FIVE LECTURES ON OPTIMAL TRANSPORTATION: GEOMETRY, REGULARITY AND APPLICATIONS

ROBERT J. MCCANN AND NESTOR GUILLEN

Abstract. In this series of lectures we introduce the Monge-Kantorovich problem of optimally transporting one distribution of mass onto another, where optimality is measured against a cost function $c(x, y)$. Connections to geometry, inequalities, and partial differential equations will be discussed, focusing in particular on recent developments in the regularity theory for Monge-Ampère type equations. An application to microeconomics will also be described, which amounts to finding the equilibrium price distribution for a monopolist marketing a multidimensional line of products to a population of anonymous agents whose preferences are known only statistically. ©2010 by Robert J. McCann. All rights reserved.

Contents

Preamble 2
1. An introduction to optimal transportation 2
1.1. Monge-Kantorovich problem: transporting ore from mines to factories 2
1.2. Wasserstein distance and geometric applications 3
1.3. Brenier's theorem and convex gradients 4
1.4. Fully-nonlinear degenerate-elliptic Monge-Ampère type PDE 4
1.5. Applications 5
1.6. Euclidean isoperimetric inequality 5
1.7. Kantorovich's reformulation of Monge's problem 6
2. Existence, uniqueness, and characterization of optimal maps 6
2.1. Linear programming duality 8
2.2. Game theory 8
2.3. Relevance to optimal transport: Kantorovich-Koopmans duality 9
2.4. Characterizing optimality by duality 9
2.5. Existence of optimal maps and uniqueness of optimal measures 10
3. Methods for obtaining regularity of optimal mappings 11
3.1. Rectifiability: differentiability almost everywhere 12
3.2. From regularity a.e. to regularity everywhere 13
3.3. Regularity methods for the Monge-Ampère equation; renormalization 13
3.4. The continuity method (schematic) 14
4. Regularity and counterexamples for general costs 15
4.1. Examples 15
4.2. Counterexamples to the continuity of optimal maps 16
4.3. Monge-Ampère type equations 17
4.4. Ma-Trudinger-Wang conditions for regularity 18
4.5. Regularity results 19
4.6. Ruling out discontinuities: Loeper's maximum principle 19
4.7. Interior Hölder continuity for optimal maps 22
5. Multidimensional screening: an application to economic theory 23
5.1. Monopolist nonlinear pricing and the principal-agent framework 23

5.2. Variational formulation using optimal transportation 24
5.3. When is this optimization problem convex? 25
5.4. Variant: maximizing social welfare 26
6. A pseudo-Riemannian and symplectic geometric afterword 27
References 28

Preamble

This survey is based on a series of five lectures by Robert McCann (of the University of Toronto), delivered at a summer school on "New Vistas in Image Processing and Partial Differential Equations" organized 7-12 June 2010 by Irene Fonseca, Giovanni Leoni, and Dejan Slepcev of Carnegie Mellon University on behalf of the Center for Nonlinear Analysis there. The starting point for the manuscript which emerged was a detailed set of notes taken during those lectures by Nestor Guillen (University of Texas at Austin).

These notes are intended to convey a flavor for the subject, without getting bogged down in too many technical details. Part of the discussion is therefore impressionistic, and some of the results are stated under the tacit requirement that the supports of the measures $\mu^\pm$ be compact, with the understanding that they extend to non-compactly supported measures under appropriate hypotheses [93] [59] [45] [46] concerning the behaviour near infinity of the measures and the costs. The choice of topics to be covered in a series of lectures is necessarily idiosyncratic. General references for these and other topics include papers of the first author posted on the website `www.math.toronto.edu/mccann` and the two books by Villani [138] [139]. Earlier surveys include the ones by Ambrosio [4], Evans [43], Urbas [136] and Rachev and Rüschendorf [115]. Many detailed references to the literature may be found there, to augment the bibliography of selected works included below.

1. An introduction to optimal transportation

1.1. **Monge-Kantorovich problem: transporting ore from mines to factories.** The problem to be discussed can be caricatured as follows: imagine we have a distribution of iron mines across the countryside, producing a total of 1000 tonnes of iron ore weekly, and a distribution of factories across the countryside that consume a total of 1000 tonnes of iron ore weekly. Knowing the *cost* $c(x, y)$ per ton of ore transported from a mine at x to a factory at y, the problem is to decide which mines should be supplying which factories so as to minimize the total transportation costs.

To model this problem mathematically, let the triples $(M^\pm, d^\pm, \omega^\pm)$ denote two complete separable metric spaces $M^\pm$ — also called *Polish* spaces — equipped with distance functions $d^\pm$ and Borel reference measures $\omega^\pm$. These two metric spaces will represent the landscapes containing the mines and the factories. They will often be assumed to be geodesic spaces, and/or to coincide. Here a *geodesic space* $M(= M^\pm)$ refers to a metric space in which every pair of points $x_0, x_1 \in M$ is connected by a curve $s \in [0, 1] \to x_s \in M$ satisfying

$$d(x_0, x_s) = s d(x_0, x_1) \qquad \text{and} \qquad d(x_s, x_1) = (1 - s) d(x_0, x_1) \qquad \forall\ s \in [0, 1]. \tag{1}$$

Such a curve is called a *geodesic segment*.

e.g. 1) Euclidean space: $M = \mathbf{R}^n$, $d(x,y) = |x-y|$, $\omega = Vol = \mathcal{H}^n$ = Hausdorff n-dimensional measure, geodesic segments take the form $x_s = (1-s)x_0 + sx_1$.

e.g. 2) Complete Riemannian manifold $(M = M^{\pm}, g_{ij})$, with or without boundary: $d\omega = dVol = d\mathcal{H}^n = (\det g_{ij})^{1/2} d^n x$,

$$\frac{1}{2}d^2(x_0,x_1) = \inf_{\{x_s | x(0)=x_0, x(1)=x_1\}} \frac{1}{2}\int_0^1 \langle \dot{x_s}, \dot{x_s} \rangle_{g(x_s)} ds.$$

A minimizing curve $s \in [0,1] \longmapsto x_s \in M$ exists by the Hopf-Rinow theorem; it satisfies (1), and is called a Riemannian *geodesic*.

The distributions of mines and factories will be modeled by Borel probability measures μ^+ on M^+ and μ^- on M^-, respectively. Any Borel map $G : M^+ \longrightarrow M^-$ defines an image or *pushed-forward* measure $\nu = G_{\#}\mu^+$ on M^- by

$$(G_{\#}\mu^+)[V] := \mu^+[G^{-1}(V)] \qquad \forall\ V \subset M^-. \tag{2}$$

A central problem in optimal transportation is to find, among all maps $G : M^+ \longrightarrow M^-$ pushing μ^+ forward to μ^-, one which minimizes the total cost

$$\text{cost}\,(G) = \int_{M^+} c(x, G(x)) d\mu^+(x). \tag{3}$$

This problem was first proposed by Monge in 1781, taking the Euclidean distance $c(x,y) = |x-y|$ as his cost function [104]. For more generic costs, some basic mathematical issues such as the existence, uniqueness, and mathematical structure of the optimizers are addressed in the second lecture below. However nonlinearity of the objective functional and a lack of compactness or convexity for its domain make Monge's formulation of the problem difficult to work with. One and a half centuries later, Kantorovich's relaxation of the problem to an (infinite-dimesional) linear program provided a revolutionary tool [67] [68].

1.2. **Wasserstein distance and geometric applications.** The minimal cost of transport between μ^+ and μ^- associated to $c(\cdot,\cdot)$ will be provisionally denoted by

$$W_c(\mu^+, \mu^-) = \inf_{G_{\#}\mu^+ = \mu^-} \text{cost}\,(G). \tag{4}$$

It can be thought of as quantifying the discrepancy between μ^+ and μ^-, and is more properly defined using Kantorovich's formulation (11), though we shall eventually show the two definitions coincide in many cases of interest. When $M = M^{\pm}$, the costs $c(x,y) = d^p(x,y)$ with $0 < p < 1$ occur naturally in economics and operations research, where it is often the case that there is an economy of scale for long trips [96]. In this case, the quantity $W_c(\mu^+, \mu^-)$ defines a metric on the space $\mathcal{P}(M)$ of Borel probability measures on M. For $p \geq 1$ on the other hand, it is necessary to extract a *p-th* root to obtain a metric

$$d_p(\mu^+, \mu^-) := W_c(\mu^+,\, u^-)^{1/p} \tag{5}$$

on $\mathcal{P}(M)$ which satisfies the triangle inequality.

Though the initial references dealt specifically with the case $p = 1$ [69] [142], the whole family of distances are now called *Kantorovich-Rubinstein* or *Wasserstein* metrics [41]. Apart from the interesting exception of the limiting case $d_\infty = \lim_{p\to\infty} d_p$ [98] [26], on a compact metric space M all these metrics give rise to the same topology, namely weak-$*$ convergence. For non-compact M, the d_p topologies differ from each other only in the number of moments of a sequence of measures which are required to converge. Moreover, $(\mathcal{P}(M), d_p)$ inherits geometric properties from (M,d), such as being a geodesic space. Notions such as Ricci curvature in the underlying space (M,d) can be characterized by

the geodesic convexity first explored in [94] of certain functionals on the larger space $\mathcal{P}(M)$ — such as Boltzmann's entropy. One direction of this equivalence was proved by Cordero-Erausquin, McCann, and Schmuckenschäger [30] and Otto and Villani [110] in projects which were initially independent (see also [28] and [31]), while the converse was established by von Renesse and Sturm [116] confirming the formal arguments of [110]. This equivalence forms the basis of Lott-Villani and Sturm's definition of lower bounds for Ricci curvature in the metric measure space setting, without reference to any underlying Riemannian structure [88] [127]. McCann and Topping used a similar idea to characterize the Ricci flow [99], which led Lott [87] and Topping [130] to simpler proofs of Perelman's celebrated monotonicity results [112]. Despite the interest of these recent developments, we shall not pursue them farther in these lectures, apart from sketching a transportation-based proof of the isoperimetric theorem whose ideas underpin many such geometric connections.

1.3. **Brenier's theorem and convex gradients.** It turns out that Monge's cost $c(x,y) = |x-y|$ is among the hardest to deal with, due to its lack of strict convexity. For this cost, the minimizer of (4) is not generally unique, even on the line $M^\pm = \mathbf{R}$. Existence of solutions is tricky to establish: the first 'proof', due to Sudakov [128], relied on an unsubstantiated claim which turned out to be correct only in the plane $M^\pm = \mathbf{R}^2$ [11]; higher dimensional arguments were given increasing generality by Evans-Gangbo [44], and then Ambrosio [4], Caffarelli-Feldman-McCann [21], and Trudinger-Wang [132] independently. Simpler approaches were proposed by Champion-DePascale [25] and Bianchini-Cavalletti [11] more recently.

The situation for the quadratic cost $c(x,y) = |x-y|^2$ is much simpler, mirroring the relative simplicity of the Hilbert geometry of L^2 among Banach spaces L^p with $p \geq 1$. Brenier [14] [15] (and others around the same time [114] [124] [33] [123] [34] [1]) showed that there is a unique [34] [1] solution [33] , and characterized it as a convex gradient [114] [124] [123].

Theorem 1.1 (A version of Brenier's theorem)**.** *If $\mu^+ \ll dVol$ and μ^- are Borel probability measures on $M^\pm = \mathbf{R}^n$, then there exists a convex function $u : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ whose gradient $G = Du : \mathbf{R}^n \to \mathbf{R}^n$ pushes μ^+ forward to μ^-. Apart from changes on a set of measure zero, G is the only map to arise in this way. Moreover, G uniquely minimizes Monge's problem* (4) *for the cost $c(x,y) = |x-y|^2$.*

Remark: In this generality the theorem was established by McCann [93], where the assumption $\mu^+ \ll dVol$ was also relaxed. A further relaxation by Gangbo-McCann [59] is shown to be sharp in Gigli [61].

1.4. **Fully-nonlinear degenerate-elliptic Monge-Ampère type PDE.** How do partial differential equations (more specifically, fully nonlinear degenerate elliptic PDE) enter the picture? Let's consider the constraint $G_\# \mu^+ = \mu^-$, assuming moreover that $\mu^\pm = f^\pm dVol^\pm$ on $\mathbf{R}^n$ or on Riemannian manifolds $M^\pm$. Then if $\phi \in C(M^-)$ is a test function, it follows that

$$\int_{M^+} \phi(G(x)) f^+(x) dVol^+(x) = \int_{M^-} \phi(y) f^-(y) dVol^-(y).$$

If G was a diffeomorphism, we could combine the Jacobian factor $d^n y = |\det DG(x)| d^n x$ from the change of variables $y = G(x)$ with arbitrariness of $\phi \circ G$ to conclude $f^+(x) = |\det DG(x)| f^-(G(x))$ for all x. We will actually see that this nonlinear equation holds f^+-a.e. as a consequence of Theorem 3.2.

In the case of Brenier's map $G(x) = Du(x)$, convexity of u implies non-negativity of the Jacobian $DG(x) = D^2u(x) \geq 0$. It also guarantees almost everywhere differentiability of G by Alexandrov's theorem (or by Lebesgue's theorem in one dimension); see Theorem 3.2 for the sketch of a proof. Thus u solves the Monge-Ampére equation [15]

$$f^+(x) = \det(D^2u(x))f^-(Du(x)) \tag{6}$$

a.e. [94] subject to the condition $Du(x) \in M^-$ for $x \in M^+$. This is known as the *2nd boundary value problem* in the partial differential equations literature. We shall see that linearization of this equation around a *convex* solution leads to a (degenerate) elliptic operator (29), whose ellipticity becomes uniform if the solution u is smooth and *strongly* convex, meaning positivity of its Hessian is strict: $D^2u(x) > 0$.

1.5. **Applications.** The Monge-Kantorovich theory has found a wide variety of applications in pure and applied mathematics. On the pure side, these include connections to inequalities [92] [131] [94] [32] [90] [52], geometry (including sectional [85] [73], Ricci [88] [87] [127] [99] and mean [75] curvature), nonlinear partial differential equations [14] [18] [16] [135] [89], and dynamical systems (weak KAM theory [9]; nonlinear diffusions [109]; gradient flows [5]). On the applied side these include applications to vision (image registration and morphing [64]), economics (equilibration of supply with demand [42] [27], structure of cities [23], maximization of profits [118] [22] [49] or social welfare [49]), physics [40] [129] [95] [51], engineering (optimal shape / material design [12] [13], reflector antenna design [63] [140] [141], aerodynamic resistance [113]), atmosphere and ocean dynamics (the semigeostrophic theory [36] [37] [35]), biology (irrigation [10], leaf growth [143]), and statistics [115]. See [138] [139] for further directions, references, and discussion.

1.6. **Euclidean isoperimetric inequality.** It was observed (independently by McCann [92] [94] and Trudinger [131]) that a solution to the second boundary value problem for the Monge-Ampère equation (6) yields a simple proof of the isoperimetric inequality (with its sharp constant): for $M^+ \subset \mathbf{R}^n$

$$Vol(M^+) = Vol(B_1) \quad \Rightarrow \quad \mathcal{H}^{n-1}(\partial M^+) \geq \mathcal{H}^{n-1}(\partial B_1). \tag{7}$$

The following streamlined argument was perfected later; it combines optimal maps with an earlier approach from Gromov's appendix to [101].

Proof. Take $f^+ = \chi_{M^+}$ and $f^- = \chi_{B_1}$ to be uniformly distributed. Brenier's theorem then gives a volume-preserving map $G = Du$ between M^+ and B_1:

$$1 = det^{1/n}(D^2u(x)).$$

The expression on the right is the geometric mean of the eigenvalues of $D^2u(x)$, which are non-negative by convexity of u, so the arithmetic-geometric mean inequality yields

$$1 \leq (arithmetic\ mean\ of\ eigenvalues) = \frac{1}{n}\Delta u \tag{8}$$

almost everywhere in M^+. (The right hand side is the absolutely continuous part of the distributional Laplacian; convexity of u allows it to be replaced by the full distributional Laplacian of u without spoiling the inequality.) Integrating inequality (8) on M^+ yields

$$Vol(M^+) \leq \frac{1}{n}\int_{M^+} \Delta u \, d^n x = \frac{1}{n}\int_{\partial M^+} Du(x) \cdot \hat{n}_{M^+}(x) d\mathcal{H}^{n-1}(x). \tag{9}$$

Now, since $G = Du \in B_1$ whenever $x \in M^+$, we have $|Du| \leq 1$, thus $Vol(M^+) = Vol(B_1)$ gives

$$Vol(B_1) \leq \frac{1}{n} \int_{\partial M^+} 1 \, d\mathcal{H}^{n-1} = \frac{1}{n} \mathcal{H}^{n-1}(\partial M^+). \tag{10}$$

In the case special $M^+ = B_1$, Brenier's map coincides with the identity map so equalities hold throughout (8)–(10), yielding the desired conclusion (7)! □

As the preceding proof shows, one of the important uses of optimal transportation in analysis and geometry is to encode non-local 'shape' information into a map which can be localized, reducing global geometric inequalities to algebraic inequalities under an integral. For subsequent developments in this direction, see works of Ambrosio, Cordero-Erausquin, Carrillo, Figalli, Gigli, Lott, Maggi, McCann, Nazaret, Otto, Pratelli, von Renesse, Schmuckenschläger, Sturm, Topping and Villani in [95] [110] [30] [31] [32] [5] [24] [116] [127] [88] [87] [99] [52].

1.7. **Kantorovich's reformulation of Monge's problem.** Now let us turn to the proof of Brenier's theorem and the ideas it involves. A significant breakthrough was made by Kantorovich [67] [68], who *relaxed* our optimization problem (the Monge problem), by dropping the requirement that all the ore from a given mine goes to a single factory. In other words:

Replace $G : M^+ \to M^-$ by a measure $0 \leq \gamma$ on $M^+ \times M^-$ whose marginals are μ^+ and μ^-, respectively, and among such measures choose γ to minimize the functional

$$\text{cost}\,(\gamma) = \int_{M^+ \times M^-} c(x,y) d\gamma(x,y).$$

Such a joint measure γ is also known as a "transport plan" (in analogy with"transport map"). This is better than Monge's original formulation for at least two reasons:

1) The functional to be minimized now depends linearly on γ.

2) The set $\Gamma(\mu^+, \mu^-)$ of admissible competitors γ is a convex subset of a suitable Banach space: namely, the dual space to continuous functions $(C(M^+ \times M^-), \|\cdot\|_\infty)$ (which decay to zero at infinity in case the compactness of $M^\pm$ is merely local).

In this context, well-known results in functional analysis guarantee existence of a minimizer γ under rather general hypotheses on c and $\mu^\pm$. Our primary task will be to understand when the solution will be unique, and to characterize it. At least one minimizer will be an extreme point of the convex set $\Gamma(\mu^+, \mu^-)$, but its uniqueness remains an issue. Necessary and sufficient conditions will come from the duality theory of (infinite dimensional) linear programming [6] [70].

2. Existence, uniqueness, and characterization of optimal maps

Let's get back to the Kantorovich problem:

$$W_c(\mu^+, \mu^-) := \min_{\gamma \in \Gamma(\mu^+, \mu^-)} \int_{M^+ \times M^-} c(x,y) d\gamma(x,y) = \min_{\gamma \in \Gamma(\mu^+, \mu^-)} \text{cost}\,(\gamma) \tag{11}$$

The basic geometric object of interest to us will be the support $\text{spt}\,\gamma := S$ of a competitor γ, namely the smallest closed subset $S \subset M^+ \times M^-$ carrying the full mass of γ.

What are some of the competing candidates for the minimizer?

eg.1) Product measure: $\mu^+ \otimes \mu^- \in \Gamma(\mu^+, \mu^-)$, for which spt $\gamma =$ spt $\mu^+ \times$ spt μ^-.

eg.2) Monge measure: if $G : M^+ \to M^-$ with $G_{\#}\mu^+ = \mu^-$ then $id \times G : M^+ \longrightarrow M^+ \times M^-$ and $\gamma = (id \times G)_{\#}\mu^+ \in \Gamma(\mu^+, \mu^-)$ has cost $(\gamma) =$ cost (G).

The second example shows in what sense Kantorovich's formulation is a relaxation of Monge's problem, and why (4) must be at least as big as (11). In this example spt γ will be the (closure of the) graph of $G : M^+ \longrightarrow M^-$, which suggests how Monge's map G might in principle be reconstructed from a minimizing Kantorovich measure γ. Before attempting this, let us recall a notion which characterizes optimality in the Kantorovich problem.

Definition 2.1 (c-cyclically monotone sets)**.** *$S \subset M^+ \times M^-$ is* c-cyclically monotone *if and only if all $k \in \mathbf{N}$ and $(x_1, y_1), ..., (x_k, y_k) \in S$ satisfy*

$$\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_i, y_{\sigma(i)}) \tag{12}$$

for each permutation σ of k letters.

The following result was deduced by Smith and Knott [125] from a theorem of Rüschendorf [120]. A more direct proof was given by Gangbo and McCann [59]; its converse is true as well.

Theorem 2.2 (Smith and Knott '92)**.** *If $c \in C(M^+ \times M^-)$, then optimality of $\gamma \in \Gamma(\mu^+, \mu^-)$ implies* spt γ *is a c-cyclically monotone set.*

The idea of the proof in [59] is that if *spt* γ is not cyclically monotone, then setting $o_i = (x_i, y_i)$ and $z_i = (x_i, y_{\sigma(i)})$ we could with some care define a perturbation

$$\gamma_\epsilon = \gamma + \epsilon(\text{near the z's}) - \epsilon(\text{near the o's})$$

in $\Gamma(\mu^+, \mu^-)$ of γ for which cost $(\gamma_\epsilon) <$ cost (γ_0), thus precluding the optimality of γ.

e.g. If $c(x, y) = -x \cdot y$ or $c(x, y) = \frac{1}{2}|x - y|^2$ then (12) becomes $\sum_{i=1}^{k} \langle y_i, x_i - x_{i-1} \rangle \geq 0$ with the convention $x_0 := x_k$. This is simply called *cyclical monotonicity*, and can be viewed as a discretization of

$$\oint \bar{y}(x) \cdot dx \geq 0,$$

a necessary and sufficient condition for the vector field $\bar{y}(x)$ to be conservative, meaning $y = Du(x)$. This heuristic underlies a theorem of Rockafellar [119]:

Theorem 2.3 (Rockafellar '66)**.** *The set $S \subset \mathbf{R}^n \times \mathbf{R}^n$ is cyclically monotone if and only if there exists a convex function $u : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ such that $S \subset \partial u$ where*

$$\partial u := \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^n \mid u(z) \geq u(x) + \langle z - x, y \rangle + o(|z - x|) \quad \forall z \in \mathbf{R}^n\}. \tag{13}$$

The *subdifferential* ∂u defined by (13) consists of the set of (*point*, *slope*) pairs for which y is the slope of a hyperplane supporting the graph of u at $(x, u(x))$.

Remark 2.4 (Special case (*monotonicity*)). Note that when $c(x, y) = -x \cdot y$ and $k = 2$, (12) implies for all $(x_1, y_1), (x_2, y_2) \in S$ that

$$\langle \Delta x, \Delta y \rangle := \langle x_2 - x_1, y_2 - y_1 \rangle \geq 0. \tag{14}$$

This condition implies that y_2 is constrained to lie in a halfspace with y_1 on its boundary and Δx as its inward normal. Should $y = Du(x)$ already be known to be conservative, the monotonicity inequalities (14) alone become equivalent to convexity of u.

2.1. **Linear programming duality.** An even more useful perspective on these linear programming problems is given by the the duality theorem discovered by Kantorovich [67] and Koopmans [76] — for which they later shared the Nobel Memorial Prize in economics. It states that our minimization problem is equivalent to a maximization problem

$$\min_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{M^+\times M^-} c\,d\gamma = \sup_{(-u,-v)\in Lip_c} -\int_{M^+} u(x)\,d\mu^+(x) - \int_{M^-} v(y)\,d\mu^-(y). \tag{15}$$

Here

$$Lip_c = \{(u^+,u^-)\ with\ u^\pm \in L^1(M^\pm, d\mu^\pm) \mid c(x,y) \geq u^+(x)+u^-(y)\ \forall (x,y)\in M^+\times M^-\}.$$

One of the two inequalities ($\geq$) in (15) follows at once from the definition of $-(u,v)\in Lip_c$ by integrating

$$c(x,y) \geq -u(x) - v(y)$$

against $\gamma\in\Gamma(\mu^+,\mu^-)$. The magic of duality is that *equality* holds in (15).

2.2. **Game theory.** Some intuition for why this magic works can be gleaned from the the theory of (two-player, zero-sum) games. In that context, Player 1 chooses strategy $x\in X$, Player 2 chooses strategy $y\in Y$, and the outcome is that Player 1 pays $P(x,y)$ to Player 2. The payoff function $P\in C(X\times Y)$ is predetermined and known in advance to both players; P1 wants to minimize the resulting payment and P2 wants to maximize it.

Now, what if one the players declares his or her strategy (x or y) to the other player in advance? If P1 declares first, the outcome is better for P2, who has a chance to optimize his response to the announced strategy x, and conversely. This implies that

$$\inf_{x\in X}\sup_{y\in Y} P(x,y) \geq \sup_{y\in Y}\inf_{x\in X} P(x,y); \tag{16}$$

(Player 1 declares first vs player 2 declares first.) Von Neumann [107] identified structural conditions on the payoff function to have a saddle point (17), in which case equality holds in (15); see also Kakutani's reference to [108] in [66].

Theorem 2.5 (convex/concave min-max)**.** *If $X\subset\mathbf{R}^m$ and $Y\subset\mathbf{R}^n$ are compact and convex, then equality holds in* (16) *provided for each $(x_0,y_0)\in X\times Y$ both functions $x\in X\longmapsto P(x,y_0)$ and $y\in Y\longmapsto -P(x_0,y)$ are convex. (In fact, convexity of all sublevel sets of both functions is enough.)*

Proof. Let

$$x_b(y)\in \arg\min_{x\in X} P(x,y) \qquad y_b(x)\in \arg\max_{y\in Y} P(x,y)$$

denote the best responses of P1 and P2 to each other's strategies y and x. Note x_b and y_b are continuous if the convexity and concavity assumed of the payoff function are both strict. In that case, Brouwer's theorem asserts the the function $y_b\circ x_b : Y\to Y$ has a fixed point y_0. Setting $x_0 = x_b(y_0)$, since $y_0 = y_b(x_0)$ we have found a saddle point

$$\inf_{x\in X}\max_{y\in Y} P(x,y) \leq \max_{y\in Y} P(x_0,y) = P(x_0,y_0) = \min_{x\in X} P(x,y_0) \leq \sup_{y\in Y}\min_{x\in X} P(x,y) \tag{17}$$

of the payoff function, which proves equality holds in (16). If the convexity and concavity of the payoff function are not strict, apply the theorem just proved to the perturbed payoff $P_\epsilon(x,y) = P(x,y) + \epsilon(|x|^2 - |y|^2)$ and take the limit $\epsilon\to 0$. □

e.g. Expected payoff [107]: von Neumann's original example of a function to which the theorem and its conclusion applies is the expected payoff $P(x,y) = \sum_{i=1}^m\sum_{j=1}^n p_{ij}x_iy_j$ of *mixed* or *randomized* strategies x and y for a game in which P1 and P2 each have only finitely many pure strategies, and the payoff corresponding to strategy $1\leq i\leq m$

and $1 \leq j \leq n$ is p_{ij}. In this case $X = \{x \in [0,1]^m \mid \sum x_i = 1\}$ and $Y = \{y \in [0,1]^n \mid \sum y_j = 1\}$ are standard simplices of the appropriate dimension, whose vertices correspond to the pure strategies.

2.3. **Relevance to optimal transport: Kantorovich-Koopmans duality.** Infinite dimensional versions of von-Neumann's theorem can also be formulated where X and Y lie in Banach spaces; they are proved using Schauder's fixed point theorem instead of Brouwer's. A payoff function germane to optimal transportation is defined on the strategy spaces $X = \{0 \leq \gamma \; on \; M^+ \times M^-\}$ and $Y = \{(u,v) \in L^1(M^+,\mu^+) \oplus L^1(M^-,\mu^-)\}$ by

$$P(\gamma,(u,v)) = \int_{M^+\times M^-} (c(x,y) + u(x) + v(y))\, d\gamma(x,y) - \int_{M^+} u d\mu^+ - \int_{M^-} v d\mu^-.$$

Note the bilinearity of P on $X \times Y$. Since

$$\inf_{\gamma\in X} P(\gamma,(u,v)) = \begin{cases} -\infty & \text{unless } (-u,-v) \in Lip_c, \\ 0 - \int u d\mu^+ - \int v d\mu^- & \text{otherwise,} \end{cases}$$

the Kantorovich-Koopmans dual problem is recovered from the version of the game in which P2 is compelled to declare his strategy first:

$$\sup_{(u,v)\in Y} \inf_{\gamma\in X} P(\gamma,(u,v)) = \sup_{(-u,-v)\in Lip_c} \int (-u) d\mu^+ + \int (-v) d\mu^-. \tag{18}$$

On the other hand, rewriting

$$P(\gamma,(u,v)) = \int_{M^+\times M^-} c d\gamma + \int u(d\gamma - d\mu^+) + \int v(y)(d\gamma(x,y) - d\mu^-(y))$$

we see

$$\sup_{(u,v)\in Y} P(\gamma,(u,v)) = \begin{cases} +\infty & \text{unless } \gamma \in \Gamma(\mu^+,\mu^-) \\ \int_{M^+\times M^-} c\, d\gamma & \text{if } \gamma \in \Gamma(\mu^+,\mu^-). \end{cases} \tag{19}$$

Thus the primal transportation problem of Kantorovich and Koopmans

$$\inf_{\gamma\in X} \sup_{(u,v)\in Y} P(\gamma,(u,v)) = \inf_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{M^+\times M^-} c\, d\gamma$$

corresponds to the version of the game in which P1 declares his strategy first. The equality between (18) and (19) asserted by an appropriate generalization of von Neumann's theorem implies the duality (15):

$$\min_{\gamma\in\Gamma(\mu^+,\mu^-)} \int_{M^+\times M^-} c d\gamma = \sup_{(-u,-v)\in Lip_c} \int_{M^+} (-u) d\mu^+ + \int_{M^-} (-v) d\mu^-.$$

2.4. **Characterizing optimality by duality.** The following theorem can be deduced as an immediate corollary of this duality. We may think of the potentials u and v as being Lagrange multipliers enforcing the constraints on the marginals of γ; in the economics literature they are interpreted as shadow prices which reflect the geographic variation in scarcity or abundance of supply and demand. The geography is encoded in the choice of cost.

Theorem 2.6 (Necessary and sufficient conditions for optimality)**.** *The existence of $-(u,v) \in Lip_c$ such that γ vanishes outside the zero set of the non-negative function $k(x,y) = c(x,y) + u(x) + v(y) \geq 0$ on $M^+ \times M^-$ is necessary and sufficient for the optimality of $\gamma \in \Gamma(\mu^+,\nu^-)$ with respect to $c \in C(M^+ \times M^-)$.*

Corollary 2.7 (First and second order conditions on potentials)**.** *Optimality of γ and (u, v) implies $Dk = 0$ and $D^2k \geq 0$ at any point $(x, y) \in \operatorname{spt} \gamma$ where these derivatives exist. In particular, $D_x[c(x, y) + u(x) + v(y)] = 0$ and $D_x^2[c(x, y) + u(x) + v(y)] \geq 0$ holds γ-a.e., and likewise for y-derivatives.*

e.g. Consider the special case of the bilinear cost: $c(x, y) = -x \cdot y$. Here the first and second order conditions of the corollary become

$$y = Du(x) \qquad \text{and} \qquad D^2u(x) \geq 0,$$

suggesting y is the graph of the gradient of a convex function. In this case, convexity of u guarantees D^2u is defined a.e. with respect to Lebesgue measure, by Alexandrov's theorem.

2.5. **Existence of optimal maps and uniqueness of optimal measures.** More generally, we claim u inherits Lipschitz and semiconvexity bounds (23)–(24) from $c(x, y)$, which guarantee the existence of x-derivatives in the preceding corollary — at least Lebesgue almost everywhere. This motivates the following theorem of Gangbo [57] and Levin [80]; variations appeared independently in Caffarelli [16], Gangbo and McCann [58] [59], and Rüschendorf [121] [122] at around the same time, and subsequently in [89].

Definition 2.8 (Twist conditions)**.** *A function $c \in C(M^+ \times M^-)$ differentiable with respect to $x \in M^+$ is said to be* twisted *if*

$$(20) \quad (\mathrm{A1})^+ \quad \forall\, x_0 \in M^+, \text{ the map } y \in M^- \longmapsto D_x c(x_0, y) \in T_x^* M^+ \text{ is one-to-one.}$$

For $(x, p) \in T^*M^+$ denote the unique $y \in M^-$ solving $D_x c(x, y) + p = 0$ by $y = Y(x, p)$ when it exists. When the same condition holds for the cost $\tilde{c}(y, x) := c(x, y)$, we denote it by $(\mathrm{A1})^-$. When both c and $\tilde{c}$ satisfy $(\mathrm{A1})^+$, we say the cost is bi-twisted*, and denote this by* (A1)*.*

Theorem 2.9 (Existence of Monge solutions; uniqueness of Kantorovich solutions)**.** *Fix Polish probability spaces $(M^\pm, \mu^\pm)$ and assume M^+ is a n-dimensional manifold and $d\mu^+ \ll d^n x$ is absolutely continuous (in coordinates). Let $c \in C(M^+ \times M^-)$ differentiable with respect to $x \in M^+$ satisfy the twist condition* (20) *and assume $D_x c(x, y)$ is bounded locally in $x \in M^+$ uniformly in $y \in M^-$. Then, there exists a locally Lipschitz (moreover, c-convex, as in Definition 2.10) function $u : M^+ \to \mathbf{R}$ such that*

a) $G(x) := Y(x, Du(x))$ pushes μ^+ forward to μ^-;

b) this map is unique, and uniquely solves Monge's minimization problem (4)*;*

c) Kantorovich's minimization (11) *has a unique solution γ;*

d) $\gamma = (id \times G)_\# \mu^+$.

Definition 2.10 (c-convex)**.** *A function $u : M^+ \longrightarrow \mathbf{R} \cup \{+\infty\}$ (not identically infinite) is* c-convex *if and only if $u = (u^{\tilde{c}})^c$, where*

$$(21) \qquad u^{\tilde{c}}(y) = \sup_{x \in M^+} -c(x, y) - u(x) \qquad \text{and} \qquad v^c(x) = \sup_{y \in M^-} -c(x, y) - v(y).$$

Remark 2.11 (Legendre-Fenchel transform and convex dual functions)*.* When $c(x, y) = -\langle x, y \rangle$, then $u^{\tilde{c}}(y)$ is manifestly convex: it is the Legendre-Fenchel transform or convex dual function of $u(x)$. In this case, $(u^{\tilde{c}})^c$ is well-known to yield the lower semicontinuous convex hull of the graph of u, so that $u = (u^{\tilde{c}})^c$ holds if and only if u is already lower semicontinuous and convex. More generally, we interpret the condition $u = u^{\tilde{c}c}$ as being the correct adaptation of the notion of convexity to the geometry of the cost function c.

Sketch of proof of Theorem 2.9. The key idea of the proof is to establish existence of a maximizer $-(u,v) \in Lip_c$ of (15) with the additional property that $(u,v) = (v^c, u^{\tilde{c}})$. Differentiability of $u = u^{\tilde{c}c}$ on a set dom Du of full $d\mu^+ \ll d^n x$ measure then follows from Rademacher's theorem and Lemma 3.1. The map $G(x) := Y(x, Du(x))$ is well-defined on dom Du by the twist condition (assuming the supremum (21) defining $(u^{\tilde{c}})^c(x)$ is attained). Corollary 2.7 shows any minimizer γ vanishes outside the graph of this map, and it then follows easily that $\gamma = (id \times G)_{\#}\mu^+$ and hence γ is uniquely determined by u [2]. Conversely any other c-convex $\tilde{u}$ for which $\tilde{G}(x) = Y(x, D\tilde{u}(x))$ pushes μ^+ forward to μ^- can be shown to maximize the dual problem by checking that $\tilde{\gamma} = (id \times \tilde{G})_{\#}\mu^+$ vanishes outside the support of $\tilde{G}$. Thus $\tilde{\gamma} = \gamma$ and $\tilde{G} = G$ holds μ^+-a.e.

To extract the desired $-(u,v) \in Lip_c$ from a maximizing sequence $-(u_k, v_k)$ requires some compactness. (This would come from the convexity of u_k and v_k in case $c(x,y) = -x \cdot y$ via the Blaschke selection theorem.) Observe $-(u,v) \in Lip_c$ implies

$$u(x) \geq \sup_{y \in M^-} -c(x,y) - v(y) =: v^c(x).$$

Moreover, $-(v^c, v) \in Lip_c$ and $-(v^c) \geq -u$ can only increase the value of the objective functional relative to $-(u,v)$. Thus $-(v^c, v)$ is a better candidate for a maximizer than $-(u,v)$. Repeating the process shows $-(v^c, v^{c\tilde{c}})$ and $-(v^{c\tilde{c}c}, v^{c\tilde{c}}) \in Lip_c$ are better still, since $(v^c)^{\tilde{c}} \leq v$ and $(v^{c\tilde{c}})^c \leq v^c$ by the same logic. On the other hand, starting from $v^{c\tilde{c}} \leq v$, the negative coefficient in definition (21) implies the opposite inequality $(v^{c\tilde{c}})^c \geq v^c$. Thus $v^{c\tilde{c}c} = v^c$ quite generally. (This is precisely analogous to the fact that the second Legendre transform u^{**} does not change a function $u = v^*$ which is already convex and lower semicontinuous; see Remark 2.11.)

Replacing a maximizing sequence $-(u_k, v_k)$ with $-(v_k^c, v_k^{c\tilde{c}})$ therefore yields a new maximizing sequence at least as good which moreover consists of c-convex functions. Lemma 3.1 shows this new family is locally equi-Lipschitz, hence we only need local boundedness for the Arzela-Ascoli theorem to yield a limiting maximizer $-(u,v)$, which will in fact be c-convex, though we can also replace it by $-(v^c, v^{c\tilde{c}})$ just to be sure. Local boundedness also follows from Lemma 3.1, after fixing $x_0 \in \text{spt } \mu^+ \subset M^+$ and replacing $-(u_k, v_k)$ by $-(u_k - \lambda_k, v_k + \lambda_k)$ with $\lambda_k = u_k(x_0)$. This replacement does not change the value of the objective functional (18), yet ensures that $u(x_0) = 0$. □

3. Methods for obtaining regularity of optimal mappings

Given mines and factories $(M^{\pm}, \mu^{\pm})$ and a cost function $c \in C(M^+ \times M^-)$, in the preceding section we found conditions which guarantee the existence and uniqueness of a map $G(x) = Y(x, Du(x))$ such that $G_{\#}\mu^+ = \mu^-$ with $u = u^{\tilde{c}c}$, ie. c-convex. Under the same conditions, the map G is the unique minimizer of Monge's problem (4). The space M^+ was assumed to be an n-dimensional manifold, and the following twist hypothesis $(A1)^+$, equivalent to (20), was crucial to specifying $Y(x, \cdot)$:

(22) $\quad \forall\ y_1 \neq y_2 \in M^-$ assume $x \in M^+ \longmapsto c(x, y_1) - c(x, y_2)$ has no critical points.

Notice, however, that (22) cannot be satisfied by any cost function which is *differentiable* throughout a compact manifold M^+. In case $M^+ = \mathbf{S}^n$, Monge solutions do not generally exist [60], but criteria are given in [27] [2] which guarantee uniqueness of the Kantorovich minimizer. On the other hand, it is an interesting open problem to find a criterion on $c \in C^1(M^+ \times M^-)$ which guarantees uniqueness of Kantorovich solutions for all $\mu^{\pm} \in L^1(M^{\pm})$ in more complicated topologies, such as the torus $M^{\pm} = \mathbf{T}^n$ for example. Here differentiability of the cost function is crucial; for costs such as Riemannian distance squared, the desired uniqueness is known [29] [97], but the cost fails to be differentiable at the cut locus.

3.1. **Rectifiability: differentiability almost everywhere.** The current section is devoted to reviewing methods for exploring the smoothness properties of the optimal map G found above, or equivalently of its c-convex potential u. The following lemma shows that all c-convex functions inherit Lipschitz and semiconvexity properties directly from the cost function c; it has already been exploited to prove Theorem 2.9.

Lemma 3.1 (Inherent regularity of c-convex functions)**.** *If $u = u^{\tilde{c}c}$ and $c(\cdot, y) \in C^k_{loc}(M^+)$ for each $y \in M^-$, then $k = 1$ implies* (23) *and $k = 2$ implies* (24)*:*

$$|Du(x)| \le \sup_{y \in M^-} |D_x c(x,y)| \qquad \text{(local Lipschitz regularity);} \tag{23}$$

$$D^2 u(x) \ge \inf_{y \in M^-} -D^2_{xx} c(x,y) \qquad \text{(semiconvexity).} \tag{24}$$

Similarly, c-cyclically monotone sets $S \subset M^+ \times M^-$ turn out to be contained in Lipschitz submanifolds of dimension $n = dim M^\pm$ when the cost function is non-degenerate (25). The following recent theorem of McCann-Pass-Warren [100] combines with Rademacher's theorem — which asserts the differentiability Lebesgue a.e. of Lipschitz functions — to give a simple tool for establishing that $f^+(x) = |\det(DG(x))| f^-(G(x))$ holds f^+-a.e.

Theorem 3.2 (Rectifiability of optimal transport [100])**.** *Assume $M^\pm$ are n-dimensional manifolds, at least in a neighbourhood U of $(x_0, y_0) \in M^+ \times M^-$, where $c \in C^2(U)$ and*

$$\text{(A2)} \quad det D^2_{x^i y^j} c(x_0, y_0) \neq 0. \tag{25}$$

If $S \subset M^+ \times M^-$ is c-cyclically monotone, then $S \cap V$ lies in an n-dimensional Lipschitz submanifold, for some neighbourhood $V \subset U$ of (x_0, y_0).

In view of Theorem 2.2, this conclusion applies either to the graph $S = Graph(G)$ of any optimal map (4) or the support $S = \text{spt}\,(\gamma)$ of any optimal measure (11) in the transportation problem.

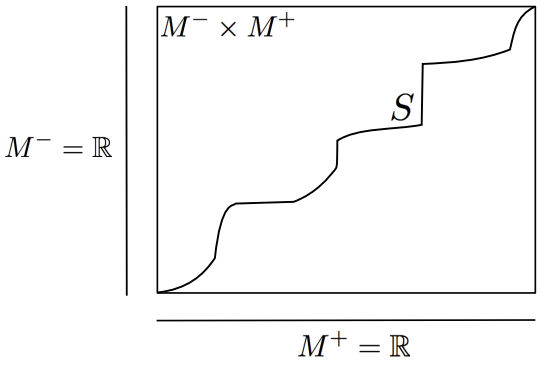

Figure 1: Optimizers have locally monotone support S in the plane.

Motivation: on the line $M^\pm = \mathbf{R}$, without further assumptions on c or $\mu^\pm$, a transport map may not exist nor be monotone, yet the theorem above says that even so all pieces of *spt* (γ) lie along along Lipschitz arcs in the plane. These curves will actually be locally monotone — non-decreasing or non-increasing depending on the sign of $D^2_{xy} c(x_0, y_0)$; see Figure 1.

Idea of proof. Introduce the notation $b = -c$. In case $c(x, y) = -x \cdot y$ on $M^\pm = \mathbf{R}^n$, monotonicity asserts for all $(x_0, y_0), (x_1, y_1) \in S$ that $\Delta x = x_1 - x_0$ and $\Delta y = y_1 - y_0$

satisfy

$$0 \leq \langle \Delta x, \Delta y \rangle = \left\langle \frac{\Delta z - \Delta w}{\sqrt{2}}, \frac{\Delta z + \Delta w}{\sqrt{2}} \right\rangle$$

where

$$(z, w) := (\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}). \tag{26}$$

This implies that $|\Delta w|^2 \leq |\Delta z|^2$ meaning $w = w(z)$ has Lipschitz constant 1 as a graph over $z \in \mathbf{R}^n$. Equivalently, S has Lipschitz constant 1 as a graph over the diagonal in $M^+ \times M^-$. This special case was established by Alberti and Ambrosio [3], using an argument of Minty [102].

For more general costs $b = -c$ and any $\epsilon > 0$, the non-degeneracy (25) implies the existence of new coordinates $\tilde{y} = \tilde{y}(y)$ on M^- in a neighbourhood of y_0 such that $\tilde{b}(x, \tilde{y}(y)) = b(x, y)$ satisfies $|D^2_{x\tilde{y}} \tilde{b}(x, \tilde{y}) - I| < \epsilon$ in a neighborhood $\tilde{V}$ of $(x_0, \tilde{y}_0)$ which is convex in coordinates.

Now, $\langle \Delta x, \Delta \tilde{y} \rangle \geq -\epsilon |\Delta x| \Delta \tilde{y}|$ follows from

$$0 \leq \tilde{b}(x_0, \tilde{y}_0) + \tilde{b}(x_1, \tilde{y}_1) - \tilde{b}(x_0, \tilde{y}_1) - \tilde{b}(x_1, \tilde{y}_0) = D^2_{x\tilde{y}} \tilde{b}(x^*, \tilde{y}^*)(x_1 - x_0)(\tilde{y}_1 - \tilde{y}_0) \tag{27}$$

and the change of variables analogous to (26) yields

$$|\Delta z|^2 - |\Delta w|^2 \geq -\epsilon |\Delta w - \Delta z||\Delta w + \Delta z| \geq -\epsilon \left(|\Delta w|^2 + |\Delta z|^2 \right).$$

Thus $(1+\epsilon)|\Delta z|^2 \geq (1-\epsilon)|\Delta w|^2$, which shows $w = w(z)$ is again a Lipschitz function of $z \in \mathbf{R}^n$ in the chosen coordinates. □

3.2. **From regularity a.e. to regularity everywhere.** The regularity results discussed so far — Lipschitz continuity of the potential u, and of $Graph(G) \subset M^+ \times M^-$ rather than of the map $G(x) = Y(x, Du(x))$ itself — required no hypotheses on the probability measures $\mu^- = G_{\#}\mu^+$. To address the continuity, differentiability, and higher regularity everywhere for the map $G : M^+ \longrightarrow M^-$ is a much more delicate issue which certainly requires further hypotheses on the data $\mu^{\pm}$ and c. For example, if spt μ^- is connected but spt μ^+ is not, then G cannot be continuous. The same reasoning makes it clear that ellipticity of the Monge-Ampère equation (6) cannot be non-degenerate for all convex solutions; regularity must propagate from boundary conditions since the *purely local* effect of the equation is insufficent to conclude $u \in C^1_{loc}$. It is often easier to work with the scalar potential u rather than the mapping G; we shall see this reduces the problem to a question in the theory of second-order, fully-nonlinear, degenerate-elliptic partial differential equations (33) generalizing the Monge-Ampère equation. However, this question was answered first in the special case $c(x, y) = -x \cdot y$ corresponding to the case (6) by Delanoë in the plane $n = 2$ [38], and by Caffarelli and Urbas in higher dimensions $M^{\pm} = \mathbf{R}^n$ [18] [17] [19] [135].

Remark 3.3. Note for $c \in C^{k+1}(M^+ \times M^-)$ that $u \in C^{k+1}$ implies $G \in C^k$ by the following remark. Whereas the twist condition $(A1)^+$ asserts that the definition $Y(x, p)$ by $D_x c(x, Y(x, p)) + p = 0$ is unambiguous, non-degeneracy (A2) allows the implicit function theorem to be applied to conclude C^k smoothness of $Y(x, p)$ (where defined).

3.3. **Regularity methods for the Monge-Ampère equation; renormalization.** There are several methods for obtaining regularity results for convex solutions of the Monge-Ampère equation. The first to be discussed here is the *continuity method*, used for example by Delanoë '91 and Urbas '97. This approach requires relatively strong assumptions on the smoothness of the measures $d\mu^{\pm} = f^{\pm} dVol$ and on the convexity and smoothness of their domains $M^{\pm} \subset \mathbf{R}^n$. When it applies, it yields global regularity

of the resulting potential up to the boundary of ∂M^+ from the same fixed point argument which shows a solution exists. The second method is the *renormalization method* pioneered by Caffarelli '92-'96, which starts from the unique (weak) solution to the 2nd boundary value problem and uses affine invariance of the equation to blow-up the solution near a putative singularity and derive a contradiction in the limit. This method is quite flexible: it has the advantage of yielding certain conclusions under weaker assumptions on the data, and has therefore proven useful for addressing such phenomena as the free boundary which arises in partial transport problems, where the densities $f^\pm$ need not be continuous and are constrained but not specified a priori [20]. Using this method, Caffarelli '92 was able to prove the following regularity result on the *interior* M^+_{int} of $M^+ \subset \mathbf{R}^n$:

Theorem 3.4 (Local regularity [18])**.** *Fix $c = -x \cdot y$ and let $M^- \subset \mathbf{R}^n$ be convex. Assume $\mu^\pm = f^\pm dx$ with $\log f^\pm \in L^\infty(M^\pm)$. If $\log f^\pm \in W^{k,\infty}_{loc}(M^\pm_{int})$, there exists $\alpha \in (0,1)$ (depending only on n, k and the bounds on $\log f^\pm$) such that $u \in C^{k+1,\alpha}_{loc}(M^+_{int})$, where u is the convex function with $Du(M^+) \subset M^-$ such that $Du_\# \mu^+ = \mu^-$.*

Remark 3.5 (Degenerate ellipticity)*.* As shown also in [18], when one drops the convexity assumption on M^- the gradient map may be discontinuous at interior points. This goes in hand with the claim made in the previous subsection that regularity (even in the interior) must propagate from the boundary.

Remark 3.6 (Local versus global regularity)*.* This is a local regularity result. Global regularity (up to ∂M^+) requires both domains $\partial M^\pm$ to be strongly convex and smooth (Caffarelli '96) [19]. Here *strong* convexity means the principal curvatures (or second fundamental form) of the domain boundaries should be positive-definite.

Remark 3.7 (Higher regularity via uniformly elliptic linearization)*.* The cases of primary relevance are $\log f^\pm$ merely bounded and measurable ($k = 0$), and $\log f^\pm$ also locally Lipschitz ($k = 1$). Once $u \in C^{2,\alpha}_{loc}$ has been deduced from these assumptions, higher regularity in the interior of $M^\pm$ follows from uniform ellipticity of the Monge-Ampère equation (6):

$$0 = \log(\det(D^2u(x) + \tau D^2 w(x))) + \log(f^-(Du(x) + \tau Dw(x))) - \log(f^+(x)). \tag{28}$$

For example, linearizing this equation at $\tau = 0$ yields the equation

$$0 = Tr(D^2u(x)^{-1} D^2 w(x)) + D \log f^-|_{Du(x)} \cdot Dw(x), \tag{29}$$

which must be satisfied by spatial derivatives $w = D_{x^i} u$ of u. Convexity combines with $u \in C^{2,\alpha}_{loc}$ and the equation (6) itself to bound $\|\Lambda^{n-1} f^-/f^+\|^{-1}_{L^\infty} \le D^2u(x) \le \Lambda$ on compact subsets of M^+_{int}. The derivatives of u thus satisfy a uniformly elliptic linear equation (29) with Hölder continuous coefficients, so Schauder estimates [62] and bootstrapping yield as much as regularity as can be expected when $k \ge 2$.

3.4. **The continuity method (schematic).** (cf. Delanoë '91, Urbas '97): To apply the continuity method, we assume $M^\pm \subset \mathbf{R}^n$ are smooth and strongly convex, and $\log f^\pm \in C^{2,\alpha}(M^\pm) \cap L^\infty$.

Choose a dilation by $\epsilon > 0$ sufficiently small and translation $M_0 = G_0(M^+)$ of M^+ by $x_0 \in \mathbf{R}^n$ such that $M_0 \subset\subset M^-$. Let $\mu_0 := (G_0)_\# \mu^+$ be the push-forward of μ^+ through the corresponding dilation and translation $G_0(x) = \epsilon x - x_0$. Notice that G_0 is the gradient of the smooth convex function $u_0(x) = \epsilon|x|^2/2 - x_0 \cdot x$, and as such gives the optimal map between μ^+ and μ_0. The idea behind the continuity method is to construct a family of target measures $d\mu_t = f_t dVol$ interpolating between $d\mu_0$ and $d\mu_1 := d\mu^-$, and to study the set T of $t \in [0,1]$ for which the optimal transportation problem of

Brenier admits a solution with a convex potential $u_t \in C^{2,\alpha}(M^+)$. The interpolating measures must be constructed so that the $C^{2,\alpha}(M^\pm) \cap L^\infty$ norms of $\log f_t$, and the strong convexity and smoothness of $M_t = \text{spt } f_t$, can be quantified independently of $t \in [0,1]$. We then hope to show $T \subset [0,1]$ is both open and closed. If so, it must exhaust the entire interval (since $0 \in T$), therefore $1 \in T$ as desired.

Closed: To show closedness of this set requires an a priori estimate of the form $\|u_t\|_{C^{2,\alpha}(M^+)} \leq C(\|\log f^\pm\|_{C^{2,\alpha}(M^\pm)}, \|\partial M^\pm\|_{C^{2,\alpha},\text{strong convexity}})$ for any smooth solution $u_t \in C^4(M^+)$ of the 2nd boundary value problem

$$\det D^2 u_t(x) = \frac{f_0(x)}{f_t(Du_t(x))} \qquad \text{with} \qquad Du_t(M^+) \subset M_t. \tag{30}$$

Such estimates are delicate, but can be obtained by differentiating the equation twice, and constructing barriers. Once obtained, they imply that if $t_k \in T$ and $t_k \to t_\infty$ then $t_\infty \in T$ also; the corresponding solutions u_{t_k} belong to $C^4(M^+)$ as in Remark 3.7.

Open: The fact that T is open is shown using an implicit function theorem in Banach spaces. This requires knowing that the linearized operator is invertible (ie. uniformly elliptic), and can be solved for the relevant boundary conditions.

For $u \in C^{2,\alpha}(M^+)$ we have already argued the uniform ellipticity of the linearization (29). To linearize the boundary conditions $Du_t(M^+) \subset M_t$, introduce a sufficiently smooth and strongly convex function $h : \mathbf{R}^n \to \mathbf{R}$ whose level sets $M_t = \{y \in \mathbf{R}^n \mid h(y) \leq t\}$ give the domains $M_t := \text{spt } f_t$, and rewrite the non-linear boundary condition in the form $h(Du_t(x)) \leq t$ with equality on ∂M^+. Linearizing this in u yields the boundary condition of the linear equation for w:

$$Dh(Du_t(x)) \cdot Dw(x) = 0 \qquad \text{on } \partial M^+. \tag{31}$$

For the linear problem (29) to be well-posed, we need a *uniformly non-tangential* prescribed gradient for w on ∂M^+. Since Dh parallels the normal $\hat{n}_{M_t}$ to M_t, this amounts to the uniform obliqueness estimate

$$\hat{n}_{M_t}(Du_t(x)) \cdot \hat{n}_{M^+}(x) \geq \delta > 0 \quad \text{(obliqueness)}$$

provided by Urbas [135], with δ depending only on coarse bounds for the data. This concludes the sketch that $T \subset [0,1]$ is open: well-posedness of the linear problem (29)–(31) when $t = t_0 \in T$ implies the existence of solutions $u_t \in C^{2,\alpha}(M^+)$ to the nonlinear problem (30) for any t close enough to t_0.

Both approaches (renormalization and continuity method) have been extended in recent years to more general costs, and this will be the topic of the next few lectures.

4. Regularity and counterexamples for general costs

4.1. **Examples.** The development of a regularity theory for general cost functions satisfying appropriate hypotheses on compact domains $M^\pm \subset \mathbf{R}^n$ began with the work of Ma, Trudinger and Wang [89]. Prior to that there were regularity results only for a few special costs, such as:

Example 4.1 (Bilinear cost)**.** *$c(x,y) = -x \cdot y$ or equivalently $c(x,y) = |x-y|^2/2$ [38] [18] [19] [135], and its restriction to $M^\pm = \partial B_1(0)$ in $\mathbf{R}^n$ (Gangbo and McCann [60]);*

Example 4.2 (Logarithmic cost; conformal geometry and reflector antenna design)**.** *$c(x,y) = -\log|x-y|$ appearing in conformal geometry (cf. Viaclovsky's review [137]), and its restriction to the Euclidean unit sphere, which is relevant to reflector antenna design (Glimm and Oliker [63], X.-J. Wang [140][141]) and helped to inspire Wang's subsequent collaborations [89] [133] [134] [83] with Trudinger, Ma, and Liu.*

In the wake of Ma, Trudinger and Wang's [89] results, many new examples have emerged of cost functions which satisfy [84] [39] [53] [55] [54] [71] [73] [74] [78] [79] [77] [81] [86] [89] — or which violate [89] [85] — their sufficient conditions (A0)–(A4) and $(A3)_s$ for regularity from §4.4 below — not to mention the subsequent variants (A3) and (B3) introduced by Trudinger and Wang [133] and Kim and McCann [73] respectively, on the crucial condition $(A3)_s$. Among the most interesting of these are the geometrical examples and counterexamples of Loeper:

Example 4.3 (Sphere)**.** *$c(x,y) = \frac{1}{2}d^2_{\mathbf{S}^n}(x,y)$ on the round sphere satisfies* $(A3)_s$ [84] *(and* (B3) [74]*);*

Example 4.4 (Saddle)**.** *$c(x,y) = \frac{1}{2}d^2_M(x,y)$ on hyperbolic space $M = \mathbf{H}^n$ violates* $(A3)_s$ *(and* (A3) [85]*) — as does the Riemannian distance squared cost on any Riemannian manifold $M^\pm = M$ which has (at least one) negative sectional curvature at some point $x \in M$.*

4.2. **Counterexamples to the continuity of optimal maps.** For any cost function which violates (A3), Loeper went further to show there are probability measures $d\mu^\pm = f^\pm d^n x$ with smooth positive densities bounded above and below — so that $\log f^\pm \in C^\infty(M^\pm)$ — for which the unique optimal map $G : M^+ \longrightarrow M^-$ is discontinuous [85]. Let's see why this is so for the quadratic cost given on either the hyperbolic plane or a saddle surface as in Example 4.4.

Consider transportation from the uniform measure μ^+ on a sufficiently small ball to a target measure consisting of three point masses $\mu^- = \frac{1}{3}\sum_{i\le 3}\delta_{y_i}$ near the center of the ball, choosing y_2 to be the midpoint of y_1 and y_3. In this case Theorem 2.9 provides constants $v_1, \dots, v_3$ and a c-convex function

$$u(x) = \max\{u_i(x) \mid i = 1,2,3\} \qquad \text{where} \qquad u_i(x) = -c(x,y_i) - v_i, \tag{32}$$

such that the optimal map $G_\#\mu^+ = \mu^-$ satisfies $G^{-1}(y_i) = \{x \in M^+ \mid u(x) = u_i(x)\}$. We interpret $-v_i$ to be the value of the good at the potential destination $y_i \in \operatorname{spt}\mu^-$; the producer at $x \in M^+$ will ship his good to whichever target point y_i provides the greatest value after transportation costs are deducted (32); here the values $v_1, \dots, v_3$ are adjusted to balance supply with demand, so that each of the three regions $G^{-1}(y_i)$ contains 1/3 of the mass of μ^+. For the Euclidean distance-squared cost these three regions are easily seen to be convex sets, while for the spherical distance-squared they remain connected. For the hyperbolic distance-squared, however, the 'middle' region $G^{-1}(y_2)$ consists of two disconnected components, near opposite sides of the ball $\operatorname{spt}\mu^+$ (see figure below, or for instance Figure 1 of [73]). This disconnectedness is the hallmark of costs for which (A3) fails, and allowed Loeper to construct counterexamples to the continuity of optimal mappings as follows.

In the preceding discussion, μ^- was not given by a smooth positive density; still it can be approximated by a sequence of measures $\mu^-_\epsilon := \mu^- * \eta_\epsilon$ which are. Now consider the reverse problem of transporting μ^-_ϵ to μ^+. Call the optimal map for this new problem $x = G^-_\epsilon(y)$. For $\delta > 0$, taking $\epsilon > 0$ sufficiently small ensures for each $1 \le i \le 3$ that nearly 1/3 of the mass of μ^-_ϵ concentrates near y_i and is mapped into a δ-neighbourhood of $G^{-1}(y_i)$. Intuitively, for δ sufficiently small, this forces a discontinuity of G^-_ϵ which tears the region near y_2 into at least two disconnected components: nearly half of the mass near this point must map to each disconnected component of $G^{-1}(y_2)$; see Figure 2. This construction shows why the distance-squared cost on a hyperbolic or saddle surface cannot generally produce smooth optimal transport maps.

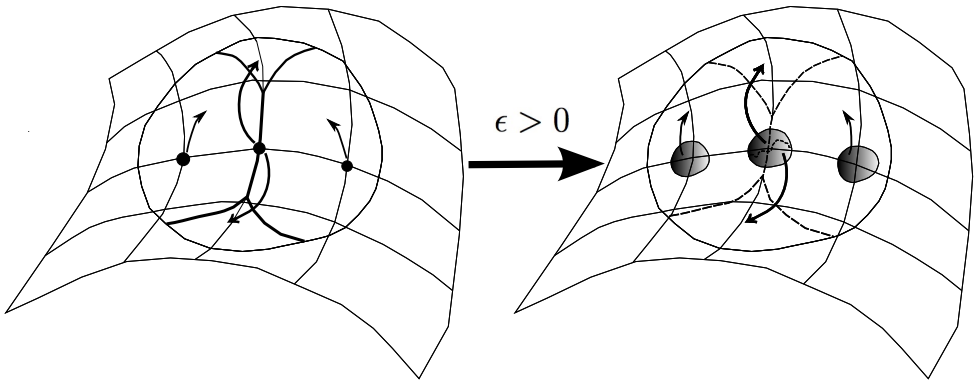

Figure 2: A tear occurs when spreading a triply peaked density uniformly over the saddle.

There is another of obstruction to the continuity of G, namely the convexity (at least when $M^{\pm} = \mathbb{R}^n$) of the support of μ^-. This was shown by Caffarelli [18] with the following elementary example: consider $u : \mathbb{R}^2 \to \mathbb{R}$ given by

$$u(x) = |x_1| + \frac{1}{2}|x|^2, \;\; x = (x_1, x_2).$$

If we consider the cost $c(x, y) = -x \cdot y$, then $y = Du(x)$ gives the optimal transport map between the unit disc (with Lebesgue measure) into two shifted half discs (Figure 3); in particular, the transport map is discontinuous across $\{x_1 = 0\}$.

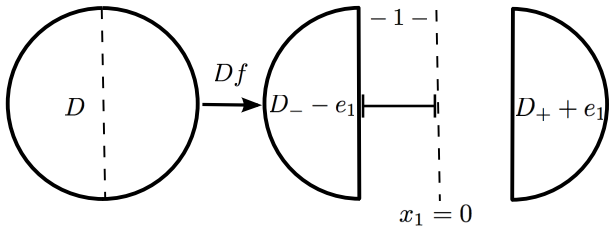

Figure 3. Disconnected targets also produce tears, as do non-convex targets (spt μ^-).

We turn now to conditions which rule out these type of examples, and lead to positive regularity results.

4.3. **Monge-Ampère type equations.** For the quadratic cost, finding a smooth optimal map was equivalent to solving the 2nd boundary value problem for the Monge-Ampere equation (6). Let us now derive the analogous equation for a more general cost, keeping in mind that whatever PDE we end up with cannot generally be better than degenerate-elliptic, since vanishing of $d\mu^- = f^- dVol$ can lead to non-smooth solutions in the interior of the support of $d\mu^+ = f^+ dVol$.

Let us see what specific PDE emerges from the local expression $|\det DG(x)| = f^+(x)/f^-(G(x))$ for $G_{\#}\mu^+ = \mu^-$. Recall from Corollary 2.7 that $D^2_{xx}c(x, G(x)) + D^2u(x) \geq 0$ and $D_x c(x, G(x)) + Du(x) = 0$. Differentiating the latter expression gives a relation

$$D^2_{xx}c(x, G(x)) + D^2_{xy}c(x, G(x))DG(x) + D^2u(x) = 0$$

which can be solved for $DG(x)$ to yield

$$\det\left(D^2u(x) + D^2_{xx}c(x, Y(x, Du(x)))\right) = \left|\det\left(D^2_{xy}c(x, y)\right) \; \frac{f^+(x)}{f^-(y)}\right|_{y=Y(x,Du(x))}. \tag{33}$$

Here we have assumed $(A1)^+$–(A2), the form $G(x) = Y(x, Du(x))$ of the optimal map is from Theorem 2.9, and the boundary condition is $Y(x, Du(x)) \subset M^-$ for all $x \in M^+$. We have arrived as before at a fully-nonlinear second-order equation, whose linearization around any c-convex solution $u = u^{\tilde{c}c}$ is degenerate-elliptic.

4.4. **Ma-Trudinger-Wang conditions for regularity.** Sufficient conditions for the c-optimal map $G : M^+ \longrightarrow M^-$ to be smooth between a pair of smooth bounded probability densities satisfying $\log f^\pm \in C^\infty(M^\pm)$ on compact domains $M^\pm \subset \mathbf{R}^n$ were found by Ma-Trudinger-Wang and Trudinger-Wang [89] [133]. The crucial condition on the cost $c(x, y)$ distinguishing Examples 4.1–4.3 from Example 4.4 above involves a quantity they identified, which other authors have variously dubbed the *Ma-Trudinger-Wang tensor* [139], *c-sectional curvature* [85], or *cross-curvature* [73]; c.f. §6 below. To define it, let us adopt their convention that subscripts such as $c_{i,j} = \partial^2 c/\partial x^i \partial y^j$ and $c_{ij,kl} = \partial^4 c/\partial x^i \partial x^j \partial y^k \partial y^\ell$ indicate iterated derivatives in coordinates, with commas separating derivatives with respect to $x \in M^+$ from those with respect to variables $y \in M^-$. Let $c^{j,i}(x, y)$ denote the inverse matrix to $c_{i,j}(x, y)$.

Definition 4.5 (Cross-curvature)**.** *Given tangent vectors $p \in T_{x_0}M^+$ and $q \in T_{y_0}M^-$, define* $\mathrm{cross}(p, q) := (-c_{ij,kl} + c_{ij,r}c^{r,m}c_{m,kl})\, p^i p^j q^k q^l$. *Here and subsequently, the Einstein summation convention is in effect.*

The conditions assumed by Ma, Trudinger and Wang were the following [89]; our designations $(A3)_s$ and (A3) correspond to their (A3) and (A3w) from [133]:

(A0) $c \in C^4(M^+ \times M^-)$, and for all (x_0, y_0) in the compact set $M^+ \times M^- \subset \mathbf{R}^n \times \mathbf{R}^n$;

(A1) $y \in M^- \longmapsto D_x c(x_0, y)$ and $x \in M^+ \longmapsto D_y c(x, y_0)$ are injective;

(A2) $\det D^2_{x^i, y^j} c(x_0, y_0) = \det(c_{i,j}) \neq 0$;

(A3) $\mathrm{cross}(p, q) \geq 0$ for all $(p, q) \in T_{(x_0,y_0)}M^+ \times M^-$ such that $p^i c_{i,j} q^j = 0$;

(A4) $M^-_{x_0} := D_x c(x_0, M^-) \subset \mathbf{R}^n$ and $M^+_{y_0} := D_y c(M^+, y_0) \subset T^*_{y_0}M^-$ are convex.

Among the variants on (A3) subsequently proposed [73] [86] [53], let us recall the *non-negative cross-curvature* condition [73]:

(B3) $\mathrm{cross}(p, q) \geq 0$ for all $(p, q) \in T_{(x_0,y_0)}M^+ \times M^-$.

The first two conditions above are familiar from Theorems 2.9 and 3.2; (A1) was proposed independently of [89] in [57] [80], while there is an antecedent for (A2) in the economics literature [91]. The last condition (A4) adapts the convexity required by Delanoë, Caffarelli (Theorem 3.4) and Urbas, to the geometry of the cost function $c(x, y)$; when $M^+_{x_0}$ and $M^-_{y_0}$ are smooth and their convexity is strong — meaning the principal curvatures of their boundaries are all strictly positive — we denote it by $(A4)_s$. When inequality (A3) or (B3) holds strictly — and hence uniformly on the compact set $M^+ \times M^-$ — we denote that fact by $(A3)_s$ or $(B3)_s$, respectively.

Remark 4.6. The quadratic cost $c(x, y) = -x \cdot y$ of Brenier satisfies (B3) but not $(A3)_s$. Since we have already seen that (A3) is necessary [85] as well as sufficient for the continuity of optimal maps, the quadratic cost is actually a delicate borderline case. The negative $-c$ of any cost c satisfying $(A3)_s$ — including those of Examples 4.2–4.3 — necessarily violates (A3).

4.5. **Regularity results.** Assume (A0)–(A4) and $\log f^{\pm} \in C^{\infty}(M^{\pm})$. Under the stronger condition $(A3)_s$, Ma, Trudinger and Wang [89] proved the interior regularity of the optimal map G and corresponding c-convex potential $u \in C^{\infty}(M^{+}_{int})$; a flaw in their argument was later repaired in [134] (see [73] for another approach). Substituting strong convexity $(A4)_s$ for $(A3)_s$, but retaining (A3), Trudinger and Wang [133] used the continuity method to establish regularity up to the boundary $u \in C^{\infty}(M^{+})$. Relaxing strong convexity to (A4) in that context is an open problem.

For densities merely satisfying $f^{+}/f^{-} \in L^{\infty}(M^{+} \times M^{-})$, under the strong condition $(A3)_s$, Loeper was able to establish local Hölder continuity of the optimal map — or equivalently $u \in C^{1,\alpha}_{loc}(M^{+}_{int})$ — with explicit Hölder exponent $\alpha = 1/(4n-1)$, using a direct argument [85] that we sketch out below. This exponent was later improved to its sharp value $\alpha = 1/(2n-1)$ by Liu [82]. For the quadratic cost $c(x,y) = -x \cdot y$, the best known estimates [56] for the Hölder exponent α are much worse, and depend on bounds for $\log(f^{+}/f^{-})$. Assuming non-negative cross-curvature (B3) and $(A4)_s$ instead of $(A3)_s$, Figalli, Kim and McCann adapted Caffarelli's renormalization techniques [47] to derive continuity and injectivity of optimal maps but without any Hölder exponent; using one of their arguments, a similar conclusion was obtained by Figalli and Loeper [50] in the special case $n = 2$ assuming only (A3) and $(A4)_s$. Liu, Trudinger and Wang showed that higher regularity then follows from further assumptions on f^{+}/f^{-} in any dimension [83].

Using this theory, regularity results have now been obtained in geometries such as the round sphere [84], perturbations [39] [53] [54], submersions [39] [74] and products [48] thereof, and hyperbolic space [81] [78]. Significant cut-locus issues arise in this context. Loeper and Villani [86] conjecture, and in some cases have proved, that condition $(A3)_s$ on the quadratic cost $c(x,y) = d^2(x,y)$ actually implies convexity of the domain of injectivity of the Riemannian exponential map $\exp_x : T_xM \longrightarrow M$.

4.6. **Ruling out discontinuities: Loeper's maximum principle.** Let us discuss how the condition (A3) rules out the tearing phenomenon which we saw on the saddle surface of Example 4.4.

Discontinuities in the optimal map $G(x) = Y(x, Du(x))$ correspond to locations $x_0 \in M^{+}$ where differentiability of the potential function $u = u^{\tilde{c}c}$ fails, such as locations where the supremum

$$u(x_0) = \sup_{y \in M^{-}} -c(x_0, y) - u^{\tilde{c}}(y) \tag{34}$$

is attained by two or more points $y_0 \neq y_1 \in M^{-}$.

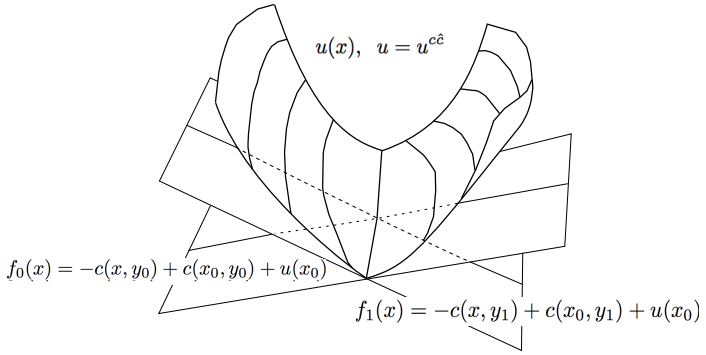

Figure 4. Discontinuous optimal maps arise from distinct supporting hyperplanes.

The set of such y is denoted by $\partial^c u(x_0)$, while the set of such pairs is denoted $\partial^c u \subset M^{+} \times M^{-}$. Unless we can find a continuous curve $t \in [0,1] \longmapsto y_t \in \partial^c u(x_0)$ which

connects y_0 to y_1, it will be possible [85] to construct probability densities satisfying $\log f^{\pm} \in C^\infty(M^{\pm}) \cap L^\infty$ with a discontinuous optimal map as in §4.2 above. But there are not many possibilities to have such a curve.

In the classical case $c(x,y) = -x \cdot y$ (see Figures 4 and 5), $\partial^c u = \partial u$ and we have a convex function u with two different supporting planes at x_0.

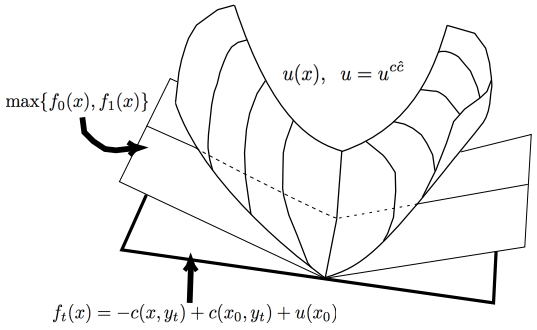

Figure 5. Can one c-affine support at x_0 be rotated to another, without exceeding u?

In particular one may continuously rotate the first plane about the point $(x_0, u(x_0))$ without ever crossing the graph of u until it agrees with the second plane giving a one-parameter family of supporting planes to u at the same point. This way one sees that $\partial u(x_0)$ contains a "segment" $\{y_t\}_{t\in(0,1)}$. In this special case $y_t = (1-t)y_0 + ty_1$ where y_0 and y_1 are the slopes of the original supporting hyperplanes. In the general case, the corresponding local picture forces the "c-segment" $\{y_t\}_{t\in(0,1)}$ given by

$$D_x c(x_0, y_t) = (1-t) D_x c(x_0, y_0) + t D_x c(x_0, y_1) \tag{35}$$

to be our only hope for a continuous path connecting y_0 to y_1 in $\partial^c u(x_0)$.

Now, were G to exhibit a discontinuity, this construction suggests G has to transport a very small mass around x_0 into a set with very large mass, which would give a contradiction given the constraint $G_\# \mu^+ = \mu^-$ and the assumptions on $\mu^{\pm}$. This is indeed the case, at least under the stronger assumption $(A3)_s$, as we shall see below (Proposition 4.12 and Theorem 4.11).

All of this is of course contingent on whether the entire family of functions $\{g_t\}_t$ lie below $u(x)$, which might not be true for an arbitrary cost c; see Figure 5. Indeed, Loeper's key observation is that the Ma-Trudinger-Wang condition (A3) is what guarantees that any family of functions $f_t(y) = -c(x, y_t) + c(x_0, y_t)$ with y_t satisfying (35) never goes above $u(x)$. More precisely, it remains below $\max\{f_0(x), f_1(x)\}$.

Theorem 4.7 (Loeper's maximum principle [85] [73])**.** *If* (A0)–(A4) *hold and* $x_0 \in M^+$ *and* $(y_t)_{t\in[0,1]} \subset M^-$ *satisfy* (35)*, then*

$$f(x,t) := -c(x, y_t) + c(x_0, y_t) \le \max\{f(x,0), f(x,1)\} \quad \forall\ (x,t) \in M^+ \times [0,1]. \tag{36}$$

Remark 4.8. ([74]) If in addition, (B3) holds, then $t \in [0,1] \longrightarrow f(x,t)$ is convex .

Loeper's original proof was quite tortuous, relying on global regularity results for optimal transportation already established by Trudinger and Wang [133]. Here we sketch instead a simple, direct proof due to Kim and McCann [73], who later added Remark 4.8. A preliminary lemma gives some insight into the relevance of the cross-curvature.

Lemma 4.9 (A non-tensorial expression for cross-curvature [73])**.** *Assuming* (A0)–(A4), *if* $(x(s))_{-1\leq s\leq 1} \subset M^+$ *and* $(y(t))_{-1\leq t\leq 1} \subset M^-$ *satisfy either*

$$\frac{d^2}{dt^2}\bigg|_{t=0} D_x c(x(0), y(t)) = 0 \qquad \text{or} \qquad \frac{d^2}{ds^2}\bigg|_{s=0} D_y c(x(s), y(0)) = 0,$$

then $\text{cross}(\dot{x}(0), \dot{y}(0)) = -\frac{\partial^4}{\partial^2 s \partial^2 t}\Big|_{s=0=t} c(x(s), y(t))$.

This lemma is precisely analogous to the formula for the distance between two arclength parameterized geodesics $x(s)$ and $y(t)$ passing through $x(0) = y(0)$ in a Riemannian manifold:

$$d^2(x(s), \bar{x}(t)) = s^2 + t^2 - 2st\cos\theta - \frac{k}{3}s^2t^2\sin^2\theta + O((s^2+t^2)^{5/2})$$

where θ is the angle between $\dot{x}(0)$ and $\dot{y}(0)$ and k is the sectional curvature of the plane which they span. Therefore, we will not give its proof.

Proof of Remark 4.8 and sketch of Theorem 4.7. Assume $(\text{A3})_s$ for simplicity. It suffices to prove the following claim.

Claim 1: if $\frac{\partial f}{\partial t}(x, t_0) = 0$ then $\frac{\partial^2 f}{\partial t^2}(x, t_0) > 0$.

Proof of Claim 1: Convexity (A4) allows us to define $s \in [0,1] \longmapsto x(s)$ by

$$D_y c(x(s), y(t_0)) = (1-s)D_y c(x_0, y(t_0)) + sD_y c(x, y(t_0)) \tag{37}$$

and $g(s) = \frac{\partial^2 f}{\partial t^2}(x(s), t_0)$. Our claim is that $g(1) > 0$. Since $f(x_0, t) = 0$ and hence $g(0) = 0$, to prove Claim 1 it suffices to establish strict convexity in Claim 2.

Claim 2: $g : [0,1] \to \mathbf{R}$ is convex, and minimized at $s = 0$.

Proof of Claim 2: Once $g(s)$ is known to be convex, we need only observe that

$$g'(0) = -\frac{\partial^3}{\partial s \partial t^2}\bigg|_{s=0, t=t_0} c(x(s), y(t))$$

vanishes by our choice (35) of $y(t) = y_t$, to conclude $g(s)$ is minimized at $s = 0$.

Why should $g(s)$ be convex? Note that $g''(s) = -\frac{\partial^4}{\partial s^2 \partial t^2}\Big|_{t=t_0} c(x(s), y(t))$ is already non-negative according to Lemma 4.9 if we assume (B3). Remark 4.8 is thereby established. Under the weaker condition (A3), we need $\dot{x}^i(s) c_{i,j} \dot{y}^j(t_0) = 0$ to conclude $g(s)$ is convex — and strictly convex if $(\text{A3})_s$ holds. But

$$0 = \frac{\partial f}{\partial t}(x, t_0) = \int_0^1 c_{i,j}(x(s), y(t_0))\dot{x}^i(s)\dot{y}^j(t_0) ds$$

and the integrand is constant by our construction (37) of $x(s)$. □

To deduce the continuity result of the next section, the following corollary is crucial.

Corollary 4.10. *Assume* (A0)–(A4) *and fix* $(x_0, y_0) \in M^+_{int} \times M^-$. *If* $u = u^{\tilde{c}c}$ *satisfies*

$$u(x) \geq -c(x, y_0) + c(x_0, y_0) + u(x_0) \tag{38}$$

in a neighbourhood of x_0, *the same equality holds for all* $x \in M^+$.

Proof. The local inequality (38) implies $p_0 := -D_x c(x_0, y_0) \in \partial u(x_0)$. If $x_0 \in \text{dom}\, Du$, the conclusion is easy. The global inequality

$$u(x) \geq -c(x, y_1) + c(x_0, y_1) + u(x_0)$$

holds for any $(x_0, y_1) \in \partial^c u$, and for $y_1 \in \arg\max_{y\in M^-} -c(x_0, y) - u^{\tilde{c}}(y)$ in particular. The twist condition (A1) then implies $y_0 = y_1$.

Even if $x_0 \notin \operatorname{dom} Du$, taking e.g. $p = -D_x c(x_0, y_1)$ yields

$$-c(x, Y(x_0, p)) + c(x_0, Y(x_0, p)) \le u(x) - u(x_0) \qquad \forall x \in M^+. \tag{39}$$

In fact, the set $P = \{p \in M^-_{x_0} \mid (39) \text{ holds}\}$ is convex, according to Theorem 4.7. On the other hand, P includes all the extreme points p of $\partial u(x_0)$, since the preceding argument can be applied to a sequence $(x_k, y_k) \in \partial^c u \cap (\operatorname{dom} Du \times M^-)$ with $(x_k, Du(x_k)) \to (x_0, p)$. Thus $P \supset \partial u(x_0)$, whence $p_0 \in P$ as desired. (In fact, $P = \partial u(x_0)$). □

4.7. **Interior Hölder continuity for optimal maps.** To conclude our discussion on regularity of optimal mappings, let us sketch Loeper's Hölder continuity result [85].

Theorem 4.11 (Loeper '09)**.** *Assume* (A0)–(A2)*and* (A4)*. (i) If* (A3) *is violated, there exist probability densities with* $\log f^\pm \in C^\infty(M^\pm) \cap L^\infty$ *and a discontinuous optimal map* $G : M^+_{int} \longrightarrow M^-$ *satisfying* $G_\#(f^+ dVol) = f^- dVol$*. (ii) Conversely, if* $(A3)_s$ *holds and* $f^+/f^- \in L^\infty(M^+ \times M^-)$*, then* $G \in C^{\frac{1}{4n-1}}_{loc}(M^+_{int}, M^-)$*.*

In one dimension $n = 1$, we see G is Lipschitz directly from the equation $G'(x) = f^+(x)/f^-(G(x))$. In higher dimensions, this theorem is a direct consequence of the following proposition, whose inequalities $\sim$ hold up to multiplicative constants depending only on the cost c, and in particular on the size of the uniform modulus of positivity in condition $(A3)_s$.

Proposition 4.12 (Sausage into ball [85])**.** *Assuming the hypotheses and notation of Theorem 4.11(ii), take* $x_0, x_1 \in M^+$ *and set* $\Delta x = x_1 - x_0$ *and* $\Delta y = y_1 - y_0$ *where* $y_i = G(x_i)$*. If* $|\Delta x| \lesssim |\Delta y|^5$*, there is a ball* $B_\epsilon(x) \supset G^{-1}(S_\delta)$ *of radius* $\epsilon \sim \sqrt{\frac{|\Delta x|}{|\Delta y|}}$ *centered on the line segment joining* x_0 *to* x_1*, containing the preimage of the "sausage"*

$$S_\delta = \{y \in M^- \mid \inf_{t\in[1/3,2/3]} |y - y_t| \le \delta\}$$

of radius $\delta \sim \epsilon|\Delta y|^2$ *around the middle third of the curve* $(y_t)_{t\in[0,1]} \subset M^-$ *satisfying* $0 = \frac{d^2}{dt^2} D_x c(x, y_t)$*.*

Proof of Theorem 4.11(ii). At pairs of points $y_i = G(x_i)$ where $\|\Delta x\| \gtrsim \|\Delta y\|^5$ we already have Hölder exponent $1/5$ — even better than claimed. At other points, using the fact that G is a transport map between $\mu^\pm = f^\pm dx$, the Proposition yields $\mu^-(S_\delta) = \mu^+(G^{-1}(S_\delta)) \lesssim \|f^+\|_\infty \epsilon^n$, but also $\delta^{n-1}|\Delta y| \inf_{M^-} f^- \lesssim \mu^-(S_\delta)$. Combining the squares of these two inequalities, our choices $\delta \sim \epsilon|\Delta y|^2$ and $\epsilon^2 \sim |\Delta x|/|\Delta y|$ yield the desired Hölder estimate:

$$\|(f^-)^{-1}\|^{-2}_\infty \epsilon^{2n-2} |\Delta y|^{4n-2} \lesssim \|f^+\|^2_\infty \epsilon^{2n-2} \frac{|\Delta x|}{|\Delta y|}.$$

Thus $G \in C^{\frac{1}{4n-1}}_{loc}$. □

The proposition relies delicately on Corollary 4.10 and the correct choice of δ and ϵ:

Proof sketch of Proposition 4.12; c.f. [72]*:* According to Theorem 2.9, the optimal map $G(x) = Y(x, Du(x))$ is given by a potential $u = u^{\tilde{c}c}$ and $Graph(G) \subset \partial^c u$. Thus $(x_i, y_i) \in \partial^c u$, meaning $f_i(x) = -c(x, y_i) + c(x_i, y_i) + u(x_i)$ satisfies $u(x) \ge \max\{f_0(x), f_1(x)\}$ with equality at x_0 and x_1. Take x to be the point on the segment joining x_0 to x_1 where $f_0(x) = f_1(x)$ $(= 0$ without loss of generality). The semiconvexity of u shown in Lemma 3.1 then yields the bound $u(x) \lesssim |\Delta x||\Delta y| + |\Delta x|^2$. Assumption $(A3)_s$ allows

Theorem 4.7 to be quantified, so that $f_t(\cdot) := -c(\cdot, y_t) + c(x, y_t) + u(x) \leq u(\cdot)$ actually satisfies

$$f_t(x') - u(x') \lesssim -t(1-t)|x' - x|^2|\Delta y|^2$$

for x' near x. For $t \in [1/3, 2/3]$, these estimates give some leeway to shift y_t up to distance δ without spoiling the inequality $g_y(x') := -c(x', y) + c(x, y) + u(x) \leq u(x')$ on the boundary $x' \in \partial B_\epsilon(x)$. Since $g_y(x) = u(x)$, this inequality does not extend to the interior of the ball $B_\epsilon(x)$, unless we subtract some non-negative constant from $g_y(\cdot)$. Subtracting the smallest such constant λ yields a function $g_y(\cdot) - \lambda \leq u(\cdot)$ on $B_\epsilon(x)$, with equality at some $x_* \in B_\epsilon(x)$. Corollary 4.10 implies $(x_*, y) \in \partial^c u$. For almost every such $y \in S_\delta$ this provides the desired preimage $x_* \in G^{-1}(y)$. □

5. Multidimensional screening: an application to economic theory

We now sketch an application [49] of the mathematics we have developed to one of the central problems in microeconomic theory: making pricing or policy decisions for a monopolist transacting business with a field of anonymous agents whose preferences are known only statistically. Economic buzzwords associated with problems of this type include "asymmetric information," "mechanism design," "incentive compatibility," "nonlinear pricing," "signalling," "screening," and the "principal / agent" framework.

5.1. Monopolist nonlinear pricing and the principal-agent framework.

To describe the problem, imagine we are given: a set of "customer" types $M^+ \subset \mathbf{R}^n$ and "product" types $M^- \subset \mathbf{R}^n$ and

$b(x, y)$= benefit of product $y \in M^-$ to customer $x \in M^+$;

$a(y)$= monopolist's cost to manufacture $y \in M^-$;

$d\mu^+(x) \geq 0$ relative frequency of different customer types on M^+.

Knowing all this data, the **Monopolist's problem** is to assign a price to each product, for which she will be willing to manufacture that product and sell it to whichever agents choose to buy it. Her task is to design the price menu $v : M^- \to \mathbf{R} \cup \{+\infty\}$ so as to maximize profits. The only constraint that prevents her from raising prices arbitrarily high is the existence of a fixed $y_\emptyset \in M^-$, called the "outside option" or "null product", which she is compelled to sell at cost $v(y_\emptyset) = a(y_\emptyset)$. Though it is not necessary, we can fix the cost of the null product to vanish without loss of generality.

The **Agent's problem** consists in computing

$$u(x) = \max_{y \in M^-} b(x, y) - v(y) \tag{40}$$

and choosing to buy that product $y_{b,v}(x)$ for which the maximum is attained. The monopolist is generally called the *principal*, while the customers are called *agents*.

Economists use this framework to model many different types of transactions, including tax policy [103] (where the government wants to decide a tax structure which encourages people both to work and report income), contract theory [126] (where a company wants to decide a salary structure which attracts and rewards effective employees without overpaying them), and the monopolist nonlinear pricing problem described above [106]. In the initial studies, the type spaces $M^\pm$ were assumed one-dimensional, with $x \in M^+$ representing the innate ability or talent of the prospective tax-payer or employee, and $y \in M^-$ the amount of work that he chooses to do or the credentials

he chooses to acquire. The basic insight of Mirrlees and Spence was that under condition (A2) (which implies (A1) in a single dimension) the variables $x, y \in \mathbf{R}$ would be monotonically correlated by the optimal solution, reducing the monopolist's problem to an ordinary differential equation. For this reason the one-dimensional versions of (A1)–(A2) are called *Spence-Mirrlees* (or *single-crossing*) conditions in the economics literature; both Mirrlees and Spence were awarded Nobel prizes for exploring the economic implications of their solution.

Of course, many types of products are more realistically modeled using several parameters $y \in \mathbf{R}^n$ — in the case of cars these might include fuel efficiency, size, comfort, safety, reliability, and appearance — while the preferences of customers for such parameters are similarly nuanced. Thus it is natural and desirable to want to solve the multidimensional version $n \geq 2$ of the problem, about which much less is known [8]. Monteiro and Page [105] and independently Carlier [22] showed only that enough compactness remains to conclude that the monopolist's optimal strategy exists. An earlier connection to optimal transportation can be discerned in the work of Rochet [117], who proved a version of Theorem 2.3 (Rockafellar) for general utility functions b ($= -c$ in our earlier notation).

Rochet and Choné [118] studied the special case $b(x, y) = x \cdot y$ on $M^{\pm} = [0, \infty[^n$. Taking $a(y) = \frac{1}{2}|y|^2$, $d\mu^+ = \chi_{[0,1]^2} d^2x$, and $y_\emptyset = (0, 0)$, they deduced that the mapping $y_{b,v} : M^+ \to M^-$ was the gradient of a convex function; it sends a positive fraction of the square to the point mass $y_\emptyset$, and a positive fraction to the line segment $y_1 = y_2$, while the remaining positive fraction gets mapped in a bijective manner, so that

$$\mu^- := (y_{b,n})_\# \mu^+ = f_0^- \delta_{y_\emptyset} + f_1^- d\mathcal{H}^1 + f_2^- d\mathcal{H}^2. \tag{41}$$

They interpreted this solution to mean that while the top end of the market gets customized vehicles f_2^-, price discrimination alone forces those customers in the next market segment to choose from a more limited set f_1^- of economy vehicles offering a compromise between attributes y_1 and y_2. A fraction $f_0^- > 0$ of consumers will be priced out of the market altogether — which had already been observed by Armstrong [7] to be a hallmark of nonlinear pricing in more than one dimension $n \geq 2$. Economists refer to this general phenomenon (41) as "bunching", and to the fact that $f_0^- > 0$ as "the desirability of exclusion."

How robust is this picture? It remains a pressing question to understand whether the bunching phenomena of Rochet and Choné is robust, or merely an accident of the particular example they explored. As we now explain, their results were obtained by reducing the monopolist's problem to the minimization of a Dirichlet energy:

$$\min_{0 \leq u \text{ convex}} \int_{[0,1]^2} \left(\frac{1}{2}|Du|^2 - \langle x, Du(x) \rangle + u(x) \right) d\mathcal{H}^2(x). \tag{42}$$

The constraint that $u : M^+ \longrightarrow \mathbf{R}$ be *convex* makes this problem non-standard: its solution satisfies a Poisson type equation only on the set where u is strongly convex ($D^2u > 0$), and there are free boundaries separating the regions where the different constraints $u \geq 0$ and $D^2u \geq 0$ begin to bind.

5.2. **Variational formulation using optimal transportation.** The principal's problem is to choose $v : M^- \longrightarrow \mathbf{R} \cup \{+\infty\}$ to maximize her profits, or equivalently to minimize her net losses:

$$\min_{\{v | v(y_\emptyset) = a(y_\emptyset)\}} \int_{M^+} [a(y_{b,v}(x)) - v(y_{b,v}(x))] d\mu^+(x). \tag{43}$$

Note that the integrand vanishes for all customers x who choose the null product $y_{b,v}(x) = y_\emptyset$.

Wherever the agent's maximum (40) is achieved, we have $Du(x) - D_x b(x, y_{b,v}(x)) = 0$, so using the twist condition (A1) from our previous lectures we can invert this relation to get $y_{b,v}(x) = Y(x, Du(x))$. Moreover, the function $u(x)$ from (40) is a b-convex function, called the *surplus* or *indirect utility* $u = u^{\tilde{b}b}$. (In our previous notation, u is a $(-b)$-convex function and $u = u^{(-\tilde{b})(-b)}$, but we suppress the minus signs hereafter.)

Since $v(Y(x, Du(x)) = b(x, Y(x, Du(x)) - u(x)$, we may reformulate the variational problem (43) as the minimization of the principal's losses

$$L(u) := \int_{M^+} [a(Y(x, Du(x))) - b(x, Y(x, Du(x))) + u(x)]\, d\mu^+(x). \tag{44}$$

over the set $\mathcal{U}_\emptyset = \{u \in \mathcal{U} \mid u \geq u_\emptyset\}$ of b-convex functions $\mathcal{U} = \{u \mid u = u^{\bar{b}b}\}$ which exceed the reservation utility $u_\emptyset(\cdot) = b(., y_\emptyset) - a(y_\emptyset)$ associated with the outside option or null product. This strange reformulation due to Carlier [22]

$$\min_{u \in \mathcal{U}_\emptyset} L(u) \tag{45}$$

reduces to (42) in the case considered by Rochet and Choné.

5.3. **When is this optimization problem convex?** From [105] and [22] we know that a minimizer exists. The contribution of Figalli, Kim and McCann is to give sufficient conditions for the variational problem to become convex — in which case it is considerably simpler to analyze, theoretically and computationally. It is very interesting that the Ma, Trudinger and Wang criteria for the regularity of optimal mappings turn out to be related to this question. The following are among the main results of [49]:

Theorem 5.1 (Convexity of the principal's strategy space [49])**.** *If $b = (-c)$ satisfies* (A0)–(A2) *and* (A4) *then the set $\mathcal{U} = \{u = u^{\tilde{b}b}\}$ is convex if and only if* (B3) *holds, ie., if and only if* $\mathrm{cross}(p, q) \geq 0$ *for all tangent vectors $(p, x_0) \in TM^+$ and $(q, y_0) \in TM^-$.*

Remark 5.2. It was pointed out subsequently by Brendan Pass [111] that the convexity of $M^-_{x_0}$ assumed in (A4) for each $x_0 \in M^+$ is also necessary for convexity of $\mathcal{U}$.

Sketch of proof. First assume (B3) holds — assuming always (A0)–(A2) and (A4). Given $u_0, u_1 \in \mathcal{U}$ and $t \in [0, 1]$ we claim $u_t := (1 - t)u_0 + tu_1$ is b-convex. This can be established by finding for each $x_0 \in M^+$ a $y_t \in M^-$ such that

$$u_t(\cdot) \geq b(\cdot, y_t) - b(x_0, y_t) + u_t(x_0) \qquad \text{throughout } M^+, \tag{46}$$

for then $u_t(\cdot)$ is the supremum of such functions. Corresponding to $t = 0, 1$ the desired points $y_0, y_1 \in M^-$ exist, by b-convexity of $u_i = u_i^{\tilde{b}b}$ for $i = 0, 1$. By (A4), we can solve the equation $D_x b(x_0, y_t) = (1-t)D_x b(x_0, y_0) + tD_x b(x_0, y_t)$; the solution $y_t \in M^-$ makes $f(\cdot, t) := b(\cdot, y_t) - b(x_0, y_t)$ a convex function of $t \in [0, 1]$, according to Remark 4.8. Inequality (46) holds at the endpoints $t = 0, 1$; taking a convex combination yields the desired inequality for intermediate values of $t \in [0, 1]$. For the converse direction, we refer to [49]. □

Theorem 5.3 (Convexity of principal's losses and uniqueness of optimal strategy [49])**.** *If* (A0)–(A4) *and* (B3) *hold and if $a = a^{b\tilde{b}}$, then the functional $u \in \mathcal{U} \longmapsto L(u)$ defined by* (44) *is convex. Furthermore, it has enough strict convexity to conclude the optimizer $u \in \mathcal{U}_\emptyset$ is uniquely determined (at least μ^+-a.e.) if $\mu^+ \ll \mathcal{H}^n$ and either (i) $y_{b,a} : M^+ \to M^-$ is continuous or else (ii) b has positive cross-curvature* $(\mathrm{B3})_s$.

Sketch of proof. To deduce convexity of $L : \mathcal{U} \longrightarrow \mathbf{R}$, recall $\tilde{b}$-convexity of a implies

$$a(Y(x, p)) - b(x, Y(x, p)) = \sup_{x_1 \in M^+} b(x_1, Y(x, p)) - b(x, Y(x, p)) - a^b(x_1).$$

For each $x \in M^+$ fixed, the functions under the supremum are convex with respect to $p \in M_x^-$, according to Remark 4.8. Thus the integrand in (44) is linear in $u(x)$ and convex with respect to $p = Du(x)$, which establishes the desired convexity of the integral $L(u)$. In case (i) the integrand is strictly convex, while in case (ii) it is strictly convex for all $x \in \operatorname{dom} Da^b$, which is a set of full $\mu^+ \ll \mathcal{H}^n$ measure. We refer to [49] for details. □

Regarding robustness: we may mention that, as in Remark 4.6, the bilinear function $b(x, y)$ lies on the borderline of costs which satisfy (B3). Thus there will be perturbations of this function which destroy convexity of the problem, and we can anticipate that under such perturbations, uniqueness and other properties of its solution may no longer persist. In fact, for $a = 0$ and $b(x, y) = -d_M^2(x, y)$ on a Riemannian ball $M^+ = M^- = B_r(y_\emptyset)$, we arrive at a problem equivalent to a fourfold symmetrized version of Rochet and Choné's in the Euclidean case, but which satisfies or violates (B3) depending on whether the metric is spherical or hyperbolic. This can used to model local delivery of a centralized resource for a town in the mountains [49]; cf. Examples 4.3–4.4. On the other hand, under the hypotheses of Theorem 5.3 we are able to show that Armstrong's "desirability of exclusion" [7] continues to hold. We give the statement only and refer to [49] for its proof.

Theorem 5.4 (The desirability of exclusion)**.** *Assume* (A0)–(A4), (B3), $a = a^{b\tilde{b}}$ *and that* $d\mu^+ = f^+ d\mathcal{H}^n$ *with* $f^+ \in W^{1,1}(M^+)$ *and the convex set* $M_{y_\emptyset}^- = D_y b(M^+, y_\emptyset) \subset \mathbf{R}^n$ *has no* $(n-1)$ *dimensional facets. Then a positive fraction of agents will be priced out of the market by the principal's optimal strategy.*

Remark 5.5. It is interesting to note that the strict convexity condition on $M_{y_\emptyset}^+$ holds neither in one dimension — where Armstrong noted counterexamples to the desirability of exclusion — nor for the example of Rochet-Choné, where convexity of $M_{y_\emptyset}^+$ is not strict.

5.4. **Variant: maximizing social welfare.** Idealistic readers may be taken somewhat aback by the model just presented, for it is the very theory which predicts, among other things, just how uncomfortable airlines ought to make their economy seating to ensure — without sacrificing too much economy-class revenue — that passengers with the means to secure a business-class ticket have sufficient incentive to do so. Such readers will doubtless be glad to know that the same mathematics is equally relevant to the more egalitarian question of how to price public services so as to maximize societal benefit.

For example, suppose the welfare $w(x, u(x))$ of agent x is a concave function of the indirect utility $u(x)$ he receives. A public service provider would like to set a price menu for which $u = v^b$ maximizes the toral welfare among all agents:

$$\max_{u \in \mathcal{U}_\emptyset,\ L(u) \le 0} \int_{M^+} w(x, u(x)) d\mu^+(x),$$

subject to the constraint $L(u) \le 0$ that the service provider not sustain losses. Introducing a Lagrange multiplier λ for this budget constraint, the problem can be rewritten in the unconstrained form

$$\max_{u \in \mathcal{U}_\emptyset} -\lambda L(u) + \int_{M^+} w(x, u(x)) d\mu^+(x),$$

(for a suitable $\lambda \ge 0$). Under the same assumptions (A0)–(A4), (B3) and $a = a^{b\tilde{b}}$ as before, we see from the results above that this becomes a concave maximization problem for which existence and uniqueness of solution follow directly, and which is therefore quite amenable to further study, both theoretical and computational.

6. A pseudo-Riemannian and symplectic geometric afterword

The conditions of Ma, Trudinger and Wang for regularity of optimal transport have a differential geometric significance uncovered by Kim and McCann [73], which led to their discovery with Warren [75] of a surprising connection of optimal transport to the theory of volume-maximizing special Lagrangian submanifolds in split geometries.

Indeed, since the smoothness of optimal maps $G : M^+ \longrightarrow M^-$ is a question whose answer is independent of coordinates chosen on M^+ and M^-, it follows that the necessary and sufficient condition (A3) for continuity in Theorem 4.11 should have a geometrically invariant description. We give this description, below, as the positivity of certain sectional curvatures of a metric tensor h induced on the product manifold $N := M^+ \times M^-$ by the cost function $c \in C^4(N)$. This motivates the appellation *cross-curvature*. The rationale for such a description to exist is quite analogous to that underlying general relativity, Einstein's theory of gravity, which can be expressed in the language of pseudo-Riemannian geometry due to the coordinate invariance that results from the equivalence principle (observer independence).

Use the cost function to define the symmetric and antisymmetric tensors

$$h = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 c}{\partial x^i \partial y^j}(dx^i \otimes dy^j + dy^j \otimes dx^i) \tag{47}$$

$$\omega = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 c}{\partial x^i \partial y^j}(dx^i \otimes dy^j - dy^j \otimes dx^i) \tag{48}$$

on $N = M^+ \times M^-$. Then condition (A2) is equivalent to non-degeneracy of the metric tensor h, which in turn is equivalent to the assertion that ω is a symplectic form. Note however that h is not positive-definite, but has equal numbers of positive and negative eigenvalues in any chosen coordinates, ie. signature (n, n). Conditions (A3) and (A4) are conveniently re-expressed in terms of the pseudo-metric h, and its pseudo-Riemannian curvature tensor R_{ijkl} [73]. Indeed, condition (A4) asserts the h-geodesic convexity of $\{x_0\} \times M^-$ and $M^+ \times \{y_0\}$, while the formula

$$\mathrm{cross}(p, q) = R_{ijkl} p^i q^j p^k q^l$$

shows the cross-curvature is simply proportional to the pseudo-Riemannian sectional curvature of the 2-plane $(p \oplus 0) \wedge (0 \oplus q)$. The restriction distinguishing (A3) from (B3) is that $p \oplus q$ be lightlike, which is equivalent to the h-orthogonality of $p \oplus 0$ and $0 \oplus q$.

Kim and McCann went on to point out that the graph of any c-optimal map is h-spacelike and ω-Lagrangian, meaning any tangent vectors $P, Q \in T_{(x,G(x))}N$ to this graph satisfy $h(P, P) \geq 0$ and $\omega(P, Q) = 0$. This is a consequence of Corollary 2.7, it is also very illustrative to check it directly by hand in the case of the quadratic cost in $\mathbb{R}^n$. When (A0)–(A4) hold they also showed the converse to be true: any diffeomorphism whose graph is h-spacelike and ω-Lagrangian is also c-cyclically monotone, hence optimal. With Warren [75], they introduced a pseudo-metric

$$h_c^{f^\pm} = \left(\frac{f^+(x) f^-(y)}{|\det c_{i,j}(x, y)|}\right)^{1/n} h \tag{49}$$

conformally equivalent to h. In this new metric, they show the graph of the c-optimal map pushing $d\mu^+(x) = f^+(x)d^n x$ forward to $\mu^-(y) = f^-(y)d^n y$ has maximal volume with respect to compactly supported perturbations. In particular, $Graph(G)$ has zero mean-curvature as a submanifold (with half the dimension) of $(N, h_c^{f^\pm})$ — yielding an unexpected connection of optimal transportation to more classical problems in geometry and geometric measure theory. (Note that the metric (49) depends only on the measures

$f^{\pm}$ and the sign of the mixed partial $D^2_{xy}c$ in dimension $n = 1$.) The preprint of Harvey and Lawson [65] contains a wealth of related information concerning special Lagrangian submanifolds in pseudo-Riemannian (= semi-Riemannian) geometry.

References

[1] T. Abdellaoui and H. Heinich. Sur la distance de deux lois dans le cas vectoriel. *C.R. Acad. Sci. Paris Sér. I Math.*, 319:397–400, 1994.

[2] N. Ahmad, H.K. Kim, and R.J. McCann. Optimal transportation, topology and uniqueness. *Preprint at* `www.math.toronto.edu/mccann`.

[3] G. Alberti and L. Ambrosio. A geometrical approach to monotone functions in $\mathbf{R}^n$. *Math. Z.*, 230:259–316, 1999.

[4] L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical Aspects of Evolving Interfaces*, volume 1812 of *Lecture Notes in Mathematics*, pages 1–52. Springer, Berlin, 2003.

[5] L.A. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lecture Notes in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[6] E.J. Anderson and P. Nash. *Linear Programming in Infinite-Dimensional Spaces*. Wiley, Chichester, 1987.

[7] M. Armstrong. Multiproduct nonlinear pricing. *Econometrica*, 64:51–75, 1996.

[8] S. Basov. *Multidimensional Screening*. Springer-Verlag, Berlin, 2005.

[9] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *J. Eur. Math. Soc. (JEMS)*, 9:85–121, 2007.

[10] M. Bernot, V. Caselles, J.-M. Morel. *Optimal Transportation Networks: Models and Theories*. Springer-Verlag, Berlin, 2009.

[11] S. Bianchini and F. Cavalletti. The Monge problem for distance cost in geodesic spaces. *Preprint at* `http://cvgmt.sns.it/papers/biacav09/Monge@problem.pdf`.

[12] G. Bouchitté and G. Buttazzo. Characterization of optimal shapes and masses through Monge-Kantorovich equation. *J. Eur. Math. Soc. (JEMS)*, 3:139–168, 2001.

[13] G. Bouchitté, W. Gangbo and P. Seppecher. Michell trusses and lines of principal action. *Math. Models Methods Appl. Sci.*, 18:1571–1603, 2008.

[14] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[15] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[16] L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editor, *Partial Differential Equations and Applications*, number 177 in Lecture Notes in Pure and Appl. Math., pages 29–35. Dekker, New York, 1996.

[17] L.A. Caffarelli. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.*, 45:1141–1151, 1992.

[18] L.A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5:99–104, 1992.

[19] L.A. Caffarelli. Boundary regularity of maps with convex potentials — II. *Ann. of Math. (2)*, 144:453–496, 1996.

[20] L.A. Caffarelli and R.J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Ann. of Math. (2)*, 171:673–730, 2010.

[21] L.A. Caffarelli, M. Feldman and R.J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.*, 15:1–26, 2002.

[22] G. Carlier. A general existence result for the principal-agent problem with adverse selection. *J. Math. Econom.*, 35:129–150, 2001.

[23] G. Carlier and I. Ekeland. Equilibrium structure of a bidimensional asymmetric city. *Nonlinear Anal. Real World Appl.*, 8:725–748, 2008.

[24] J.A. Carrillo, R.J. McCann, and C. Villani. Contractions in the 2-wasserstein length space and thermalization of granular media. *Arch. Rational Mech. Anal.*, 179:217–263, 2006.

[25] T. Champion and L. De Pascale. The Monge problem in $\mathbf{R}^d$. *Preprint at* `http://cvgmt.sns.it/papers/chadep09/champion-depascale.pdf`.

[26] T. Champion, L. De Pascale, and P. Juutinen. The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40, 2008.

[27] P.-A. Chiappori, R.J. McCann, and L. Nesheim. Hedonic price equilibria, stable matching and optimal transport: equivalence, topology and uniqueness. *Econom. Theory*, 42:317–354, 2010.

[28] D. Cordero-Erausquin. Inégalité de Prékopa-Leindler sur la sphère. *C.R. Acad. Sci. Paris Sér. I Math.*, 329:789–792, 1999.

[29] D. Cordero-Erausquin. Sur le transport de mesures périodiques. *C.R. Acad. Sci. Paris Sér. I Math.*, 329:199–202, 1999.
[30] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146:219–257, 2001.
[31] D. Cordero-Erausquin, R.J. McCann and M. Schmuckenschläger. Prekopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Ann. Fac. Sci. Toulouse Math. (6)*, 15:613–635, 2006.
[32] D. Cordero-Erausquin, B. Nazaret, and C. Villani. A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities. *Adv. Math.*, 182:307–332, 2004.
[33] J.A. Cuesta-Albertos and C. Matrán. Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.*, 17:1264–1276, 1989.
[34] J.A. Cuesta-Albertos and A. Tuero-Díaz. A characterization for the solution of the Monge-Kantorovich mass transference problem. *Statist. Probab. Lett.*, 16:147–152, 1993.
[35] M.J.P. Cullen. *A Mathematical Theory of Large Scale Atmosphere/Ocean Flows*. Imperial College Press, London, 2006.
[36] M.J.P Cullen and R.J. Purser. An extended Lagrangian model of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497, 1984.
[37] M.J.P Cullen and R.J. Purser. Properties of the Lagrangian semi-geostrophic equations. *J. Atmos. Sci.*, 46:2684–2697, 1989.
[38] P. Delanoë. Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère operator. *Ann. Inst. H. Poincarè Anal. Non Linèaire*, 8:443–457, 1991.
[39] P. Delanoë and Y. Ge. Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds. To appear in *J. Reine Angew. Math.*
[40] R. Dobrushin. Definition of a system of random variables by means of conditional distributions (Russian). *Teor. Verojatnost. i Primenen.*, 15:469–497, 1970.
[41] R.M. Dudley. *Probabilities and metrics - Convergence of laws on metric spaces, with a view to statistical testing*. Universitet Matematisk Institut, Aarhus, Denmark, 1976.
[42] I. Ekeland. An optimal matching problem. *ESAIM: Control, Optimisation and Calculus of Variations*, 11:57–71, 2005.
[43] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 1998.
[44] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137:1–66, 1999.
[45] A. Fathi and A. Figalli. Optimal transportation on non-compact manifolds. *Israel J. Math.*, 175:1–59, 2010.
[46] A. Figalli and N. Gigli. Local seimconvexity of Kantorovich potentials on noncompact manifolds. To appear in *ESAIM Control Optim. Calc. Var.*
[47] A. Figalli, Y.-H. Kim, and R.J. McCann. Continuity and injectivity of optimal maps for non-negatively cross-curved costs. Preprint at `www.math.toronto.edu/mccann`.
[48] A. Figalli, Y.-H. Kim, and R.J. McCann. Regularity of optimal transport maps on multiple products of spheres. Preprint at `www.math.toronto.edu/mccann`.
[49] A. Figalli, Y.-H. Kim, and R.J. McCann. When is multidimensional screening a convex program? Preprint at `www.math.toronto.edu/mccann`.
[50] A. Figalli and G. Loeper. C^1 regularity of solutions of the Monge-Ampère equation for optimal transport in dimension two. *Calc. Var. Partial Differential Equations*, 35:537–550, 2009.
[51] A. Figalli and F. Maggi. On the shape of liquid drops and crystals in the small mass regime. *Preprint at* `http://cvgmt.sns.it/papers/figmag10/fm-small4.pdf`.
[52] A. Figalli, F. Maggi and A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. To appear in *Invent. Math.*
[53] A. Figalli and L. Rifford. Continuity of optimal transport maps on small deformations of $\mathbb{S}^2$. *Comm. Pure Appl. Math.*, 62: 1670–1706, 2009.
[54] A. Figalli, L. Rifford and C. Villani. Nearly round spheres look convex. *Preprint.*
[55] A. Figalli, L. Rifford and C. Villani. On the Ma-Trudinger-Wang curvature on surfaces. *Preprint.*
[56] L. Forzani and D. Maldonado. Properties of the solutions to the Monge-Ampère equation. *Nonlinear Anal.*, 57:815–829, 2004.
[57] W. Gangbo. *Habilitation thesis*. Université de Metz, 1995.
[58] W. Gangbo and R.J. McCann. Optimal maps in Monge's mass transport problem. *C.R. Acad. Sci. Paris Sér. I Math.*, 321:1653–1658, 1995.
[59] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[60] W. Gangbo and R.J. McCann. Shape recognition via Wasserstein distance. *Quart. Appl. Math.*, 58:705–737, 2000.
[61] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(P_2(M), W_2)$. *Preprint at* `http://cvgmt.sns.it/papers/gigc/Inverse.pdf`.
[62] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, New York, 1983.
[63] T. Glimm and V. Oliker. Optical design of single reflector systems and the Monge-Kantorovich mass transfer problem. *J. Math. Sci.*, 117:4096–4108, 2003.
[64] S. Haker, L. Zhu, A. Tannenbaum and S. Angenent. Optimal mass transport for registration and warping. *Int. J. Computer Vision*, 60:225–240, 2004.
[65] F.R. Harvey and H.B. Lawson, Jr. Split special Lagrangian geometry. *Preprint*.
[66] S. Kakutani. A generalization of brouwer's fixed point theorem. *Duke Math. J.*, 8:457–459, 1941.
[67] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
[68] L. Kantorovich. On a problem of Monge (In Russian). *Uspekhi Math. Nauk.*, 3:225–226, 1948.
[69] L.V. Kantorovich and G.S. Rubinstein. On a space of completely additive functions. *Vestnik Leningrad. Univ.*, 13:52–59, 1958.
[70] H.G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67:399–432, 1984.
[71] Y.-H. Kim. Counterexamples to continuity of optimal transportation on positively curved Riemannian manifolds. *Int. Math. Res. Not.*, Art. ID rnn120:1–15, 2008.
[72] Y.-H. Kim and R.J. McCann. Appendices to original version of Continuity, curvature, and the general covariance of optimal transportation. Preprint at `arXiv:math/0712.3077v1`.
[73] Y.-H. Kim and R.J. McCann. Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc. (JEMS)*, 12:1009–1040, 2010.
[74] Y.-H. Kim and R.J. McCann. Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular). Preprint at `arXiv:math/0806.0351v1` To appear in *J. Reine Angew. Math.*
[75] Y.-H. Kim, R.J. McCann and M. Warren. Pseudo-Riemannian geometry calibrates optimal transportation. Preprint at `www.math.toronto.edu/mccann`. To appear in *Math. Res. Lett.*
[76] T.C. Koopmans. Optimum utilization of the transportation system. *Econometrica (Supplement)*, 17:136–146, 1949.
[77] P.W.Y. Lee. New computable necessary conditions for the regularity theory of optimal transportation. Preprint at `arXiv:0912.4795`. To appear in *SIAM J. Math. Anal.*
[78] P.W.Y. Lee and J. Li. New examples on spaces of negative sectional curvature satisfying Ma-Trudinger-Wang conditions. Preprint at `arXiv:0911.3978`.
[79] P.W.Y. Lee and R.J. McCann. The Ma-Trudinger-Wang curvature for natural mechanical actions. Preprint at `www.math.toronto.edu/mccann`. To appear in *Calc. Var. Partial Differential Equations*.
[80] V.L. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-valued Anal.*, 7:7–32, 1999.
[81] J. Li. Smooth optimal transportation on hyperbolic space. Master's thesis, University of Toronto, 2009.
[82] J. Liu. Hölder regularity of optimal mappings in optimal transportation. *Calc Var. Partial Differential Equations*, 34:435–451, 2009.
[83] J. Liu, N.S. Trudinger, X.-J. Wang. Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation. *Comm. Partial Differential Equations*, 35:165–184, 2010.
[84] G. Loeper. On the regularity of maps solutions of optimal transportation problems II. The sphere case and the reflector antenna. To appear in *Arch. Ration. Mech. Anal.*
[85] G. Loeper. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202:241–283, 2009.
[86] G. Loeper and C. Villani. Regularity of optimal transport in curved geometry: the non-focal case. *Duke Math. J.*, 151:431–485, 2010.
[87] J. Lott. Optimal transport and Perelman's reduced volume. *Calc. Var. Partial Differential Equations*, 36:49–84, 2009.
[88] J. Lott and C. Villani. Ricci curvature for metric measure spaces via optimal transport. *Annals Math. (2)*, 169:903–991, 2009.
[89] X.-N. Ma, N. Trudinger and X.-J. Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Rational Mech. Anal.*, 177:151–183, 2005.
[90] F. Maggi and C. Villani. Balls have the worst best Sobolev inequalities. II. Variants and extensions. *Calc. Var. Partial Differential Equations*, 31:47–74, 2008.

[91] R.P. McAfee and J. McMillan. Multidimensional incentive compatibility and mechanism design. *J. Econom. Theory*, 46:335–354, 1988.
[92] R.J. McCann. *A Convexity Theory for Interacting Gases and Equilibrium Crystals*. PhD thesis, Princeton University, 1994.
[93] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.
[94] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.
[95] R.J. McCann. Equilibrium shapes for planar crystals in an external field. *Comm. Math. Phys.*, 195:699–723, 1998.
[96] R.J. McCann. Exact solutions to the transportation problem on the line. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 455:1341–1380, 1999.
[97] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.
[98] R.J. McCann. Stable rotating binary stars and fluid in a tube. *Houston J. Math.*, 32:603–632, 2006.
[99] R.J. McCann and P. Topping. Ricci flow, entropy, and optimal transportation. *Amer. J. Math.*, 132:711–730, 2010.
[100] R.J. McCann, B. Pass and M. Warren. Rectifiability of optimal transportation plans. *Preprint at* `www.math.toronto/mccann`.
[101] V.D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov*. Springer-Verlag, Berlin, 1986.
[102] George J. Minty. Monotone (nonlinear) operators in hilbert space. *Duke Math. J.*, 29:341–346, 1962.
[103] J.A. Mirrlees. An exploration in the theory of optimum income taxation. *Rev. Econom. Stud.*, 38:175–208, 1971.
[104] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.
[105] P.K. Monteiro and F.H. Page Jr. Optimal sellling mechanisms for multiproduct monopolists: incentive compatibility in the presence of budget constraints. *J. Math. Econom.*, 30:473–502, 1998.
[106] M. Mussa and S. Rosen. Monopoly and product quality. *J. Econom. Theory*, 18:301–317, 1978.
[107] J. von Neumann. Zur Theorie der Gesellschaftsspiele. (German). *Math. Ann.*, 100:295–320, 1928.
[108] J. von Neumann. Über ein Okonomisches Gleichungssystems und eine Verallgemeinerung die Brouwerschen Fixpunktsatzes. *Ergebuisse eines Mathematischen Seminars*, 8:73–83, 1937.
[109] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations*, 26:101–174, 2001.
[110] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173:361–400, 2000.
[111] B. Pass. Personal communication. 2010.
[112] G. Perelman. The entropy formula for the Ricci flow and its geometric applications. Preprint at `http://arxiv.org/abs/math/0211159`.
[113] A.Yu. Plakhov. Newton's problem of the body of minimal averaged resistance (Russian). *Mat. Sb.*, 195:105–126, 2004.
[114] J. Purser and M.J.P. Cullen. *J. Atmos. Sci*, 44:3449–3468, 1987.
[115] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems*. Probab. Appl. Springer-Verlag, New York, 1998.
[116] M.-K. von Renesse and K.-T. Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.*, 58:923–940, 2005.
[117] J.-C. Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *J. Math. Econom.*, 16:191–200, 1987.
[118] J.-C. Rochet and P. Choné. Ironing, sweeping and multidimensional screening. *Econometrica*, 66:783–826, 1998.
[119] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.
[120] L. Rüschendorf. Bounds for distributions with multivariate marginals. In K. Mosler and M. Scarsini, editors, *Stochastic Orders and Decision Under Risk*, number 19 in IMS Lecture Notes - Monograph Series, pages 285–310. Institute of Mathematical Statistics, Hayward, CA, 1991.
[121] L. Rüschendorf. Optimal solutions of multivariate coupling problems. *Appl. Math. (Warsaw)*, 23:325–338, 1995.

[122] L. Rüschendorf. On c-optimal random variables. *Statist. Probab. Lett.*, 37:267–270, 1996.
[123] L. Rüschendorf and S.T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32:48–54, 1990.
[124] C. Smith and M. Knott. On the optimal transportation of distributions. *J. Optim. Theory Appl.*, 52:323–329, 1987.
[125] C. Smith and M. Knott. On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.*, 40:328–334, 1992.
[126] M. Spence. Competitive and optimal responses to signals: An analysis of efficiency and distribution. *J. Econom. Theory*, 7:296–332, 1974.
[127] K.-T. Sturm. On the geometry of metric measure spaces, I and II. *Acta Math.*, 196:65–177, 2006.
[128] V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, 141:1–178, 1979.
[129] H. Tanaka. An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 27:47–52, 1973.
[130] P. Topping. $\mathcal{L}$-optimal transportation for Ricci flow. *J. Reine Angew. Math.*, 636:93–122, 2009.
[131] N.S. Trudinger. Isoperimetric inequalities for quermassintegrals. *Ann. Inst. H. Poincar Anal. Non Linaire*, 11:411–425, 1994.
[132] N.S. Trudinger and X.-J. Wang. On the Monge mass transfer problem. *Calc. Var. Paritial Differential Equations*, 13:19–31, 2001.
[133] N.S. Trudinger and X.-J. Wang. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8:1–32, 2009.
[134] N.S. Trudinger and X.-J. Wang. On strict convexity and C^1-regularity of potential functions in optimal transportation. *Arch. Rational Mech. Anal.*, 192:403–418, 2009.
[135] J. Urbas. On the second boundary value problem for equations of Monge-Ampère type. *J. Reine Angew. Math.*, 487:115–124, 1997.
[136] J. Urbas. *Mass transfer problems*. Lecture Notes, University of Bonn, 1998.
[137] J.A. Viaclovsky. Conformal geometry and fully nonlinear equations. In Phillip A. Griffiths, editor, *Inspired by S.S. Chern.*, volume 11, pages 435–460. World Scientific Publishing Co., Hackensack, 2006.
[138] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.
[139] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, New York, 2009.
[140] X.-J. Wang. On the design of a reflector antenna. *Inverse Problems*, 12:351–375, 1996.
[141] X.-J. Wang. On the design of a reflector antenna II. *Calc. Var. Partial Differential Equations*, 20:329–341, 2004.
[142] L.N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission*, 5:47–52, 1969.
[143] Q. Xia. The formation of a tree leaf. *ESAIM Control Optim. Calc. Var.*, 13:359–377, 2007.