A glimpse into the differential topology and geometry of optimal transport*

Robert J. McCann†

July 11, 2012

Abstract

This note exposes the differential topology and geometry underlying some of the basic phenomena of optimal transportation. It surveys basic questions concerning Monge maps and Kantorovich measures: existence and regularity of the former, uniqueness of the latter, and estimates for the dimension of its support, as well as the associated linear programming duality. It shows the answers to these questions concern the differential geometry and topology of the chosen transportation cost. It also establishes new connections — some heuristic and others rigorous — based on the properties of the cross-difference of this cost, and its Taylor expansion at the diagonal.

1 Introduction

What is optimal transportation? This subject, reviewed by Ambrosio and Gigli [5], McCann and Guillen [58], Rachev and Ruschendorf [69], and Villani [81] [82] among others, has become a topic of much scrutiny in recent years, driven by applications both within and outside mathematics. However, the problem has also lead to the development of its own theory, in which a number of challenging questions arise and some fascinating answers have

*The author is pleased to acknowledge the support of Natural Sciences and Engineering Research Council of Canada Grants 217006-08. ©2012 by the author.

†Department of Mathematics, University of Toronto, Toronto Ontario M5S 2E4 Canada, mccann@math.toronto.edu

been discovered. The present manuscript is intended to reveal some of the differential topology and geometry underlying these questions, their solution and variants, and give some novel and simple yet powerful heuristics for a few highlights from the literature that we survey. It attempts to frame the phenomenology of the subject, without delving deeply into many of the methodologies — both novel and standard — which are used to pursue it. The new heuristics are largely based on the properties of the cross-difference (4), and its Taylor expansion (6) at the diagonal.

Given Borel probability measures $\mu^\pm$ on complete separable metric spaces $M^\pm$, and a continuous bounded function $c(x,y)$ representing the cost per unit mass transported from $x \in M^+$ to $y \in M^-$, the basic question is to correlate the measures μ^+ and μ^- so as to minimize the total transportation cost. In Monge's 1781 formulation [63], we seek to minimize

$$\mathit{cost}(G) := \int_{M^+} c(x, G(x))\, d\mu^+(x) \tag{1}$$

among all Borel maps $G : M^+ \longrightarrow M^-$ pushing μ^+ forward to $\mu^- = G_\# \mu^+$, where the pushed-forward measure is defined by $G_\#\mu^+(Y) = \mu^+(G^{-1}(Y))$ for each $Y \subset M^-$. This question is interesting, because it leads to canonical ways to reparameterize one distribution of mass with another. When the probability measures are given by densities $d\mu^\pm(x) = f^\pm(x)dx$ on manifolds $M^\pm$, we can expect G to satisfy the Jacobian equation $\pm \det[\partial G^i/\partial x^j] = f^+(x)/f^-(G(x))$. Additional desirable properties of G can sometimes be guaranteed by a suitable choice of transportation cost; for example, G will be irrotational for the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$ on Euclidean space [9]. For subsequent purposes, we will often assume the cost $c(x,y)$ and manifolds $M^\pm$ to be smooth, but quite general otherwise.

In Kantorovich's 1942 formulation, we seek to minimize

$$\mathit{cost}(\gamma) := \int_{M^+ \times M^-} c(x,y)\, d\gamma(x,y) \tag{2}$$

over all joint measures $\gamma \geq 0$ on $M^+ \times M^-$ having μ^+ and μ^- as marginals. The form of the latter problem — minimize the linear functional $\mathit{cost}(\gamma)$ on the convex set $\Gamma(\mu^+,\mu^-) := \{\gamma \geq 0 \mid \pi^\pm_\# \gamma = \mu^\pm\}$, where $\pi^+(x,y) = x$ and $\pi^-(x,y) = y$ — makes it easy to show the Kantorovich infimum is attained. A result of Pratelli [67] following Ambrosio and Gangbo asserts that its value coincides with the Monge infimum

$$\min_{\gamma \in \Gamma(\mu^+,\mu^-)} \mathit{cost}(\gamma) = \inf_{\mu^- = G_\#\mu^+} \mathit{cost}(G) \tag{3}$$

if c is continuous and μ^+ is free of atoms. However it is not straightforward to establish uniqueness of the Kantorovich minimizer, nor whether the Monge infimum is attained, and if so, whether the mapping G which attains it is continuous. A sufficient condition **(A1)**$'_+$ for existence (and uniqueness) of optimizers G (and γ) was found by Gangbo [30] and Levin [44], building on work of many authors, including Brenier, Caffarelli, Gangbo, McCann, Rachev and Rüschendorf. When $M^\pm \subset \mathbf{R}^n$ are the closures of open domains, sufficient conditions for the existence of a *smooth* minimizer $G : M^+ \longrightarrow M^-$ were provided by Ma, Trudinger and Wang [53], building on work of Delanoë, Caffarelli, Urbas and Wang, and later refined through work of Delanoë, Figalli, Ge, Kim, Liu, Loeper, McCann, Rifford, Trudinger, Villani and Wang, among others. See Appendix A for a statement of their conditions **(A0)**$'$-**(A4)**$'$. At the same time, we introduce a new but equivalent formulation of conditions **(A0)**$'$-**(A4)**$'$ in terms of the cross-difference (4), which emphasizes their purely topological **(A0)**-**(A2)** and geometric **(A3)**-**(A4)** nature, exposing their naturality and relevance. This process of reformulation, begun with Kim in [38], is completed here, as part of a series of questions and responses.

2 Why do Kantorovich minimizers concentrate on low-dimensional sets?

Abstractly, one expects a linear functional $cost(\gamma)$ on a convex set $\Gamma(\mu^+, \mu^-)$ to attain its infimum at one of the extreme points. So it is interesting to understand the extreme points of $\Gamma(\mu^+, \mu^-)$. Such extreme points are sometimes called *simplicial* measures. Despite much progress, surveyed in [2], a characterization of simplicial measures in terms of their support has long remained elusive and is probably too much to hope for. Recall that a measure $\gamma \geq 0$ is *simplicial* if it is not the midpoint of any segment in $\Gamma(\pi^+_\# \gamma, \pi^-_\# \gamma)$. Ahmad, Kim and McCann [2] showed each simplicial measure γ vanishes outside the union of a graph $\{(x, G(x)) \mid x \in M^+\}$ and an antigraph $\{(H(y), y) \mid y \in M^-\}$, generalizing Hestir and Williams [35] result from the special case of Lebesgue measure $\mu^\pm$ on the unit interval $M^\pm = [0, 1]$. This shows γ concentrates on a set whose topological dimension should not exceed $\max\{n^+, n^-\}$, where $n^\pm = \dim M^\pm$. Taking $n^+ \leq n^-$ without loss of generality, if the measure μ^- fills the space M^-, then γ cannot concentrate on any subset of lower dimension than n^-, so it would seem we have identified the topological dimension of the set on which γ concentrates to be precisely $n^- = \max\{n^+, n^-\}$. Unfortunately, this simple argument is somewhat deceptive. Although the graph and antigraph of [35] and [2] enjoy further structure, they are not generally γ-measurable; a priori it is conceivable that their closures might actually fill the product space $N = M^+ \times M^-$.

With some assumptions on the topology of the cost function c and spaces $M^\pm$, it is possible to better estimate the size of support of the particular extreme points of interest using the more robust notion of Hausdorff dimension. The basic object of geometrical relevance will be the support $\operatorname{spt}\gamma$ of the Kantorovich optimizer, defined as the smallest closed subset $S \subset M^+ \times M^+$ carrying the full mass of γ. If the Monge infimum (3) is attained by a map $G : M^+ \longrightarrow M^-$ and the Kantorovich minimizer is unique, it will turn out that $\operatorname{spt}\gamma$ agrees (γ-a.e.) with the graph of G; when this map is a diffeomorphism, then γ concentrates on a subset of dimension $n^+ = \dim M^+$ in $M^+ \times M^-$. We shall show why this might be expected more generally, assuming $M^\pm$ to be (smooth) manifolds henceforth.

Setting $N := M^+ \times M^-$ and $S = \operatorname{spt}\gamma$, consider the cross-difference [56]

$$\delta(x, y; x_0, y_0) := c(x, y_0) + c(x_0, y) - c(x, y) - c(x_0, y_0) \tag{4}$$

defined on N^2. An observation — special cases of which date back to Monge — asserts $\delta(x, y; x_0, y_0) \geq 0$ on $S^2 \subset N^2$; in other words, we cannot lower the cost by exchanging partners between (x, y) and (x_0, y_0); for a modern proof, see Gangbo and McCann [31]. This fact is called the *c-monotonicity* of S.

If $c \in C^2$, then $(x_0, y_0) \in N$ is a critical point for the function $\delta^0(x, y) := \delta(x, y; x_0, y_0)$, whose Hessian

$$h = \tfrac{1}{2} \operatorname{Hess} \delta^0(x_0, y_0) \tag{5}$$

is then well-defined (though it need not be at points $(x, y) \neq (x_0, y_0)$ which are non-critical). Now for $(x_0, y_0) \in S$, we have $\delta^0(x, y) \geq 0$ on S, with equality at (x_0, y_0). On the other hand, the symmetries of the cross-difference δ ensure that the Hessian h contributes the only non-vanishing term in the second order Taylor expansion of δ^0: more explicitly

$$\begin{aligned} \delta^0 &\quad (x_0 + \Delta x, y_0 + \Delta y) \\ &= \quad h((\Delta x, \Delta y), (\Delta x, \Delta y)) + o(|\Delta x|^2 + |\Delta y|^2) \\ &= \quad -\sum_{i=1}^{n^+} \sum_{j=1}^{n^-} D^2_{x^i y^j} c(x_0, y_0) \Delta x^i \Delta y^j + o(|\Delta x|^2 + |\Delta y|^2) \end{aligned} \tag{6}$$

as $(\Delta x, \Delta y) \to 0$. It is then not so surprising to discover that the Hessian h controls the geometry and dimension of the support of any Kantorovich optimizer γ near (x_0, y_0) in various ways, as we now make precise following Pass [66] and my joint works with Kim [38], Pass and Warren [60].

Let (k_+, k_0, k_-) be the $(+, 0, -)$ signature of h, meaning $k_+, k_0, k_- \in \mathbf{N}$ count the number of positive, zero, and negative eigenvalues of h in one and hence any choice of coordinates.

Claim 2.1 (Signature and rank) *For each $(x_0, y_0) \in N$, the signature of* (5) *is given by $(k_+, k_0, k_-) = (k, n_+ + n_- - 2k, k)$ where $k \leq \min\{n^+, n^-\}$ and $n^\pm = \dim M^\pm$. We may henceforth refer to the integer $2k$ as the* rank *of $c \in C^2$ (and of h) at (x_0, y_0); it depends lower-semicontinuously on (x_0, y_0).*

Proof. The sum $k_+ + k_0 + k_- = n_+ + n_-$ must agree with the total dimension of $N = M^+ \times M^-$. Since any perturbation direction $(\Delta x, \Delta y)$ in which δ^0 grows, corresponds to another direction $(-\Delta x, \Delta y)$ in which δ^0 shrinks (6), it follows that $k_+ = k_-$. Thus $(k_+, k_0, k_-) = (k, n_+ + n_- - 2k, k)$.

In fact, since the matrix h is symmetric, in any coordinate system we can find a basis of orthogonal eigenvectors for h. The preceding argument shows that if $(\Delta x, \Delta y)$ is an eigenvector with eigenvalue $\lambda > 0$ then $(-\Delta x, \Delta y)$ is an eigenvector with eigenvalue $-\lambda$. In this case $\Delta x = -\frac{1}{2\lambda} \sum_{j=1}^{n^-} D^2_{xy^j} c(x_0, y_0) \Delta y^j$ is determined by Δy and vise versa, so at most $k \leq \min\{n^+, n^-\}$ eigenvectors can correspond to positive eigenvalues [66].

Lowersemicontinuity of $k = k(x_0, y_0)$ follows from the fact that $c \in C^2$. ■

The Hessian h of the cross-difference also determines the *spacelike, timelike,* and *lightlike* cones Σ^+, Σ^- and $\Sigma^0 \subset T_{(x_0,y_0)}N$ according to the definitions $\Sigma^\pm = \{V \in T_{(x_0,y_0)}N \mid \pm h(V, V) \geq 0\}$ and $\Sigma^0 = \Sigma^+ \cap \Sigma^-$.

Definition 2.2 (Spacelike, timelike, lightlike) *A subset $S \subset N$ is* spacelike *if each (not necessarily continuous) curve $t \in [-1, 1] \longmapsto z(t) \in S$ differentiable at $t = 0$ satisfies $h(\dot z(0), \dot z(0)) \geq 0$, where $\dot z(0)$ is the tangent vector and h denotes the Hessian* (5) *at $(x_0, y_0) = z(0)$. Similarly, S is timelike (or lightlike) if the inequality is reversed (or if both inequalities hold).*

Since we want to allow sets S which are rough and potentially incomplete, it is important to permit curves in the definition above whose continuity at $t = 0$ may not extend to any neighbourhood of $t = 0$.

Lemma 2.3 (c-monotone implies spacelike) *Any c-monotone set $S \subset N$ is spacelike.*

Proof. Take any curve $(x(t), y(t)) \in S$, not necessarily continuous but differentiable at $(x_0, y_0) = (x(0), y(0))$, with tangent vector $V = (\dot{x}(0), \dot{y}(0))$. Since $\delta^0(x(t), y(t)) \geq 0$, from (5)–(6) we see $h(V, V) \geq 0$, as desired. ■

Corollary 2.4 (Dimensional bounds) *If c is C^2 and has rank $2k$ at a point (x_0, y_0) where S has a well-defined tangent space T, then c-monotonicity of $S \subset N$ implies the dimension of this tangent space satisfies* $\dim T \leq n_+ + n_- - k$.

Proof. Fix coordinates on N. As a consequence of the (Courant-Fischer) min-max formula for eigenvalues of h at (x_0, y_0), the signature $(k_+, k_0, k_-) = (k, n_+ + n_- - 2k, k)$ of h limits the maximal number of linearly independent tangent vectors to N which are not timelike to $k_+ + k_0 = n_+ + n_- - k$. Since the preceding lemma shows the tangent space T of S to be spanned by such a set of vectors, its dimension satisfies the asserted bound. ■

The following much stronger result of Pass [66] asserts S is contained in a spacelike Lipschitz submanifold of the prescribed dimension — hence implies differentiability a.e. as a consequence instead of a hypothesis. The case $k = n_+ = n_-$ was worked out earlier by McCann, Pass and Warren [60], by adapting an idea of Minty [61] [3] from the special case $c(x, y) = -x \cdot y$.

Theorem 2.5 (Rectifiability [66]) *If c has rank $2k$ at (x_0, y_0) and is C^2 nearby, then on a (possibly smaller) neighbourhood $N_0 \subset M^+ \times M^-$ of (x_0, y_0), c-monotonicity of $S \subset N_0$ implies $S \subset L$ where $L \subset N_0$ is a spacelike Lipschitz submanifold of dimension* $\dim L \leq n_+ + n_- - k$ *with* $n_\pm = \dim M^\pm$.

Idea of proof. A kernel of the proof can be apprehended already in the one-dimensional case $n^\pm = 1$. When c has rank zero, taking $L = N_0$ implies the result, so assume c has full rank ($2k = 2$), meaning either $\partial^2 c / \partial x \partial y < 0$ or $\partial^2 c / \partial x \partial y > 0$ near (x_0, y_0). In the first case, c-monotonicity of S implies $S \cap R$ is contained in a non-decreasing subset of any sufficiently small two-dimensional rectangle $R = B_\epsilon(x_0) \times B_\epsilon(y_0)$. This monotonicity is well-known in both mathematical [51] and economic contexts [76] [62]. Rotating coordinates by setting $u = (x + y)/\sqrt{2}$ and $v = (y - x)/\sqrt{2}$, the monotonicity is equivalent to asserting that S is contained in the graph of $\{(u, V(u))\}$ of a function $v = V(u)$ with Lipschitz constant one. In the second case,

c-monotonicity would imply $S \cap R$ is non-increasing, hence contained in a 1-Lipschitz graph of u over v.

The same argument carries over immediately to the bilinear cost $c(x, y) = -x \cdot y$ in higher dimensions $n^+ = n^-$ [61]. For other costs with rank $2k = 2n^+ = 2n^-$, one can make a similar argument after a linear change of coordinates $\tilde{x} = x - x_0$ and $\tilde{y} = \Lambda(y - y_0)$ chosen so that in the new coordinates the cost takes the form $\tilde{c}(\tilde{x}, \tilde{y}) = -\tilde{x} \cdot \tilde{y} + o(\Delta\tilde{x}^2 + \Delta\tilde{y}^2) + f(\tilde{x}) + f(\tilde{y})$ [60]. The cases $k < \min\{n^+, n^-\}$ and $n^+ \neq n^-$ are worked out in [66]. ■

When the rank of c is maximal (i.e. $k = \min\{n^+, n^-\}$), then the dimensional bound is $\dim L \leq \max\{n^+, n^-\}$. Taking $n^+ \leq n^-$ without loss of generality, if the measure μ^- fills M^- (say, by being mutually absolutely continuous with respect to Lebesgue measure in any coordinate patch), the dimension of the Lipschitz submanifold L on which γ concentrates cannot be less than n^-, in which case we see the bound given by the theorem is sharp: $\dim L = n^-$.

Example 2.6 (Submodular costs on the line) *If $M^\pm = \mathbf{R}$ there is a unique measure in $\Gamma(\mu^+, \mu^-)$ whose support $S = \operatorname{spt}\gamma$ forms a non-decreasing subset of the plane. This measure is the unique minimizer of Kantorovich's problem* (3) *for each cost $c \in C^1(\mathbf{R}^2)$ satisfying $\partial^2 c/\partial x \partial y < 0$; see e.g. [56]. Apart from at most countably many vertical segments, the set S is contained in the graph of some $G : \mathbf{R} \longrightarrow \mathbf{R} \cup \{\pm\infty\}$ non-decreasing. Unless μ^+ has atoms, the vertical segments in S are γ negligible, in which case $\gamma = (id \times G)_\# \mu^+$ and Monge's infimum is attained uniquely by G.*

Example 2.7 (Transporting mass between spheres) *Transporting mass on the surface of the earth has lead to consideration of the cost function $c(x, y) = \frac{1}{2}|x - y|^2$ restricted to the boundary of the unit sphere $x, y \in \partial B_1^{n+1}(\mathbf{0}) \subset \mathbf{R}^{n+1}$ so that $0 \leq c \leq 2$ [14][2][26][59], a problem considered earlier in the context of shape recognition [32][1]. The restricted cost has rank $2n$ except on the degenerate set $c = 1$, where it has rank $2n - 2$. Thus any c-cyclically monotone subset S of the $2n$-dimensional product space $\partial B_1^{n+1}(\mathbf{0}) \times \partial B_1^{n+1}(\mathbf{0})$ has dimension at most n except along the degenerate set, where it has dimension at most $n + 1$ (in spite of the fact that the degenerate set is $2n - 1$ dimensional). Since the degenerate set separates the orientation preserving and orientation reversing parts S^+ and S^- of S, this means that S^+ cannot intersect S^- transversally (except in dimension $n = 1$); instead, if S^+ meets S^- at a point where both have n-dimensional tangent* *spaces, these spaces must have $n-1$ directions in common. For example, if $n=3$, and both S^+ and S^- are generically 3-dimensional, but their union is contained in a 4-dimensional Lipschitz submanifold, whereas the cost degenerates on a smooth 5-dimensional hypersurface.*

In summary, c-monotonicity implies rectifiability of $S = \operatorname{spt}\gamma \subset N = M^+ \times M^-$ in a dimension determined locally by the rank of the Hessian h of the cross-difference $\delta^0 \in C^2$; moreover S must be spacelike with respect this Hessian (5). If h is non-degenerate, we will eventually see that h can be viewed as a pseudo-metric on N whose Riemannian sectional curvatures combine with $\mu^\pm$ to determine smoothness of S.

3 When do optimal maps exist?

We now turn to the more classical question of attainment of the infimum (3). To expect existence of Monge maps, we generally need μ^+ to be more than atom-free. We need μ^+ not to concentrate positive mass on any lower dimensional submanifold of M^+, or more precisely on any hypersurface parameterized locally in coordinates as the graph of a difference of convex functions. This condition, proposed by Gangbo and McCann [31], is sharp in a sense made precise by Gigli [33], and implies Lipschitz continuity and C^2-rectifiability of the hypersurfaces in question. Absolute continuity of μ^+ in coordinates — i.e. the existence of a density f^+ such that $d\mu^+(x) = f^+(x)dx$ — is more than enough to guarantee this. However, we also require further structure of the transportation cost.

For $c \in C^1(N)$, the Gangbo [30] and Levin [44] criterion for existence of Monge solutions $G : M^+ \longrightarrow M^-$ given in Appendix A is equivalent to:
(A1)$_+$ *For each $x_0 \in M^+$ and $y_0 \neq y_1 \in M^-$, assume $x \in M^+ \longmapsto \delta^0(x, y_1)$ has no critical points, where $\delta^0(x, y_1) = \delta(x, y_1; x_0, y_0)$ is from* (4).
Naturally, this implies $n^+ \geq n^-$, due to the fact we cannot generally hope to use a (rectifiable) map G on a low dimensional space to spread a measure over a higher dimensional space. In fact, **(A1)$_+$** implies something stronger: namely that every solution of the Kantorovich problem is a Monge solution. This in turn implies uniqueness of the Kantorovich (and hence Monge) solution, for the following reason. Suppose two Kantorovich solutions exist, and both correspond to Monge solutions: $\gamma_0 = (id \times G_0)_\# \mu^+$ and $\gamma_1 = (id \times G_1)_\# \mu^+$. Linearity of the Kantorovich problem implies

$\gamma_2 := (\gamma_0 + \gamma_1)/2$ is again a solution, hence by **(A1)**$_+$ must concentrate on the graph of a map $G : M^+ \longrightarrow M^-$. It is then easy to argue $\gamma_i = (id \times G)_\# \mu^+$ for $i = 0, 1, 2$ as in e.g. [2]. This implies $\gamma_0 = \gamma_1$; moreover $G_0 = G = G_1$ μ-a.e. Thus we arrive at the following theorem [30] [44] [2] [33]:

Theorem 3.1 (Existence and uniqueness of optimal maps) *Let $\mu^\pm$ be probability measures on manifolds $M^\pm$, with a cost $c \in C^1(M^+ \times M^-)$ which is bounded and satisfies* **(A1)**$_+$. *If μ^+ assigns zero mass to each Lipschitz hypersurface in M^+, then Kantorovich's minimum is uniquely attained, and the minimizer $\gamma = (id \times G)_\# \mu^+$ vanishes outside the graph of a map G solving Monge's problem. In fact, not all Lipschitz hypersurfaces are required: it is enough that μ^+ vanish on each hypersurface locally parameterizable in coordinates as the graph of a difference of two convex functions.*

Notice **(A1)**$_+$ asserts the restriction of δ^0 to each horizontal fibre $M^+ \times \{y_1\}$ has no critical points, except on the fibre $y_1 = y_0$ where δ^0 vanishes identically. To guarantee invertibility of the map G, we need the same condition to hold for the reflected cost $c^*(y, x) := c(x, y)$, meaning the roles of M^+ and M^- are interchanged. If both c and c^* satisfy **(A1)**$_+$, we say **(A1)** holds. Thus **(A1)** is equivalent to asserting that (x_0, y_0) is the only critical point of $\delta^0(x, y)$.

Many interesting costs, such as $c(x, y) = h(x - y)$ with h strictly convex or concave on $M^\pm = \mathbf{R}^n$ satisfy these hypotheses. The most classical of these is the Euclidean distance squared [8] [16] [75] [73]. Regularity of the convex gradient map it induces, generalizing Example 2.6, was established by Delanoë [17] for $n = 2$ and Caffarelli [10] [11] and Urbas [80] for $n \geq 3$.

Example 3.2 (Euclidean distance squared) *If $M^\pm \subset \mathbf{R}^n$ and μ^+ vanishes on all hypersurfaces, there is a unique measure in $\Gamma(\mu^+, \mu^-)$ concentrated on the graph of the gradient of a convex function $u : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{+\infty\}$. This measure is the unique minimizer of Kantorovich's problem* (3) *for the cost $c(x, y) = \frac{1}{2}|x - y|^2$ [9] [54]. If $d\mu^\pm = f^\pm d\mathcal{H}^n$ are both absolutely continuous with respect to Lesbesgue, the Monge-Ampère equation $f^+(x) = f^-(Du(x)) \det D^2 u(x)$ holds μ^+-a.e. [55]. If moreover, $\log f^\pm \in L^\infty(M^\pm)$ with M^- convex and $\mathcal{H}^n(\partial M^+) = 0$, then $u \in C^{1,\alpha}_{loc}(M^+)$ for some $\alpha > 0$ [10] estimated in [29]. If, in addition $f^\pm \in C^{1,\bar\beta}$ and M^+ and M^- are both smooth and strongly convex — meaning the principle curvatures of their boundaries are all strictly positive — then $u \in C^{2,\beta}(\bar{M}^+)$ for all $0 < \beta < \bar\beta < 1$ [17] [11] [80]. Higher regularity follows from smoothness of $f^\pm$.*

On the other hand, $\mathbf{(A1)}_+$ also fails for many interesting geometries. We mention two such examples. In the first — the cost function of interest to Monge [63] — optimal maps turn out to exist but are not unique. Their non-uniqueness was quantified with Feldman [22]. In the second, Monge's infimum turns out not to be attained, despite the fact that the Kantorovich minimizer is unique.

Example 3.3 (Uniqueness fails for Monge's cost) *Let open sets $M^\pm \subset \mathbf{R}^n$ have finite volume and $c(x,y) = |x-y|$. Monge was originally interested in transporting the uniform measure $\mu^\pm = \frac{1}{\mathcal{H}^n(M^\pm)}\mathcal{H}^n$ from one domain to the other, when $n = 3$ and $\mathcal{H}^n$ denotes the n-dimensional Hausdorff measure, and coincides with Lebesgue measure in this case [63]. Taking M^+ disjoint from M^- ensures smoothness of c. Notice that when $n = 1$ and M^+ and M^- are disjoint intervals, every $\gamma \in \Gamma(\mu^+,\mu^-)$ has the same total cost cost(γ). In this case the solution to Kantorovich's problem is badly non-unique. Clearly $\mathbf{(A1)}_+$ also fails in this case. In higher dimensions, the situation is slightly less degenerate since the cost takes a range of values on $\Gamma(\mu^+,\mu^-)$, but it remains true that its extrema are not uniquely attained. In this setting, it can be a difficult problem to show that Monge's infimum is attained. This problem was first solved by Sudakov in the plane $n = 2$; he asserted a result in all dimensions but it was later discovered that one of his claims sometimes fails if $n > 2$. This existence result was extended to higher dimensions by Evans and Gangbo, assuming $\mu^\pm$ to be given by Lipschitz continuous densities on $\mathbf{R}^n$ [21], and for general absolutely continuous densities $\mu^\pm$ by Ambrosio [4], Trudinger-Wang [78] and Caffarelli-Feldman-McCann [12] simultaneously and independently. The last group also considered costs given by non-Euclidean norms, but with smooth and strongly convex unit balls, restrictions removed in a seqeunce of papers by different teams of authors including Ambrosio, Bernard, Buffoni, Bianchini, Caravenna, Kirchheim, and Pratelli, and culminating in work of Champion and DePascale [13].*

On the other hand, if M^+ is a compact manifold without boundary, it is evident that $x \in M^+ \longmapsto \delta^0(x, y_1)$ must attain at least one maximum and one minimum so that — as long as the cost is assumed differentiable — it is clear that $\mathbf{(A1)}_+$ cannot be satisfied. In this case, it will not always be true that Monge's infimum (3) is attained, as my examples with Gangbo [32] show:

Example 3.4 (Transporting mass between spheres, revisited) *Restrict $c(x,y) = \frac{1}{2}|x-y|^2$ to $M^\pm = \partial B_1(\mathbf{0}) \subset \mathbf{R}^{n+1}$ so that $0 \le c \le 2$, as in Example 2.7. Take $\mu^\pm$ to be mutually absolutely continuous with respect to surface area $\mathcal{H}^n$ on their respective spheres, but take most of the mass of μ^+ to be concentrated near the north pole and most of the mass of μ^- to be concentrated near the south pole. Then Monge's infimum* (3) *will not be attained, despite the fact that the Kantorovich minimizer γ is unique. The intersection of $S = \operatorname{spt}\gamma$ with the set $c \le 1$ is contained in the graph of a map $G : M^+ \longrightarrow M^-$, while the intersection $S \cap \{c \ge 1\}$ is contained in the graph of a map $H : M^- \longrightarrow M^+$ — sometimes called an* antigraph. *If the densities $f^\pm = d\mu^\pm/d\mathcal{H}^n$ are both bounded, so that $\log f^\pm \in L^\infty$, then G is a homeomorphism of ∂B_1 and H may be taken to be continuous [32]; both maps enjoy a local Hölder exponent of continuity $\alpha = 1/(4n-1)$ except possibly where their graphs touch the set $\{c = 1\}$ where the rank of c drops from $2n$ to $2n-2$ [59]. It may be possible to improve this Hölder exponent to $\alpha = 1/(2n-1)$ using techniques of Liu [46], but even when $f^\pm$ are smooth we have no idea how to prove G will be smoother, nor how to extend Hölder continuity of G up to the degenerate set $\{c = 1\}$.*

Notice that global differentiability of the cost is crucial to this discussion. For costs whose differentiability fails — even on a small set such as the Riemannian cut locus — the theorem which follows gives many natural examples where existence and uniqueness both hold.

Theorem 3.5 (Minimizing Riemannian distance squared) *Let $c(x,y) = d^2(x,y)/2$ be the square distance induced by some Riemannian metric on a compact manifold $M^+ = M^-$. If μ^+ is absolutely continuous (with respect to Riemannian volume) then the Kantorovich minimizer is unique in* (3)*, and takes the form $\gamma = (id \times G)_\#\mu^+$ for a map solving Monge's problem [57]. In case $M^\pm$ are round spheres [49] (or quotients [18], submersions [37] or products thereof [25]), and both $\mu^\pm$ are given by smooth positive densities with respect to surface area, then the map G will be a smooth diffeomorphism.*

Notice that the existence and uniqueness asserted in Theorem 3.5 is not quite a corollary of Theorem 3.1, since compactness of the manifold $M^\pm$ forces the cut-locus to be non-trivial. Here the cut-locus is defined as (the closure of) the set of points where differentiability of the cost $c = d^2/2$ fails.

4 When are optimal measures unique?

The preceding section shows that if the cross-difference $\delta^0(x,y) = \delta(x,y;x_0,y_0)$ has no critical points unless $x = x_0$ or $y = y_0$, then Monge's problem is soluble and the Kantorovich problem admits a unique solution. Although very useful when it applies, this criterion is not satisfied in all cases of interest. — for example, when trying to minimize the restriction of the quadratic cost $c(x,y) = |x-y|^2/2$ to the Euclidean unit sphere $M^\pm = \partial B_1(\mathbf{0}) \subset \mathbf{R}^{n+1}$. In such situations, my results with Chiappori, Nesheim [14], Ahmad and Kim [2] may be useful:

Theorem 4.1 (Uniqueness of minimizer for subtwisted costs) *Fix Borel probability measures $\mu^\pm$ on manifolds $M^\pm$, with μ^+ vanishing on each hypersurface in M^+, and a bounded cost $c \in C^1(M^+ \times M^-)$. Suppose for each $x_0 \in M^+$ and $y_0 \neq y_1 \in M^-$, the cross-difference $\delta^0(x,y) := \delta(x,y;x_0,y_0)$ from* (4) *satisfies*

$$x \in M^+ \longmapsto \delta^0(x,y_1) \quad \text{has at most two critical points, namely, a unique global minimum and a unique global maximum.} \tag{7}$$

Then the Kantorovich problem has a unique solution, and it takes the form $\gamma = (id \times G)_\#\mu + (H \times id)_\#(\mu^- - G_\#\mu)$ for some maps $G : M^+ \longrightarrow M^-$ and $H : M^- \longrightarrow M^+$ and non-negative measure $\mu \leq \mu^+$ such that $\mu^- - G_\#\mu$ vanishes on the range of G.

In other words, the unique Kantorovich solution concentrates on the union of the graph and an antigraph, of $G : M^+ \longrightarrow M^-$ and of $H : M^- \longrightarrow M^+$ respectively. Notice that if the manifold M^+ is compact, hypothesis (7) restricts its Morse structure to be that of the sphere, so the theorem generalizes of Example 3.4: However apart from the continuity results of [32] [59] and [**?**], it is not known when G and H can be expected to be smooth. It is even more shocking that no criterion analogous to Theorem 4.1 is known which guarantees uniqueness of Kantorovich minimizer on the torus — or indeed on any other compact manifolds $M^\pm$ apart from the sphere.

5 When are optimal maps continuous? Smooth?

Examples 3.2, 3.4 and Theorem 3.5 complement Theorems 2.5 and 3.1 by providing a variety of settings where the optimal map G is continuous and/or support of the optimal measure can actually be shown to be smooth. In each case, we need the cost to be suitable, the domain geometry to be favorable, and the measures to be positive, bounded and possibly smooth.

Following the analysis of a number of such examples, including the restriction of $c(x, y) = -\log|x-y|$ to the unit sphere [83] [84], a general theory for addressing such questions has begun to be developed, starting from the pioneering work of Ma, Trudinger and Wang [53], who identified conditions on the transportation cost c which are close to being necessary and sufficient for smoothness of G. Their work is set on bounded domains $M^\pm \subset \mathbf{R}^n$, and as we now explain, each of their conditions can be reformulated in terms of the topology and geometry of the cross-difference $\delta^0(x, y) = \delta(x, y; x_0, y_0)$ from (4) and its Hessian $h = \frac{1}{2}\mathrm{Hess}_{(x_0,y_0)}\delta^0$.

Where c has full rank $2n$, the Hessian h is non-degenerate and can be understood as a pseudo-Riemannian metric tensor on the product space. According to Claim 2.1, this pseudo-metric tensor is not positive definite, but instead has the same number of spacelike and timelike dimensions. At each point point $(x_0, y_0) \in N$, the light-cone separating these spacelike from timelike directions consists of the tangent spaces to $\{x_0\} \times M^-$ and $M^+ \times \{y_0\}$. However, just as in Riemannian (and Lorentzian) geometry, the pseudo-metric tensor h induces a geometry on the product space $N = M^+ \times M^-$, including geodesics and a pseudo-Riemannian curvature tensor $R_{i'j'k'l'}$, which assigns sectional curvature

$$\sec^{(\bar{N},h)}_{(x_0,y_0)} P \wedge Q = \sum_{1 \le i',j',k',l' \le 2n} R_{i'j'k'l'} P^{i'} Q^{j'} P^{k'} Q^{l'}$$

to each pair of vectors $P, Q \in T_{x_0,y_0}N$. The explicit formulae expressing geodesics and the curvature tensor (12) in terms of h can be found in [38] or deduced from Appendix A; they are precisely analogous to the Riemannian case.

In terms these notions, we may now state conditions equivalent to those of Ma, Trudinger and Wang **(A1)′**–**(A4)′** found in Appendix A below:

(A0) $c \in C^4(\bar{N})$, and for each $(x_0, y_0) \in \bar{N} = \bar{M}^+ \times \bar{M}^- \subset \mathbf{R}^n \times \mathbf{R}^n$:

(A1) $(x, y) \in \bar{N} \mapsto \delta^0(x, y)$ from (4) has no critical points save (x_0, y_0);

(A2) c has rank $2n$, so $h =\mathrm{Hess}_{(x_0,y_0)}\delta^0$ defines a pseudo-metric tensor;

(A3) $\sec^{(\bar{N},h)}_{(x_0,y_0)}(p \oplus \mathbf{0}) \wedge (\mathbf{0} \oplus q) \ge 0$ for each lightlike $(p, q) \in T_{(x_0,y_0)}\bar{N}$;

(A4) the sets $\{x_0\} \times M^-$ and $M^+ \times \{y_0\}$ are h-geodesically convex.

Here a subset $Z \subset \bar{N}$ is said to be *h-geodesically convex* if each pair of points (x_0, y_0) and $(x_1, y_1) \in Z$ can be joined by an geodesic in $\bar{N}$ lying entirely within Z, geodesics being defined relative to the pseudo-metric h.

The most intriguing of these conditions is the curvature condition **(A3)**. A large body of example costs which satisfy [53] [49] [18] [38] [27] [28] [43] [41] [42] [19] or violate it [53] [48] have now been established. Among the former we may mention the restriction of the Euclidean distance squared to the graphs $M^\pm \subset \mathbf{R}^{n+1}$ of any pair of 1-Lipschitz convex functions [53], as well as the Riemannian distance squared on the round sphere [49], and any products [38], submersions [38] or perturbations [18] [27] [28] thereof. Among the latter we may mention the Riemannian distance squared on any manifold (M, g_{ij}) with a non-negative sectional curvature somewhere [48], and the restriction of the Euclidean distance squared to the graphs of two functions in $\mathbf{R}^{n+1}$, one of which is convex and the other non-convex [53]. Thus the distance squared in hyperbolic space $c = d^2_{\mathbf{H}^n}$ violates **(A3)**, though $c = -\cosh d_{\mathbf{H}^n}$ satisfies it [45] [42].

To conclude continuity or higher regularity of G at present requires a slight strengthening of one of the geometric conditions **(A3)** or **(A4)**. If the inequality in **(A3)** holds strictly whenever the h-orthogonal vectors $p \oplus \mathbf{0}$ and $\mathbf{0} \oplus q$ are non-vanishing, we denote that by $(\mathbf{A3})_\mathbf{s}$. If instead the geodesic convexity of the sets in **(A4)** is strong (i.e. 2-uniform, in the sense of Example 3.2 or Appendix A), we denote that by $(\mathbf{A4})_\mathbf{s}$. Under these assumptions the following extensions of Theorem 3.1 and Example 3.2 have been proved, in works of Ma, Trudinger, Wang, Loeper, Liu, Figalli, Kim and myself.

Theorem 5.1 (Continuity and smoothness of optimal maps) *Assume* **(A0)**-**(A4)** *hold, and* $d\mu^\pm = f^\pm d\mathcal{H}^n$ *are given by densities satisfying* $\log f^\pm \in L^\infty(U^\pm)$ *with* $U^- = M^-$ *and* $U^+ \subset M^+$ *open. (a) If* $(\mathbf{A3})_\mathbf{s}$ *holds, the map* $G \in C^\alpha_{loc}(U^+, M^-)$ *is Hölder continuous [48], with an exponent* $\alpha = 1/(2n-1)$ *known to be sharp [46]. (b) If* $(\mathbf{A3})_\mathbf{s}$ *fails but* $(\mathbf{A4})_\mathbf{s}$ *holds, the same conclusion persists but with an unknown exponent* α *independent of* c*, but presumed to depend on* $\|\log(f^+/f^-)\|_{L^\infty(U^+\times M^-)}$. *Either way, higher interior regularity of* G *follows from smoothness of* $f^\pm$ *[47]. If,* $U^\pm = M^\pm$ *and* $f^\pm$ *are smooth in case (b), the smoothness of* G *shown in [53] extends up to the boundary [79].*

It is possible to construct smooth bounded $f^\pm$ for which continuity of G fails in the absence of either **(A3)** or **(A4)** as was done by Loeper [48] and by Ma, Trudinger and Wang [53] respectively. Still, there are few results quantifying the discontinuities of G, except for the cost $c(x,y) = \frac{1}{2}|x-y|^2$ of Example 3.2 [85] [23] [24], for which examples of discontinuous maps go back to Caffarelli [10].

6 Closed forms and c-cyclical monotonicity

The sections above have discussed many necessary conditions for optimality of γ, but few sufficient conditions. In fact, for bounded continuous $c \in C(M^+ \times M^-)$, a condition on the support $S = \operatorname{spt}\gamma$ well-known to be necessary and sufficient for optimality in $\Gamma(\pi^+_\#\gamma, \pi^-_\#\gamma)$ is given by:

Definition 6.1 (c-cyclical monotonicity) *A set $S \subset M^+ \times M^-$ is c-*cyclically monotone *if and only if each $k \in \mathbf{N}$, sequence $(x_1, y_1), \ldots, (x_k, y_k) \in S$, and permutation τ on k letters satisfy the following inequality:*

$$\sum_{i=1}^{k} c(x_i, y_i) \leq \sum_{i=1}^{k} c(x_{\tau(i)}, y_i). \tag{8}$$

This result can be found in Pratelli [68] or Schachermayer-Teichmann [74], building on earlier works of Knott-Smith, Gangbo-McCann, Rüschendorf, and Ambrosio-Pratelli. The case $k = 2$ corresponds to the c-monotonicity condition which implies that S is h-spacelike. The result quoted above shows the cross-difference $\delta(x, y; x_0, y_0)$ is just the first in an infinite sequence of functions whose non-negativity on S^k for each $k \in \mathbf{N}$ characterizes optimality of γ. In fact, since all permutations are made up of cycles, for each k it is enough to check (8) for the cyclic permutation $\tau(i) = i + 1$ if $i < k$ with $\tau(k) = 1$. This family of conditions has a differential topological content whose relevance we now try to make clear.

Choose any map $G : U^+ \subset M^+ \longrightarrow M^-$ defined on a subset $U^+ \subset M^-$, whose graph lies inside S. Any differentiable loop $\sigma : S^1 \longrightarrow M^+$ may be approximated by $x_i = \sigma(\theta_i)$ for a partition $0 < \theta_1 < \cdots < \theta_k \leq 2\pi$ as fine as we please. The non-negative sums (8) then approximate Riemann sums for the integral

$$0 \leq \int_0^{2\pi} D_x c(\sigma(\theta), G(\sigma(\theta))) \cdot \sigma'(\theta) d\theta$$

arbitrarily closely. If the form $x \in U^+ \longmapsto D_x c(x, G(x))$ is continuous on an open set $U^+ \subset M^+$ containing σ, then the Riemann integral exists. Since the curve can be traversed in either direction, the non-negative integral must actually vanish, hence the form must be closed: for U^+ simply connected, there would exist $u \in C^1_{loc}(U^+)$ such that $D_x c(x, G(x)) = Du(x)$. Similarly, if G could be continuously inverted on a simply connected domain $U^- \subset M^-$, there would exist $v \in C^1_{loc}(U^-)$ such that $D_y c(G^{-1}(y), y) = Dv(y)$. These suppositions are not so implausible when **(A1)**–**(A2)** hold, since S at least coincides with the graph of a map G and has a well-defined tangent space $\mathcal{H}^n$-almost everywhere.

However, despite the fact that neither G nor its inverse will be continuous in general, some vestige of this integrability persists. If c is Lipschitz continuous for example, then (8) implies the existence of Lipschitz u, v such that $c(x, y) - u(x) - v(y) \geq 0$ on $N = M^+ \times M^-$ with equality holding throughout S. This fact, which goes back to [72] [71], is in many senses better than mere integrability of a form: it requires no topology restriction on the domains, and not only do we get the first-order condition $Du(x) = D_x c(x, y)$ for those points $(x, y) \in S$ with x in the set of $\mathcal{H}^n$ full measure $\mathrm{Dom}\, Du$ where u is differentiable; as a second-order condition we get positive-definiteness of the matrix $D^2_{xx} c(x, y) - D^2 u(x) \geq 0$ if $x \in \mathrm{Dom}\, D^2 u$, and analogous conditions for v. Verily is S contained in the gradient of a convex function when $c(x, y) = -x \cdot y$ or $c(x, y) = \frac{1}{2}|x - y|^2$ on $U^\pm \subset \mathbf{R}^n$.

As Gangbo and McCann argue [31], this rough integrability result of Rockafellar and Rochet implies the famous duality of Kantorovich [36], Koopmans and Beckmann [40]:

$$\min_{\Gamma(\mu^+, \mu^-)} \int_{M^+ \times M^-} c(x, y) d\gamma(x, y) = \sup_{(u^+, u^-) \in Lip_c} \int_{M^+} u^+ d\mu^+ + \int_{M^-} u^- d\mu^- \quad (9)$$

with the supremum over

$$Lip_c := \{u^\pm \in L^1(d\mu^+) \mid c(x, y) \geq u^+(x) + u^-(y) \text{ throughout } N\} \quad (10)$$

being attained at $(u^+, u^-) = (u, v)$. Indeed, for any $(u^+, u^-) \in Lip_c$, integrating the inequality (10) against $\gamma \in \Gamma(\mu^+, \mu^-)$ yields

$$\int_{M^+ \times M^-} c d\gamma \geq \int_{M^+} u^+ d\mu^+ + \int_{M^-} u^- d\mu^-. \quad (11)$$

Thus the min dominates the sup in (9). Starting from $\gamma \in \Gamma(\mu^+, \mu^-)$ with c-cyclically monotone support, Rochet's generalization of Rockafellar's theorem provides $(u^+, u^-) = (u, v) \in Lip_c$ — bounded and Lipschitz if c is — such that equality holds in (11), and hence in (9) as desired.

7 Connections to differential geometry

We have already seen that the pseudo-Riemannian geometry induced on the product space $N = M^+ \times M^-$ by the metric tensor $h = \frac{1}{2}\operatorname{Hess}\delta^0(x_0, y_0)$ plays a key role in determining whether or not maps $y = G(x)$ which solve Monge's problem (1) are smooth. Here h is the Hessian of the cross-difference (4)–(5) associated to the cost c. The antisymmetry

$$\delta(x, y; x_0, y_0) = \delta(x_0, y_0; x, y) = -\delta(x, y_0; x_0, y)$$

ensures that h vanishes on $n\times n$ diagonal blocks. The involution $U(\Delta x, \Delta y) = (\Delta x, -\Delta y)$ on $T_{(x_0,y_0)}N$ allows us to define an antisymmetrized analog of h by

$$\omega(P, Q) = h(P, U(Q)).$$

Here ω turns out to be a symplectic form if and only if h has the full rank $2n = 2n^\pm$ that we often assume. Notice the similarity to Kähler geometry, with the splitting $T_{(x_0,y_0)}N = T_{x_0}M^+ \oplus T_{y_0}M^-$ of the tangent space associated to U playing the role of the almost complex structure J, and the cost c playing the role of the Kähler potential. For geometric measure theory in such geometries see Harvey and Lawson [34].

Kim and McCann showed that any c-optimal diffeomorphism $G : M^+ \longrightarrow M^-$ has a graph which is ω-Lagrangian in addition to being h-spacelike. Conversely, when **(A0)**–**(A4)** hold, then any diffeomorphism with an ω-Lagrangian and h-spacelike graph is necessarily c-optimal [38]. Here a submanifold $S \subset N$ is called ω-Lagrangian if $\omega(P, Q) = 0$ for every pair of tangent vectors $P, Q \in T_{(x_0,y_0)}N$. Being ω-Lagrangian is essentially the integrability condition which asserts closure of the form $D_x c|_{(x,G(x))}$ on M^+; it amounts to equality of the cross-derivatives $\partial G^i/\partial x^j = \partial G^j/\partial x^i$ which imply the existence of u such that $G(x) = Du(x)$ in case $c(x, y) = -x \cdot y$.

So far these geometric structures — the pseudo-metric h, symplectic form ω, c-cyclical monotonicity, and c-optimality — reflect only the cost function $c(x, y)$, and not the densities $d\mu^\pm(x) = f^\pm(x)dx$. Remarkably, however, there is a conformally equivalent pseudo-metric

$$\tilde{h}(x_0, y_0) = \left(\frac{f^+(x_0)f^-(y_0)}{|\det \partial^2 c/\partial x^i \partial y^j|}\right)^{1/n} h(x_0, y_0)$$

for which the graph Graph(G) of an optimal mapping $G_\# \mu^+ = \mu^-$ turns out to be a zero mean curvature surface — and in fact $\tilde{h}$-volume maximizing among homologous surfaces. This surprising connection of optimal transportation to geometric measure theory was discovered with Kim and Warren [39].

Thus the properties of optimal maps relate to both sectional and mean curvatures with respect to $\tilde{h}$. On the other hand, in the special case of the quadratic cost $c = d^2$ on a Riemannian manifold $M = M^{\pm}$, several surprising connections relate optimal transportation to the Riemannian geometry of (M, g_{ij}). For example, in this case Loeper and Villani conjecture [50] — and in some cases have proved — $\mathbf{(A3)_s}$ implies *convexity* of the tangent injectivity locus, which is to say the cut locus of each given point $x_0 \in M$, lifted to the tangent space $T_{x_0}M^+$ by the Riemannian exponential $\exp_{x_0}^{-1}$.

An earlier development involved lifting the metrical distance d from M to the space $P(M)$ of Borel probability measures $\mu^{\pm} \in P(M)$ using the minimal transportation cost $d_2(\mu^+, \mu^-) = \sqrt{cost(\gamma)}$ with respect to distance squared $c = d^2$ [6] [20] [64]. Geodesic convexity of various entropy functionals on $P(M)$ turns out to be equivalent to Ricci non-negativity of (M, g). This was shown by von Renesse and Sturm [70], building on work of myself [55], Cordero-Erausquin, Schmuckenschläger and I [15], and Otto and Villani [65]. This idea was turned on its head by Lott-Villani [52] and independently Sturm [77], who used geodesic convexity of the same entropies to *define* Ricci non-negativity in (not necessarily smooth) metric-measure spaces. This non-negativity is stable under measured Gromov-Hausdorff convergence, and has significant consequences.

A Ma-Trudinger-Wang conditions

The conditions **(A0)**-**(A4)** above have been synthesized in a language selected to manifest their topological and geometric invariance — aspects not readily apparent [7] from the original formulation by Ma, Trudinger, and Wang [53] in coordinates on the bounded sets $M^{\pm} \subset \mathbf{R}^n$, as we now recall.

Use subscripts such as i and j to denote derivatives with respect to x^i and y^j, and commas to separate derivatives in M^+ from those in M^-, so that $c_{i,j} = \partial^2 c/\partial x^i \partial y^j$ and $c_{ij,kl} = \partial^4 c/\partial x^i \partial x^j \partial y^k y^l$, etc. Also let $c^{k,l}$ denote the matrix inverse of $c_{i,j}$, and let $D_x c(x, y) = (c_1, c_2, \ldots, c_n)(x, y)$. Then the original conditions of Ma, Trudinger and Wang were formulated as the existence of a constant $C_0 > 0$ such that:

(A0)′ $c \in C^4(\bar{N})$, and for each $(x_0, y_0) \in \bar{N} = \bar{M}^+ \times \bar{M}^- \subset \mathbf{R}^n \times \mathbf{R}^n$:

(A1)$'_+$ the map $y \in \bar{M}^- \longmapsto D_x c(x_0, y) \in T^*_{x_0} M^+$ is injective;

(A1)$'$ both $c(x,y)$ and $c^*(y,x) := c(x,y)$ satisfy **(A0)**$'$ and **(A1)**$'_+$;

(A2)$'$ $\det c_{i,j}(x_0, y_0) \neq 0$;

(A3)$'_{\mathbf{s}}$ $(-c_{ij,kl} + c_{ij,m} c^{m,n} c_{kl,n}) p^i q^j p^k q^l \geq C_0 |p|^2 |q|^2$ whenever $p^i c_{i,j} q^j = 0$;

(A4)$'$ the sets $D_x c(x_0, M^-) \subset \mathbf{R}^n$ and $D_y c(M^+, y_0) \subset \mathbf{R}^n$ are convex.

Here the Einstein summation convention is in effect, and $|p|$ and $|q|$ denote the Euclidean norm on $p \in T_{x_0} M^+$ and $q \in T_{y_0} M^- \subset \mathbf{R}^n$ respectively.

Their method is heavily based on a priori C^2 estimates, which require a maximum principle for the directional second derivatives $D^2_{pp} u := u_{ij} p^i p^j$ of the unknown maximizers $u^\pm \in C(M^\pm)$ for the dual problem (9). A second-order linear elliptic equation satisfied by $D^2_{pp} u$ is obtained by twice differentiating the prescribed Jacobian equation for the map G, which is a fully nonlinear Monge-Ampère type equation for the potential $u = u^+$. Condition $(\mathbf{A3})_{\mathbf{s}}'$ ensures the zeroth order term in the elliptic equation satisfied by $D^2_{pp} u$ has a coefficient with the correct sign to admit a maximum principle.

The relaxation **(A3)**$'$ of $C_0 > 0$ to $C_0 = 0$ and strengthening $(\mathbf{A4})_{\mathbf{s}}'$ which requires all principal curvatures of $D_x c(x_0, M^-)$ and $D_y c(M^+, y_0)$ to be positive was introduced in the subsequent investigation of boundary regularity by Trudinger and Wang [79]. We leave it as an exercise to the reader to confirm the equivalence of each primed hypothesis **(A0)**$'$-**(A4)**$'$ and their variants to the corresponding unprimed hypothesis in the text. The connection of these conditions to the Riemann curvature tensor

$$\sec^{(\bar{N},h)}_{(x_0,y_0)} (p \oplus \mathbf{0}) \wedge (\mathbf{0} \oplus q) = (-c_{ij,kl} + c_{ij,m} c^{m,n} c_{kl,n}) p^i q^j p^k q^l \qquad (12)$$

and geodesic equations for the pseudo-metric $h = \frac{1}{2} \mathrm{Hess}_{(x_0,y_0)} \delta^0$ was first discovered in my joint work with Kim [38]. However, the link to the cross-difference $\delta^0(x,y)$ originates in the present work.

References

[1] N. Ahmad. *The geometry of shape recognition via a Monge-Kantorovich optimal transport problem* `http://www.math.toronto.edu/mccann/ahmad.pdf`. PhD thesis, Brown University, 2004.

[2] N. Ahmad, H.K. Kim, and R.J. McCann. Optimal transportation, topology and uniqueness. *Bull. Math. Sci.*, 1:13–32, 2011.

[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 and N. Gigli. *A user's guide to optimal transport. Preprint.*

[6] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[7] J.-P. Bourguignon. Ricci curvature and measures. *Japan. J. Math.*, 4:25–47, 2009.

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

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

[10] L.A. Caffarelli. Some regularity properties of solutions of Monge-Ampère equation. *Comm. Pure Appl. Math.*, 64:965–969, 1991.

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

[12] 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.

[13] T. Champion and L. De Pascale. The Monge problem in $\mathbf{R}^d$. *Duke Math. J.*, 157:551–572, 2011.

[14] 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.

[15] 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.

[16] M.J.P Cullen and R.J. Purser. An extended Lagrangian model of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497, 1984.

[17] 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.

[18] P. Delanoë and Y. Ge. Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds. *J. Reine Angew. Math.*, 646:65–115, 2010.

[19] P. Delanoë and F. Rouvière. Positively curved Riemannian locally symmetric spaces are positively square distance curved. To appear in *Canad. J. Math.*

[20] R.M. Dudley. *Probabilities and metrics - Convergence of laws on metric spaces, with a view to statistical testing.* Universitet Matematisk Institut, Aarhus, Denmark, 1976.

[21] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137:1–66, 1999.

[22] M. Feldman and R.J. McCann. Uniqueness and transport density in Monge's transportation problem. *Calc. Var. Partial Differential Equations*, 15:81–113, 2002.

[23] A. Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35:465–479, 2010.

[24] A. Figalli and Y.-H. Kim. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Contin. Dyn. Syst.*, 28:559–565, 2010.

[25] A. Figalli, Y.-H. Kim, and R.J. McCann. Regularity of optimal transport maps on multiple products of spheres. To appear in *J. Euro. Math. Soc. (JEMS).*

[26] A. Figalli, Y.-H. Kim, and R.J. McCann. When is multidimensional screening a convex program? *J. Econom Theory*, 146:454–478, 2011.

[27] 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.

[28] A. Figalli, L. Rifford and C. Villani. Nearly round spheres look convex. *Amer. J. Math.*, 134:109, 2012.

[29] L. Forzani and D. Maldonado. Properties of the solutions to the Monge-Ampère equation. *Nonlinear Anal.*, 57:815–829, 2004.

[30] W. Gangbo. *Habilitation thesis.* Université de Metz, 1995.

[31] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[32] W. Gangbo and R.J. McCann. Shape recognition via Wasserstein distance. *Quart. Appl. Math.*, 58:705–737, 2000.

[33] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(P_2(M), W_2)$. *Methods Appl. Anal.*, 18:127, 2011.

[34] F.R. Harvey and H.B. Lawson, Jr. Split special Lagrangian geometry. *Preprint.*

[35] K. Hestir and S.C. Williams. Supports of doubly stochastic measures. *Bernoulli*, 1:217–243, 1995.

[36] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[37] 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.*

[38] 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.

[39] Y.-H. Kim, R.J. McCann and M. Warren. Pseudo-Riemannian geometry calibrates optimal transportation. *Math. Res. Lett.*, 17:1183–1197, 2010.

[40] T.C. Koopmans and M. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957.

[41] P.W.Y. Lee. New computable necessary conditions for the regularity theory of optimal transportation. *SIAM J. Math. Anal.*, 42:3054–3075, 2010.

[42] 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`.

[43] P.W.Y. Lee and R.J. McCann. The Ma-Trudinger-Wang curvature for natural mechanical actions. *Calc. Var. Partial Differential Equations*, 41:285–299, 2011.

[44] V.L. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-valued Anal.*, 7:7–32, 1999.

[45] J. Li. Smooth optimal transportation on hyperbolic space `www.math.toronto.edu/mccann/papers/li.pdf`. Master's thesis, University of Toronto, 2009.

[46] J. Liu. Hölder regularity of optimal mappings in optimal transportation. *Calc Var. Partial Differential Equations*, 34:435–451, 2009.

[47] 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.

[48] G. Loeper. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202:241–283, 2009.

[49] G. Loeper. Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna. *Arch. Ration. Mech. Anal.*, 199:269–289, 2011.

[50] G. Loeper and C. Villani. Regularity of optimal transport in curved geometry: the non-focal case. *Duke Math. J.*, 151:431–485, 2010.

[51] G.G. Lorentz. An inequality for rearrangements. *Amer. Math. Monthly*, 60:176–179, 1953.

[52] J. Lott and C. Villani. Ricci curvature for metric measure spaces via optimal transport. *Annals Math. (2)*, 169:903–991, 2009.

[53] 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.

[54] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.

[55] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

[56] 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.

[57] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

[58] R.J. McCann and N. Guillen. Five lectures on optimal transportation: geometry, regularity, and applications. `http://arxiv.org/abs/1011.2911` *To appear in Proceedings of the Sminaire de Mathmatiques Suprieure (SMS) held in Montral, QC, June 27-July 8, 2011.*

[59] R.J. McCann and M. Sosio. Hölder continuity of optimal multivalued mappings. *SIAM J. Math. Anal.*, 43:1855–1871, 2011.

[60] R.J. McCann, B. Pass and M. Warren. Rectifiability of optimal transportation plans. *Canad. J. Math*, 64:924–934, 2012.

[61] G.J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29:341–346, 1962.

[62] J.A. Mirrlees. An exploration in the theory of optimum income taxation. *Rev. Econom. Stud.*, 38:175–208, 1971.

[63] 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.

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

[65] 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.

[66] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calc. Var. Partial Differential Equations*, 43:529–536, 2012.

[67] A. Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Ann. Inst. H. Poincaré Probab. Statist.*, 43:113, 2007.

[68] A. Pratelli. On the sufficiency of c-cyclical monotonicity for optimality of transport plans. *Math. Z.*, 258:677–690, 2008.

[69] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems.* Probab. Appl. Springer-Verlag, New York, 1998.

[70] M.-K. von Renesse and K.-T. Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.*, 58:923–940, 2005.

[71] J.-C. Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *J. Math. Econom.*, 16:191–200, 1987.

[72] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.

[73] L. Rüschendorf and S.T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32:48–54, 1990.

[74] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137:519–529, 2009.

[75] C. Smith and M. Knott. On the optimal transportation of distributions. *J. Optim. Theory Appl.*, 52:323–329, 1987.

[76] M. Spence. Job market signaling. *Quarterly J. Econom.*, 87:355–374, 1973.

[77] K.-T. Sturm. On the geometry of metric measure spaces, I and II. *Acta Math.*, 196:65–177, 2006.

[78] N.S. Trudinger and X.-J. Wang. On the Monge mass transfer problem. *Calc. Var. Paritial Differential Equations*, 13:19–31, 2001.

[79] 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.

[80] J. Urbas. On the second boundary value problem for equations of Monge-Ampère type. *J. Reine Angew. Math.*, 487:115–124, 1997.

[81] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, 2003.

[82] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer, New York, 2009.

[83] X.-J. Wang. On the design of a reflector antenna. *Inverse Problems*, 12:351–375, 1996.

[84] X.-J. Wang. On the design of a reflector antenna II. *Calc. Var. Partial Differential Equations*, 20:329–341, 2004.

[85] Y. Yu. Singular set of a convex potential in two dimensions. *Comm. Partial Differential Equations*, 32:1883–1894, 2007.