GAFA, Geom. funct. anal.
Vol. 11 (2001) 1 – 0
1016-443X/01/030001-0 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

POLAR FACTORIZATION OF MAPS ON RIEMANNIAN MANIFOLDS

ROBERT J. MCCANN

Abstract

Let (M, g) be a connected compact manifold, C^3 smooth and without boundary, equipped with a Riemannian distance $d(x, y)$. If $s : M \to M$ is merely Borel and never maps positive volume into zero volume, we show $s = t \circ u$ factors uniquely a.e. into the composition of a map $t(x) = \exp_x[-\nabla\psi(x)]$ and a volume-preserving map $u : M \to M$, where $\psi : M \to \mathbf{R}$ satisfies the additional property that $(\psi^c)^c = \psi$ with $\psi^c(y) := \inf\{c(x, y) - \psi(x) \mid x \in M\}$ and $c(x, y) = d^2(x, y)/2$. Like the factorization it generalizes from Euclidean space, this non-linear decomposition can be linearized around the identity to yield the Hodge decomposition of vector fields.

The results are obtained by solving a Riemannian version of the Monge–Kantorovich problem, which means minimizing the expected value of the cost $c(x, y)$ for transporting one distribution $f \geq 0$ of mass in $L^1(M)$ onto another. Parallel results for other strictly convex cost functions $c(x, y) \geq 0$ of the Riemannian distance on non-compact manifolds are briefly discussed.

1 Introduction

The purpose of this note is twofold: to announce a solution of the Monge–Kantorovich transportation problem in curved geometries, and to derive from it a factorization of maps which extends Brenier's polar decomposition theorem from Euclidean space to Riemannian manifolds.

The author gratefully acknowledges the support of an American Mathematical Society Centennial Fellowship and National Science Foundation grant DMS 9622997 held at Brown University, where much of this work was performed. He is also grateful for the hospitality of the Institut des Hautes Etudes Scientifiques and the Max-Planck Institut für Mathematik in den Naturwissenschaften during various stages of the work. These results were first announced in the July 1997 NSF-CBMS workshop on the Monge–Ampère equation at Boca Raton, FL.

Given two spatial distributions of mass, the problem of Monge [Mo] and Kantorovich [K] is to transport the mass from one distribution to the other as efficiently as possible. Here efficiency is measured against a cost function $c(x, y)$ specifying the transportation tariff per unit mass. The problem has a long history including applications to physics, economics, and statistics among other fields – partly chronicled by Evans [E] and Rachev and Rüschendorf [RR]. Moreover, the significance of the problem increases when the cost $c(x, y)$ is linked to an underlying geometry – such as geodesic distance on a Riemannian manifold. While manifestations of this theme are already apparent in the work of Monge on the Euclidean distance $c(x, y) = |x - y|$ in $\mathbf{R}^n$, it has reemerged with new vitality since the work of Brenier on Euclidean distance squared $c(x, y) = |x - y|^2$ marked a turning point in the flow of recent developments concerning partial differential equations, inequalities, and applications, surveyed in Evans [E] and the forthcoming lecture notes of Villani [V].

In its original form, Brenier's theorem [Br1] factored each $s : \Omega \to \mathbf{R}^n$ in $L^1(\Omega; \mathbf{R}^n)$ uniquely (on a bounded smooth domain $\Omega \subset \mathbf{R}^n$) into the composition $s = t \circ u$ of a volume preserving map $u : \Omega \to \Omega$ with the gradient $t = \nabla\psi$ of a convex function $\psi : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$. Here s is assumed not to map positive volume into zero volume. His proof exploited the Monge–Kantorovich problem to produce the factor $t = \nabla\psi$, which best approximates the identity map (in the L^2 sense) among all maps pushing the volume forward to the same measure as $s : \Omega \to \mathbf{R}^n$. The hypotheses were subsequently relaxed, [Br2], [BuD], [M1], and the theorem inspired a whole line of subsequent work and some fascinating applications. For example, the data s prescribes the Jacobian determinant of the the map $t = \nabla\psi$, so the theorem unexpectedly yields a variational solution to the Monge–Ampère equation [Br1] with regularity addressed by Caffarelli [Ca2]. Linearizing the factorization around the identity map leads (formally) to the Hodge–Helmholtz decomposition for vector fields, as Brenier himself pointed out [Br1]. Finally, the mapping can be combined with various change-of-variables formulas to localize geometric inequalities under an integral, leading to elementary proofs and far-reaching generalizations by the present author [M2], Barthe [B], and Alesker, Dar, and Milman [ADV], of inequalties due to Brunn–Minkowski, Brascamp–Lieb, and Alexandrov and Fenchel. These convexity inequalities [M2] have in turn been coupled with the gradient flow ideas of Jordan, Kinderlehrer, and Otto [JKO] to derive rates of convergence for nonlinear diffusion processes in work of

Otto [O]. Many other developments are reviewed in [E] and [V]. However, it has remained unclear how to formulate such a theorem in curved geometries, where convex functions (and their gradients) hardly make sense, let alone their connection with mappings. In the meantime, the optimal transportation problem of Monge [Mo] and Kantorovich [K], central to Brenier's proof, has been studied intensively; see Rachev and Rüschendorf for a review [RR]. An understanding of its geometry obtained with Gangbo [GM2] for a large class of costs on Euclidean space sets the stage for exploring its Riemannian structure. (For the Euclidean case see also [GM1] and Caffarelli [Ca1]

For simplicity, let (M, g) be a Riemannian manifold without boundary, connected, compact, and C^3 smooth – meaning the metric tensor components $g_{ij}(x)$ are twice continuously differentiable functions of local coordinates. The geodesic distance between x and $y \in M$ is denoted by $d(x, y)$, and the volume element by $d\,\mathrm{vol}(x)$ $(= \sqrt{\det[g_{ij}(x)]}\, d^n x$ in local coordinates). We set $c(x, y) = d^2(x, y)/2$ throughout, mentioning only briefly in section §5 some parallel results which may be obtained when $d^2/2$ is replaced by another strictly convex increasing function (25)–(26) of distance. Given finite Borel measures $\mu, \nu \geq 0$ on M with the same total mass $\mu[M] = \nu[M]$, Monge's problem [Mo] is to find the map $t : M \to M$ minimizing the transportation cost

$$\mathcal{C}(s) = \int_M c\big(x, s(x)\big) d\mu(x) \tag{1}$$

among all Borel maps $s \in S(\mu, \nu)$ which *push μ forward* to ν, meaning

$$\nu[V] = \mu[s^{-1}(V)] \tag{2}$$

holds for each measurable $V \subset M$. Strictly speaking, the elements of $S(\mu, \nu)$ consist of equivalence classes of maps which agree up to sets on which μ vanishes. When μ is absolutely continuous with respect to Riemannian volume, denoted $\mu \ll \mathrm{vol}$, a unique optimal map t shall be shown to exist in $S(\mu, \nu)$ and be characterized geometrically:

Given a function $\psi : M \to \mathbf{R} \cup \{\pm\infty\}$, its *infimal convolution* ψ^c with c is defined by

$$\psi^c(y) = \inf_{x \in M} c(x, y) - \psi(x)\,. \tag{3}$$

This transformation acts as an involution $\psi^{cc} := (\psi^c)^c = \psi$ if and only if ψ is itself an infimal convolution of some function with c, in which case ψ is said to be *c-concave* [RR, §3.3]. The unique $t \in S(\mu, \nu)$ of the form

$$t(x) = \exp_x[-\nabla\psi(x)] \tag{4}$$

with $\psi : M \to \mathbf{R}$ c-concave yields the optimal map. Thus the existence of a potential $\psi = \psi^{cc}$ whose gradient specifies which direction – $\nabla\psi(x) \in TM_x$ – and how far – $|\nabla\psi(x)|_x$ – to move the mass geodesically along M from [μ-a.e.] x to its destination characterizes optimality. In the Euclidean case, one recovers $t(x) = x - \nabla\psi(x)$ from (4), while for $c(x,y) = |x-y|^2/2$ the condition $\psi = \psi^{cc}$ reduces to convexity of $|x|^2/2 - \psi$. Verily is $t(x)$ the gradient of a convex function.

Now fix a finite non-negative Borel measure $\mu \ll \text{vol}$ on M and a Borel map $s : M \to M$. To polar factorize s, *define* a second Borel measure $\nu := s_{\#}\mu$ using (2), called the ***push-forward*** of μ through s. Clearly $\nu[M] = \mu[M]$. The optimal map $t(x) = \exp_x[-\nabla\psi(x)]$ pushing μ forward to ν with $\psi = \psi^{cc}$ is the first factor decomposing s. To find the second factor, assume $\nu \ll \text{vol}$, meaning s collapses no set with positive μ measure onto a set of zero volume; s is *non-degenerate* in the terminology of Brenier. It is easy to guess that the optimal map $t^* \in S(\nu,\mu)$ must be the inverse to t in the sense that $t(t^*(y)) = y$ [ν-a.e.]. Setting $u = t^* \circ s$ ensures $t \circ u = s$ holds μ-a.e., while $u_{\#}\mu = t^*_{\#}(s_{\#}\mu) = t^*_{\#}\nu = \mu$ so the measure μ is preserved under u. Volume is preserved if we started out with $\mu = \text{vol}$. Apart from sets of measure zero, there is only one map in $S(\mu,\nu)$ of the form (4) with $\psi = \psi^{cc}$, implying t and u are uniquely determined μ-a.e.

The main part of our work will be devoted to proving existence and uniqueness of a map $t \in S(\mu,\nu)$ minimizing Monge's transportation cost (1), and establishing (4) with $\psi = \psi^{cc}$. Since the measures μ and ν are both finite and non-negative with the same total mass, it costs no generality to normalize them so that $\mu[M] = \nu[M] = 1$; i.e. to restrict our attention to (Borel) probability measures. As in the Euclidean case, our departure point is a dual problem [RR] of Kantorovich type:

$$\sup_{(\psi,\phi)\in Lip_c} J(\psi,\phi) = \inf_{s\in S(\mu,\nu)} \int_M c\big(x, s(x)\big),\, d\mu(x) \tag{5}$$

where

$$J(\psi,\phi) = \int_M \psi(x)\, d\mu(x) + \int_M \phi(y)\, d\nu(y) \tag{6}$$

is a linear functional defined on a convex subset

$$Lip_c := \big\{\psi,\phi : M \longrightarrow \mathbf{R} \text{ continuous } \big|\ \psi(x) + \phi(y) \le c(x,y)\big\} \tag{7}$$

of continuous functions. In fact, the supremum of $J(\psi,\phi)$ turns out to be attained [RR, §2.3.12] – by a function $\psi = \psi^{cc}$ (with $\phi = \psi^c$) from which we shall construct a map minimizing Monge's cost. Compactness of the manifold facilitates a direct proof of the existence of (ψ, ϕ), and the duality relation (5), following an approach of Gangbo [G] also developed in Caffarelli [Ca3,1], and Gangbo and McCann [GM1]. We give this proof after some preliminaries on infimal convolutions with c. A key ingredient is the change of variables formula for pushed-forward measures $s_{\#}\mu = \nu$:

$$\int_M h\, d(s_{\#}\mu) = \int_M h(s(x))\, d\mu(x) \tag{8}$$

holds for all Borel $s : M \to M$ and $h : M \to \mathbf{R} \cup \{\pm\infty\}$, as is readily verified from definition (2) by approximating h using simple functions. Alternately, the Riesz representation theorem allows one to take (8) as the definition of $s_{\#}\mu$. Before entering into further details, a few remarks are necessary concerning the history of this manuscript.

This paper has been in gestation for quite a long time. It is the author's pleasure to recall that the question of how to polar factorize vector fields on Riemannian manifolds was first put to him by Dennis Sullivan (at a time when he was ill-prepared to solve it, especially since the Euclidean example initially misled him into trying to factor vector fields rather than maps) and again a year later by Tudor Ratiu (at a time when he found himself better equipped). He is pleased to acknowledge both of them, along with Stephen Semmes, for providing stimulating conversations during the course of the work. He is also grateful to Michael Cullen and Robert Douglas, who included a statement of the result in their announcement of its first application [CuD]: finding simplifying variables for the semigeostrophic model of atmospheric dynamics on a sphere; see also [CuDRS], [BeB]. Their setting is actually a non-compact manifold – the Northern hemisphere – with a conformally round metric proportional to the Coriolis force, which degenerates on the equator. As indicated in §5 below, such manifolds require the additional hypothesis that (unbroken) minimal geodesics exist between every $x \in \operatorname{spt} \mu$ and $y \in \operatorname{spt} \nu$. Interestingly enough, near the equator where this hypothesis fails, their model breaks down and weather patterns change drastically because the atmosphere has no preferred direction to swirl!

An unrelated application was discovered by Cordero-Erausquin, who used the map (4) to formally derive a Prekopa–Leindler inequality for spherical and hyperbolic geometries [Co1]. A rigorous proof of the analogous inequalities on more general manifolds is being developed in joint work with Cordero-Erausquin and Schmuckenschläger [CoMS]. It must also be noted that Cordero-Erausquin obtained an independent solution to the transportation problem on the flat torus $M = \mathbf{T}^n$ before becoming aware of the present manuscript [Co2]. Finally, if the measures are given by $L^1(M, d\text{vol})$ densities $d\mu(x) = m(x)d\text{vol}(x)$ and $d\nu(y) = n(y)d\text{vol}(y)$, then for μ-a.e. x the Jacobian of the optimal map $t_{\#}\mu = \nu$ can be shown [CoMS] to satisfy the expected equation:

$$n(t(x))\det[Dt(x)] = m(x) .$$

The form (3)–(4) of the optimal map precisely ensures that this Monge–Ampère like relation falls into the class of elliptic equations explored on manifolds by Cabré [C].

We now proceed to state a series of standard lemmas and introduce some non-smooth analysis as a prelude to our first substantial remarks: Proposition 6 and Lemma 7. These are used to prove our main results: Theorems 8, 9 and 11 of section §3. This is followed by section §4 highlighting how the formal relationship between the polar factorization of maps and the Hodge decomposition of vector fields extends to the Riemannian setting. A final section §5 indicates (without proof) some adaptations of these results to other strictly convex costs (25)–(26) in place of $c = d^2/2$ and to non-compact manifolds (M, g).

2 Preliminaries

The results of this section, though well-known to part of our readership, are included for ease of reference and completeness.

LEMMA 1 (Lipschitz cost). *Let (M, d) be a metric space whose diameter $|M| := \sup\{d(x, z) \mid x, z \in M\}$ is finite. For each $y \in M$, the function $\psi(x) = d^2(x, y)/2$ is Lipschitz continuous:*

$$\big|\psi(x) - \psi(z)\big| \le |M| d(x, z) . \tag{9}$$

Proof. The triangle inequality shows that $\phi(x) := d(x, y)$ has Lipschitz constant one:

$$\phi(x) - \phi(z) = d(x, y) - d(z, y) \le d(x, z) , \tag{10}$$

for all $x, z \in M$. Also, $\phi(x) = d(x, y) \le |M| < \infty$ is bounded. The desired estimate (9) then follows easily for $\psi(x) = \phi^2(x)/2$:

$$\begin{aligned} 2\big|\psi(x) - \psi(z)\big| &= \big|\phi(x)(\phi(x) - \phi(z)) + \phi(z)(\phi(x) - \phi(z))\big| \\ &\le |M| d(x, z) + |M| d(x, z) . \end{aligned}$$

□

LEMMA 2 (Infimal convolutions are Lipschitz). *Fix a metric space (M, d) having finite diameter. Any $\psi : M \to \mathbf{R} \cup \{\pm\infty\}$ given by an infimal* *convolution* $\psi = \psi^{cc}$ *with* $c(x,y) = d^2(x,y)/2$ *is either identically infinite* $\psi = \pm\infty$ *or Lipschitz continuous throughout* M. *Indeed, it satisfies* (9).

Proof. More generally, suppose $\psi = \phi^c$ for some $\phi : M \to \mathbf{R} \cup \{\pm\infty\}$, meaning

$$\psi(x) = \inf_{y \in M} c(x,y) - \phi(y) . \tag{11}$$

Observe $0 \leq c(x,y) \leq |M|^2/2$ is bounded. Either ϕ is unbounded above, in which case (11) yields $\psi = -\infty$ and the lemma holds trivially, or else ψ is bounded below. Fix $z \in M$, and note $\psi(z) = +\infty$ in (11) occurs only if $\phi := -\infty$ everywhere, in which case $\psi = +\infty$ again holds trivially. Thus we may assume that ψ is finite everywhere. Given any $\epsilon > 0$, there exists $y \in M$ such that $\psi(z) + \epsilon \geq c(z,y) - \phi(y)$, while $\psi(x) \leq c(x,y) - \phi(y)$ holds because of (11). Subtracting these two inequalities yields

$$\begin{aligned} \psi(x) - \psi(z) &\leq c(x,y) - c(z,y) + \epsilon \\ &\leq |M| d(x,z) + \epsilon \end{aligned}$$

by Lemma 1. Since the last inequality holds for all $\epsilon > 0$, the Lipschitz estimate (9) has been proved. □

PROPOSITION 3 (Dual potentials). *Fix Borel probability measures* μ *and* ν *on a separable metric space* (M,d) *having finite diameter* $|M|$. *Set* $c(x,y) = d^2(x,y)/2$. *The supremum* (5) *is attained by some* $\psi : M \to \mathbf{R}$ *satisfying* $\psi = \psi^{cc}$ *and its infimal convolution* $\phi = \psi^c$ *with* c.

Proof. Claim #1: If $(u,v) \in Lip_c$ then $(v^c, v) \in Lip_c$ and $J(u,v) \leq J(v^c, v)$. Moreover, $(v^c, v^{cc}) \in Lip_c$ and $J(u,v) \leq J(v^c, v^{cc})$.

Proof of claim: Fixing $(u,v) \in Lip_c$, note that

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

Thus the infimal convolution (3) in x which defines u^c must be finite-valued: $u^c(y) \geq v(y)$, hence Lipschitz continuous by Lemma 2. Now $(u, u^c) \in Lip_c$ follows immediately from (3) and (7). The inequality $u^c \geq v$ also implies $J(u,v) \leq J(u,u^c)$. Symmetry under $u \leftrightarrow v$ shows the first half of the claim is established. Applying what we just proved to (v^c, v) yields $(v^c, v^{cc}) \in Lip_c$ and $J(v^c, v) \leq J(v^c, v^{cc})$, from which the remaining claim follows immediately.

To complete the proposition, choose a sequence $(\psi_n, \phi_n) \in Lip_c$ for which $J(\psi_n, \phi_n)$ tends to its maximum value on Lip_c. According to Claim #1, the sequence (ϕ_n^c, ϕ_n^{cc}) also maximizes J on Lip_c. Fix $z \in M$

and define a new sequence $(u_n, v_n) := (\phi_n^c - \lambda_n, \phi_n^{cc} + \lambda_n)$ in Lip_c, again maximizing J since $\mu[M] = \nu[M]$; the normalization constants $\lambda_n := \phi_n^c(z)$ are selected to make $u_n(z) = 0$. Now Lemma 2 provides a uniform bound $|M|$ for the Lipschitz constants of u_n and v_n. Moreover, this equicontinuous family of functions is uniformly bounded throughout M:

$$\bigl|u_n(x)\bigr| = \bigl|u_n(x) - u_n(z)\bigr| \leq |M| d(x,z) \leq |M|^2 , \tag{12}$$

while $|v_n(y)| \leq 3|M|^2/2$ follows from (3) since $v_n = u_n^c$. The Ascoli–Arzela argument extracts a pointwise convergent subsequence, also denoted (u_n, v_n), with Lipschitz continuous limit $(u,v) \in Lip_c$. Since μ and ν are finite measures, the dominated convergence theorem yields $J(u,v) = \lim_n J(u_n, v_n)$. The supremum (5) is attained at (u,v) and hence (by Claim #1) at $(\psi, \phi) := (v^c, v^{cc})$. Finally, $v^{ccc} = v^c$ [RR, §3.3.5], so $\psi^{cc} = \psi$ to complete the proof. □

LEMMA 4 (Rademacher's theorem). *Any function $\psi : M \to \mathbf{R}$ locally Lipschitz with respect to the geodesic distance $d(x,y)$ on a connected, C^1 smooth Riemannian manifold must be differentiable outside a set $Z \subset M$ of zero volume. Its gradient gives a Borel map from* $\operatorname{dom} \nabla\psi := M \setminus Z$ *into the tangent bundle TM.*

Proof. Fix $z \in M$ and normal coordinates $\xi : U \to \mathbf{R}^n$ centered at $z = \xi^{-1}(\mathbf{0})$ so that the metric is diagonalized there: $g_{ij}(z) = \delta_{ij}$. Since the coefficients $g_{ij}(x)$ of the quadratic form $g\langle \cdot , \cdot \rangle_x$ depend continuously on these coordinates, there is a smaller neighbourhood of z, also denoted by $U \subset M$, on which its largest eigenvalue does not vary by more than a finite factor:

$$g\langle \mathbf{v}, \mathbf{v} \rangle_x \leq k^2 \sum_{i=1}^{n} (v^i)^2 \tag{13}$$

for all $x \in U$ and $\mathbf{v} \in TM_x$. Choose $\epsilon > 0$ small enough so that $\mathbf{B}^n(\mathbf{0}, \epsilon) \subset \xi(U)$ and replace U by the preimage $\xi^{-1}(\mathbf{B}^n(\mathbf{0}, \epsilon))$ of this ball. We shall show ψ to be differentiable vol-a.e. on U. Since a connected Riemannian manifold is locally compact and second countable from Kobayashi and Nomizu [KoN, Appendix 2], it is σ-compact, hence covered by countably many such neighbourhoods $U \subset M$. The differentiability of ψ will therefore follow vol-a.e. on M.

The geodesic distance between $x, y \in U$ is bounded by the length of the path $\sigma(\tau) := \xi^{-1}((1-\tau)\xi(x) + \tau\xi(y))$ through U from x to y. Computing its arclength in local coordinates yields

$$d(x,y) \leq \int_0^1 \sqrt{g\langle\,\dot\sigma,\,\dot\sigma\rangle_\sigma} d\tau \leq k\big|\xi(y)-\xi(x)\big| \tag{14}$$

after combining $\dot\sigma^i(\tau) = \xi^i(y) - \xi^i(x)$ with (13). Since ψ was assumed Lipschitz with respect to the Riemannian distance, it follows from (14) that $f := \psi \circ \xi^{-1}$ is Lipschitz with respect to the Euclidean metric on $\mathbf{B}^n(\mathbf{0},\epsilon)$. Rademacher's theorem then asserts differentiability of f Lebesgue a.e. on $\mathbf{B}^n(\mathbf{0},\epsilon)$. But Riemannian volume can be expressed locally by a bounded multiple $\sqrt{\det g_{ij}(x)}$ of Lebesgue measure, so differentiability of ψ has been established vol-a.e. on U to complete the first claim.

To establish the remaining claim, it is necessary to recall the proof of Rademacher's theorem from Evans and Gariepy [EG, §3.1.2]. After extending $f = \psi \circ \xi^{-1}$ continuously to all of $\mathbf{R}^n$, they observe that its upper derivative

$$\overline{D}_{\mathbf{v}} f(\mathbf{x}) = \lim_{k\to\infty} \sup_{|t|\in(0,1/k)\cap\mathbf{Q}} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t}$$

in direction $\mathbf{v} \in \mathbf{R}^n$ is expressed as a limit of suprema of continuous functions of $\mathbf{x}$, hence Borel. The lower derivative $\underline{D}_{\mathbf{v}} f$ defined by the analogous limit infimum is also Borel, so the directional derivative $D_{\mathbf{v}} f$ is a Borel function on the set of full measure where $\overline{D}_{\mathbf{v}} f = \underline{D}_{\mathbf{v}} f$. In particular, the n partial derivatives $\partial_{x^j} f$ are Borel, as is

$$F_{\mathbf{v}} = \Big\{\mathbf{x} \in \mathbf{B}^n(\mathbf{0},\epsilon) \;\big|\; \overline{D}_{\mathbf{v}} = \underline{D}_{\mathbf{v}} f = \sum v^j \partial_{x^j} f\Big\}.$$

As Evans and Gariepy show, f is differentiable on any countable intersection $\bigcap F_{\mathbf{v}_i}$ of these sets over a dense set of directions $\{\mathbf{v}_i\} \in \partial\mathbf{B}^n(\mathbf{0},1)$. Outside $F_{\mathbf{v}_i}$ differentiability fails, so ∇f must be Borel. Clearly $g^{kj}\partial_{x^j} f$ is also Borel on $\bigcap F_{\mathbf{v}_i}$ – and gives the coordinates of $\nabla\psi$ on $U \setminus Z$. We conclude that both $\nabla\psi$ and $Z \subset M$ are Borel. □

The preceding proof used only that the metric tensor $g_{ij}(x)$ was bounded (13) and Borel – but not necessarily continuous. Indeed, with the definitions of De Cecco and Palmieri [DP], the lemma can be extended immediately to Lipschitz Riemannian manifolds.

3 Method and Results

The chief technical complications arising in the Riemannian setting stem from non-uniqueness of minimal geodesics, and hence a lack of smoothness in the cost and the distance function. However, all singularities come in the form of upward pointing creases (and conical points) along the cut locus. Remarkably, this one-sidedness turns out to permit the problem to be finessed by introducing appropriate notions from non-smooth analysis. Indeed, on smooth manifolds, one may expect the cost $c(x, y) = d^2(x, y)/2$ to have a *semi-concavity* property [GM2], though for present purposes it suffices to show it is superdifferentiable in the following sense:

Fix $x \in M$. A function $\phi : M \to \mathbf{R}$ is said to be *superdifferentiable* at x with *supergradient* $\mathbf{p} \in TM_x$ if

$$\phi(\exp_x \mathbf{v}) \leq \phi(x) + g\langle \mathbf{p}, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x) \tag{15}$$

holds for small $\mathbf{v} \in TM_x$, where $o(\lambda)/\lambda$ tends to zero with λ. Such (supergradient, point) pairs $(\mathbf{p}, x)$ form a subset $\overline{\partial}\phi \subset TM$ of the tangent bundle; we also express their relationship (15) by writing $\mathbf{p} \in \overline{\partial}\phi_x$. If the opposite inequality $\phi(\exp_x \mathbf{v}) \geq \phi(x) + g\langle \mathbf{q}, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x)$ holds, ϕ is said to be *subdifferentiable* with *subgradient* $\mathbf{q} \in \underline{\partial}\phi_x \subset TM_x$. When both inequalities hold, then ϕ is differentiable at x and its super- and sub-gradients coincide: $\mathbf{p} = \mathbf{q} = \nabla\phi(x)$. This observation, which proves crucial later, motivates a chain rule for supergradients used to establish the connection between minimal geodesics and superdifferentiability of the cost.

LEMMA 5 (Chain rule). *Let $\phi : M \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$ have supergradients $\mathbf{p} \in \overline{\partial}\phi_x$ and $\tau \in \overline{\partial}h_{\phi(x)}$ at some $x \in M$. If h is non-decreasing then $\tau\mathbf{p} \in \overline{\partial}(h \circ \phi)_x$.*

Proof. Applied to h, definition (15) yields

$$h\big(\phi(x) + \epsilon\big) \leq h(\phi(x)) + \tau\epsilon + o(\epsilon)\,.$$

Since h is non-decreasing, setting $\epsilon = g\langle \mathbf{p}, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x)$ and invoking (15) yields

$$\begin{aligned} h\big(\phi(\exp_x \mathbf{v})\big) &\leq h\big(\phi(x) + g\langle \mathbf{p}, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x)\big) \\ &\leq h(\phi(x)) + \tau g\langle \mathbf{p}, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x) \end{aligned}$$

to complete the proof. □

PROPOSITION 6 (Superdifferentiability of geodesic distance squared). *Let (M, g) be a C^3-smooth Riemannian manifold, possibly with boundary. Suppose $\sigma : [0, 1] \to M$ has minimal length among piecewise C^1 curves joining $y = \sigma(0)$ to $x = \sigma(1) \notin \partial M$, parameterized with constant speed. Then $\psi(\,\cdot\,) = d^2(\,\cdot\,, y)/2$ has supergradient $\dot\sigma(1) \in \overline{\partial}\psi_x$ at x.*

Proof. Since x lies in the interior of M, there is some $\epsilon > 0$ and neighbourhood $X \subset M$ of x such that: at each $z \in X$, the exponential map $\exp_z$ maps the ball $\mathbf{B}(\mathbf{0},\epsilon) \subset TM_z$ diffeomorphically onto some open set $U_z \supset X$, as in Milnor [Mi, §10.3]. The proposition will first be established when $y = \sigma(0) \in X$, in which case ψ is actually differentiable at $\exp_y \dot\sigma(0) = x$. We compute its derivative by linearizing $\exp_x \mathbf{v} \in X$ around the origin and $\exp_y$ around $\dot\sigma(0)$:

$$\begin{aligned}\psi(\exp_x \mathbf{v}) &= d^2\big(y, \exp_y(\exp_y^{-1}\exp_x \mathbf{v})\big)/2\\ &= \big|\exp_y^{-1}(\exp_x \mathbf{v})\big|_y^2/2\\ &= \big|\dot\sigma(0) + D(\exp_y^{-1})_x D(\exp_x)_{\mathbf{0}}\mathbf{v} + o(|\mathbf{v}|_x)\big|_y^2/2\\ &= |\dot\sigma(0)|_y^2/2 + g\langle\, \dot\sigma(0),\, (D\,\exp_y)^{-1}_{\dot\sigma(0)} I\mathbf{v}\rangle_y + o(|\mathbf{v}|_x)\\ &= d^2(x,y)/2 + g\langle\, \dot\sigma(1),\, \mathbf{v}\rangle_x + o(|\mathbf{v}|_x)\,,\end{aligned}$$

so that $\nabla\psi(x) = \dot\sigma(1)$. Here the last equation follows from $\dot\sigma(1) = D(\exp_y)_{\dot\sigma(0)}\dot\sigma(0)$ and Gauss' lemma, see do Carmo [Car, §3.3.5] or Milnor [Mi, §10.5]. (Note that C^3 smoothness of (M,g) ensures that the coefficients Γ^k_{ij} in the geodesic equation are continuously differentiable; the Picard theorem for ODEs set forth in Graves [Gr, §IX.3.13 and §V.2.16] then guarantees differentiability of the exponential map and the equality of mixed partials [Mi, §8.7] needed to establish Gauss' lemma.) Since $d(x,y) = \sqrt{2\psi(x)}$ the chain rule yields $\nabla_x\, d(x,y) = \dot\sigma(1)/|\dot\sigma(1)|_x$ as long as $x \neq y$.

Now we return to the possibility that $y \notin X$. In this case, take $z \in X$ to be any point lying on the geodesic σ near the endpoint $x = \sigma(1)$. Applying the foregoing argument to $z \neq x$ instead of y yields $\nabla_x\, d(x,z) = \dot\sigma(1)/|\dot\sigma(1)|_x$. The triangle inequality then gives

$$\begin{aligned}d(y, \exp_x \mathbf{v}) &\le d(y,z) + d(z,x) + g\langle\, \dot\sigma(1),\, \mathbf{v}\rangle_x/|\dot\sigma(1)|_x + o(|\mathbf{v}|_x)\\ &= d(y,x) + g\big\langle \dot\sigma(1)/|\dot\sigma(1)|_x, \mathbf{v}\big\rangle_x + o(|\mathbf{v}|_x)\,,\end{aligned}$$

so $d(y,\,\cdot\,) = \sqrt{2\psi(\,\cdot\,)}$ is superdifferentiable at x. Applying the one-sided chain rule, Lemma 5, with $h(d) = \frac{d^2}{2}$ and $\phi = \sqrt{2\psi}$ yields $\frac{d(y,x)\dot\sigma(1)}{|\dot\sigma(1)|_x} = \dot\sigma(1) \in \overline{\partial}\psi_x$ to complete the proof. □

The next lemma establishes a Young-like inequality [RR, §3.3.3] together with conditions for equality. These conditions determine the form of the optimal map, and play a critical role in a uniqueness argument based on the original idea of Brenier [Br1,2].

LEMMA 7 (Tangency). *Let (M,g) be a connected, compact Riemannian manifold, C^3-smooth and without boundary. Suppose $\psi = \psi^{cc}$, meaning*

$\psi : M \to \mathbf{R}$ is an infimal convolution (3) *with* $c(x,y) = d^2(x,y)/2$. *Then*

$$c(x,y) - \psi(x) - \psi^c(y) \geq 0 \tag{16}$$

for all $x, y \in M$. *If a point* $x \in M$ *is selected where* ψ *happens to be differentiable, then equality holds in* (16) *if and only if* $y = \exp_x[-\nabla\psi(x)]$.

Proof. The inequality (16) follows immediately from the definition (3) of ψ^c; it appears also in Rachev and Rüschendorf [RR, §3.3.3]. Now select a point $x \in M$ where ψ happens to be differentiable. To establish the *only if* statement, assume some $y \in M$ can be found for which (16) becomes an equality. Then for all $z \in M$ one has

$$\begin{aligned} c(z,y) - \psi(z) - \psi^c(y) &\geq 0 \\ &= c(x,y) - \psi(x) - \psi^c(y)\,. \end{aligned}$$

Defining $\phi(z) := c(z,y)$ and $z := \exp_x \mathbf{v}$ yields

$$\begin{aligned} \phi(\exp_x \mathbf{v}) = c(z,y) &\geq c(x,y) - \psi(x) + \psi(z) \\ &= \phi(x) - \psi(x) + \psi(x) + g\langle \nabla\psi(x), \mathbf{v}\rangle_x + o(|\mathbf{v}|_x)\,, \end{aligned}$$

so $\phi(z)$ has subgradient $\nabla\psi(x) \in \underline{\partial}\phi_x$ at x. On the other hand, the Hopf–Rinow theorem [Mi, §10.19] [KoN, §IV.4.1–4] assures us of the existence of a minimal geodesic $\sigma : [0,1] \to M$ from y to x given by $\sigma(1-\tau) = \exp_x[-\tau\dot{\sigma}(1)]$. Like the manifold, this geodesic will be C^3-smooth, whence Proposition 6 yields $\dot{\sigma}(1) \in \overline{\partial}\phi_x$. Thus ϕ is both super- and subdifferentiable at x, so it is differentiable and its super- and subgradients coincide: $\dot{\sigma}(1) = \nabla\psi(x)$, implying $y = \sigma(0) = \exp_x[-\nabla\psi(x)]$ as desired.

For our fixed $x \in M$, we have now established that at most one point $y \in M$ produces equality in (16). We have not yet shown that the equality must be achieved by some $y \in M$, nor used the condition $\psi = \psi^{cc}$ (except to know $\psi^c(y)$ is finite). We now do both. Indeed, since $\psi = \psi^{cc}$ is real-valued, Lemma 2 yields Lipschitz continuity of $\psi^c : M \to \mathbf{R}$. Also, the manifold is compact, so the infimum defining

$$\psi^{cc}(x) := \inf_{y \in M} c(x,y) - \psi^c(y)$$

is attained at some $y \in M$. The same point produces equality in (16). But then our previous argument shows $y = \exp_x[-\nabla\psi(x)]$, so the *if* statement of the lemma is satisfied. □

Theorem 8 (Unique optimal maps). *Fix a connected, compact Riemannian manifold, C^3-smooth and without boundary, and a Borel probability measure $\mu \ll$ vol on (M,g). If $\psi : M \to \mathbf{R}$ is an infimal convolution*

$\psi = \psi^{cc}$ with $c(x,y) = d^2(x,y)/2$, then $t(x) = \exp_x[-\nabla\psi(x)]$ minimizes (1) on $S(\mu, t_\#\mu)$. Any other map $s \in S(\mu, t_\#\mu)$ minimizing $\mathcal{C}(s)$ must coincide with $t(x)$ μ-almost everywhere.

Proof. Taking $\psi = \psi^{cc}$ and t as above, Lemmas 2 and 4 show that the potential ψ is Lipschitz and differentiable μ-a.e., while its gradient (and hence t) are Borel. Now $t \in S(\mu, t_\#\mu)$ and $(0,0) \in Lip_c$ so both domains are non-empty. For any $(u,v) \in Lip_c$ and $s \in S(\mu, t_\#\mu)$ one has

$$J(u,v) = \int_M u(x)d\mu(x) + \int_M v(s(x))d\mu(x) \tag{17}$$

$$\leq \int_M c\big(x, s(x)\big)d\mu(x) \tag{18}$$

$$= \mathcal{C}(s) \tag{19}$$

from (6)–(8) and a change of variables applied to $s_\#\mu$. Since both sides are finite, this shows

$$\sup_{(u,v)\in Lip_c} J(u,v) \leq \tau \leq \inf_{s\in S(\mu,t_\#\nu)} \int_M c\big(x, s(x)\big)d\mu(x) \tag{20}$$

for some $\tau \in \mathbf{R}$. However Lemma 7 shows $(\psi, \psi^c) \in Lip_c$ satisfies $\psi(x) + \psi^c(t(x)) = c(x, t(x))$ [μ-a.e.]. Choosing $(u,v) = (\psi, \psi^c)$ and $s = t$ therefore leads to equality in (18), and hence (20). Moreover $J(\psi, \psi^c) = \tau = \mathcal{C}(t)$ proves optimality of the map t.

Conversely, any map $s \in S(\mu, t_\#\mu)$ which achieves optimality must also satisfy $\mathcal{C}(s) = \tau = J(\psi, \psi^c)$, so equality continues to hold in (18). From (16) we see $\psi(x) + \psi^c(s(x)) = c(x, s(x))$ must hold pointwise μ-a.e. On the set of full measure where ψ is differentiable, we conclude $s(x) = \exp_x[-\nabla\psi(x)] = t(x)$ from Lemma 7. □

Note that the Kantorovich duality (5) was established for $\nu = t_\#\mu$ in the preceding proof. It must therefore continue to hold under the hypotheses of Theorem 9.

Theorem 9 (Existence and characterization of optimal maps). *Let (M, g) be a connected, compact Riemannian manifold, C^3-smooth and without boundary. Fix $c(x,y) = d^2(x,y)/2$ and two Borel probability measures $\mu \ll$ vol and ν arbitrary on M. Then for some potential $\psi : M \to \mathbf{R}$ satisfying $\psi = \psi^{cc}$ the map $t(x) = \exp_x[-\nabla\psi(x)]$ pushes μ forward to ν. Modulo discrepancies on sets of μ-measure zero, only one $t \in S(\mu,\nu)$ can arise in this way.*

Proof. To begin choose $(\psi, \phi) = (\psi^{cc}, \psi^c)$ which maximize (6) on Lip_c. These exist by Proposition 3. Both functions are Lipschitz according to Lemma 2, so $\nabla\psi : M \to TM$ and hence $t : M \to M$ are Borel maps defined μ-a.e. in view of Lemma 4. One way to prove that the map $t(x) = \exp_x[-\nabla\psi(x)]$ pushes μ forward to ν is to integrate each continuous function $h \in C(M)$ against the measures ν and $t_{\#}\mu$ and show the two integrals coincide. We do this following Gangbo [G], Gangbo and McCann [GM1] and Caffarelli [Ca1]. For $x, y \in M$ and $|\epsilon| < 1$ define perturbations $\phi_\epsilon(y) := \psi^c(y) + \epsilon h(y)$ and $\psi_\epsilon := (\phi_\epsilon)^c$ by

$$\psi_\epsilon(x) := \inf_{y\in M} c(x,y) - \phi(y) - \epsilon h(y)\,. \tag{21}$$

Fix a point x where ψ_0 is differentiable. Continuity and compactness of M ensure the infimum (21) is attained. For $\epsilon = 0$ it is attained uniquely at $y = t(x)$ in view of Lemma 7, so for small ϵ, it must be attained at some nearby point $y_\epsilon = t(x) + o(1)$. Thus

$$c\big(x, t(x)\big) - \phi(t(x)) - \epsilon h(y_\epsilon) \le \psi_\epsilon(x) \le c(x,y) - \phi(y) - \epsilon h(y)$$

for all $y \in M$. Choosing $y = t(x)$ yields $\psi_\epsilon(x) = \psi_0(x) - \epsilon h(t(x)) + o(\epsilon)$, where the error term satisfies a bound $\epsilon^{-1}o(\epsilon) \le 2\|h\|_\infty$ uniform in x in addition to vanishing as $\epsilon \to 0$. Because $J(\psi_\epsilon, \phi_\epsilon)$ attains its maximum at $\epsilon = 0$, we deduce

$$\begin{aligned}
\lim_{\epsilon\to 0} \frac{J(\psi_\epsilon,\phi_\epsilon) - J(\psi_0,\phi_0)}{\epsilon} &= \lim_{\epsilon\to 0}\int_M \frac{\psi_\epsilon(x) - \psi_0(x)}{\epsilon}\, d\mu(x) + \int_M h(y)\, d\nu(y)\\
&= \int_M -h(t(x))\, d\mu(x) + \int_M h(y)\, d\nu(y)\\
&= -\int_M h\, d(t_{\#}\mu) + \int_M h\, d\nu\\
&= 0\,,
\end{aligned}$$

where the dominated convergence theorem was used in the second equality and the change of variables formula (8) in the third. But now the existence claim is established: $t_{\#}\mu = \nu$ by the Riesz–Markov theorem since $h \in C(M)$ was arbitrary.

To prove that we have characterized t uniquely in $S(\mu,\nu)$, let $u = (u^c)^c$ denote any other potential on M for which the map $s(x) := \exp_x[-\nabla u(x)]$ pushes μ forward to ν. According to Theorem 8, both s and t minimize (1) on $S(\mu,\nu)$, hence they coincide μ-almost everywhere. □

COROLLARY 10 (Invertibility). *Take M, g, μ, ν, $c = d^2/2$, $\psi = \psi^{cc}$ and t from Theorem 9. If $\nu \ll$* vol *also, then the inverse map $t^*(y) :=$*

$\exp_y[-\nabla\psi^c(y)]$ *belongs to* $S(\nu,\mu)$ *and satisfies* $t^*(t(x)) = x$ [μ-a.e.] *and* $t(t^*(y)) = y$ [ν-a.e.].

Proof. We shall exploit the symmetry $\mu \leftrightarrow \nu$ and $(\psi,\psi^c) \leftrightarrow (\phi,\phi^c)$, where $\phi := \psi^c$ is the infimal convolution of ψ with $c(x,y)$, so that $\phi^c = \psi^{cc} = \psi$ and $\phi^{cc} = \psi^c = \phi$.

Denote the set where ϕ is differentiable by $\mathrm{dom}\,\nabla\phi \subset M$. Now $\nu[\mathrm{dom}\,\nabla\phi] = 1$ by Lemmas 2 and 4, so $t_{\#}\mu = \nu$ implies $U := \mathrm{dom}\,\nabla\psi \cap t^{-1}(\mathrm{dom}\,\nabla\phi)$ is a Borel set of full measure $\mu[U] = 1$. For each $x \in U$ Lemma 7 yields

$$\begin{aligned} 0 &= c\big(x, t(x)\big) - \psi^c(t(x)) - \psi(x) \\ &= c\big(t(x), x\big) - \phi(t(x)) - \phi^c(x)\,. \end{aligned}$$

Since $t(x) \in \mathrm{dom}\,\nabla\phi$, we conclude $x = \exp_{t(x)}[-\nabla\phi(t(x))]$ from the same lemma applied to ϕ. Thus $x = t^*(t(x))$ on U and hence μ-a.e., which we'll use to prove $t^* \in S(\nu,\mu)$.

Given any continuous function $h \in C(M)$ we have

$$\int_M h\,d(t^*_{\#}\nu) = \int_M h(t^*(y))\,d\nu(y) = \int_M h(t^*(t(x)))\,d\mu(x) = \int_M h(x)\,d\mu(x)$$

from the change of variables formula (8) applied to $t^*_{\#}\nu$ and $\nu = t_{\#}\mu$, and the fact $t^*(t(x)) = x$ [μ-a.e.] established above. Since $h \in C(M)$ was arbitrary this shows $t^*_{\#}\nu = \mu$ as desired.

Now it follows by symmetry that $V := \mathrm{dom}\,\nabla\phi \cap (t^*)^{-1}(\mathrm{dom}\,\nabla\psi)$ is a Borel set with $\nu[V] = 1$, while $t^*(t(y)) = y$ holds for each $y \in V$ to conclude the corollary. □

Theorem 11 (Polar factorization of maps). *Let (M,g) be a connected, compact Riemannian manifold, C^3-smooth and without boundary. Fix $c(x,y) = d^2(x,y)/2$, a Borel map $s : M \to M$ and a Radon measure $\mu \ll \mathrm{vol}$ on M. If $s_{\#}\mu \ll \mathrm{vol}$, meaning s maps no set with positive μ measure to a set of zero volume, then $s = t \circ u$ [μ-a.e.] for some map $t(x) = \exp_x[-\nabla\psi(x)]$ with $\psi = \psi^{cc}$ and a measure preserving map $u_{\#}\mu = \mu$. The two factoring maps $u, t : M \to M$ are unique μ-a.e. and Borel.*

Proof. Use (2) to define $\nu := s_{\#}\mu$ as the push-forward of the measure μ through the map $s : M \to M$. Compactness of M combines with the local finiteness of Radon measures to provide a normalization constant $\tau^{-1} := \mu[M] = \nu[M] < \infty$ which makes $\tau\mu$ and $\tau\nu$ probability measures, both absolutely continuous with respect to the volume by hypothesis. Theorem 9 provides a potential $\psi = \psi^{cc}$ on M for which the Borel map $t(x) = \exp_x[-\nabla\psi(x)]$ pushes $\tau\mu$ forward to $\tau\nu$, or equivalently $t \in S(\mu,\nu)$.

Its corollary provides a Borel inverse $t^* \in S(\nu,\mu)$ such that $t(t^*(y)) = y$ holds on a Borel set $V \subset M$ of full ν-measure: $\tau\nu[V] = 1$. Setting $u := t^* \circ s$, one immediately verifies $s(x) = t(u(x))$ holds on $s^{-1}(V)$, hence μ-a.e. The existence proof is completed by noting that $t^* \circ s \in S(\mu,\mu)$ follows from $s \in S(\mu,\nu)$ and $t^* \in S(\nu,\mu)$: three changes of variables (8) yield

$$\int_M h\, d(u_\# \mu) = \int_M h(t^*(s(x)))\, d\mu(x) = \int_M h(t^*(y))\, d\nu(y) = \int_M h\, d\mu$$

for each $h \in C(M)$, proving u measure-preserving, $u_\#\mu = \mu$.

To establish the uniqueness of this decomposition, suppose $s = t' \circ u'$ holds on a subset $U' \subset M$ of full μ-measure, where $t'(x) = \exp_x[-\nabla\phi(x)]$ with $\phi = \phi^{cc}$ and $u' \in S(\mu,\mu)$. Clearly $t' \in S(\mu, s_\#\mu)$, so $t' = t$ holds on a set $U \subset M$ of full μ-measure by Theorem 9. For $x \in s^{-1}(V) \cap u'^{-1}(U) \cap U'$, which is to say μ-a.e., one has $s(x) = t'(u'(x)) = t(u'(x))$ hence $u(x) := t^*(s(x)) = u'(x)$, to complete the uniqueness proof. □

4 Polar Factorization as Nonlinear Hodge Theorem

It is interesting to expose in the Riemannian setting the formal link between the polar factorization of maps and the Hodge/Helmholtz decomposition of vector fields uncovered in the Euclidean context by Brenier [Br1]. On a compact manifold, this decomposition asserts that any smooth vector field $\mathbf{v} : M \to TM$ decomposes as $\mathbf{v} = \mathbf{w} + \nabla\zeta$ where $\mathbf{w}$ is divergence free and $\zeta : M \to \mathbf{R}$. Just as a vector field generates (and linearly approximates) a one-parameter family of diffeomorphisms, the Hodge/Helmholtz decomposition arises from the *linear* approximation of the polar factors in Theorem 11. Indeed, let $s(x,\tau) \in M$ be the flow along the vector field $\mathbf{v}$ obtained by integrating

$$\dot{s}(x,\tau) = \mathbf{v}\big(s(x,\tau)\big) \tag{22}$$

from the initial condition $s(x,0) = x$; the dot denotes differentiation with respect to τ. At each instant τ, Theorem 11 factorizes the diffeomorphism $s(x,\tau) = t(u(x,\tau),\tau)$ as the composition of a volume-preserving map $u(\,\cdot\,,\tau)$ with $t(\,\cdot\,,\tau) := \exp[-\nabla\psi(\,\cdot\,,\tau)]$, where $\psi(x,\tau)$ is a $d^2/2$-concave function of x for each τ. As we shortly derive, the Hodge theorem follows formally from the relation

$$\dot{s}(z,0) = \tfrac{\partial}{\partial\tau}\big[t(u(x,\tau),\tau)\big]_{(z,0)} \tag{23}$$

and the identifications $\mathbf{w} = \partial u/\partial\tau$ and $\zeta = -\partial\psi/\partial\tau$ evaluated at $\tau = 0$.

Since the derivation is formal, we simply assume the map $u(x,\tau)$ and potential $\psi(x,\tau)$ provided by Theorem 11 are C^2 smooth in both space and time; it follows that u is a diffeomorphism. Now the unique polar factors of $s(x,0)=x$ are $u(x,0)=x=t(x,0)$. As long as τ is small, our smoothness assumption ensures all three maps stay close to the identity, so we may work in a small neighbourhood of any $(\mathbf{0},z)\in TM$ in the tangent bundle, where we have normal coordinates $(p^1,\dots,p^n,x^1,\dots,x^n)$ vanishing at $(\mathbf{0},z)$. Letting $\gamma(\mathbf{p},x):=\exp_x\mathbf{p}$ so that $s(x,\tau)=\gamma(-\nabla\psi(u(x,\tau),\tau),u(x,\tau))$, we compute

$$\begin{aligned}\frac{\partial s^i}{\partial\tau}\Big|_{(z,0)} &= \frac{\partial\gamma^i}{\partial x^j}\Big|_{(\mathbf{0},z)}\dot{u}^j(z,0)+\frac{\partial\gamma^i}{\partial p^j}\Big|_{(\mathbf{0},z)}\frac{\partial}{\partial\tau}\Big[-g^{jk}\big(u(x,\tau)\big)\frac{\partial\psi}{\partial x^k}\big(u(x,\tau),\tau\big)\Big]_{(z,0)}\\ &=\Big[\dot{u}^i-\frac{\partial g^{ik}}{\partial x^m}\Big|_z\frac{\partial\psi}{\partial x^k}\dot{u}^m-g^{ik}(z)\Big(\frac{\partial^2\psi}{\partial x^m\partial x^k}\dot{u}^m+\frac{\partial^2\psi}{\partial\tau\partial x^k}\Big)\Big]_{(z,0)}\\ &=\Big[\dot{u}^i-g^{ik}(z)\frac{\partial\dot{\psi}}{\partial x^k}\Big]_{(z,0)}\,,\end{aligned}\tag{24}$$

with summation on like indices. Here the expressions $\gamma^i(\mathbf{0},x)=x^i$ and $\gamma^i(\mathbf{p},z)=p^i$ in normal coordinates were used to obtain the second equality, while $\psi(x,0)=\mathit{const}$ throughout M (since $t(x,0)=x$) has been invoked to eliminate purely spatial derivatives of $\psi(x,0)$ from the third equality. Identifying $\mathbf{w}(z)=\dot{u}(z,0)$ and $\zeta(z)=-\dot{\psi}(z,0)$, we recover the Hodge/Helmholtz decomposition $\mathbf{v}=\mathbf{w}+\nabla\zeta$ globally from (22)–(24); here $\mathbf{w}$ is divergence free since the C^2 diffeomorphism $u(z,\tau)$ of M preserves volumes at each τ.

5 Other Cost Functions on Riemannian Manifolds

As in the Euclidean case [Ca1], [GM1,2], the techniques developed above can be adapted to other strictly convex functions $c(x,y)=\Lambda(d(x,y))$ of the geodesic distance on non-compact Riemannian manifolds. Without describing the technicalities, we give a flavor for the sort of result which may be obtained, by stating (without proof) a theorem which can be proved under the simplifying hypothesis

$$c(x,y)=\int_0^{d(x,y)}\lambda(\tau)d\tau\,,\tag{25}$$

with

$$\lambda:(0,\infty)\to\mathbf{R}\text{ continuously increasing from }0=\lim_{\tau\to0}\lambda(\tau)\,.\tag{26}$$

Here continuity of $\lambda(\tau)$ implies the cost $c(x,y)$ remains superdifferentiable.

Compactness of the manifold may also be relaxed, provided minimal geodesics join each pair of points in the supports of the mass distributions μ and ν. Here *support* $\operatorname{spt}\nu$ refers to the smallest closed set in M carrying the full mass of ν. Recall [GM2]:

DEFINITION 12 (c-transform of a function on Y). *Fix $c : M \times M \to \mathbf{R} \cup \{+\infty\}$. Given $Y \subset M$, define a special subset $\mathcal{I}^c_Y$ of the infimal convolutions* (3) *by*

$$\mathcal{I}^c_Y := \left\{\phi^c \;\middle|\; \phi : M \to \mathbf{R} \cup \{-\infty\} \text{ and } \phi(y) = -\infty \text{ unless } y \in Y\right\}. \quad (27)$$

In particular, taking $Y = M$ yields the class $\mathcal{I}^c_M$ of all c-concave functions on M, also known as infimal convolutions with c. The opposite extreme – Y consisting of finitely many (say k) points – makes $\mathcal{I}^c_Y$ a k-dimensional manifold with boundary.

Theorem 13 (Strictly convex costs on non-compact manifolds). *Let (M, g) be a C^3-smooth Riemannian manifold. Choose $c(x, y) = \Lambda(d(x, y))$ satisfying* (25)–(26), *and compactly supported Borel probability measures $\mu \ll$ vol and ν on M. Fix a compact $Y \supset \operatorname{spt}\nu$, and assume each pair of points $x \in \operatorname{spt}\mu$ and $y \in Y$ are connected by an unbroken minimal geodesic. Then there exists $\psi \in \mathcal{I}^c_Y$ such that the map*

$$t(x) = \begin{cases} \exp_x\left[-\frac{\lambda^{-1}(|\nabla\psi|_x)}{|\nabla\psi|_x}\nabla\psi(x)\right] & \text{where } \nabla\psi \neq 0 \\ x & \text{otherwise} \end{cases} \quad (28)$$

pushes μ forward to ν. This map is uniquely characterized in $S(\mu, \nu)$ by the formula (28) *with $\psi \in \mathcal{I}^c_Y$. Moreover, $s = t$ minimizes* (1) *uniquely on $S(\mu, \nu)$.*

Since $Y \subset Z$ implies $\mathcal{I}^c_Y \subset \mathcal{I}^c_Z$ in definition (27), the strongest existence result is obtained by taking $Y = \operatorname{spt}\nu$ in the theorem. The strongest uniqueness result is obtained by taking $Y \subset M$ as large as possible.

Compactness of Y makes it easy to see $\psi \in \mathcal{I}^c_Y$ must be locally Lipschitz throughout M, with Lipschitz constant strictly smaller than $\lambda(\infty) := \lim_{\tau\to+\infty}\lambda(\tau)$ locally. Together with (26), this ensures the map $t(x)$ is well-defined almost everywhere by (28). As in Lemma 7, differentiating $c(x, y) - \psi(x) - \psi^c(y) \geq 0$ with respect to x at a minimizing pair (x_0, y_0) yields both the distance $\lambda^{-1}(|\nabla\psi|_{x_0})$ and direction $\nabla\psi(x_0)$ of optimal transport, thus dictating the form $y_0 = t(x_0)$ of the optimal map (28).

References

[ADV] S. ALESKER, S. DAR, V. MILMAN, A remarkable measure preserving diffeomorphism between two convex bodies in $\mathbf{R}^n$, Geom. Dedicata 74 (1999), 201–212.

[B] F. BARTHE, On a reverse form of the Brascamp-Lieb inequality, Invent. Math. 134 (1998), 335–361.

[BeB] J.-D. BENAMOU, Y. BRENIER, Weak existence for the semi-geostrophic equations formulated as a coupled Monge–Ampère/transport problem, SIAM J. Appl. Math. 58 (1998), 1450–1461.

[Br1] Y. BRENIER, Décomposition polaire et réarrangement monotone des champs de vecteurs, C.R. Acad. Sci. Paris Sér. I Math. 305 (1987), 805–808.

[Br2] Y. BRENIER, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (1991), 375–417.

[BuD] G.R. BURTON, R.J. DOUGLAS, Rearrangements and polar factorisation of countably degenerate functions, Proc. Roy. Soc. Edinburgh Sect. A, 128 (1998), 671–681.

[C] X. CABRÉ, Nondivergent elliptic equations on manifolds with nonnegative curvature, Comm. Pure Appl. Math. 50 (1997), 623–655.

[Ca1] L. CAFFARELLI, Allocation maps with general cost functions, in "Partial Differential Equations and Applications," (P. Marcellini, et al., eds.), Lecture Notes in Pure and Appl. Math. 177 Dekker, New York (1996), 29–35.

[Ca2] L.A. CAFFARELLI, The regularity of mappings with a convex potential, J. Amer. Math. Soc. 5 (1992), 99–104.

[Ca3] L.A. CAFFARELLI, Boundary regularity of maps with convex potentials, Comm. Pure Appl. Math. 45 (1992), 1141–1151.

[Car] M.P. DO CARMO, Riemannian Geometry, Mathematics: Theory and Applications, Birkhäuser, Boston, 1992.

[Co1] D. CORDERO-ERAUSQUIN, Inégalité de Prékopa–Leindler sur la sphère, C.R. Acad. Sci. Paris Sér. I Math. 329 (1999), 789–792.

[Co2] D. CORDERO-ERAUSQUIN, Sur le transport de mesures périodiques, C.R. Acad. Sci. Paris Sér. I Math. 329 (1999), 199–202.

[CoMS] D. CORDERO-ERAUSQUIN, R.J. MCCANN, M. SCHMUCKENSCHLÄGER, A Riemannian interpolation inequality à la Borell, Brascamp and Lieb, Invent. Math. (2001), to appear.

[CuD] M.J.P. CULLEN, R.J. DOUGLAS, Applications of the Monge–Ampère equation and Monge transport problem to meteorology and oceanography, in "Monge Ampère Equation: Applications to Geometry and Optimization" (L.A. Caffarelli, M. Milman, eds.), Contemp. Math. 226, American Mathematical Society, Providence (1998), 33–53.

[CuDRS] M.J.P. CULLEN, R.J. DOUGLAS, I. ROULSTONE, M.J. SEWELL, Generalised semigeostrophic theory on a sphere, preprint.

[DP] G. DE CECCO, G. PALMIERI, Length of curves on Lip manifolds, Atti Accad. Naz. Lincei Cl. Sci. Fis. Math. Natur. Rend. (9) Mat. Appl. 1 (1990), 215–221.

[E] L.C. EVANS, Partial differential equations and Monge–Kantorovich mass transfer, in "Current Developments in Mathematics" (R. Bott, et al., eds.), International Press, Cambridge (1997), 26–78.

[EG] L.C. EVANS, R.F. GARIEPY, Measure Theory and Fine Properties of Functions, Stud. Adv. Math. CRC Press, Boca Raton, 1992.

[G] W. GANGBO, An elementary proof of the polar factorization of vector-valued functions, Arch. Rational Mech. Anal. 128 (1994), 381–399.

[GM1] W. GANGBO, R.J. MCCANN, Optimal maps in Monge's mass transport problem, C.R. Acad. Sci. Paris Sér. I Math. 321 (1995), 1653–1658.

[GM2] W. GANGBO, R.J. MCCANN, The geometry of optimal transportation, Acta Math. 177 (1996), 113–161.

[Gr] L.M. GRAVES, The Theory of Functions of Real Variables, International Series in Pure and Applied Mathematics, McGraw Hill, New York, 2nd edition, 1956.

[JKO] R. JORDAN, D. KINDERLEHRER, F. OTTO, The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. 29 (1998), 1–17.

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

[KoN] S. KOBAYASHI, K. NOMIZU, Foundations of Differential Geometry, Wiley Classics Lib. John Wiley & Sons, New York, 1996.

[M1] R.J. MCCANN, Existence and uniqueness of monotone measure-preserving maps, Duke Math. J. 80 (1995), 309–323.

[M2] R.J. MCCANN, A convexity principle for interacting gases, Adv. Math. 128 (1997), 153–179.

[Mi] J. MILNOR, Morse Theory, Ann. of Math. Stud. 15, Princeton University Press, Princeton, 1969.

[Mo] 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, (1781), 666–704.

[O] F. OTTO, The geometry of dissipative evolution equations: The porous medium equation, Comm. Partial Differential Equations, to appear.

[RR] S.T. RACHEV, L. RÜSCHENDORF, Mass Transportation Problems, Probab. Appl. Springer-Verlag, New York, 1998.

[V] C. VILLANI, Topics in mass transportation, in preparation.

ROBERT J. MCCANN, Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 3G3, mccann@math.toronto.edu

Submitted: January 2000
Revised version: July 2000