $\sigma - d - (|\beta|+1)q \geq \sigma - d - kq > 0$ as here we are assuming $|\beta| \leq k-1$

we have $D^\beta P_r \to P$ uniformly on compacta as $r \to 0$. Since these are analytic, also all derivatives converge and hence we get[5]

$$\lim_{r\to 0} \frac{D^\beta P_r(x_0 + re_i) - D^\beta P_r(x_0)}{r} = D^\beta \frac{\partial P}{\partial x_i}\bigg|_{x=x_0} = v_{\beta+e_i}(x_0).$$

We now show that

$$\lim_{r\to 0} \frac{v_\beta(x_0 + re_i) - v_\beta(x_0)}{r} = \lim_{r\to 0} \frac{a_\beta(x_0 + e_i r, 2|r|) - a_\beta(x_0, 2|r|)}{r}$$

which would establish the result. From Proposition 2.16,

$$\left|\frac{v_\beta(x_0) - a_\beta(x_0, 2|r|)}{r}\right| \leq C|r|^{\left(\frac{\sigma-d-q(|\beta|+1)}{q}\right)}.$$

Since $|\beta| \leq k-1$ the $|r|$ exponent is positive so this quantity vanishes with r. An identical statement also holds replacing x_0 with $x_0 + re_i$. Finally, we note that

$$\begin{aligned}\frac{v_\beta(x_0 + re_i) - v_\beta(x_0)}{r} =& \frac{a_\beta(x_0 + e_i r, 2|r|) - a_\beta(x_0, 2|r|)}{r} \\ &+ \frac{v_\beta(x_0 + re_i) - a_\beta(x_0 + re_i, 2|r|)}{r} \\ &+ \frac{a_\beta(x_0, 2|r|) - v_\beta(x_0)}{r},\end{aligned}$$

which gives the equivalence of limits by the above discussion. □

2.2.3 Campanato ≅ Hölder

Equipped with the above, we can now finally prove both directions of Theorem 2.10, corresponding to [11, Thrm. 5.III]

Proof that Campanato functions are Hölder. The previous sections give $v_{(0)} \in C^{k,\alpha}(\Omega)$, we now prove that $v_0 = u$ a.e. where $(0) \in \mathbb{R}^d$ is the zero multi-index. We show equality holds at each Lebesgue point of u, points x_0 such that

$$\lim_{r\to 0} ⨍_{\Omega(x_0,r)} |u(x) - u(x_0)|^q \mathrm{d}x = 0.$$

For such an x_0, we have

$$\begin{aligned}&|a_{(0)}(x_0, r) - u(x_0)|^q \\ &\qquad \leq C(q)\Big(|P(x, x_0, r) - a_{(0)}(x_0, r)|^q + |P(x, x_0, r) - u(x)|^q + |u(x) - u(x_0)|^q\Big).\end{aligned}$$

[5]If $f_r \to f$ uniformly and $x_r \to x$ then $f_r(x_r) \to f(x)$. In this case, $f_r(x)$ is the difference quotient.

Average integrating over $\Omega(x_0, r)$ and using the regularity property (2.1) gives

$$
\begin{aligned}
|a_{(0)}(x_0,r) - u(x_0)|^q \leq & \frac{C(q)}{C_\Omega r^d} \int_{\Omega(x_0,r)} |P(x,x_0,r) - a_{(0)}(x_0,r)|^q \mathrm{d}x \\
& + \frac{C(q)}{C_\Omega r^d} \int_{\Omega(x_0,r)} |P(x,x_0,r) - u(x)|^q \mathrm{d}x \\
& + \frac{C(q)}{|\Omega(x_0,r)|} \int_{\Omega(x_0,r)} |u(x) - u(x_0)|^q \mathrm{d}x \\
& := (I) + (II) + (III).
\end{aligned}
$$

We claim all terms vanish as $r \to 0$, giving $\lim_{r\to 0} a_{(0)}(x_0, r) = v_{(0)}(x_0)$. (III) vanishes by the definition of a Lebesgue point. For (II), by the definition of $\mathcal{L}_k^{q,\sigma}$

$$
(II) \leq C(q,\Omega)\|u\|_{k,q,\sigma}^q r^{\sigma-d},
$$

which also is infinitesimal with r. Finally for (I), using Proposition 2.16

$$
\begin{aligned}
(I) = & \frac{C(q,\Omega)}{r^d} \int_{\Omega(x_0,r)} \Big| \sum_{1\leq|\beta|\leq k} a_\beta(x_0,r)(x-x_0)^\beta \Big|^q \mathrm{d}x \\
\leq & C(q,k,d,\Omega) \sum_{1\leq|\beta|\leq k} \big|a_\beta(x_0,r)\big| r^{|\beta| q} \to 0,
\end{aligned}
$$

as the limits of $a_\beta(x_0, r)$ are controlled uniformly as $r \to 0$.

Proof that Hölder functions are Campanato. Let $u \in C^{k,\alpha}(\Omega)$, fix x_0 and consider the degree $k-1$ Taylor polynomial of u at x_0. For each x there exists some $y_x \in \mathbb{R}^n$ on the line segment between x and x_0 with

$$
u(x) = \sum_{|\beta|\leq k-1} \frac{D^\beta u(x_0)}{\beta!}(x-x_0)^\beta + \sum_{|\beta|=k} \frac{D^\beta u(y_x)}{\beta!}(x-x_0)^\beta.
$$

by the Lagrange form of the remainder. Thus the distance of u from its degree k Taylor polynomial is controlled by

$$
\begin{aligned}
\Big| u(x) - \sum_{|\beta|\leq k} \frac{D^\beta u(x_0)}{\beta!}(x-x_0)^\beta \Big| = & \Big| \sum_{|\beta|=k} \frac{D^\beta u(y_x) - D^\beta u(x_0)}{\beta!}(x-x_0)^\beta \Big| \\
\leq & [u]_{k,\alpha} \sum_{|\beta|=k} \frac{1}{\beta!} |x-x_0|^{\frac{\sigma-d-kq}{q}} |x-x_0|^k \\
= & \, C(k)[u]_{k,\alpha}|x-x_0|^{\frac{\sigma-n}{q}},
\end{aligned}
$$

where we used that $(x-x_0)^\beta \leq |x-x_0|^k$ for any $|\beta| = k$. Integrating over $\Omega(x_0, r)$ gives

$$
\min_{P\in\mathscr{P}_k} \int_{\Omega(x_0,r)} |u(x) - P(x)|^q \mathrm{d}x \leq C(k)[u]_{k,\alpha}^q r^\sigma,
$$

so that $u \in \mathcal{L}_k^{q,\sigma}(\Omega)$, and Theorem 2.10 is proven. □

3 Regularity for elliptic equations

Let $\Omega \subset \mathbb{R}^d$ be a bounded, open connected set. In this chapter, we discuss the regularity of solutions to scalar linear elliptic PDE of the form

$$\nabla \cdot (A\nabla u) = \nabla \cdot F \quad \text{in } \Omega, \tag{3.1}$$

for a vector field $F : \Omega \to \mathbb{R}^d$ and a matrix-valued $A : \Omega \to \mathbb{R}^{d\times d}$ which is always assumed to be uniformly positive definite and bounded; in the sense that there exists $0 < \lambda \leq \Lambda$ such that for each $x \in \Omega$,

$$\lambda|y|^2 \leq y^T A(x) y \leq \Lambda |y|^2 \text{ for all } y \in \mathbb{R}^d.$$

We call u a "solution" to (3.1) if $u \in H^1(\Omega)$ and

$$\int_\Omega (A\nabla u) \cdot \nabla\varphi \mathrm{d}x = \int_\Omega f \cdot \nabla\varphi \mathrm{d}x \quad \text{for all } \varphi \in H_0^1(\Omega),$$

and we do not consider any boundary data on $\partial\Omega$, choosing to focus on the interior regularity. Where it is clear throughout the chapter, we take $B_R(x_0) = B_R$ as well as ignoring integration measures where clear, to ease some of the notation.

3.1 Interior Sobolev regularity for L^2 coefficients

Our goal in this section is to establish Sobolev regularity of solutions to our PDE (3.1). The idea is to consider a difference quotient

$$D_i^h u(x) := \frac{u(x + he_i) - u(x)}{h} \text{ for } x \in \Omega_{i,h} := \{x \in \Omega : x + he_i \in \Omega\},$$

for a unit vector e_i. This acts as a sort of discrete derivative, which we will show is uniformly bounded over h in L^2 norm for solutions to our PDE. Then we use that bounded sequences in reflexive Banach spaces have a weakly convergent subsequence to deduce a limit as $h \to 0$. We show that the limit of this satisfies the required properties to be a weak derivative of u, thus giving higher regularity (we actually do this process for ∇u, not u). This section follows [14, Ch. 4.2, 4.3], as well as [26, Ch. 5.2].

3.1.1 The method of difference quotients

We begin by gathering some results regarding difference quotients, as found in [14, Ch. 4.3].

Proposition 3.1. *Let $u \in H^1(\Omega)$, then $D_i^h u \in H^1(\Omega_{i,h})$ and the commutativity with weak derivatives $D(D_i^h u) = D_i^h(Du)$ holds. If either $u, v \in H^1(\Omega)$ are compactly supported in Ω then for suitably small h,*

$$\int_\Omega u D_i^h v \mathrm{d}x = -\int_\Omega D_i^{-h} u v \mathrm{d}x.$$

We also have the Leibniz rule

$$D_i^h(uv)(x) = u(x + he_i) D_i^h v(x) + D_i^h u(x) v(x).$$

Proof. The first two claims follow directly from the linearity of Sobolev spaces and weak derivatives. For the second, direct calculation yields

$$\int_\Omega u D_i^h v + D_i^{-h} u v \mathrm{d}x = \frac{1}{h} \int_\Omega u(x) v(x + he_i) - u(x - he_i) v(x) \mathrm{d}x. \tag{3.2}$$

Since one of the functions is compactly supported the infimum distance of the support from the boundary is positive, so for h small, the change of variables on the first term $y = x + he_i$ keeps the support of $uD_i^h v$ and $D_i^{-h} uv$ inside Ω, and thus the two terms in (3.2) cancel out. The Leibniz rule is just a direct computation. □

Lemma 3.2. *Let $1 < p < \infty$. For every $\Omega_0 \Subset \Omega$, $u \in W^{1,p}(\Omega)$, $i = 1, \dots, n$ and $|h| < \mathrm{Dist}(\Omega_0, \Omega)/2$ we have*

$$\|D_i^h u\|_{L^p(\Omega_0)} \leq \|\nabla u\|_{L^p(\Omega)}.$$

Proof. since $C^\infty(\Omega)$ is dense in $W^{1,p}(\Omega)$ and the inequality is stable under L^p, hence Sobolev convergence, we need only prove it for $u \in C^\infty(\Omega)$. The fundamental theorem of calculus gives

$$D_i^h u(x) = \fint_0^h \frac{\partial}{\partial x_i} u(x + te_i) \mathrm{d}t,$$

then we use Jensen's inequality followed by Fubini's theorem,

$$\begin{aligned} \|D_i^h u\|_{L^p(\Omega_0)}^p &= \int_{\Omega_0} \left| \fint_0^h \frac{\partial}{\partial x_i} u(x + te_i) \mathrm{d}t \right|^p \mathrm{d}x \\ &\leq \int_{\Omega_0} \fint_0^h |\nabla u(x + te_i)|^p \mathrm{d}t \mathrm{d}x \\ &\leq \fint_0^h \int_\Omega |\nabla u|^p \mathrm{d}x \mathrm{d}t = \|\nabla u\|_{L^p(\Omega)}^p. \end{aligned}$$

□

Proposition 3.3. *Let $1 < p < \infty$ and $\Omega_0 \Subset \Omega$. Let $u \in L^p(\Omega)$ be such that there exists $M > 0$ with*

$$\|D_i^h u\|_{L^p(\Omega_0)} \leq M \text{ for all } |h| < \mathrm{Dist}(\Omega_0, \Omega) \text{ and each } i = 1, \dots n.$$

Then $u \in W^{1,p}(\Omega_0)$ with $\|\partial_{x_i} u\|_{L^p(\Omega_0)} \leq M$, and $D_i^h u \to \partial_{x_i} u$ in $L^p(\Omega_0)$ as $h \to 0$ for each i.

Proof. Fix i, by the reflexivity of $L^p(\Omega)$ and that $D_i^h u$ is uniformly bounded in h, we have a weakly converging subsequence h_m with limit we call g. By definition of weak convergence in L^p as a duality product with L^q with q the Hölder conjugate, and that $C_c^\infty(\Omega_0) \subset L^q(\Omega_0)$, we have for all $\varphi \in C_c^\infty(\Omega_0)$,

$$\begin{aligned}\int_{\Omega_0} g\varphi \mathrm{d}x &= \lim_{h_m \to 0} \int_{\Omega_0} D_i^{h_m} u\varphi \mathrm{d}x \\ &= \lim_{h_m \to 0} - \int_{\Omega_0} u D_i^{-h_m} \varphi \mathrm{d}x \\ &= - \int_{\Omega_0} u \frac{\partial \varphi}{\partial x_i} \mathrm{d}x.\end{aligned}$$

Thus $g = \frac{\partial u}{\partial x_i}$ in the sense of distributions, so the distributional derivative has an L^p representative, hence $u \in W^{1,p}(\Omega_0)$. For the strong convergence, taking an arbitrary $v \in C^\infty(\Omega_0)$,

$$D_i^h u - \frac{\partial u}{\partial x_i} = D_i^h(u - v) + \frac{\partial}{\partial x_i}(u - v) + (D_i^h v - \frac{\partial v}{\partial x_i}),$$

so that

$$\left\| D_i^h u - \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega_0)} \leq \left\| D_i^h(u - v) \right\|_{L^p(\Omega_0)} + \left\| \frac{\partial}{\partial x_i}(u - v) \right\|_{L^p(\Omega_0)} + \left\| (D_i^h v - \frac{\partial v}{\partial x_i}) \right\|_{L^p(\Omega_0)}.$$

Taking v arbitrarily close to u by density of $C^\infty(\Omega_0)$ in $W^{1,p}(\Omega_0)$ makes the first two terms arbitrarily small. Then since we have strong convergence of the difference quotient, the third term goes to 0 as $h_m \to 0$. □

The following inequality acts as a sort of "reverse-Poincaré" inequality and will be crucial for uniformly bounding the difference quotient. This is adapted from the case for systems found in [14, Ch. 4.2].

Theorem 3.4 (Caccioppoli inequality). *Let u solve (3.1) with A having eigenvalues between $0 < \lambda \leq \Lambda$ for each x. Then there exists a constant $C(\lambda, \Lambda) > 0$ such that for any $r < R$,*

$$\int_{B_r(x_0)} |\nabla u|^2 \leq C \left(\frac{1}{(R - r)^2} \int_{B_R(x_0) \setminus B_r(x_0)} (u - a)^2 + \int_{B_R(x_0)} |F|^2 \right)$$

for every $x_0 \in \Omega$ and $a \in \mathbb{R}$.

Proof. Take a cut-off function $\eta \in C_c^\infty(\Omega)$ with the following properties:

1. $\eta \equiv 1$ on B_r and $\eta \equiv 0$ on $\Omega \setminus B_R$
2. $0 \leq \eta \leq 1$
3. $|\nabla \eta| \leq \frac{2}{R-r}$.[1]

Considering the test function $\varphi = \eta^2(u-a)$ with $\nabla\varphi = \eta^2\nabla u + 2\eta(u-a)\nabla\eta$ gives

$$\begin{aligned}\int_{B_R} \eta^2 (A\nabla u)\cdot\nabla u &= -\int_{B_R} 2\eta(u-a)(A\nabla u)\cdot\nabla\eta \\ &\quad + \int_{B_R} 2\eta(u-a)\nabla\eta\cdot F + \int_{B_R}\eta^2\nabla u\cdot F \\ &:= (I) + (II) + (III). \end{aligned} \tag{3.3}$$

By uniform positive definiteness and properties of η we also have

$$\int_{B_r} |\nabla u|^2 \leq \int_{B_R} \eta^2|\nabla u|^2 \leq \frac{1}{\lambda}\int_{B_R}\eta^2(A\nabla u)\cdot\nabla u. \tag{3.4}$$

So if we can bound the three terms (3.3) suitably then we are done. These can all be estimated using Young's inequality $2AB \leq \varepsilon A^2 + \frac{B^2}{\varepsilon}$ for $\varepsilon > 0$. For the first integral,

$$\begin{aligned}(I) &= -\sum_{i,j=1}^n \int_{B_R} A^{ij}(x)2\eta(u-a)\frac{\partial\eta}{\partial x_i}\frac{\partial u}{\partial x_j} \\ &\leq \sum_{i,j=1}^n \int_{B_R} \frac{A^{ij}(x)}{\varepsilon}\left((u-a)\frac{\partial\eta}{\partial x_i}\right)^2 + \varepsilon A^{ij}(x)\left(\eta\frac{\partial u}{\partial x_j}\right)^2 \\ &\leq \frac{\Lambda}{\varepsilon}\int_{B_R}(u-a)^2|\nabla\eta|^2 + \varepsilon\Lambda\int_{B_R}\eta^2|\nabla u|^2 \\ &\leq \frac{4\Lambda}{\varepsilon(R-r)^2}\int_{B_R\setminus B_r}(u-a)^2 + \varepsilon\Lambda\int_{B_R}\eta^2|\nabla u|^2,\end{aligned}$$

where we used that $\nabla\eta = 0$ on the smaller ball as well as the bound $|\nabla\eta| \leq \frac{2}{R-r}$ on the final line. Identical estimates give

$$\begin{aligned}(II) &\leq \frac{4}{\varepsilon(R-r)^2}\int_{B_R\setminus B_r}(u-a)^2\mathrm{d}x + \varepsilon\int_{B_R}|F|^2 \\ (III) &\leq \frac{\varepsilon}{2}\int_{B_R}\eta^2|\nabla u|^2\mathrm{d}x + \frac{1}{2\varepsilon}\int_{B_R}\eta^2|F|^2.\end{aligned}$$

Then choose ε such that $\lambda > \varepsilon(\Lambda + 1/2)$ so that the coefficient of the $\eta^2|\nabla u|^2$ term on the RHS of (3.3) is less than λ so we can rearrange considering (3.4), thus giving the required estimate. □

[1] We can choose such an η with this bound since to interpolate linearly down from 1 to 0 as we move out radially from r to R would give a gradient of $1/(R-r)$, so we allow the factor of 2 to smooth out either end.

3.1.2 Sobolev regularity

Equipped with the above, we now establish Sobolev regularity, as in [14, Thrm. 4.9].

Theorem 3.5 (Interior regularity)**.** *Let u solve (3.1) with the additional assumption of the coefficients that $F \in H^1(\Omega)$ and A is Lipschitz. Then $u \in H^2_{\text{loc}}(\Omega)$.*

Proof. By density, for a weak solution it is sufficient to only test against $\varphi \in C_c^\infty(\Omega)$. Since Ω is open, the support of each φ is bounded away from the boundary. Thus for $h \ll 1$, we can also test against the translated $\varphi(x - he_k)$ as its support lies in Ω. Then after a change of variables, we ascertain

$$\int_\Omega (A(x + he_i)\nabla u(x + he_i)) \cdot \nabla\varphi(x) = \int_\Omega F(x + he_i) \cdot \nabla\varphi(x).$$

Subtracting the un-transposed weak formulation from the above and dividing by h gives

$$\int_\Omega D_i^h(A\nabla u) \cdot \nabla\varphi = \int_\Omega D_i^h F \cdot \nabla\varphi$$
$$\int_\Omega (A\nabla D_i^h u) \cdot \nabla\varphi + \int_\Omega (D_i^h A\nabla u) \cdot \nabla\varphi = \int_\Omega D_i^h F \cdot \nabla\varphi,$$

where D_i^h acts component-wise on vectors/ matrices, and we used the Leibniz rule for the commutativity of the difference quotient with the derivative from Proposition 3.1. Now restricting to any ball $B_{2R}(x_0) \subset \Omega$, this tells us that $D_i^h u$ is a weak solution to the PDE in v

$$\nabla \cdot (A\nabla v) = \nabla \cdot (D_i^h F - D_i^h A\nabla u), \tag{3.5}$$

where here we considered u as a fixed term. The Caccioppoli inequality gives

$$\begin{aligned}\int_{B_R} |D_i^h \nabla u|^2 \leq& \frac{C}{R^2}\int_{B_{2R}} (D_i^h u)^2 + C\int_{B_{2R}} |D_i^h F|^2 + C\int_{B_{2R}} |D_i^h A|^2 |\nabla u|^2 \\ :=& (I) + (II) + (III).\end{aligned}$$

Since u and F are both Sobolev, the L^2 norms of their derivatives are finite and provide an upper bound for that of the difference quotients, so by Proposition 3.2 (I) and (II) are uniformly bounded in h. The Lipschitz assumption on A gives a uniform bound to $|D_i^h A(x)|$, so again using Proposition 3.2 combined with $\nabla u \in L^2$, the third term is also uniformly bounded in h. Thus $D_i^h \nabla u$ is uniformly bounded in L^2 norm over h, so applying Proposition 3.3 to each component followed by a covering argument establishes the regularity $u \in H^2_{\text{loc}}(\Omega)$. □

We can extend the above to equations with higher-order coefficients by induction, à la [14, Thrm. 4.11].

Theorem 3.6 (Higher order regularity)**.** *Let u solve (3.1), and assume the additional regularity of the coefficients*

1. $F \in H^{k+1}(\Omega)$,
2. $A^{ij} \in C^{k,1}$, so that $D^\beta A^{ij}$ is Lipschitz for any multi-index $|\beta| = k$.

Then $u \in H^{k+2}_{\text{loc}}(\Omega)$.

Proof. The case $k = 0$ is precisely Theorem 3.5, so we proceed by induction assuming the result holds for $k-1$. Take $\psi \in C_c^\infty(\Omega)$ and test against $\varphi := \frac{\partial \psi}{\partial x_i}$, then integrating by parts gives

$$\int_\Omega \frac{\partial}{\partial x_i}(A\nabla u) \cdot \nabla \psi = \int_\Omega \frac{\partial F}{\partial x_i} \cdot \nabla \psi$$
$$\int_\Omega A\left(\nabla \frac{\partial u}{\partial x_i}\right) \cdot \nabla \psi = \int_\Omega \left(-\frac{\partial A}{\partial x_i}\nabla u + \frac{\partial F}{\partial x_i}\right) \cdot \nabla \psi.$$

The above gives $\frac{\partial u}{\partial x_i}$ as a weak solution to the PDE $\nabla \cdot (A(x)\nabla u) = \nabla \cdot \tilde{F}$, where

$$\tilde{F} := -\frac{\partial A}{\partial x_i}\nabla u + \frac{\partial F}{\partial x_i}.$$

Note that $\tilde{F} \in H^k_{\text{loc}}(\Omega)$, by the induction assumptions and that any kth derivative of A is Sobolev[2]. Thus the induction assumption gives $\frac{\partial u}{\partial x_i} \in H^{k+1}_{\text{loc}}(\Omega_0)$ so that $u \in H^{k+2}_{\text{loc}}(\Omega_0)$. The estimate also follows inductively from the computations. □

Corollary 3.6.1. *Let u be a weak solution of $\nabla \cdot (A(x)\nabla u) = \nabla \cdot F$, with A uniformly positive definite and bounded, with $F, A \in C^\infty(\Omega)$. Then $u \in C^\infty(\Omega)$.*

Proof. The above theorem gives $u \in H^k_{\text{loc}}(\Omega)$ for all $k \geq 0$, so the result follows by Sobolev embeddings. □

3.2 Schauder estimates: interior Hölder regularity

In this section, we discuss Hölder regularity of solutions based on that of the coefficients. It turns out that this is best attacked obliquely, making use of the theory of Campanato and Hölder spaces developed in the previous chapter. This approach is well-motivated since the weak formulation of the PDE is an integral condition, so showing energy bounds seems tractable.

[2]Lipschitz functions are always locally Sobolev, and Sobolev on bounded domains by Rademacher's theorem.

3.2.1 Energy growth estimates

To prove that solutions u to our PDE lie in $L^{q,\sigma}$ or $\mathcal{L}_0^{q,\sigma}$ spaces, we need to show that the growth of energies of the form $\phi : [0, \mathbb{R}) \to [0, \mathbb{R})$

$$\phi(r) = \int_{B_r} |u - u_{x_0,r}|^2 \mathrm{d}x \quad \text{or} \quad \phi(r) = \int_{B_r} |u|^2 \mathrm{d}x$$

satisfy σ growth conditions, being controlled globally by constants $\phi(r) \leq Cr^\sigma$. To help establish this we will need some estimates and the following lemma, found in both [14, Ch. 5] and [26, Ch. 5].

Lemma 3.7 (Campanato iteration)**.** *Let $\phi : [0, R_0] \to \mathbb{R}^+$ be an increasing function, which for $r < R$ satisfies*

$$\phi(r) \leq A\left(\varepsilon + \left(\frac{r}{R}\right)^\alpha\right)\phi(R) + BR^\beta$$

for some constants $A, B, \varepsilon \geq 0$ and $\alpha > \beta > 0$. Then there exists $\varepsilon_0 = \varepsilon_0(A, \alpha, \beta)$ such that if $\varepsilon < \varepsilon_0$ then $\phi(r) \leq Cr^\beta$ for all $r \in [0, R_0]$ for some $C = C(A, B, R_0, \phi(R_0), \alpha, \beta)$.

Proof. For $0 < \tau < 1$ we have $\phi(\tau R) \leq (\varepsilon + \tau^\alpha)\phi(R) + BR^\beta$. We assume without loss of generality that $2A > 1$, and take $\gamma \in (\beta, \alpha)$. Then we can choose $\tau \in (0, 1)$ such that $2A\tau^\alpha = \tau^\beta$. Choosing $\varepsilon_0 = \tau^\alpha$, we have for $\varepsilon < \varepsilon_0$

$$\phi(\tau R) \leq \tau^\gamma \phi(R) + BR^\beta.$$

Iterating this estimate gives

$$\begin{aligned}
\phi(\tau^k R) \leq& \tau^\gamma \phi(\tau^{k-1}R) + B\tau^{(k-1)\beta}R^\beta \\
\leq& \tau^{k\gamma}\phi(R) + BR^\beta \tau^{(k-1)\beta} \sum_{j=0}^{k-1} \tau^{j(\gamma-\beta)} \\
\leq& \tau^{k\beta}\phi(R) + BR^\beta \tau^{(k-1)\beta} \sum_{j=0}^{\infty} \tau^{j(\gamma-\beta)} \\
\leq& \left(\tau^{-\beta} + \tau^{-2\beta} \sum_{j=0}^{\infty} \tau^{j(\gamma-\beta)}\right)(\phi(R_0) + BR_0^\beta)\tau^{(k+1)\beta} \\
\leq& C(A, B, \alpha, \beta, R_0, \phi(R_0))\tau^{(k+1)\beta}
\end{aligned}$$

since $|\tau| < 1$ and $\gamma > \beta$ so the series is convergent and thus bounded, and we also used the monotonicity of ϕ. Now for any r choose $k \in \mathbb{N}$ such that $\tau^{k+1}R \leq r \leq \tau^k R$ then

$$\phi(r) \leq \phi(\tau^k R) \leq C\tau^{(k+1)\beta} \leq Cr^\beta,$$

by the monotonicity of ϕ and the fact that $\tau^{k+1} \leq \frac{r}{R} \leq r$. □

The following gives very strong controls on the decay of $\fint_{B_r} |u|^2$ in terms of r, proved as in [26, Prop. 5.7].

Proposition 3.8. *Let u be a weak solution of $\nabla \cdot (A\nabla u) = 0$ in $B_R(x_0)$, where A is a constant positive definite matrix with eigenvalues bounded between $0 < \lambda \leq \Lambda$. Then there exists a constant $C = C(\lambda, \Lambda)$ such that for all $R > r > 0$ and $x_0 \in \Omega$ we have*

$$\int_{B_r(x_0)} |u|^2 \mathrm{d}x \leq C \left(\frac{r}{R}\right)^d \int_{B_R(x_0)} |u|^2 \mathrm{d}x \tag{3.6}$$

and

$$\int_{B_r(x_0)} |u - u_{x_0,r}|^2 \mathrm{d}x \leq C \left(\frac{r}{R}\right)^{d+2} \int_{B_R(x_0)} |u - u_{x_0,R}|^2 \mathrm{d}x, \tag{3.7}$$

where $u_{x_0,r} := \fint_{B_r} u \mathrm{d}x$ is the average of u on $B_r(x_0)$.

Proof. For (3.6), we begin by noting that if $\nabla \cdot (A\nabla u) = 0$ then $u_A(x) := u(A^{1/2}x)$ is harmonic[3]. Harmonic functions stay harmonic after an orthogonal change of basis so we need only consider a diagonal matrix. For $u_D(x) := u(D^{1/2}x)$ with diagonal matrix D, the expression for Δu_D is precisely $\nabla \cdot (D\nabla u)$.

Since $(\cdot)^2$ is convex, $|u_A|^2$ is subharmonic, so mean value formulas give

$$\fint_{B_r} |u_A|^2 \mathrm{d}x \leq \fint_{B_R} |u_A|^2 \mathrm{d}x,$$
$$\int_{B_r} |u_A|^2 \mathrm{d}x \leq \left(\frac{r}{R}\right)^d \int_{B_R} |u_A|^2 \mathrm{d}x.$$

Applying the change of variables $y = A^{1/2}x$ and noting that $B_{\lambda r} \subset A^{1/2}B_r$ and $A^{1/2}B_R \subset B_{\Lambda R}$ gives

$$\int_{B_{\lambda r}} |u|^2 \mathrm{d}x \leq \left(\frac{r}{R}\right)^d \int_{B_{\Lambda R}} |u|^2 \mathrm{d}x.$$

Now with $r' = \lambda r$ and $R' = \Lambda R$, this gives

$$\int_{B'_r} |u|^2 \mathrm{d}x \leq C \left(\frac{r'}{R'}\right)^d \int_{B'_R} |u|^2 \mathrm{d}x$$

for the constant $C = \left(\frac{\Lambda}{\lambda}\right)^d$. Here we are still under the condition $r < R$ for our application of the mean value formula, so we have proven the inequality in the case $r' \leq \frac{\lambda}{\Lambda}R'$. Alternatively, if we were to restrict ourselves to the case $r \geq aR$ for some a, we can take $C = a^{-d}$, since

$$\int_{B_r} |u|^2 \mathrm{d}x \leq \left(\frac{R}{r}\right)^d \left(\frac{r}{R}\right)^d \int_{B_R} |u|^2 \mathrm{d}x \leq C \left(\frac{r}{R}\right)^n \int_{B_R} |u|^2 \mathrm{d}x.$$

[3] Here, $A^{1/2}$ is the square root defined for positive matrices as $A^{1/2} := P^T D^{1/2} P$ where $A = P^T DP$ is the diagonalisation of A with P orthogonal and D diagonal with positive entries, and $D^{1/2}$ is given by taking the square root of each diagonal entry.

So in fact for any $r < R$ without restriction, we can take $C = \left(\frac{\Lambda}{\lambda}\right)^d$.

For (3.7), we use the Poincaré-Wirtinger inequality (Theorem B.4) combined with the first inequality, and then Caccioppoli's inequality,

$$\begin{aligned}\int_{B_r} |u - u_{x_0,r}|^2 \mathrm{d}x \leq & Cr^2 \int_{B_r} |\nabla u|^2 \mathrm{d}x \\ \leq & Cr^2 \left(\frac{r}{R}\right)^d \int_{B_R} |\nabla u|^2 \mathrm{d}x \\ \leq & Cr^2 \left(\frac{r}{R}\right)^d \left(\frac{1}{R}\right)^2 \int_{B_{2R}} |u - u_{x_0,2R}|^2 \mathrm{d}x.\end{aligned}$$

This gives us a global constant for the case $r < R/2$. Now for the cases $r \geq aR$, noting that $t \mapsto \int_{B_r} (u-t)^2 \mathrm{d}x$ is minimised[4] by taking $t = u_{x_0,r}$;

$$\begin{aligned}\int_{B_r} |u - u_{x_0,r}|^2 \mathrm{d}x \leq & \int_{B_r} |u - u_{x_0,R}|^2 \mathrm{d}x \\ & \leq \int_{B_R} |u - u_{x_0,r}|^2 \mathrm{d}x \leq \left(\frac{r}{R}\right)^{d+2} \left(\frac{1}{a}\right)^{d+2} \int_{B_R} |u - u_{x_0,r}|^2 \mathrm{d}x.\end{aligned}$$

So taking $C = 2^{d+2}$ works for all pairs of radii with $r \geq R/2$, and taking the max of the two estimates gives the global bound as before. □

Remark 3.9. Since the coefficients are constant and the equation is linear, if we knew higher regularity of u we could differentiate the PDE to get any derivative of u satisfying

$$\int_{B_r(x_0)} |\nabla u|^2 \mathrm{d}x \leq C \left(\frac{r}{R}\right)^d \int_{B_R(x_0)} |\nabla u|^2 \mathrm{d}x,$$

and similar for the second inequality or with any higher-order derivatives we know to exist.

Remark 3.10. In both inequalities, the difficulty was bounding the energies for arbitrarily small ratios of r/R, with the estimates for large values of r/R not using any of the data from the PDE. So the important part of this result is that the local energy of solutions to elliptic equations grows at most sublinearly as we increase the radius, one cannot have local sharp spikes in the gradient without these affecting the global behaviour.

Finally, we will need the following control in terms of the flux [26, Lemma 5.9].

Lemma 3.11. *Let $F \in L^2(B_R)$ and A be uniformly positive definite and bounded with eigenvalues between $0 < \lambda < \Lambda$. Then there exists a unique solution $w \in H_0^1(B_R)$ of (3.1), and this solution satisfies*

$$\lambda^2 \int_{B_R} |\nabla w|^2 \mathrm{d}x \leq \int_{B_R} |F - F_{x_0,R}|^2 \mathrm{d}x \leq \int_{B_R} |F|^2 \mathrm{d}x,$$

[4] Simply expand the square and use linearity then optimise in t.

where $F_{x_0,R} \in \mathbb{R}^d$ is the componentwise averaged vector, s.t. $(F_{x_0,R})_i = \fint F_i \mathrm{d}x$.

Proof. The existence and uniqueness are a result of minimising the strictly convex functional $w \mapsto \int(\frac{1}{2}A(x)|\nabla w|^2 - F \cdot \nabla\varphi)\mathrm{d}x$. For the bound, we test the weak formulation against w giving

$$\int_{B_R} A\nabla w \cdot \nabla w \mathrm{d}x = \int_{B_R} F \cdot \nabla w \mathrm{d}x = \int_{B_R} (F - F_{x_0,R}) \cdot \nabla w \mathrm{d}x,$$

where we smuggled in the additional term since $w \in H_0^1(B_R)$ so $\int \nabla w \cdot y \mathrm{d}x = 0$ for each constant $y \in \mathbb{R}^d$. Similar to the proof of Caccioppoli's inequality, we use the lower bound of the LHS integrand $\lambda|\nabla w|^2$:

$$\begin{aligned}
\lambda \int_{B_R} |\nabla w|^2 \mathrm{d}x &\leq \int_{B_R} (F - F_{x_0,R}) \cdot \nabla w \mathrm{d}x \\
&\leq \left(\int_{B_R} |F - F_{x_0,R}|^2 \mathrm{d}x\right)^{\frac{1}{2}} \left(\int_{B_R} |\nabla w|^2 \mathrm{d}x\right)^{\frac{1}{2}} \\
\lambda^2 \int_{B_R} |\nabla w|^2 \mathrm{d}x &\leq \int_{B_R} |F - F_{x_0,R}|^2 \mathrm{d}x.
\end{aligned}$$

□

3.2.2 Hölder regularity

We can now establish the regularity of solutions to (3.1). We do this in three stages, first for constant coefficients, then continuous coefficients, then a general statement for Hölder continuous coefficients. These correspond to [14, Ch. 5] and [26, Ch. 5].

Theorem 3.12 (Regularity with constant coefficients)**.** *Let u solve (3.1) with A constant and positive definite. If $F \in \mathcal{L}_0^{2,\sigma}(\Omega)$ with $\sigma < d + 2$. Then $\nabla u \in \mathcal{L}_{0,\mathrm{loc}}^{2,\sigma}(\Omega)$.*

Proof. Take an arbitrary $B_R \subset \Omega$, and define $\phi(r) := \int_{B_r} |\nabla u - (\nabla u)_{x_0,r}|^2 \mathrm{d}x$. We want a global C such that $\phi(r) \leq Cr^\sigma$. We begin by decomposing into homogeneous and inhomogeneous parts $u = w + v$, where w and v are the unique solutions to

$$\begin{cases} \nabla \cdot (A\nabla w) = \nabla \cdot F & \text{in } B_R \\ w = 0 & \text{on } \partial B_R, \end{cases} \qquad \begin{cases} \nabla \cdot (A\nabla v) = 0 & \text{in } B_R \\ v = u & \text{on } \partial B_R. \end{cases}$$

Note that $(\nabla u)_{x_0,R} = (\nabla v)_{x_0,R}$ since $(\nabla w)_{x_0,R} = 0$. We then estimate

$$\begin{aligned}
\phi(r) = \int_{B_r} |\nabla u - (\nabla u)_{x_0,r}|^2 \mathrm{d}x &= \int_{B_r} |\nabla v - (\nabla v)_{x_0,r} + \nabla w|^2 \mathrm{d}x \\
&\le 2 \int_{B_r} |\nabla v - (\nabla v)_{x_0,r}|^2 \mathrm{d}x + 2 \int_{B_r} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d+2} \int_{B_R} |\nabla v - (\nabla v)_{x_0,R}|^2 \mathrm{d}x + 2 \int_{B_R} |\nabla w|^2 \mathrm{d}x \\
&= C \left(\frac{r}{R}\right)^{d+2} \int_{B_R} |\nabla u - (\nabla u)_{x_0,R} - \nabla w|^2 \mathrm{d}x + 2 \int_{B_R} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d+2} \int_{B_R} |\nabla u - (\nabla u)_{x_0,R}|^2 \mathrm{d}x + C \int_{B_R} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d+2} \phi(R) + C \int_{B_R} |F - F_{x_0,R}|^2 \mathrm{d}x \le C \left(\frac{r}{R}\right)^{d+2} \phi(R) + C R^\sigma,
\end{aligned}$$

where we used Proposition 3.8 on the third line[5], then Lemma 3.11 followed by $F \in \mathcal{L}_0^{2,\sigma}(\Omega)$ on the final line.

Thus the conditions for a Campanato iteration (Lemma 3.7) apply, giving C with $\phi(r) \le Cr^\sigma$, and hence an upper bound for the Campanato seminorm on an arbitrary ball of radius R inside Ω. A covering argument gives $\nabla u \in \mathcal{L}_0^{2,\sigma}(K)$ for each $K \Subset \Omega$. □

Theorem 3.13 (Regularity with continuous coefficients)**.** *Let u solve (3.1) with A continuous, uniformly positive definite and bounded. If $F \in L^{2,\sigma}(\Omega)$ for some $\sigma < d$ then $\nabla u \in L^{2,\sigma}_{\mathrm{loc}}(\Omega)$.*

Proof. Fix a ball $B_R = B_R(x_0) \subset \Omega$, and decompose $u = v + w$, with w and v the unique solutions to

$$\begin{cases}
\nabla \cdot (A(x_0)\nabla w) = \nabla \cdot \big(F + (A(x_0) - A)\nabla u\big) & \text{in } B_R \\
w = 0 & \text{on } \partial B_R, \\
\nabla \cdot (A\nabla v) = 0 & \text{in } B_R \\
v = u & \text{on } \partial B_R.
\end{cases} \tag{3.8}$$

Set $G := (F + (A(x_0) - A))\nabla u$ and $\phi(r) := \int_{B_R} |\nabla u|^2 \mathrm{d}x$, then as before

$$\begin{aligned}
\phi(r) &\le 2 \int_{B_r} |\nabla v|^2 \mathrm{d}x + 2 \int_{B_r} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d} \int_{B_R} |\nabla v|^2 \mathrm{d}x + 2 \int_{B_R} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d} \int_{B_R} |\nabla u|^2 \mathrm{d}x + C \int_{B_R} |\nabla w|^2 \mathrm{d}x \\
&\le C \left(\frac{r}{R}\right)^{d} \phi(R) + C \int_{B_R} |G|^2 \mathrm{d}x.
\end{aligned} \tag{3.9}$$

[5]Formally, we are using Remark 3.9 and the additional 1 step regularity we have from Theorem 3.5.

The final term is bounded using

$$\int_{B_R} |G|^2 \mathrm{d}x \leq 2 \int_{B_R} |F|^2 \mathrm{d}x + 2\omega(R)^2 \int_{B_R} |\nabla u|^2 \mathrm{d}x$$

where $\omega(R) := \sup_{x,y \in B_R} |A(x) - A(y)|$ is the modulus of continuity of A. Then since $F \in L^{2,\sigma}$,

$$\phi(r) \leq C \left(\frac{r}{R}\right)^d \phi(R) + CR^\sigma + C\varepsilon\phi(R).$$

Taking $\omega(R)$ suitably small for the ε in Lemma 3.7 gives $\phi(r) \leq Cr^\sigma$, hence $\nabla u \in L^{2,\sigma}(B_R)$, and the result once again follows from a covering argument. □

Theorem 3.14 (Regularity with Hölder coefficients)**.** *Let u solve (3.1) with A uniformly positive definite and bounded and $A, F \in C^{0,\alpha}(\Omega)$. Then $\nabla u \in C^{0,\alpha}_{\mathrm{loc}}(\Omega)$.*

Proof. Fix $B_R \subset \Omega$, define $\phi(r) := \int_{B_R} |\nabla u - (\nabla u)_{x_0,r}|^2 \mathrm{d}x$ and decompose $u = v + w$ as in (3.8). then the computations up to (3.9) give

$$\phi(r) \leq C \left(\frac{r}{R}\right)^{n+2} \phi(R) + C \int_{B_R} |\nabla w|^2 \mathrm{d}x. \tag{3.10}$$

The second term is bounded this time using the sharper estimate of Lemma 3.11, giving

$$\begin{aligned} \int_{B_R} |\nabla w|^2 \leq& \frac{1}{\lambda} \int_{B_R} |G - G_{x_0,R}|^2 \mathrm{d}x \leq \frac{1}{\lambda} \int_{B_R} |G - F_{x_0,R}|^2 \mathrm{d}x. \\ \leq& C \int_{B_R} |F - F_{x_0,R}|^2 \mathrm{d}x + C\omega(R)^2 \int_{B_R} |\nabla u|^2 \mathrm{d}x. \end{aligned} \tag{3.11}$$

Since A is α-Hölder continuous we have $\omega(R)^2 \leq CR^{2\alpha}$. By Remark 2.7 we know

$$F \in C^{0,\alpha} \cong \mathcal{L}^{2,d+2\alpha}_0 \subseteq \mathcal{L}^{2,d-\varepsilon}_0 \cong L^{2,d-\varepsilon}$$

for each $\varepsilon > 0$. Theorem 3.13 gives $\nabla u \in L^{2,d-\varepsilon}_{\mathrm{loc}}$. These estimates combine to give

$$\omega(R)^2 \int_{B_r} |\nabla u|^2 \leq CR^{2\alpha+d-\varepsilon}.$$

Combining (3.10) and (3.11) gives

$$\phi(r) \leq C \left(\frac{r}{R}\right)^{d+2} \phi(R) + CR^{2\alpha+d} + CR^{2\alpha+d-\varepsilon}.$$

Then a Campanato iteration (Lemma 3.7) gives[6] $\phi(r) \leq Cr^{2\alpha+d-\varepsilon}$, which in turn gives $\nabla u \in \mathcal{L}^{2,2\alpha+d-\varepsilon}_{0,\mathrm{loc}}(\Omega) \cong C^{0,\alpha-\varepsilon/2}_{\mathrm{loc}}(\Omega)$. In particular, since ∇u is continuous it is locally L^∞, so restricting to any $B_r \Subset \Omega$,

$$\int_{B_r} |\nabla u|^2 \mathrm{d}x \leq |B_1| \|\nabla u\|_{L^\infty(B_r)} r^d,$$

[6] Here we can replace the term with $2\alpha + d$ with one with an ε if we make the constant large enough as there is an upper bound on R.

which gives $\omega(R^2)\int_{B_r}|\nabla u|^2 \leq CR^{2\alpha+d}$. Finally, returning to (3.10) and (3.11) we ascertain

$$\phi(r) \leq C\left(\frac{r}{R}\right)^{d+2}\phi(R) + CR^{2\alpha+d},$$

which gives the required regularity by another Campanato iteration. □

Corollary 3.14.1. *Let u solve (3.1) with A uniformly positive definite and bounded. If we further have $A, F \in C^{k,\alpha}$ then $u \in C^{k+1,\alpha}_{\text{loc}}(\Omega)$.*

Proof. We proceed by induction, the case $k = 0$ is the above theorem. Assuming the result for k, Sobolev regularity Theorem 3.5 gives $\nabla u \in H^k_{\text{loc}}(\Omega)$ and so we can differentiate the equation. Denoting $\frac{\partial}{\partial x_i} = (\cdot)'$ and differentiating, we have

$$\nabla \cdot (A\nabla u') = \nabla \cdot (F' - A'\nabla u) =: \nabla \cdot G,$$

with $G \in C^{k,\alpha}$ since $F, A \in C^{k+1,\alpha}$ and $\nabla u \in C^{k,\alpha}_{\text{loc}}$ by the induction assumptions. Thus $u' \in C^{k+1,\alpha}_{\text{loc}}$, giving $u \in C^{k+2,\alpha}$ as required. □

4 Regularity of optimal transport maps

In this chapter, we give an exposition of some of the recent work in optimal transport regularity due to M. Goldman and F. Otto [16, 15], where one of the key tools is Campanato's theorem. In most of this chapter, constants and controls are global, not depending on any data other than the dimension and sometimes a Hölder exponent. In this case, we write $\lesssim$ for the existence of a global constant. An assumption of the form $f(x) \ll g(x)$ with $f, g \geq 0$ indicates that $f(x)$ is much smaller than $g(x)$, and there is a small constant $0 < C < 1$ such that $f(x) \leq Cg(x)$ uniformly. As in the previous chapter, we will ignore differentials where the integration measure is clear to ease some notation.

4.1 Statement and sketch of the proof

The quadratic case $c(x, y) = |x-y|^2$ is the most explored for optimal transport regularity, due to its connections with the Monge-Ampère equation. We only discuss this case. For transport between measures $\mu = f\mathrm{d}x$ and $\nu = g\mathrm{d}x$, Brenier's theorem gives the quadratic optimiser of the form $T = \nabla u$ for some u convex, which combined with the Jacobian equation (1) gives the Monge-Ampère equation

$$\det D^2u(x) = \frac{f(x)}{g(\nabla u(x))}.$$

There has been a wealth of study on Monge-Ampère, with the regularity theory being studied extensively by Caffarelli in the 90s, as found in [10] as well as later contributions by De Philippis and Figalli [12, 22]. These regularity results are usually based on forms of maximum principles which Monge-Ampère equations satisfy, as they are elliptic equations in a certain non-linear sense, which we will not discuss.

Some simple examples show that in general, we cannot hope for a global regularity result, indeed if we are transporting between a connected and unconnected domain, then every transport map should be discontinuous somewhere. One can also construct examples of connected domains where the map must have a discontinuity, considering some limiting case which approaches disconnectedness - if the codomain consists of two large "islands" with a very thin bridge between them. Caffarelli [9] proved that convexity of the target domain was sufficient for numerous global regularity results.

Even in the case of a non-convex codomain, we might still hope for some partial regularity result, and this is what we discuss in this chapter. In particular, we show there exist open sets of full measure on which the optimal map is well behaved. In particular, this chapter is devoted to proving the following theorem:

Theorem 4.1. *Let E and F be two bounded open subsets of $\mathbb{R}^d$ with $|E| = |F|$ and let T be a minimiser of quadratic cost Monge problem*

$$\min \left\{ \int_E |T - x|^2 \quad : \quad T_\# \chi_E = \chi_F \right\}, \tag{4.1}$$

where by a slight abuse of notation $T_\# \chi_E$ denotes the push forward of the measure $\chi_E \mathrm{d}x$. Then there exist open sets $E' \subseteq E$ and $F' \subseteq F$ of full measure such that T is a $C^{1,\alpha}$ diffeomorphism between E' and F'.

This result had been originally attained by Figalli and Kim[13], but the proof we present here is due to Goldman and Otto, as presented in [16, 15]. This result is a result of a bootstrap in two stages. The first is a classical result due to Alexandrov [1].

Theorem 4.2 (Alexandrov Theorem)**.** *Let $\Omega \subset \mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ be convex. Then u is Lebesgue a.e. twice Fréchet differentiable, in the sense that for a.e. $x_0 \in \Omega$, there exists a vector $v_{x_0} \in \mathbb{R}^d$ and a matrix $H_{x_0} \in \mathbb{R}^{d\times d}$ such that for all x,*

$$u(x) = u(x_0) + v_{x_0} \cdot (x - x_0) + \frac{1}{2}(x - x_0)^T H_{x_0}(x - x_0) + o(|x - x_0|^2).$$

Proof of this result is omitted but can be found in [4]. The immediate consequence of this is clear for quadratic transport. Brenier's theorem gives an optimal T as the gradient of a convex function, so we have a.e. differentiability of T up to null sets. We focus on the second part of the bootstrap, showing $C^{1,\alpha}$ diffeomorphicity of T on an open set of full measure, as is the focus of [16]. We sketch the proof below to motivate what is to come.

Sketch of the proof of Theorem 4.1.

It is most instructive to deconstruct the proof in reverse. First, since both $\chi_E \mathrm{d}x$ and $\chi_F \mathrm{d}x$ are sufficiently regular measures, an inverse T^{-1} exists a.e by Remark 1.17, and this is the optimal map from F to E. The differentiability given by Alexandrov's theorem precisely means that locally the distance of both T and T^{-1} from being an affine function[1] on some ball vanishes as we shrink radius.

For almost all pairs $x_0 \in E$, $y_0 = T(x_0) \in F$ we will show that there exists an affine change of coordinates so that the map has a fixed point $T(0) = 0$ (corresponding to the former x_0 and y_0) and is still optimal. Thus it is sufficient to establish the regularity in some neighbourhood of 0 under the assumption that $B_R \subseteq E \cap F$[2]. The previous condition of being close to an affine function at x_0 now becomes one of being close to the identity on B_1. Explicitly, this will be quantified in terms of the energy

$$\mathcal{E}(T, R) := \frac{1}{R^2} \fint_{B_R} |T(x) - x|^2 + |T^{-1}(x) - x|^2 \mathrm{d}x. \tag{4.2}$$

[1]In some sense to be defined. Think à la Campanato here, the sense in which we mean is not as strong as Campanato a priori, but we will use this to establish Campanato.

[2]Here we use that the original sets were open to give some nbhd. of 0 for E, F after the change of variables.

Without the R^2, this is the average displacement squared on B_R of T and T^{-1}. With the R^2 this becomes unitless, a local measure of how close T is to the identity. Due to the affine invariance, we can assume $R = 1$ throughout most of the results. Thus we focus on establishing regularity locally in $B_1 \subseteq E \cap F$ by considering the Campanato seminorm. We aim to establish $C^{1,\alpha}$ regularity, which amounts to comparing T to affine maps. To do this we construct a harmonic function φ whose gradient's first order Taylor series is a competitor to the Lagrangian displacement $T - x$, superlinearly in terms of $\mathcal{E}(T, 1)$. Due to the various averaging factors in the energy (4.2), a superlinear control as R vanishes corresponds to some $\sigma > d+2$ Campanato seminorm control on the integral, which is precisely what we need for $C^{1,\alpha}$ regularity by Campanato's theorem.

The construction of φ is actually performed at the Eulerian level using the Benamou-Brenier formulation

$$\min \left\{ \int_0^1 \int_{\mathbb{R}^d} \frac{1}{\rho}|j|^2 \,:\, \partial_t \rho + \operatorname{div} j = 0,\ \rho(\,\cdot\,, 0) = \chi_E, \rho(\,\cdot\,, 1) = \chi_F \right\}. \tag{4.3}$$

We take the minimising pair (ρ, j) and define a harmonic function on B_1 with Neumann boundary condition $\nabla\varphi \cdot \vec{n} := \int_0^1 j \cdot \vec{n}$ the average flux, where $\vec{n}$ is the outward normal of ∂B_1. We then pass this to the Lagrangian setting, making use of some interior L^∞ controls on the transport in terms of the energy $\mathcal{E}$.

We will show that, given a superlinear control of the distance $T - x$ from $\nabla\varphi$ in terms of $\mathcal{E}(T, R)$, up to a suitable change of coordinates $\mathcal{E}(\hat{T}, \theta) \leq \theta^{2\alpha}\mathcal{E}(T, R)$ is much smaller, for any $\alpha \in (0, 1)$ and constant $\theta(\alpha, d) < 1$. This is quite similar to Lemma 3.7. Iterating this procedure will allow us to control the Campanato seminorm locally around our fixed point $T(0) = 0$, which gives the regularity.

Before beginning the detail, we first note (without proof) the following two lemmas on estimates for harmonic functions as well as some controls of functions with mixed Sobolev differentiability conditions. These correspond to [16, Lem. 3.1&3.2]:

Lemma 4.3. *For $f \in L^2(\partial B_1)$, let φ be the solution of*

$$\begin{cases} -\Delta\varphi = 0 & \text{in } B_1 \\ \nabla\varphi \cdot \vec{n} = f & \text{on } \partial B_1, \end{cases}$$

with $\vec{n}$ here denoting the outward normal of ∂B_1. Then

$$\int_{B_1} |\nabla\varphi|^2 \lesssim \int_{\partial B_1} f^2, \tag{4.4}$$

$$\sup_{B_{1/2}} (|\nabla^3\varphi|^2 + |\nabla^2\varphi|^2 + |\nabla\varphi|^2) \lesssim \int_{B_1} |\nabla\varphi|^2, \tag{4.5}$$

and for the annulus $A_r := B_1 \setminus B_{1-r}$ we have for every $r \leq 1$

$$\int_{A_r} |\nabla\varphi|^2 \lesssim r \int_{\partial B_1} f^2. \tag{4.6}$$

Lemma 4.4. *Let $A_r := B_1 \setminus B_{1-r}$ for $r \le 1$. For any $\psi \in L^1(A_r \times [0,1])$ which admit weak derivatives satisfying the mixed integrability*[3] $\partial_t \psi \in L^1$ *and* $\nabla\varphi \in L^2$ *we have*

$$\left(\int_0^1 \int_{\partial B_1} (\psi - \overline{\psi})^2\right)^{\frac{1}{2}} \lesssim r^{1/2}\left(\int_0^1 \int_{A_r} |\nabla\psi|^2\right)^{\frac{1}{2}} + \frac{1}{r^{(d+1)/2}} \int_0^1 \int_{A_r} |\partial_t \psi|, \tag{4.7}$$

where $\overline{\psi}(x) := \int_0^1 \psi(x,t)\mathrm{d}t$.

4.2 Superlinear approximation by a harmonic gradient

In this section, we prove the two harmonic approximation results, first in the Eulerian framework, then translating this into the Lagrangian setting.

4.2.1 Approximation of the Eulerian velocity

We need the following lemma [15, Lem. 2.4], which will control the excess flux on an annulus.

Lemma 4.5. *Let $f \in L^2(\partial B_1 \times (0,1))$ be s.t. for a.e. $x \in \partial B_1$, $\int_0^1 f(x,t)\mathrm{d}t = 0$. Set $A_r := B_1 \setminus B_{1-r}$ and define $\mathcal{Q}$ as all density-flux pairs $(s,q) : A_r \times [0,1] \to \mathbb{R} \times \mathbb{R}^d$ with $|s| \le 1/2$ and q L^2 in spacetime, which weakly solve the continuity equation*[4]

$$\begin{cases} \partial_t s + \operatorname{div} q = 0 & \text{in } A_r \\ q \cdot \vec{n} = f & \text{on } \partial B_1 \times [0,1] \\ q \cdot \vec{n} = 0 & \text{on } \partial B_{1-r} \times [0,1] \\ s(x,0) = 0 & s(x,1) = 0, \end{cases}$$

Then provided that $\left(\int_0^1 \int_{\partial B_1} f^2 \mathrm{d}S\mathrm{d}t\right)^{1/(d+1)} \ll r$, we have

$$\inf_{(s,q)\in\mathcal{Q}} \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 \lesssim r \int_0^1 \int_{\partial B_1} f^2.$$

Proof. We first show that the class $\mathcal{Q}$ is non-empty by giving an explicit construction. Define $u : A_r \times [0,1] \to \mathbb{R}$ as the mean free solution of the Poisson problem with Neumann boundary conditions

$$\begin{cases} -\Delta u = \frac{1}{|A_r|} \int_{\partial B_1} f(x,t)\mathrm{d}S(x) & \text{in } A_r \times (0,1) \\ \nabla u \cdot \vec{n} = f & \text{on } \partial B_1 \times (0,1) \\ \nabla u \cdot \vec{n} = 0 & \text{on } \partial B_{1-r} \times (0,1). \end{cases}$$

[3] the class of all such functions is a Sobolev type space with norm given by summing these the L^1 and the L^2 norms.

[4] In the sense that for $\psi \in C^1(\bar{A}_r \times [0,1])$, we have $\int_0^1 \int_{A_r} \partial_t \psi s + \nabla\psi \cdot q = \int_0^1 \int_{\partial B_1} f\psi$.

We set $q(x,t) := \nabla u(x,t)$ and $s(x,t) := -\int_0^t \operatorname{div} q(x,z)\mathrm{d}z$ so that s, q solve the continuity equation in $A_r \times (0,1)$ with the required boundary conditions. For $\left(\int_0^1 \int_{\partial B_1} f^2\right)^{1/(d+1)} \ll r$ we have $|s| \leq \frac{1}{2}$, we need to check q is square summable. As the minimiser of some suitable energy, $\nabla u_t \in L^2(B_1)$ for each fixed t. Multiplying the PDE by u_t and integrating by parts, we have for each fixed t [5].

$$\int_{B_1} |\nabla u_t|^2 = \int_{\partial B_1} u_t f_t + \int_{B_1} c_t u_t := (I) + (II),$$

where $c_t = \int_{\partial B_1} f_t$. We claim we can control both (I) and (II) by $\|\nabla u_t\|_{L^2(B_1)} \|f_t\|_{L^2(\partial B_1)}$, then Young's inequality would give

$$\int_{B_1} |\nabla u_t|^2 \lesssim \varepsilon \int_{B_1} |\nabla u_t|^2 + \frac{1}{\epsilon} \int_{\partial B_1} f_t^2,$$

So choosing ε suitably small to rearrange and integrating over t we would have finite spacetime L^2 norm. To control (I), with the trace inequality followed by Poincaré inequality (recall u_t is mean free). we deduce

$$(I) \leq \|u_t\|_{L^2(\partial B_1)} \|f_t\|_{L^2(\partial B_1)} \lesssim \|u_t\|_{H^1(B_1)} \|f_t\|_{L^2(\partial B_1)} \lesssim \|\nabla u_t\|_{L^2(B_1)} \|f_t\|_{L^2(\partial B_1)}.$$

For (II), again using mean free Poincaré, we have

$$(II) \lesssim |c_t| \|u_t\|_{L^2(B_1)} \lesssim \|f_t\|_{L^2(\partial B_1)} \|\nabla u_t\|_{L^2(B_1)},$$

so that $\mathcal{Q}$ is non-empty. Turning towards the control, we use duality à la [3] to write the continuity equation condition in its weak form as a sup, giving

$$\begin{aligned}
&\inf_{(s,q)\in\mathcal{Q}} \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 \\
&\qquad = \inf_{(s,q),|s|\leq\frac{1}{2}} \sup_{\psi} \left\{ \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 - \int_0^1 \int_{A_r} s\partial_t \psi + q\cdot\nabla\psi + \int_0^1 \int_{\partial B_1} f\psi \right\} \\
&\qquad = \sup_{\psi} \inf_{(s,q),|s|\leq\frac{1}{2}} \left\{ \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 - \int_0^1 \int_{A_r} s\partial_t \psi + q\cdot\nabla\psi + \int_0^1 \int_{\partial B_1} f\psi \right\}
\end{aligned}$$

where the sup is over all square ψ in the mixed Sobolev space from Lemma 4.4, and we interchange the order using Theorem B.3[6]. We now optimise the integrand pointwise in (s,q) à la [25, 3] using that $|s| \leq 1/2$. This amounts to optimising

$$q \mapsto \frac{1}{2}|q|^2 - q\cdot\nabla\psi \text{ for all } q; \quad \text{and} \quad s \mapsto -s\partial_t\psi \text{ for } |s| \leq \frac{1}{2}.$$

[5]Ignoring the factor of $|A_r|$ which is not relevant to the finiteness of the norm.

[6]Whose application is justified by $\psi = 0$ combined with the above construction to show the class is non-empty.

which gives

$$\inf_{(s,q)\in\mathcal{Q}} \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 = \sup_{\psi} \left\{ -\int_0^1 \int_{A_r} \frac{1}{2}(|\nabla\psi|^2 + |\partial_t\psi|) + \int_0^1 \int_{\partial B_1} f\psi \right\}.$$

We now use that $\int_0^1 f \mathrm{d}t = 0$ to introduce a $\overline{\psi} := \int_0^1 \psi \mathrm{d}t$ term, giving

$$\begin{aligned}
&= \sup_{\psi} \left\{ -\int_0^1 \int_{A_r} \frac{1}{2}(|\nabla\psi|^2 + |\partial_t\psi|) + \int_0^1 \int_{\partial B_1} f(\psi - \overline{\psi}) \right\} \\
&\leq \sup_{\psi} \left\{ -\int_0^1 \int_{A_r} \frac{1}{2}(|\nabla\psi|^2 + |\partial_t\psi|) + \left(\int_0^1 \int_{\partial B_1} f^2\right)^{1/2} \left(\int_0^1 \int_{\partial B_1} (\psi - \overline{\psi})^2\right)^{1/2} \right\}.
\end{aligned}$$

Setting $F := \left(\int_0^1 \int_{\partial B_1} f^2\right)^{1/2}$, using Lemma 4.4 where we denote the constant attained by $C > 0$, followed by Young's inequality gives

$$\begin{aligned}
&\leq \sup_{\psi} \left\{ F \left(\int_0^1 \int_{\partial B_1} (\psi - \overline{\psi})^2\right)^{1/2} - \frac{1}{2}\int_0^1 \int_{A_r} (|\nabla\psi|^2 + |\partial_t\psi|) \right\} \\
&\leq \sup_{\psi} \left\{ CFr^{1/2} \left(\int_0^1 \int_{\partial A_r} |\nabla\psi|^2\right)^{1/2} + C\frac{F}{r^{(d+1)/2}} \int_0^1 \int_{A_r} |\partial_t\psi| - \frac{1}{2}\int_0^1 \int_{A_r} (|\nabla\psi|^2 + |\partial_t\psi|) \right\} \\
&\leq \sup_{\psi} \left\{ \frac{1}{2}CF^2 r + \left(C\frac{F}{r^{(d+1)/2}} - \frac{1}{2}\right) \int_0^1 \int_{A_r} |\partial_t\psi| \right\}.
\end{aligned}$$

Finally using the assumption $F^{2/(d+1)} \ll r$ (specifically, we need $C\frac{F}{r^{(d+1)/2}} \leq \frac{1}{2}$) gives

$$\inf_{(s,q)\in\mathcal{Q}} \int_0^1 \int_{A_r} \frac{1}{2}|q|^2 \lesssim F^2 r = r \int_0^1 \int_{\partial B_1} f^2.$$

□

We can now prove the Eulerian approximation result, [16, Prop. 4.1].

Proposition 4.6. *Let (ρ, j) be a minimiser to the Benamou-Brenier formulation (4.3), and assume that $B_1 \subset E \cap F$. If the Eulerian energy $\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2 \ll 1$ then there exists $\varphi : B_{1/2} \to \mathbb{R}$ harmonic in satisfying*

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j - \rho\nabla\varphi|^2 \lesssim \left(\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2\right)^{\frac{d+2}{d+1}} \tag{4.8}$$

and

$$\int_{B_{1/2}} |\nabla\varphi|^2 \lesssim \int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2.$$

The proof is split into 5 parts as in [15, Prop. 3.3].
Step 1. We choose a radius $R \in (1/2, 1)$ (depending on the pair ρ, j) such that the flux through ∂B_R is controlled in L^2 (independently of ρ, j) by the overall Eulerian energy inside B_1.
Step 2. We define φ, explicitly as the harmonic function whose outward normal derivative $\nabla\varphi\cdot\vec{n}$ is precisely the average flux through ∂B_R, making it a good candidate to compare to the velocity $\frac{\mathrm{d}j}{\mathrm{d}\rho}$.
Steps 3 and 4. We show the superlinear energy control (4.8) by first showing a quasi-orthogonality property which separates the Eulerian energy from the Dirichlet energy term. We then construct a competitor $\tilde{\rho}, \tilde{j}$ who has the harmonic flux inside a smaller ball and only disagrees on an annulus, allowing us to only consider the energy difference on this annulus, which is controlled by Lemma 4.5.
Step 5. We show the control on the Dirichlet energy.

Proof. [Step 1: Choice of Radius.] We begin by noting that by McCann's displacement convexity, Proposition 1.23 we have $\rho \le 1$ and hence

$$\int_{1/2}^1\int_{\partial B_r}\int_0^1 |j(x,t)|^2\mathrm{d}t\mathrm{d}S\mathrm{d}r = \int_{B_1\setminus B_{1/2}}\int_0^1 |j(x,t)|^2 \le \int_0^1\int_{B_1}\frac{1}{\rho}|j|^2.$$

We have for any $f(x)$ that

$$\int_{B_1\setminus B_{1/2}} f(x) = \int_{1/2}^1\int_{\partial B_R} f(x) \ge \frac{1}{2}\inf_{\substack{R\in(1/2,1)\\ R \text{ a Leb. point of } r\to\int_{\partial B_r} f(x)}}\left(\int_{\partial B_R} f(x)\right),$$

so that we can find $R \in (1/2, 1)$ (depending of ρ, j but the control is uniform) satisfying

$$\int_{\partial B_R}\int_0^1 |j|^2 \lesssim \int_0^1\int_{B_1}\frac{1}{\rho}|j|^2, \tag{4.9}$$

and which is a Lebesgue point of $r \mapsto j \in L^2(\partial B_r)$ in the sense that

$$\lim_{\varepsilon\to 0}\int_{R-\varepsilon}^{R+\varepsilon}\int_{\partial B_1}\int_0^1 |j(rx,t) - j(Rx,t)|^2\mathrm{d}t\mathrm{d}S(x)\mathrm{d}r = 0.$$

We claim for this R, the continuity equation inside B_R can be well understood with boundary conditions, such that for all $\phi \in H^1(B_R \times (0,1))$,

$$\int_0^1\int_{B_R}\rho\partial_t\phi + j\cdot\nabla\phi = \int_0^1\int_{\partial B_R}\phi f + \int_{B_R}\phi(\cdot,1) - \phi(\cdot,0), \tag{4.10}$$

where $f = j\cdot\vec{n}$ denotes the flux out through ∂B_R. To prove this, we introduce the cut-off function

$$\eta_\varepsilon(x) := \begin{cases} 1 & \text{if } |x| \le R-\varepsilon \\ \frac{R-|x|}{\varepsilon} & \text{if } R-\varepsilon \le |x| \le R \\ 0 & \text{otherwise,}\end{cases}$$

and by the weak formulation we have for any $\phi \in H^1(B_R \times (0,1))$

$$\int_{\mathbb{R}^d} (\phi(\cdot,1) - \phi(\cdot,0))\eta_\varepsilon = \int_0^1 \int_{\mathbb{R}^d} \partial_t[\phi\eta_\varepsilon]\rho + \nabla[\phi\eta_\varepsilon] \cdot j$$
$$= \int_0^1 \int_{\mathbb{R}^d} \eta_\varepsilon(\partial_t\phi)\rho + \eta_\varepsilon \nabla\phi \cdot j - \frac{1}{\varepsilon}\int_0^1 \int_{B_R \setminus B_{R-\varepsilon}} \phi j \cdot \vec{n}.$$

Now taking $\varepsilon \to 0$ and using the Lebesgue point property we obtain (4.10).

[Step 2: Definition of φ.] We define $\varphi : B_{1/2} \subset B_R \to \mathbb{R}$ as a solution of

$$\begin{cases} -\Delta\varphi = 0 & \text{in } B_R \\ \nabla\varphi \cdot \vec{n} = \overline{f} & \text{on } \partial B_R, \end{cases} \tag{4.11}$$

where $\overline{f}(x) = \int_0^1 f(x,t)$. Note that all solutions are the same up to an additive constant, and we are only interested in the gradient $\nabla\varphi$ which is then unique. By construction, this is a good time-independent candidate to compare to the velocity $\frac{\mathrm{d}j}{\mathrm{d}\rho}$.

[Step 3: The quasi-orthogonality.] By suitably rescaling all of the data, we can assume $R = 1/2$, and we do so to ease notation. In two steps, we now show the control (4.8). First, we establish what Goldman and Otto refer to as a quasi-orthogonality property

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j - \rho\nabla\varphi|^2 \le \int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 - \int_{B_{1/2}} |\nabla\varphi|^2. \tag{4.12}$$

Here the left-hand side is still interpreted in the sense of (1.6), so that if $\rho = 0$ then $j = 0$ a.e. in $t \in [0,1]$ for the quantity to be finite and so also $j - \rho\nabla\varphi = 0$, and everything is well defined. We compute

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j - \rho\nabla\varphi|^2$$
$$= \int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 - 2\int_0^1 \int_{B_{1/2}} j \cdot \nabla\varphi + \int_0^1 \int_{B_{1/2}} \rho|\nabla\varphi|^2$$
$$= \int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 - 2\int_0^1 \int_{B_{1/2}} (1 - \frac{\rho}{2})|\nabla\varphi|^2 - 2\int_0^1 \int_{B_{1/2}} (j - \nabla\varphi) \cdot \nabla\varphi$$
$$\le \int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 - \int_0^1 \int_{B_{1/2}} |\nabla\varphi|^2 - 2\int_0^1 \int_{B_{1/2}} (j - \nabla\varphi) \cdot \nabla\varphi,$$

where we used that $\rho \le 1$ by the displacement convexity Prop 1.23. Now using (4.10) (recalling we took $R = 1/2$ for convenience) testing against φ and noting most terms vanish as φ has no time component, we get

$$\int_{B_{1/2}} j \cdot \nabla\varphi = \int_{\partial B_{1/2}} \Big(\varphi \int_0^1 f\Big) = \int_{\partial B_{1/2}} \varphi\nabla\varphi \cdot \vec{n} = \int_{B_{1/2}} |\nabla\varphi|^2$$

where the second equality is due to (4.11) tested against φ. Consequently,

$$\int_0^1 \int_{B_{1/2}} (j - \nabla\varphi) \cdot \nabla\varphi = 0,$$

and (4.12) is proven.

[Step 4: The main estimate.] We now establish

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 - \int_{B_{1/2}} |\nabla\varphi|^2 \lesssim \left(\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2\right)^{\frac{d+2}{d+1}}, \tag{4.13}$$

which combined with (4.12) would give us (4.8). In order to show this, we construct a competitor $\tilde{\rho}, \tilde{j}$ which agrees with ρ, j outside of $B_{1/2} \times (0,1)$ and satisfies the upper bound in (4.13). Then the minimality of ρ, j for the Benamou-Brenier problem gives the result. Let $r > 0$ and we set $A_r := B_{1/2} \setminus B_{1/2(1-r)}$. We choose

$$(\tilde{\rho}, \tilde{j}) := \begin{cases} (1, \nabla\varphi) & \text{in } B_{1/2(1-r)} \times (0,1) \\ (1+s, \nabla\varphi + q) & \text{in } A_r \times (0,1), \end{cases}$$

Where $(s,q) \in \mathcal{Q}$ is the minimiser from Lemma 4.5 with f replaced by $f - \overline{f}$ and radius $1/2$ rather than 1. By construction $(\tilde{\rho}, \tilde{j})$ is admissible for the Benamou-Brenier formulation, as the flux out through $\partial B_{1/2}$ is precisely the same as that of j for a.e. x and t, so that the net flux through $B_{1/2}$ is zero and this retains its constant density $\equiv 1$ at $t = 0,1$ (since $B_1 \subseteq E \cap F$), then setting $(\tilde{\rho}, \tilde{j})$ as equal to (ρ, j) outside of $B_{1/2}$, everything else is also preserved.

Now by Lemma 4.5 if $\left(\int_0^1 \int_{\partial B_{1/2}} (f - \overline{f})^2\right)^{1/d+1} \ll r$ (which is possible for some $r \in (0,1)$ as the former is controlled by (4.9) and that $\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2 \ll 1$) then we can choose $(s,q) \in \mathcal{Q}$ such that

$$\int_0^1 \int_{A_r} |q|^2 \lesssim r \int_0^1 \int_{\partial B_{1/2}} (f - \overline{f})^2 \tag{4.14}$$

Now using that $|s| \leq \frac{1}{2}$ we have

$$\begin{aligned} \int_0^1 \int_{B_{1/2}} \frac{1}{\tilde{\rho}}|\tilde{j}|^2 &= \int_0^1 \int_{B_{1/2(1-r)}} |\nabla\varphi|^2 + \int_0^1 \int_{A_r} \frac{1}{1+s}|\nabla\varphi + q|^2 \\ &\leq \int_0^1 \int_{B_{1/2(1-r)}} |\nabla\varphi|^2 + 4\int_0^1 \int_{A_r} |\nabla\varphi|^2 + |q|^2. \end{aligned}$$

Then using this expression in the LHS of (4.13) gives

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\tilde{\rho}}|\tilde{j}|^2 - \int_{B_{1/2}} |\nabla\varphi|^2 \lesssim \int_0^1 \int_{A_r} |\nabla\varphi|^2 + |q|^2.$$

We now use the trace estimate (4.6) from Lemma 4.3 (here $\overline{f}$ plays the role of f) as well as (4.14) to give

$$\begin{aligned}\int_0^1 \int_{B_{1/2}} \frac{1}{\tilde{\rho}}|\tilde{j}|^2 - \int_{B_{1/2}} |\nabla\varphi|^2 &\lesssim r \int_{\partial B_{1/2}} \overline{f}^2 + r \int_0^1 \int_{\partial B_{1/2}} (f-\overline{f})^2 \\ &= r \int_0^1 \int_{\partial B_{1/2}} f^2.\end{aligned}$$

Note we cannot now pass any limit $r \to 0$ giving a contradiction, as we need to maintain $\big(\int_0^1 \int_{\partial B_{1/2}} (f-\overline{f})^2\big)^{1/d+1} \ll r$. But we can choose r to be a large but linear multiple of $\big(\int_0^1 \int_{\partial B_{1/2}} (f-\overline{f})^2\big)^{1/d+1} \leq \big(\int_0^1 \int_{\partial B_{1/2}} f^2\big)^{1/d+1}$ corresponding to the requirement for the lemma[7]. This gives

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\tilde{\rho}}|\tilde{j}|^2 - \int_{B_{1/2}} |\nabla\varphi|^2 \lesssim \left(\int_0^1 \int_{\partial B_{1/2}} f^2\right)^{d+2/d+1} \lesssim \left(\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2\right)^{\frac{d+2}{d+1}}$$

by (4.9), and combining with that ρ, j are a minimiser of the Benamou-Brenier formulation and agree with $\tilde{\rho}, \tilde{j}$ away from the ball, we have

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 \leq \int_0^1 \int_{B_{1/2}} \frac{1}{\tilde{\rho}}|\tilde{j}|^2,$$

which in turn gives (4.13).

[Step 5: Control on the Dirichlet energy.]

Finally, we show the control on the energy of φ. We have

$$\int_{B_{1/2}} |\nabla\varphi|^2 \leq \int_{B_R} |\nabla\varphi|^2 \lesssim \int_{\partial B_R} |\overline{f}|^2 \lesssim \int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2,$$

where we used the trace estimate in terms of the energy from Lemma 4.3 followed by (4.9). This completes the proof. □

4.2.2 Approximation of the Lagrangian displacement

We now pass to the Lagrangian framework. We will first need an L^∞ interior control on the transport displacement in terms of the energy (4.2). This is somewhat reminiscent of interior L^∞ estimates for harmonic functions as in Lemma 4.3. There the main tool is the mean value theorem, whereas here the control is a consequence of the monotonicity of any quadratic optimal map. Using the monotonicity for L^∞ controls is established global case in [5], here we prove a localised version as in [15, Lem. 3.1].

[7] Here we use $\int_0^1 \int_{B_1} \frac{1}{\rho}|j|^2$ suitably small to justify such an r existing.

Lemma 4.7. *Let T be the minimiser of (4.1) and assume $\mathcal{E}(T,1) \ll 1$, then*

$$\sup_{B_{3/4}} |T-x| \lesssim \left(\int_{B_1} |T-x|^2 \right)^{\frac{1}{d+2}} \quad \text{and} \quad \sup_{B_{3/4}} |T^{-1}-x| \lesssim \left(\int_{B_1} |T^{-1}-x|^2 \right)^{\frac{1}{d+2}}, \quad (4.15)$$

with the sup *in the L^∞ sense. Consequently for $T_t(x) = tT(x) + (1-t)x$, $t \in [0,1]$,*

$$T^{-1}(B_{1/16}) \subseteq B_{1/8}; \quad T_t(B_{1/8}) \subseteq B_{3/16} \text{ and } T_t^{-1}(B_{1/2}) \subseteq B_1. \quad (4.16)$$

Proof. Let $u(x) := T(x) - x$ denote the displacement. By monotonicity of T as the quadratic optimiser, we have for a.e. $x, y \in B_1$ that

$$(u(x) - u(y)) \cdot (x-y) = (T(x) - T(y)) \cdot (x-y) - (x-y) \cdot (x-y) \geq -|x-y|^2. \quad (4.17)$$

Let $y \in B_{3/4}$ be s.t. the above holds for a.e. $x \in B_1$. If we can show

$$u(y) \lesssim \left(\int_{B_{1/4}(y)} |u(x)|^2 \right)^{\frac{1}{d+2}},$$

we will have established the result. Since we show a bound independent of E and F, then by translating these it is sufficient to establish this just for $y = 0$. Furthermore, by rotating E, F, we only need to show the bound for a single component of u. Thus it is sufficient to prove

$$u_1(0) \lesssim \left(\int_{B_{1/4}} |u|^2 \right)^{\frac{1}{d+2}}.$$

Take $y = 0$ in (4.17), then for a.e. $x \in B_{1/4}$, we have

$$u(0) \cdot x \leq u(x) \cdot x + |x|^2 \lesssim |u(x)|^2 + |x|^2$$

For $r \leq 1/8$ we can integrate the above over $B_r(e_1 r)$ to obtain

$$2r^2 r^{d-1} u(0) \cdot e_1 \leq \left(\int_{B_r(re_1)} |u|^2 \right) + Cr^{d+2}$$

which in turn gives (still for $r \leq 1/8$)

$$u_1(0) \lesssim \frac{1}{r^{d+1}} \int_{B_{1/4}} |u|^2 + r =: \frac{a}{r^{d+1}} + r.$$

First-order optimisation in r on the RHS gives a minimiser

$$r' = a^{\frac{1}{d+2}} (d+1)^{\frac{1}{d+2}},$$

Which is admissable under our assumption $\mathcal{E}(T,1) \ll 1$ since $a \leq \mathcal{E}(T,1)$. Plugging in this r gives

$$u_1(0) \lesssim \left(\int_{B_{1/4}} |T-x|^2 \right)^{\frac{1}{d+2}},$$

giving half the estimate. For T^{-1} we can simply interchange the roles of E and F so that we have (4.15).

For the ball containments, the first two follow immediately from (4.15), as if the L^∞ displacement is very small, then no point can go far beyond the ball so the image is contained in any ball slightly bigger than it if $\mathcal{E}(T,1)$ is suitably small.

For the third containment $T_t^{-1}(B_{1/2}) \subseteq B_1$, let $x \in E$ such that $T_t(x) \in B_{1/2}$. Then since $|T_t(0)| \to 0$ as $\mathcal{E}(T,1) \to 0$ by the control above, we have

$$\begin{aligned}|T_t(0) - T_t(x)|^2 \leq& |T_t(0)|^2 + 2|T_t(0)||T_t(x)| + |T_t(x)|^2 \\ \leq& o(1) + 1/4\end{aligned}$$

where $o(1) \to 0$ as $\mathcal{E}(T,1) \to 0$. On the other hand,

$$\begin{aligned}|T_t(0) - T_t(x)|^2 =& t^2|T(x) - T(0)|^2 + 2t(1-t)(T(x) - T(0)) \cdot x + (1-t)^2|x|^2 \\ \geq& t^2|T(x) - T(0)|^2 + (1-t)^2|x|^2 \\ \geq& \frac{1}{2}\min(|T(x) - T(0)|^2, |x|^2).\end{aligned}$$

Combining the estimates, for $\mathcal{E}(T,1) \ll 1$, we have either x or $T(x)$ in $B_{\frac{1}{\sqrt{2}}+o(1)} \subseteq B_{3/4}$. In the first case we are done, in the second by the L^∞ control we have $x = T^{-1}(T(x)) \in T^{-1}(B_{\frac{1}{\sqrt{2}}+o(1)} \subseteq B_{3/4}) \subseteq B_1$ which gives the final containment of 4.16. ☐

Remark 4.8. The key part of the above is the ball containments, they will allow us to control the Eulerian energy on a smaller ball by the Lagrangian energy on a larger one. In general, the above controls give estimates of type $T(B_r) \subseteq T(B_{r+\delta})$ for any $\delta > 0$, if we $\mathcal{E}(T,1) \ll 1$.

These displacement controls now allow us to pass from the Eulerian framework to a competitor to the displacement at the Lagrangian level, so that we can approximate the displacement with a harmonic gradient superlinearly in terms of $\mathcal{E}(T,1)$. This corresponds to [16, Prop. 4.4].

Proposition 4.9. *Let T be the minimiser of (4.1) and assume $B_1 \subset E \cap F$. Then there exists harmonic $\varphi : B_{1/16} \to \mathbb{R}$ such that*

$$\int_{B_{1/16}} |T - x - \nabla\varphi|^2 + \int_{B_{1/16}} |T^{-1} - x + \nabla\varphi|^2 \lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}} \tag{4.18}$$

and

$$\int_{B_{1/16}} |\nabla\varphi|^2 \lesssim \mathcal{E}(T,1).$$

Proof. Let $T_t(x) = tT(x) + (1-t)x$, and define the measures

$$\rho(\,\cdot\, t) := T_{t\#\chi_E} \quad \text{and} \quad j(\,\cdot\,,t) := T_{t\#}((T - \mathrm{Id})\chi_E),$$

which solve the Benamou-Brenier formulation (4.3). The velocity field $v = \frac{\mathrm{d}j}{\mathrm{d}\rho}$ satisfies $v(T_t(x), t) = T(x) - x$ for a.e. $x \in E$, so we have

$$\int_0^1 \int_{B_{1/2}} \frac{1}{\rho}|j|^2 = \int_0^1 \int_{B_{1/2}} |v|^2 \mathrm{d}\rho = \int_0^1 \int_{T_t^{-1}(B_{1/2})} |T - x|^2 \leq \int_{B_1} |T - x|^2.$$

By Proposition 4.6 on $B_{1/2}$, there exists $\varphi : B_{1/4} \to \mathbb{R}$ harmonic such that

$$\int_0^1 \int_{B_{1/4}} \frac{1}{\rho}|j - \rho\nabla\varphi|^2 \lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}} \quad \text{and} \quad \int_{B_{1/4}} |\nabla\varphi|^2 \lesssim \mathcal{E}(T,1). \tag{4.19}$$

To establish (4.18) we first show

$$\int_{B_{1/8}} |(T - x) - \nabla\varphi|^2 \lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}}. \tag{4.20}$$

By the triangle inequality, the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, followed by integrating both sides over $t \in [0,1]$,

$$\int_{B_{1/8}} |(T - x) - \nabla\varphi|^2 \lesssim \int_0^1 \int_{B_{1/8}} |(T - x) - \nabla\varphi \circ T_t|^2 + \int_0^1 \int_{B_{1/8}} |\nabla\varphi - \nabla\varphi \circ T_t|^2$$
$$:= (I) + (II).$$

We claim (I) $\lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}}$ and (II) $\lesssim \mathcal{E}(T,1)^2$ which would give (4.20) since $\mathcal{E}(T,1)^2 \leq \mathcal{E}(T,1)^{\frac{d+2}{d+1}}$ as soon as $\mathcal{E}(T,1) \leq 1$.

For (I), recalling that $v = \frac{\mathrm{d}j}{\mathrm{d}\rho}$ satisfies $v(T_t(x), t) = T(x) - x$ and again using the uniform in t ball containments, we can rewrite the integrand as

$$\begin{aligned}\int_0^1 \int_{B_{1/8}} |v(T_t, t) - \nabla\varphi \circ T_t|^2 &= \int_0^1 \int_{T_t(B_{1/8})} |v - \nabla\varphi|^2 \mathrm{d}\rho \\ &\leq \int_0^1 \int_{B_{1/4}} \frac{1}{\rho}|j - \rho\nabla\varphi|^2 \lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}}.\end{aligned}$$

For (II), using $\nabla\varphi$ is Lipschitz with constant $\sup |D^2\varphi|$ combined with $T_t(B_{1/8}) \subset B_{3/16}$, we have

$$\begin{aligned}(II) &\leq \sup_{B_{3/16}} |D^2\varphi| \int_0^1 \int_{B_{1/8}} |T_t - x|^2 \\ &\leq \int_{B_{1/4}} |D^2\varphi| \int_{B_{1/8}} |T - x|^2 \lesssim \mathcal{E}(T,1)^2\end{aligned}$$

where we used that $|T_t - x| \leq |T - x|$, a general form of Lemma 4.3 that the L^∞ norm of harmonic functions on any interior ball is controlled by the Dirichlet energy and that the energy itself is controlled by $\mathcal{E}(T,1)$ from (4.19).

Making use of the above, we now show the control for T^{-1} on $B_{1/16}$ which would complete the proof of (4.18). By The Triangle inequality and the L^∞ ball controls,

$$\begin{aligned}\int_{B_{1/16}} |(T^{-1}-x)+\nabla\varphi|^2 &= \int_{T^{-1}(B_{1/16})} |(x-T)+\nabla\varphi\circ T|^2 \\ &\leq \int_{B_{1/8}} |(T-x)-\nabla\varphi|^2 + \int_{B_{1/8}} |\nabla\varphi - \nabla\varphi\circ T|^2 \lesssim \mathcal{E}(T,1)^{\frac{d+2}{d+1}},\end{aligned}$$

where the first term is precisely (4.20) and the second term is controlled as (II) above. □

4.3 Partial regularity for the optimal map

4.3.1 Local $C^{1,\alpha}$ behaviour: geometric control of $\mathcal{E}(T,R)$ in terms R

Here we establish the local regularity, as a consequence of what is referred to as an epsilon regularity result, originally appearing to prove regularity for minimal surfaces. Effectively what the result says is that if on some scale the map T is close to the identity in the sense of $\mathcal{E}(T,R)$, then there exists a change of coordinates on a locally small scale such that T is even closer to the identity. This corresponds to [16, Prop.4.5].

Proposition 4.10. *Let T be the minimiser of (4.1). For every $\alpha \in (0,1)$, there exists $0 < \theta(d,\alpha) < 1$ and $\varepsilon(d,\alpha) > 0$ such that for each $R > 0$ with $B_R \subset E\cap F$, if $\mathcal{E}(T,R) \leq \varepsilon$, there exists a symmetric matrix Q with $\det Q = 1$, and a vector $b \in \mathbb{R}^d$ satisfying*

$$|Q - \mathrm{Id}\,|^2 \lesssim \mathcal{E}(T,R) \quad \textit{and} \quad |y|^2 \lesssim R^2\mathcal{E}(T,R), \tag{4.21}$$

such that map $\hat{T}(x) := Q(T(Qx)-b)$ has a geometrically smaller energy on a scale θ smaller;

$$\mathcal{E}(\hat{T},\theta R) \leq \theta^{2\alpha}\mathcal{E}(T,R).$$

Note that Q and b depend on the specific map, but their controls in terms of the energy are global.

Proof. If we can show the result for $R = 1$, for a general R we can re-scale E, F and T to the $R = 1$ case, which would correspond to a change of variables $\tilde{x} = R^{-1}x$, a redefinition $\tilde{T}(\tilde{x}) = R^{-1}T(R\tilde{x})$ and $\tilde{b} = R^{-1}b$. Since this scaling preserves optimality (by Brenier's theorem and that the new map is simply the gradient of a suitably scaled convex potential from the original), it is sufficient for us to prove only the case $R = 1$.

Let $\varphi : B_{1/16} \to \mathbb{R}$ be the harmonic function given by the Lagrangian competitor Proposition 4.9. We set $b = \nabla\varphi(0)$ and $A = D^2\varphi(0)$ the Hessian. Then we choose $Q := e^{-A/2}$ as the exponential of the matrix, see appendix. Since A is symmetric and $\mathrm{Tr}\,A = 0$ as φ is harmonic, then Q is symmetric and $\det Q = 1$. By the harmonic function controls Lemma 4.3, we have

$$|b|^2 \lesssim \int_{B_{1/16}} |\nabla\varphi|^2 \lesssim \mathcal{E}(T,1),$$

and also

$$|Q - \mathrm{Id}\,|^2 = |\sum_{i=0}^{\infty} \frac{(-1)^i}{2^i i!} A^i - \mathrm{Id}\,|^2 \leq \sum_{i=1}^{\infty} \frac{1}{i!} |A|^i \lesssim |A| \lesssim \mathcal{E}(T,1),$$

so that our matrix and vector satisfy (4.21).

Now turning towards our map $\hat{T}(x) := Q(T(Qx) - b)$, for any $\theta > 0$, if $\mathcal{E}(T,1) \ll \theta^2 \ll 1$ then $|Q - \mathrm{Id}\,| \ll \theta$ which means $Q(B_\theta) \subset B_{2\theta}$ and also $|Qy| = |Qy - y + y| \leq |Q - \mathrm{Id}\,||y| + |y|$ so that $|Qy|^2 \lesssim |y|^2$. Consequently, we have

$$\begin{aligned} \theta^{-2} \fint_{B_\theta} |\hat{T} - x|^2 &= \theta^{-2} \int_{Q(B_\theta)} |Q(T - b) - Q^{-1}x|^2 \\ &= \theta^{-2} \int_{Q(B_\theta)} |Q((T - b) - Q^{-2}x)|^2 \lesssim \theta^{-2} \fint_{B_{2\theta}} |T - (Q^{-2}x + b)|^2 \end{aligned}$$

Now applying the triangle inequality to the above,

$$\begin{aligned} &\theta^{-2} \fint_{B_\theta} |\hat{T} - x|^2 \\ &\lesssim \theta^{-2} \fint_{B_{2\theta}} |(T - x) - \nabla\varphi|^2 + \theta^{-2} \fint_{B_{2\theta}} |(Q^{-2} - \mathrm{Id} - A)x|^2 + \theta^{-2} \fint_{B_{2\theta}} |\nabla\varphi - b - Ax|^2 \\ &:= (I) + (II) + (III). \end{aligned}$$

We have (I) $\lesssim \theta^{-(d+2)} \mathcal{E}(T,1)^{\frac{d+2}{d+1}}$ by Proposition 4.9. For (II) we have

$$\begin{aligned} \theta^{-2} \fint_{B_{2\theta}} |(Q^{-2} - \mathrm{Id} - A)x|^2 &\lesssim \theta^{-(d+2)} \int_{B_{2\theta}} |Q^{-2} - \mathrm{Id} - A|^2 |2\theta|^2 \\ &\lesssim |Q^{-2} - \mathrm{Id} - A|^2 = |e^A - A - \mathrm{Id}\,|^2 \lesssim |D^2\varphi(0)|^4 \lesssim \mathcal{E}(T,1)^2. \end{aligned}$$

Finally for (III) we have

$$\begin{aligned} \theta^{-2} \fint_{B_{2\theta}} |\nabla\varphi - b - Ax|^2 \leq& \theta^{-2} \sup_{B_{2\theta}} |\nabla\varphi - b - Ax|^2 \\ =& \theta^{-2} \sup_{B_{2\theta}} |\nabla\varphi(x) - (\nabla\varphi(0) - D(\nabla\varphi)(0)x|^2. \end{aligned}$$

This is the error or the first order Taylor series of the vector field $f = \nabla\varphi$, so by the Lagrange form of the remainder this is equal to some second derivative of f inside $B_{2\theta}$. This is a third-order derivative term $D^3\varphi$, with a second-order power θ^2 since we view it at the Taylor series in f. Hence by Lemma 4.3,

$$\fint_{B_{2\theta}} |\nabla\varphi - b - Ax|^2 \lesssim \theta^{-2} \sup_{B_{2\theta}} |D^3\varphi||\theta^2|^2 = \theta^2 \sup_{B_{2\theta}} |D^3\varphi| \lesssim \theta^2 \mathcal{E}(T,1).$$

Combining the three estimates gives, for $\mathcal{E}(T,1) \ll 1$ and $\theta < 1/2$,

$$\begin{aligned} \theta^{-2} \fint_{B_\theta} |\hat{T} - x|^2 &\lesssim \theta^{-(d+2)} \mathcal{E}(T,1)^{\frac{d+2}{d+1}} + \mathcal{E}(T,1)^2 + \theta^2 \mathcal{E}(T,1) \\ &\lesssim \theta^{-(d+2)} \mathcal{E}(T,1)^{\frac{d+2}{d+1}} + \theta^2 \mathcal{E}(T,1). \end{aligned}$$

We now show a similar control for $\hat{T}^{-1}(x) = Q^{-1}T^{-1}(Q^{-1}x + b)$. For $\mathcal{E}(T,1) \ll \theta^2 \ll 1$,

$$\theta^{-2}\fint_{B_\theta} |\hat{T}^{-1} - x|^2 = \theta^{-2}\fint_{Q^{-1}B_\theta + b} |Q^{-1}T^{-1} - Q(x-b)|^2 \lesssim \theta^{-2}\fint_{2\theta} |T^{-1} - Q^2(x-b)|^2.$$

By repeated application of the triangle inequality, we have

$$\begin{aligned}|T^{-1} - Q^2(x-b)| \leq& |T^{-1} - (x - \nabla\varphi)| + |x - \nabla\varphi - Q^2(x-b)| \\ \leq& |T^{-1} - (x - \nabla\varphi)| + |x - Q^2x - Ax| + |\nabla\varphi - Q^2 b - Ax| \\ \leq& |T^{-1} - (x - \nabla\varphi)| + |x - Q^2x - Ax| + |\nabla\varphi - Ax - b| + |b - Q^2 b|,\end{aligned}$$

so that

$$\begin{aligned}\theta^{-2}\fint_{B_\theta} |\hat{T}^{-1} - x|^2 \lesssim& \left(\theta^{-2}\int_{B_{2\theta}} \big|(T^{-1} - x) + \nabla\varphi)\big|^2\right) + \big|\operatorname{Id} - Q^2 - A\big| \\ &+ \theta^{-2}\sup_{B_{2\theta}} \big|\nabla\varphi - Q^2 b - Ax\big|^2 + \theta^{-2}\big|(\operatorname{Id} - Q^2)b\big|^2 \\ &:= (I) + (II) + (III) + (IV).\end{aligned}$$

As before (I) $\lesssim \theta^{-(d+2)}\mathcal{E}(T,1)^{\frac{d+2}{d+1}}$ and (III) $\lesssim \theta^2\mathcal{E}(T,1)$. For (II) we have

$$\big|\operatorname{Id} - Q^2 - A\big| = \big|e^{-A} + A - \operatorname{Id}\big| \lesssim \big|A^2\big|^2 \leq \big|D^2\varphi(0)\big|^4 \lesssim \mathcal{E}(T,1)^2,$$

and for (IV),

$$\theta^{-2}\big|(\operatorname{Id} - Q^2)b\big|^2 \leq \theta^{-2}|D^2\varphi(0)|^2|\nabla\varphi(0)|^2 \lesssim \theta^{-2}\mathcal{E}(T,1)^2.$$

Overall, this gives

$$\theta^{-2}\fint_{B_\theta} |\hat{T}^{-1} - x|^2 \lesssim \theta^{-d+2}\mathcal{E}(T,1)^{\frac{d+2}{d+1}} + \theta^2\mathcal{E}(T,1).$$

Combining with the estimate for T, there exists a constant $C(d) > 0$ for which

$$\mathcal{E}(\hat{T},\theta) = \theta^{-2}\fint_{B_\theta} |\hat{T} - x|^2 + |\hat{T}^{-1} - x|^2 \leq C\left(\theta^{-d+2}\mathcal{E}(T,1)^{\frac{d+2}{d+1}} + \theta^2\mathcal{E}(T,1)\right)$$

for $\mathcal{E}(T,1) \ll \theta^2 \ll 1$. Note that our calculations show that C is independent of any of the data, it is just that the controls only hold for when the data is suitably small. Finally, we now fix $\theta < 1/2$ such that $C\theta^2 \leq \frac{1}{2}\theta^{2\alpha}$, which is possible since $\alpha < 1$. Then if $\mathcal{E}(T,1) \ll \theta^2$ suitably, we have $C\theta^{-(d+2)}\mathcal{E}(T,1)^{\frac{d+2}{d+1}} \leq \frac{1}{2}\theta^{2\alpha}\mathcal{E}(T,1)$ and in this case

$$\mathcal{E}(\hat{T},\theta) \leq \frac{1}{2}\theta^{2\alpha}\mathcal{E}(T,1) + \frac{1}{2}\theta^{2\alpha}\mathcal{E}(T,1) = \theta^{2\alpha}\mathcal{E}(T,1).$$

In other words, what we have shown is that there exists $0 < \theta(d,\alpha) < 1$ and a threshold $\varepsilon(d,\alpha,\theta) = \varepsilon(d,\alpha) > 0$ (corresponding to the requirement $\mathcal{E}(T,1) \ll \theta^2$) such that if

$$\mathcal{E}(T,1) \leq \varepsilon, \quad \text{then} \quad \mathcal{E}(\hat{T},\theta) \leq \theta^{2\alpha}\mathcal{E}(T,1),$$

which is precisely what we were trying to show. □

We now prove the local regularity at a fixed point, [16, Prop. 4.6].

Proposition 4.11. *Let T be the minimiser of (4.1), and assume that for some $R > 0$, we have $B_{2R} \subset E \cap F$. Then if $\mathcal{E}(T, 2R) \ll 1$ we have $T \in C^{1,\alpha}(B_R)$ for any $\alpha \in (0,1)$, with the Hölder seminorm on B_R controlled by*

$$[T]_{1,\alpha} \lesssim R^{-\alpha}\mathcal{E}(T, 2R)^{\frac{1}{2}}.$$

Proof. We will show that $T \in \mathcal{L}_1^{2,2+2\alpha}(B_R)$, thus establishing the result by Campanato's theorem, Theorem 2.10. This means comparing T to degree 1 polynomials; affine functions. If $\mathcal{E}(T, 2R) \ll 1$, then for any $x_0 \in B_R$,

$$\mathcal{E} := R^{-2} \fint_{B_R(x_0)} |T - x|^2 + |T^{-1} - x|^2 \leq 2^{d+2}\mathcal{E}(T, 2R) \ll 1.$$

We will show that for $r \leq \frac{R}{2}$ (we only need to check this, c.f. (2.7))

$$\min_{A,b} \frac{1}{r^2} \fint_{B_r(x_0)} |T - (Ax + b)|^2 \lesssim r^{2\alpha}\mathcal{E}, \tag{4.22}$$

which would establish the result by Campanato's theorem. By translating E and F and that the new optimal map is simply the affine translation, we can assume w.l.o.g. that $x_0 = 0$. We apply the previous result Proposition 4.10 to obtain a symmetric, unit determinant matrix Q_1 and a vector b_1 such that $T_1(x) := Q_1(T(Q_1 x) - b_1)$ satisfies

$$\mathcal{E}(T_1, \theta R) \leq \theta^{2\alpha}\mathcal{E}(T, R),$$

where $\theta < 1$ is the constant given by the proposition. Given that T minimises the transport between E and F, Brenier's theorem gives that T_1 minimises the transport between the sets $E_1 := Q_1^{-1}E$ and $F_1 := Q_1(F - b_1)$, as T_1 is the gradient of the same convex potential as T under a suitable change of coordinates.

The conditions from the proposition that $|Q_1 - \mathrm{Id}\,|$ and $R^{-1}|b_1|$ very small for $\mathcal{E}(T, R) \ll 1$ are now used, to deduce that

$$B_{\theta R} \subseteq E_1 \cap F_1.$$

We now iterate Proposition 4.10, giving a sequence of symmetric matrices Q_n, vectors b_n and (optimal) maps $T_n(x) = Q_n(T_n(Q_n x) - b_n)$ between sets $E_n := Q_n^{-1}E_{n-1}$ and $F_n := Q_n(F_{n-1} - b_n)$, which satisfy

$$\mathcal{E}(T_n, \theta^n R) \leq \theta^{2\alpha n}\mathcal{E}(T, R), \tag{4.23}$$

$$|Q_n - \mathrm{Id}\,|^2 \lesssim \theta^{2\alpha n}\mathcal{E}(T, R) \quad \text{and} \quad |b_n|^2 \lesssim \theta^{2(\alpha+1)n}R^2\mathcal{E}(T, R)$$

To rigorously justify this iteration, we must note two things. The first is that the threshold for the energy ε is independent of any of the data, this holds since $\theta < 1$ so that $\mathcal{E}(T_n, \theta^n\mathbb{R}) < \mathcal{E}(T, R)$ and the original energy is below the threshold to reapply the Proposition each time independently of n.

The second thing we must justify is that each time the balls $B_{\theta^n R}$ are indeed in $E_n \cap F_n$ for all n - that is, we do not need to make the original energy $\mathcal{E}(T, R)$ smaller once n gets large to guarantee this, otherwise we could not iterate indefinitely. Note that the control on $|b_n|$ means that it decays faster with n than θ^n does as the power of θ here is greater than 1 - so if the initial energy $\mathcal{E}(T, R)$ is small enough, all the following translations will be on a scale much smaller than scaling the radius by a factor of θ. The control on $|Q_n|$ does not have a decay factor as strong as θ^n, but since this quantity is itself a scale factor of an operator rather than the displacement $|b_n|$, this is not necessary, and so long as $\mathcal{E}(T, R) \ll \theta$ initially, each Q_n will be much closer to the identity than the required factor of θ. Thus we have $B_{\theta^n R} \subseteq E_n \cap F_n$ for all n without needing to tweak $\mathcal{E}(T, R)$ so long as it starts small enough. This justifies being able to iterate the Proposition indefinitely.

We now set

$$A_n := Q_n Q_{n-1} \dots Q_1 \quad \text{ and } d_n := \sum_{i=1}^{n} Q_n Q_{n-1} \dots Q_i b_i,$$

noting still that $\det A_n = 1$ and also that $T_n(x) = A_n T(A_n^* x) - d_n$. We claim that

$$|A_n - \mathrm{Id}\,|^2 \lesssim \theta^{2\alpha} \mathcal{E}(T, R). \tag{4.24}$$

Proceeding inductively, if we had this control for n with some constant $C_n > 0$, then

$$\begin{aligned} |A_{n+1} - \mathrm{Id}\,| \leq& |Q_{n+1} A_n - A_n| + |A_n - \mathrm{Id}\,| \leq |Q_{n+1} - \mathrm{Id}\,||A_n| + |A_n - \mathrm{Id}\,| \\ \leq& |Q_{n+1} - \mathrm{Id}\,|(1 + C_n)\mathcal{E}(T, R) + C_n \mathcal{E}(T, R) \\ \leq& \left(\theta^{2\alpha(n+1)}(1 + C_n) + C_n\right) \mathcal{E}(T, R), \end{aligned}$$

so we would have the control for $n+1$ with $C_{n+1} = \theta^{2\alpha(n+1)}(1+C_n)+C_n$. In particular, the term-to-term increase is geometrically vanishing to zero at rate θ, so that there exists a control uniformly independent of n, giving (4.24). Specifically, this gives (for suitably small initial energy $\mathcal{E}(T, R) \ll 1$) that $B_{\frac{1}{2}\theta^k R} \subseteq A_k^*(B_{\theta^k R})$. We then deduce that

$$\begin{aligned} \min_{A,b} \frac{1}{(\frac{1}{2}\theta^k R)^2} \fint_{B_{\frac{1}{2}\theta^k R}} |T - (Ax + b)|^2 \lesssim& \frac{1}{(\theta^k R)^2} \fint_{A_k^*(B_{\theta^k R})} |T - (A_k^{-1}(A_k^*)^{-1} x + A_k^{-1} d_k)|^2 \\ =& \frac{1}{(\theta^k R)^2} \fint_{B_{\theta^k R}} |T(A_k^* x) - (A_k^{-1} x + A_k^{-1} d_k)|^2 \\ =& \frac{1}{(\theta^k R)^2} \fint_{B_{\theta^k R}} |A_k^{-1}(T_k - x)|^2 \\ \overset{(4.24)}{\lesssim}& \frac{1}{(\theta^k R)^2} \fint_{B_{\theta^k R}} |T_k - x|^2 \\ =& \mathcal{E}(T_k, \theta^k R) \lesssim \theta^{2\alpha k} \mathcal{E}(T, R). \end{aligned}$$

where we used that $T(A_k^* x) = A_k^{-1}(T_k(x) + d_k)$.

This establishes that distance from a polynomial on concentric circles related by a geometric factor of θ is uniformly controlled by the above power of θ times the energy on the largest circle. However, for the Campanato seminorm, we must take the sup overall r, it is certainly not sufficient only to control the decay of the energy along one specific subsequence of radii $\theta^k R$ converging to zero, one can easily construct a candidate energy for which this fails. We thus must control the energy uniformly on some annulus whose radii are different by a geometric factor of at least θ. Iterating down along any of these gives the control for all r. For any $r \in [\theta R, R]$, we have[8]

$$\mathcal{E}(T,R) = \frac{1}{R^{n+2}} \frac{R^{n+2}}{r^{n+2}|B_1|} \left(\int_{B_r} \cdots \right) \leq \frac{1}{\theta^{n+2}} \left(\int_{B_R} \cdots \right) = \frac{1}{\theta^{n+2}} \mathcal{E}(T,R).$$

The above combined with the decay control on geometric factor circles, gives (4.22), and on account of Campanato's theorem, the proof is complete. □

4.3.2 T is a $C^{1,\alpha}$ diffeomorphism on an open set of full measure

We now have all the tools we need to establish the partial regularity result. The previous results have built up to giving local regularity in a neighbourhood of 0 when $B_1 \subseteq E \cap F$ when $\mathcal{E} \ll 1$ so that T is close to being the identity. We now make use of a change of variables at a.e. point in E, which preserves optimality of the map and turns the new map into having this fixed point. Then the differentiability of Alexandrov's theorem gives $\mathcal{E} \to 0$, establishing the regularity on an open neighbourhood of each point. This is [16, Thrm. 4.7].

Theorem 4.12. *Let E and F be two bounded open sets with $|E| = |F|$, and let T be the minimiser of the quadratic transport (4.1). There exist open sets $E' \subseteq E$ and $F' \subseteq F$ of full measure such that T is a $C^{1,\alpha}$ diffeomorphism between E' and F' for any $\alpha \in (0,1)$.*

Proof. By Brenier's theorem T is a.e. the gradient of a convex function, hence by the Alexandrov theorem (Theorem 4.2) we have two sets of full measure $E_1 \subseteq E$ and $F_1 \subseteq F$ such that T and T^{-1} are Fréchet differentiable (minus negligible sets), so that for $x_0, y_0 \in E_1 \times F_1$, there exit symmetric matrices A, B s.t. for a.e. $(x,y) \in E \times F$,

$$T(x) = T(x_0) + A(x - x_0) + o(|x - x_0|) \quad \text{and} \quad T^{-1}(y) = T^{-1}(y_0) + B(y - y_0) + o(|y - y_0|). \tag{4.25}$$

On account of Remark 1.17 we can assume that T and T^{-1} are inverses of one another everywhere on $E_1 \times F_1$. Then by the Jacobian equation (1), since both densities are constant $\equiv 1$, we have $\det A = \det B = 1$, and we also have $A = B^{-1}$. We now set $E' = E_1 \cap T^{-1}(F_1)$ and $F' = T(E') = F_1 \cap T(E_1)$. Since both measures are absolutely continuous, T sends measure zero sets to measure zero sets so consequently E' and F' are of full measure, that is $|E \setminus E'| = |F \setminus F'| = 0$. We now prove the sets E' and F' are open and T is a $C^{1,\alpha}$ diffeomorphism between them.

[8]This is reminiscent of the similar control we established for Proposition 3.8.

Fix some $x_0 \in E'$ and set $y_0 = T(x_0) \in F'$. By the differentiability (4.25), we have

$$\lim_{R\to 0} \frac{1}{R^2} \fint_{B_R(x_0)} |T - y_0 - A(x - x_0)|^2 + \frac{1}{R^2} \fint_{B_R(y_0)} |T^{-1} - x_0 - A^{-1}(y - y_0)|^2 = 0. \quad (4.26)$$

We make a change of variables on both domain and codomain $x = A^{-1/2}\hat{x} + x_0$ and $y = A^{1/2}\hat{y} + y_0$, and define the new map

$$\hat{T}(\hat{x}) := A^{1/2}(T(A^{-1/2}\hat{x} + x_0) - y_0) \text{ with inverse } \hat{T}^{-1}(\hat{y}) = A^{1/2}(T^{-1}(A^{1/2}\hat{y} + y_0) - x_0).$$

This map transports between the sets $\hat{E} := A^{1/2}(E - x_0)$ and $\hat{F} := A^{-1/2}(F - y_0)$, and by construction has a fixed point $T(0) = 0$. By Breiner's theorem $T = \nabla u$ for some convex u, then

$$\hat{T}(\hat{x}) = A^{-1/2}(\nabla u(A^{1/2}\hat{x} + x_0) - y_0) = \hat{\nabla}\big(u(A^{1/2}\hat{x} + x_0) - y_0 \cdot \hat{x}\big),$$

where $\hat{\nabla}$ is the gradient in new coordinates, so $\hat{T}$ is the gradient of a convex function also (as convexity is preserved under affine changes of coordinates in the domain, and by adding an affine function to the codomain). Noting that the Jacobians are $|\det A^{1/2}| = |\det A^{1/2}| = 1$, this change of variables transforms (4.26) into

$$\lim_{R\to 0} \frac{1}{R^2} \fint_{B_R} |A^{1/2}(\hat{T} - \hat{x})|^2 + \frac{1}{R^2} \fint_{B_R} |A^{-1/2}(\hat{T}^{-1} - \hat{y})|^2 = 0.$$

Using that $A^{1/2}, A^{-1/2}$ are invertible by their determinant and thus positive definite (not just semi-definite) by convexity we have a smallest positive eigenvalue for each and by bounding in terms of these we ascertain

$$\lim_{R\to 0} \frac{1}{R^2} \fint_{B_R} |\hat{T} - \hat{x}|^2 + \frac{1}{R^2} \fint_{B_R} |\hat{T}^{-1} - \hat{y}|^2 = 0.$$

Now for small enough R, $B_R(x_0) \subseteq E$ and $B_R(y_0) \subseteq F$ such that $B_R \subseteq \hat{E} \cap \hat{F}$. Thus we invoke Proposition 4.11 to give both $\hat{T}$ and $\hat{T}^{-1}$ as $C^{1,\alpha}$ in a neighbourhood U of zero. Returning to the original map, we have T is a $C^{1,\alpha}$ diffeomorphism in a nbhd. U of x_0 to $T(U)$ a nbhd. of $T(x_0)$. We thus have $U \times T(U) \subseteq E' \times F'$, giving E' and F' as open sets and on account of Remark 1.17, T is a $C^{1,\alpha}$ diffeomorphism between E' and F'. $\square$

Remark 4.13. As a consequence of the Evans-Krylov theorem, letting $T = \nabla u$ for u convex by Brenier's theorem, we have u as $C^{2,\alpha}$ and hence a classical solution to the Monge-Ampère equation $\det D^2 u = \chi_{E'}$ so that we actually get $T \in C^\infty$ on E'. Unfortunately, we do not have room to discuss this here.

A Notation

χ_A	indicator of set A
$\|A\|$	Lebesgue measure of a set
$T_\# \mu$	push forward measure $T_\# \mu(A) := \mu(T^{-1}(A))$
$\mathcal{P}(X)$	probability measures on X, probability measures on X
$\mathcal{P}_2(X)$	probability measures with finite second moments
$\mathcal{M}^d(\mathbb{R}^d)$	$\mathbb{R}^d$ valued vector measures
$\Pi(\mu, \nu)$	prob. measures $\gamma \in \mathcal{P}(X \times Y)$ with marginals $P_x \# \gamma = \mu$, $P_y \# \gamma = \nu$
$W_2(\mu, \nu)$	the 2-Wasserstein distance
l.s.c./ u.s.c	lower/upper semi-continuous
$\ll$	Numbers: sublinearly much smaller than. Measures: absolutely continuous with respect to
$\lesssim$	globally controlled by (constant depending only on d/ Hölder exponent)
$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$	for measures with $\mu \ll \nu$, the Radon-Nikodym derivative
$⨍$	average integral, $\frac{1}{\|A\|} \int_A$
C_c	continuous with compact support
C_b	continuous and bounded
$C^{k,\alpha}$	k times differentiable functions which kth order derivatives α-Hölder
$H^k; W^{p,k}$	Sobolev spaces of k times weak differentiable functions with 2 or p-summable derivatives
$\mathscr{P}_k$	polynomials of degree k or less
$\Subset$	compactly contained
∇	gradient in the space variables, not including any time variables
A^*	for a square matrix A, the adjoint/transpose
$o(f(x))$	standard little "o" notation - something which decays to zero when divided by $f(x)$.

B Appendix

Definition B.1. *A collection of probability measures $\mathcal{Q} \subset \mathcal{P}(X)$ are called tight if for each $\varepsilon > 0$ there exits a compact $K_\varepsilon \Subset X$ such that $\mu(X \setminus K_\varepsilon) < \varepsilon$ for all $\mu \in \mathcal{Q}$, so that up to an arbitrarily small amount of mass, the support of all the measures lie inside the same compact set.*

Theorem B.2 (Prokhorov's Theorem)**.** *Let (X, d) be a Polish space, then a collection of measures is tight if and only if they are relatively compact w.r.t. the narrow topology.*

We use an infinite dimensional saddle problem result from [7].

Theorem B.3 (min-max)**.** *Let X, Y be topological vector spaces over $\mathbb{R}$ and let $A \subset X$ and $B \subset Y$ be closed convex sets. Let $F : A \times B \to \mathbb{R}$ be a functional that is convex and l.s.c. in the first variable, and concave and u.s.c. in the second. Suppose that there exists $y \in B$ and $\lambda \in \mathbb{R}$ such that $\{x \in A; F(x, y) \le \lambda\}$ is non-empty and compact. Then we can allow the* $\inf - \sup$ *exchange*

$$\sup_{y\in B} \inf_{x\in A} F(x, y) = \inf_{x\in A} \sup_{y\in B} F(x, y).$$

Theorem B.4 (Poincaré-Wirtinger inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with* $\operatorname{Diam}\Omega < l$. *There exists a constant $C(n, p) > 0$ such that for $u \in W^{1,p}(\Omega)$ and $u_\Omega = \fint_\Omega u$,*

$$\int_\Omega |u - u_\Omega|^p \mathrm{d}x \le Cl^p \int_\Omega |Du|^p \mathrm{d}x.$$

Definition B.5. *For a square matrix A, the matrix exponential is defined as*

$$e^A := \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$

Some properties we use:

- *If A is symmetric, then e^A is also symmetric.*
- *If* $\operatorname{Tr} A = 0$ *then* $\det e^A = 1$
- *If A is invertible, $(e^A)^k = e^{kA}$ for any $k \in \mathbb{Z}$.*

Lemma B.6. *The function* $\det(\cdot)^{\frac{1}{d}} : \mathbb{R}^{d\times d} \to \mathbb{R}$ *is concave.*

Bibliography

[1] AD Alexandrov. The existence almost everywhere of the second differential of a convex function and some associated properties of convex surfaces. *Ucenye Zapiski Leningrad. Gos. Univ. Ser. Math*, 37(6):3–35, 1939.

[2] Luigi Ambrosio, Alberto Bressan, Dirk Helbing, Axel Klar, Enrique Zuazua, Luigi Ambrosio, and Nicola Gigli. A user's guide to optimal transport. *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*, pages 1–155, 2013.

[3] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[4] Gabriele Bianchi, Andrea Colesanti, and Carlo Pucci. On the second differentiability of convex surfaces. *Geometriae Dedicata*, 60:39–48, 1996.

[5] Guy Bouchitté, Chloé Jimenez, and M Rajesh. A new l^∞ estimate in optimal mass transport. *Proceedings of the American Mathematical Society*, 135(11):3525–3535, 2007.

[6] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[7] Haim Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Elsevier, 1973.

[8] T. Editors of Encyclopaedia. Britannica. Gaspard monge, count de péluse.

[9] Luis A Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.

[10] Luis A Caffarelli and Xavier Cabré. *Fully nonlinear elliptic equations*, volume 43. American Mathematical Soc., 1995.

[11] Sergio Campanato. Proprieta di una famiglia di spazi funzionali. *Annali della Scuola Normale Superiore di Pisa-Scienze Fisiche e Matematiche*, 18(1):137–160, 1964.

[12] Guido De Philippis and Alessio Figalli. The monge–ampère equation and its link to optimal transportation. *Bulletin of the American Mathematical Society*, 51(4):527–580, 2014.

[13] Alessio Figalli and Young-Heon Kim. Partial regularity of brenier solutionsof the monge-ampère equation. *Discrete and Continuous Dynamical Systems*, 28(2):559–565, 2010.

[14] Mariano Giaquinta and Luca Martinazzi. *An introduction to the regularity theory for elliptic systems, harmonic maps and minimal graphs.* Springer Science & Business Media, 2013.

[15] Michael Goldman and F Otto. A variational proof of partial regularity for optimal transportation maps. In *Annales Scientifiques de l'École Normale Supérieure*, 2019.

[16] Michael Goldman and Felix Otto. A variational proof of partial regularity for optimal transportation maps between sets, 2017. cvgmt preprint.

[17] Barbro Grevholm. On the structure of the spaces $L_k^{p,\lambda}$. *Mathematica Scandinavica*, 26(2):241–254, 1970.

[18] Leonid V Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[19] Francesco Maggi. *Sets of finite perimeter and geometric variational problems: an introduction to Geometric Measure Theory*, volume 135. Cambridge University Press, 2012.

[20] Robert J McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.

[21] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[22] Guido Philippis. *Regularity of optimal transport maps and applications*, volume 17. Springer Science & Business Media, 2013.

[23] Ralph Tyrell Rockafellar. *Convex analysis:(pms-28).* Princeton University Press, 1992.

[24] Serik Sagitov. Weak convergence of probability measures, 2020. lecture notes.

[25] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Springer, Birkhäuser, 2015.

[26] Filippo Santambrogio. *A Course in the Calculus of Variations: Optimization, Regularity, and Modeling.* Springer Nature, 2023.

[27] Thomas Staudt, Shayan Hundrieser, and Axel Munk. On the uniqueness of kantorovich potentials. *arXiv preprint arXiv:2201.08316*, 2022.

[28] Luis Caicedo Torres, Luiz Manella Pereira, and M Hadi Amini. A survey on optimal transport for machine learning: Theory and applications. *arXiv preprint arXiv:2106.01963*, 2021.

[29] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.