Alan Weinstein*

Almost invariant submanifolds
for compact group actions

Abstract. We define a C^1 distance between submanifolds of a riemannian manifold M and show that, if a compact submanifold N is not moved too much under the isometric action of a compact group G, there is a G-invariant submanifold C^1-close to N. The proof involves a procedure of averaging nearby submanifolds of riemannian manifolds in a symmetric way. The procedure combines averaging techniques of Cartan, Grove/Karcher, and de la Harpe/Karoubi with Whitney's idea of realizing submanifolds as zeros of sections of extended normal bundles.

1. Introduction

A compact (not necessarily connected) Lie group G carries a (unique) bi-invariant probability measure. Using this measure, one can average orbits of actions of G on affine convex sets to obtain fixed points. In particular, if G acts on a manifold M, G leaves invariant a riemannian metric on M, and this metric can sometimes be used to obtain fixed points for the nonlinear action of G on M itself. By this method, Élie Cartan proved that a compact group G acting by isometries on a simply-connected manifold M of nonpositive sectional curvature always has a fixed point.[1] Cartan's result was extended by Grove and Karcher [14] to arbitrary manifolds under the assumption that the G-action has an orbit which is sufficiently small relative to a distance scale provided by the geometry of M.

In this paper, we develop a method for averaging nearby submanifolds in a riemannian manifold. This enables us to extend the Grove-Karcher theorem from points to submanifolds; i.e. we establish that, if a riemannian G-manifold has a compact submanifold N whose images under the G-action are sufficiently C^1-close to one another, relative to the geometry

A. Weinstein: Department of Mathematics, University of California, Berkeley, CA 94720, USA, e-mail: alanw@math.berkeley.edu

[1] In Note III, n⁰ 19 of [8], Cartan proved the fixed point theorem for a finite group; he then remarked on p. 19 of [7] that the same argument works for a compact group. See Theorem 13.5 in Chap. 1 of Helgason's book [16] for a full proof in the compact case.

of the pair (M, N), then there is a G-invariant submanifold near N. Extending the Cartan-Grove-Karcher nonlinear averaging method to submanifolds requires new estimates on the geometry of tubular neighborhoods. Our construction is based in an essential way on Whitney's idea[2] [31] of realizing submanifolds as zeros of sections of vector bundles.

Our result fits nicely into the "stability" framework promoted by Ulam (see for example page 584 of [28], as well as Anderson's interesting discussion using nonstandard analysis in [3]). Ulam asks when a mathematical object which "almost solves" a certain problem can be shown to be near an exact solution. In fact, our interest in the invariant submanifold problem was motivated by some cases of the following "almost-homomorphism problem": given a map $\phi : G \to H$ between groups such that $\phi(gh)$ is close in some sense to $\phi(g)\phi(h)$ for all g and h in G, can one conclude that there is a homomorphism near ϕ? This problem was posed by Ulam and solved by Hyers [17] for $G = H = \mathbb{R}$. It was solved for compact Lie groups G and H, with effective estimates, by Grove, Karcher, and Ruh [15] using the Grove-Karcher fixed-point theorem. When G is compact and H is the unitary group of a Hilbert space, it was solved by de la Harpe and Karoubi [9].

Our invariant submanifold theorem implies an almost-homomorphism theorem when G is finite and H is the group of diffeomorphisms of a compact manifold. Our ultimate aim is to extend the almost-homomorphism theorem to cases where G is a compact group and H is the group of smooth bisections of a Lie groupoid. (See [6].) Such an extension would imply very strong structure theorems for certain groupoids. We refer to the lecture [30] for an overview of these potential applications and their connection with symplectic geometry. Details of these applications will appear in subsequent papers.

An important step in our proof requires averaging in a Grassmann manifold. This can be accomplished by Cartan-Grove-Karcher averaging, but we get estimates better suited to our purposes by identifying subspaces with projections and averaging the projections. The averaging of projections has already been treated, in finite and infinite dimensions, by a number of authors over several decades, including de la Harpe and Karoubi in the paper cited above. We review some of this work and derive the estimates we need in an Appendix.

The body of the paper is organized as follows. In Sect. 2, we define a C^1-distance between submanifolds of a riemannian manifold and, in terms of this distance, state our main theorem on invariant manifolds, Theorem 2.2, which follows immediately from the averaging theorem, Theorem 2.3.

---

[2] See Chap. V of the cited article. After finishing this paper, I looked again at this reference and discovered that Whitney used a center of mass construction, too!

Section 3 develops estimates on geometric objects in tubular neighborhoods; the reader may wish to skim this section and return to it after reading Sect. 4, where we first outline the proof of the averaging theorem and then fill in the details. In Sect. 5, we discuss some alternate approaches to the averaging of submanifolds which may be interesting in their own right, though we have not yet succeeded in implementing any of them. We conclude in Sect. 6 with a discussion possible extensions of our results, and the relation of our work to other areas.

In view of potential applications to almost actions of compact Lie groups, we note that most of our work extends without modification to the case where M is a Hilbert manifold, though the submanifold N must remain finite-dimensional. There is just one issue, the application of degree theory, which now restricts our main result to finite-dimensional M.

We have not tried to find optimal values for the many constants in the estimates in Sects. 3 and 4. Rather than denoting these constants by letters, though, we have chosen numerical values which are valid and which may be recognized when they reappear. The resulting constants in the main theorems are therefore very far from optimal. The reader is invited to find better ones.

A word on notation: we will write $d(\cdot, \cdot)$ to denote all of the many distance functions used in this paper, decorating it with subscripts only when necessary to avoid ambiguity. We will use bars $|\cdot|$ for lengths of paths and vectors, and double bars $\|\cdot\|$ for operator norms.

Many people have offered useful suggestions in the course of this work. I would particularly like to thank David Aldous, Eugenio Calabi, Marc Chamberland, Jost Eschenburg, Robert Greene, Mikhail Gromov, Karsten Grove, Hermann Karcher, Yi Ma, Yoshiaki Maeda, and Jamie Sethian.

2. Bounded geometry and distance between submanifolds

Let N be a closed submanifold of the riemannian manifold M. The **normal bundle** $\nu N \subset TM$ consists of those tangent vectors along N which are orthogonal to TN. The open ball bundle of radius r in νN will be denoted by $\nu^r N$.

We define the **normal exponential map** $\exp_N$ to be the restriction to νN of the exponential map $\exp : TM \to M$. (If M is not complete, $\exp_N$, like $\exp$, is defined only on a proper open neighborhood of the zero section.) The **normal injectivity radius** i_N of N is

$$\sup\{r \mid \exp_N \text{ (is defined and) is an embedding on } \nu^r N\}.$$

(When N is a point, its normal injectivity radius is just the usual injectivity radius of M at that point.) The restriction of $\exp_N$ to $\nu^{i_N} N$ is still an

embedding; we call its image $\mathcal{T}N$ *the* tubular neighborhood of N. For any $r \in (0, i_N]$, the set

$$\mathcal{T}^r N = \{x \in M | d(x, M) < r\} = \exp(\nu^r N)$$

will be called the r-**tube** about N.

The bundle retraction in νN is transferred by $\exp_N$ to a retraction $\pi_N : \mathcal{T}N \to N$ which maps each point of $\mathcal{T}N$ to the nearest point in N. The fibres of π_N will be called the **normal slices** for N.

2.1. The Grassmann bundle

We will denote by GM the Grassmann bundle over M, whose fibre over each point x is the (disconnected) manifold of all finite-codimensional subspaces of the tangent space $T_x M$. We give each fibre the Finsler metric described in the Appendix.

These fibre metrics define a supremum distance on the space of sections of GM over any subset of M.

The **Gauss map** of a submanifold N is defined as the section of GM over N whose value at each point is the normal space to N. This section extends in a natural way to a section $\Gamma_N : \mathcal{T}N \to GM$ defined over the tubular neighborhood of N: we take the normal spaces to N and parallel translate them along all the geodesics normal to N. The image of Γ_N, when thought of as a vector bundle over the tubular neighborhood, will be called the **vertical bundle** and its orthogonal complement (obtained by translating the tangent spaces of N) the **horizontal bundle**.

The vertical bundle contains the tangent vectors to the normal geodesics but does not in general coincide with the tangent bundle along the normal slices. We will refer to the latter as the **quasi-vertical bundle**. (The two bundles are compared in Sect. 3.3.) The orthogonal complement to the quasi-vertical bundle will be called the **quasi-horizontal bundle**.

2.2. Bounded geometry and C^1 distance

To define a geometrically meaningful C^1 distance between submanifolds, we should measure distances and angles in the same units. Since angles are dimensionless, we do this by choosing a positive number c which functions as a "unit of inverse length." Given c, we say that a pair (M, N) consisting of a manifold M and its submanifold N has **geometry bounded by** c if: (i) $i_N > 1/c$; (ii) the sectional curvatures of M in the $1/c$-tube about N are bounded in absolute value by c^2; (iii) the injectivity radius of each point in

the tube is at least $1/c$. If N is compact, (M, N) necessarily[3] has geometry bounded by some c.

By rescaling the metric, we can always convert geometry bounded by c to geometry bounded by 1. (Note that rescaling the metric does not change the metric on the Grassmann manifold.) We will therefore restrict the statement of our results to this case, leaving it to the reader who so desires to put the c's back in. A pair (M, N) with geometry bounded by 1 will be called a **gentle** pair.

Remark 2.1 Condition (iii) in the definition of bounded geometry is used in only a couple of places below. Perhaps our results can be obtained without this condition. One example where the condition is violated is a geodesic circle N in a very thin cylinder M.

If N and N' are submanifolds of M such that N' lies in $\mathcal{T}N$ and is the image under $\exp_N$ of a section $\sigma(N, N')$ of νN, we define their (nonsymmetric! but see Remark 3.18) C^1-distance $d(N, N')$ as follows. We assign two numbers to each to each $x' \in N'$: the length of the geodesic segment τ from x' to the nearest point x on N and the distance between $T_{x'}N'$ and the parallel translate of $T_x N$ along τ. The C^1-distance $d(N, N')$ is defined as the supremum of all these numbers as x' ranges over N'. Since the distance between two subspaces equals the distance between their orthogonal complements (Corollary A.6), we can use normal spaces instead of tangent spaces and interpret $d(N, N')$ as the maximum of the supremum norm of $\sigma(N, N')$ and the distance between the extended Gauss map Γ_N and the Gauss map $\Gamma_{N'}$ as sections of GM over N'.

We will occasionally use a C^0 distance as well, defined simply of the maximum over N' of the distance to the nearest point of N. This distance is defined for any pair of compact subsets of a metric space.

We can now state the main result of this paper.

Theorem 2.2 *Let M be a riemannian G-manifold for a compact group G, and let (M, N) be a gentle pair. If $d(N, gN) < \epsilon < \frac{1}{20000}$ for all $g \in G$, then there is a G-invariant submanifold $\overline{N}$ with $d(N, \overline{N}) < 136\sqrt{\epsilon}$.*

This theorem follows immediately from a more general statement, analogous to Grove and Karcher's center of mass construction [14] for points.

Theorem 2.3 *Let M be a riemannian manifold and $\{N_g\}$ a family of compact submanifolds of M parametrized in a measurable way[4] by elements*

---

[3] Even if M has infinite dimension!

[4] We will not attempt to give a precise definition of this notion, which should involve measurable dependence of derivatives on parameters to permit the exchange of integration and differentiation. In the applications we have in mind, the parameter space is a manifold and the dependence is smooth.

g of a probability space G, such that all the pairs (M, N_g) are gentle. If $d(N_g, N_h) < \epsilon < \frac{1}{20000}$ for all g and h in G, there is a well defined **center of mass** *submanifold $\overline{N}$ with $d(N_g, \overline{N}) < 136\sqrt{\epsilon}$ for all g in G. The center of mass construction is equivariant with respect to isometries of M and measure preserving automorphisms of G.*

Remark 2.4 When G consists of two points with equal masses, we obtain a natural midpoint construction for pairs of nearby submanifolds in a riemannian manifold. We know of no simpler construction, even for pairs of embedded circles in euclidean space.

The numbers 20000 and 136 in the theorems above can certainly be reduced if more care is taken in the many estimates used in the proof. On the other hand, it is not at clear whether the factor $\sqrt{\epsilon}$ can be replaced by a multiple of ϵ itself. Our proof does show that the C^0 distance from N (or N_g) to $\overline{N}$ is bounded by 100ϵ, but we do not see how to get an estimate linear in ϵ for the derivative. (See Remark 3.14.)

3. Geometry in tubular neighborhoods

In this section, we will establish some geometric estimates needed for proving the main theorem.

3.1. Estimate on the second fundamental form

Recall that the second fundamental form B of N is defined for a normal vector v and tangent vectors w_i at $p \in N$ by

$$B_v(w_1, w_2) = -\langle v, \mathbf{D}_{w_1} W_2 \rangle,$$

where W_2 is a vector field tangent to N with $W_2(p) = w_2$, $\mathbf{D}$ is the riemannian connection, and $\langle\ ,\ \rangle$ is the riemannian inner product.[5] The norm $\|B\|$ is then defined as the supremum of $|B_v(w, w)|$ as v and w range over unit normal and tangent vectors respectively, and their basepoint ranges over N.

Since focal points along geodesics normal to a submanifold cannot occur within the normal injectivity radius, gentleness of (M, N) implies that all focal points to N occur at a distance greater than 1 from N. This fact and

---

[5] We follow the sign convention used by Warner [29] (following [4]) and Eschenburg [10]. This convention is consistent with the interpretation of the second fundamental form as the "first variation of the metric," but most authors use the opposite sign, consistent with the direction of principal curvature vectors of curves in a submanifold. Lang [21] introduces second fundamental forms with *both* signs as the off-diagonal entries in a single 2×2 block matrix.

Proposition 3.1 below will imply a bound for the second fundamental form of N (Corollary 3.2).

Proposition 3.1 *Let v and w be unit normal and tangent vectors respectively at $p \in N$. Suppose that $B_v(w, w) = -b$ for some $b > 0$ and that the sectional curvatures of M are bounded above by -1. Then the geodesic segment γ with initial velocity v and length L contains at least one focal point to N as long as $b \geq 1/L + L/2$.*

Proof. We use a standard argument to show that the index form of γ (with appropriate boundary conditions) is not positive definite.

Using parallel translation and a slight abuse of notation, we represent any vector field X along γ by a function $X : [0, L] \rightarrow T_p M$. The index form on the space of such X with $X(t) \in T_p N$ and $X(L) = 0$ is defined by the integral:

$$I(X, X) = B_v(X(0), X(0)) + \int_0^L \left(\langle X'(t), X'(t) \rangle - \langle R(t)X(t), X(t) \rangle \right) dt,$$

where $R(t)$ is the operator given via the Riemann curvature tensor as

$$X \mapsto R_{\gamma(t)}(\gamma'(t), X)\gamma'(t).$$

For the "linearly decreasing" vector field $X(t) = (1 - t/L)w$,

$$I(X, X) = B_v(w, w) + \int_0^L ((1/L^2) - (1 - t/L)\langle R(t)w, w \rangle)dt$$

$$= -b + 1/L - \int_0^L (1 - t/L)\langle R(t)w, w \rangle dt.$$

Since the sectional curvatures are at least -1,

$$I(X, X) \leq -b + 1/L + \int_0^L (1 - t/L)dt = -b + 1/L + L/2.$$

When $b \geq 1/L + L/2$, $I(X, X) \leq 0$, so there must be at least one focal point on γ. $\qquad\square$

Corollary 3.2 *If (M, N) is a gentle pair, then the second fundamental form of N satisfies the bound $\|B\| \leq \frac{3}{2}$.*

Proof. The distance from N to any focal point is at least 1. If $|B_v(w, w)| > \frac{3}{2}$ for some unit vectors v and w, we can change the sign of v if necessary to assure that $B_v(w, w) < -\frac{3}{2}$, and Proposition 3.1 leads to a contradiction.

$\qquad\square$

3.2. Comparison of curves on and near N

The following lemma shows that the retraction $\pi_N : \mathcal{T}N \to N$ does not increase lengths too much.

Lemma 3.3 *Let (M, N) be a gentle pair. Let σ be a path in the r-tube about N for some $r < \frac{1}{2}$. Then*

$$(\cos r - \tfrac{3}{2} \sin r)|\pi_N \circ \sigma| \le |\sigma|. \tag{1}$$

(The function in parentheses decreases with r and takes the values 1 at $r = 0$, $.59\ldots$ at $r = \frac{1}{4}$, and $.15\ldots$ at $r = \frac{1}{2}$.)

Proof. Assume that σ is defined on the interval $[0, 1]$ and construct a "rectangle" $\Sigma : [0, 1] \times [0, 1] \to M$ as follows: for each $s \in [0, 1]$, $s \mapsto \Sigma(s, 1)$ is the minimizing geodesic from $\pi_N(\sigma(s))$ to $\sigma(s)$, parametrized proportionally to arc length. Corresponding tangent vectors to σ and $\pi_N \circ \sigma$ are thus the terminal and initial values of N-Jacobi fields along normal geodesics. By Warner's generalization of the Rauch comparison theorem ([29], Theorem 4.3(b)), using the upper bound 1 for the sectional curvature and the bound $B_{\sigma'(0)} \ge -\frac{3}{2}I$ (Corollary 3.2) on the second fundamental form of N, we obtain the inequality $|\sigma'(s)| \ge (\cos r - \frac{3}{2}\sin r)|(\pi_N \circ \sigma)'(s)|$ for all s, which gives (1). $\qquad\square$

Corollary 3.4 *Let (M, N) be a gentle pair, x and y points of the $\frac{1}{4}$-tube about N such that $d(x, y) < \frac{1}{2}$. Then*

$$d(\pi_N(x), \pi_N(y)) < 7d(x, y).$$

Proof. A path of length less than $\frac{1}{2}$ from x to y remains in the $\frac{1}{2}$-tube about N. Lemma 3.3 gives the desired inequality since $1/.15 < 7$. $\qquad\square$

A similar argument gives:

Corollary 3.5 *For a gentle pair (M, N), let x and y be points of N such that $d_M(x, y) = \alpha < \frac{1}{2}$. Then*

$$d_M(x, y) \le d_N(x, y) \le \frac{1}{\cos \frac{1}{2}\alpha - \frac{3}{2}\sin \frac{1}{2}\alpha} d_M(x, y) \le 7d_M(x, y).$$

Remark 3.6 This estimate and the ones to follow can surely be sharpened, but the versions which we give suffice for our purposes.

3.3. Vertical and quasi-vertical bundles

The quasi-vertical bundle, tangent to the normal slices for $N \subset M$, is generally not parallel along normal geodesics and thus does not coincide with the vertical bundle. We will estimate the distance between the two bundles by estimating the covariant derivative of the quasi-vertical bundle along normal geodesics.

Proposition 3.7 *For $x \in \mathcal{T}N$, the distance between the quasi-vertical and vertical spaces is at most $\frac{1}{4}d(x, N)^2$. The distance between the projections on this spaces is at most $\frac{1}{5}d(x, N)^2$.*

Proof. Tangent vectors to normal slices are the values of those Jacobi fields along normal geodesics vanishing on N and with derivative there in νN. Let $X(t)$ be such a field along the geodesic $\gamma(t)$, parametrized by arc length. The initial derivative $X'(0)$ belongs to $\nu_{\gamma(0)}N$; for convenience we take it to be a unit vector. As in the proof of Proposition 3.1, we identify all the tangent spaces along γ with $T_{\gamma(0)}M$ by parallel translation.

Since $X(0) = 0$, we write $X(t) = tX'(0) + Z(t)$. Then $Z(0) = Z'(0) = 0$, and $Z''(t) = X''(t) = -R(t)X(t)$ by the Jacobi equation. The curvature estimate $\|R(t)\| \leq 1$ gives first of all the estimate $\|X(t)\| \leq \sinh t$, so $\|Z''(t)\| \leq \sinh t$. Integration gives $\|Z'(t)\| \leq \cosh t - 1$ and then $\|Z(t)\| \leq \sinh t - t \leq t^3/6$. Then $\|X(t)/t - X'(0)\| \leq t^2/6$ and $\|X(t)/t\| \geq 1 - t^2/6$.

The typical quasi-vertical unit vector $(X(t)/t)/\|X(t)/t\|$ is thus within distance

$$\frac{t^2/6}{1 - t^2/6} \leq \frac{t^2}{5}$$

of the vertical vector $X'(0)/\|X(t)/t\|$, from which it follows (for $t \leq 1$) that the distance between the vertical and quasi-vertical spaces at $x = \gamma(t)$ is at most

$$\sin^{-1}(t^2/5) \leq t^2/4 = d(x, N)^2/4,$$

while the difference between the projections on these spaces is at most $\frac{1}{5}d(x, N)^2$. $\square$

An estimate for the covariant derivative of the quasi-vertical bundle in non-vertical directions would simplify our later considerations, but we have not been able to find such an estimate.[6] Instead, we will estimate the covariant derivatives of the vertical bundle and then use Proposition 3.7 to shuttle between the two bundles.

---

[6] The difficulty seems to be related to the fact that a bound on curvature does not imply a bound on the derivatives of Christoffel symbols in normal coordinates; see the example on p. 34 of [18].

Proposition 3.8 *Let (M, N) be a gentle pair, Γ_N its extended Gauss map, and $\mathbb{P}_{\Gamma_N}$ the field of orthogonal projections onto the image of Γ_N (the vertical bundle). Then the covariant derivative $\mathbf{D}\mathbb{P}_{\Gamma_N}$, a field of maps from TM to* End *TM, has operator norm bounded pointwise by 11 on the tube $\mathcal{T}^{\frac{1}{2}}N$.*

Proof. We estimate $\mathbf{D}\mathbb{P}_{\Gamma_N}$ along N in terms of the second fundamental form of N, then we extend the estimate into the tubular neighborhood by differentiating along the normal geodesics.

Along N, $\mathbf{D}_v\mathbb{P}_{\Gamma_N}$ is zero if v is vertical, by the definition of Γ_N. If w is horizontal, $\mathbf{D}_w\mathbb{P}_{\Gamma_N}$ is a symmetric operator which exchanges horizontal and vertical spaces. For vertical v and horizontal h, it follows easily from the definition of the second fundamental form that $\langle(\mathbf{D}_w\mathbb{P}_{\Gamma_N})(v), h\rangle = B_v(w, h)$. Along N, then, $\|\mathbf{D}\mathbb{P}_{\Gamma_N}\| = \|B\|$, which is at most $\frac{3}{2}$ by Corollary 3.2.

To estimate $\mathbf{D}\mathbb{P}_{\Gamma_N}$ at a point $x \in \mathcal{T}N$, we let γ be the unit speed geodesic segment from $\pi_N(x)$ to x and estimate $\mathbf{D}_{\gamma'(s)}\mathbf{D}\mathbb{P}_{\Gamma_N}$. Let Y be a unit length vector field on a neighborhood of γ such that γ is an integral curve of Y. For any unit vector $v \in T_xM$, we can find a vector field Z near γ such that $[Y, Z] = 0$ and $Z(x) = v$.

Now $\mathbf{D}_Y\mathbf{D}_Z\mathbb{P}_{\Gamma_N} = \mathbf{D}_Z\mathbf{D}_Y\mathbb{P}_{\Gamma_N} + [R(Y, Z), \mathbb{P}_{\Gamma_N}]$. But $\mathbf{D}_Y\mathbb{P}_{\Gamma_N} = 0$, $\|Y\| = 1$, and $\|\mathbb{P}_{\Gamma_N}\| = 1$, so $\|\mathbf{D}_Y\mathbf{D}_Z\mathbb{P}_{\Gamma_N}\| \leq 2\|R\| \|Z\|$, and hence

$$\|\mathbf{D}_v\mathbb{P}_{\Gamma_N}\| \leq \frac{3}{2} + d(x, N) \cdot 2\|R\| \sup_{0 \leq s \leq d(x,N)} \|Z(\gamma(s))\|.$$

Since the sectional curvatures are bounded in absolute value by 1, $\|R\| \leq \frac{4}{3}$ (see §6.1 in [5]), so it remains only to estimate the size of Z along γ. For this purpose, we construct a particular choice of Z by forming the rectangle (as used in Lemma 3.3) consisting of minimizing geodesics from N to the points on the geodesic with initial vector $Z(x) = v$ and letting Z be the variational vector field. Z is an N-Jacobi field along γ, so we can estimate it using Warner's comparison theorem [29]. As in Lemma 3.3, we conclude from Warner's theorem that $\|Z(\gamma(s))\| \geq (\cos s - \frac{3}{2} \sin s)\|Z(\gamma(0))\|$ for all s. Setting $s = d(x, N)$, we get: $\|Z(\gamma(0))\| \leq (\cos d(x, N) - \frac{3}{2} \sin d(x, N))^{-1}$, which is less than 7 if $d(x, N) < \frac{1}{2}$.

Now the maximum of $\|Z(\gamma(s))\|$ might occur at an interior point $\gamma(s_0)$ of the segment. If that happens, we must have $\mathbf{D}_{\gamma'(s_0)}Z \perp Z(s_0)$. In that case, we can use Warner's theorem again, comparing with the situation where the initial manifold is totally geodesic, to get an estimate even better than before.

Our conclusion is that $\sup_{0 \leq s \leq d(x,N)} \|Z(\gamma(s))\| \leq 7$, so that $\|\mathbf{D}\mathbb{P}_{\Gamma_N}\| \leq \frac{3}{2} + \frac{1}{2} \cdot 2 \cdot \frac{4}{3} \cdot 7 < 11$.

$\square$

3.4. The potential function of a submanifold

On the tubular neighborhood $\mathcal{T}N$, the distance function from N, $\rho_N(x) = d(x, \pi_N(x)) = \| \exp_N^{-1}(x) \|$, is singular along N, but $P_N = \frac{1}{2}\rho_N^2$ is smooth and has N as a nondegenerate critical manifold (of absolute minima). We call P_N the **potential function** of N, thinking of N as the attracting manifold of a "linear" attracting force field. The points of N can also be recovered from P_N as the nondegenerate minimum points of its restrictions to the normal slices, to each normal slice, or equivalently as the zeros of the projection of the gradient ∇P_N into the quasi-vertical bundle. This description of N will be most useful to us, since the projected gradient is transverse to the zero section in the quasi-vertical bundle. We will be proving Theorem 2.3 by identifying the invariant submanifold $\overline{N}$ as the zero set of an invariant object obtained by averaging ∇P_N over G and projecting into an averaged vertical bundle.

We will need estimates on the hessian $H_N = \mathbf{D}\nabla P_N$, a field of symmetric bilinear forms on the tangent spaces of $\mathcal{T}N$ defined by $H_N(X, Y) = \langle \mathbf{D}_X \nabla P_N, Y \rangle$. Estimates for this hessian when N is a point are given in [5] (Proposition 6.4.6) and play an important role in applications of the center of mass construction. When $\dim N \geq 1$, the estimates are more complicated, since the hessian behaves quite differently in horizontal and vertical directions.

To calculate the hessian of P_N, we adapt the method of Gray [13], who used a Riccati equation to study the second fundamental forms of the boundaries of the r-tubes, i.e. of the level hypersurfaces of P_N. His analysis used the gradient U_N of the function ρ_N, which has the same level hypersurfaces. Since U_N is the unit outward normal to these hypersurfaces, the hessian $S_N = \mathbf{D}U_N$ gives the second fundamental form of the tube boundaries. The covariant derivative of the tensor field S_N in the radial direction satisfies the Riccati equation

$$\mathbf{D}_{U_N} S_N = S_N^2 + R_N,$$

where R_N is the field of operators obtained (as one does for the Jacobi equation) by contracting the radial vector field twice with the Riemann curvature tensor R of M; i.e. $R_N(X) = R(U_N, X)U_N$. The Riccati equation determines S_N in the tubular neighborhood once we know its behavior along N. But S_N, like U_N, is not even defined along N, so this initial behavior must be described asymptotically. This point is dealt with only briefly by Gray, so we paraphrase here the following result of Eschenburg ([10], §6.1).

Proposition 3.9 *Along a unit speed geodesic γ emanating normally from N, the solution S_N of the Riccati equation has the asymptotic expansion*

$S_N(\gamma(t)) = t^{-1}S_{-1} + S_0 + \mathrm{O}(t)$, *where S_{-1} and S_0 are in block matrix form with respect to the decomposition into horizontal and vertical subspaces as follows:*

$$S_{-1} = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right] \qquad S_0 = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & B_{\gamma'(0)} \end{array}\right].$$

Unfortunately, the proof of this Proposition seems to belong to folklore rather than literature. Nevertheless, we will not write a proof here, either. (The proof is a nice exercise in the use of Fermi coordinates or Fermi frames [13].)

We return now to the hessian $H_N = \mathbf{D}\nabla P_N$. Differentiating the relation $\nabla P_N = \rho_N \nabla \rho_N$ yields

$$H_N = U_N \otimes U_N + \rho_N S_N,$$

where $U_N \otimes U_N$ is the operator of orthogonal projection onto the unit vector field U_N. As the hessian of a smooth function, H_N is defined throughout the tubular neighborhood, and its values along N are quite easy to compute. In fact, since P_N is critical at all points of N, we can compute the hessian at any such point as the second derivative matrix in any coordinate system for which the coordinate vector fields form an orthonormal basis at that point. From this observation, it is easy to check (see below) that H_N is block diagonal with respect to the normal–tangent splitting of TM along N. On the tangent bundle TN, H_N is zero, while it equals the riemannian metric on the normal bundle νN.

Since it is not H_N but rather the singular field

$$S_N = (1/\rho_N)(H_N - U_N \otimes U_N)$$

which satisfies the Riccati equation, the initial conditions required for determining or estimating a solution also include the normal (covariant) derivative of H_N along N. It is here that the second fundamental form of N appears, as we saw in Proposition 3.9 and one may easily check in the simple example where N is a circle in the euclidean plane.

3.5. Estimates for the hessian

We will use the following estimate of Eschenburg and Heintze [11] on solutions of Riccati equations to estimate the hessian of the potential function.

Theorem 3.10 (Eschenburg–Heintze [11]) *Let $S_1(t)$ and $S_2(t)$ be maximal solutions of the Riccati equations*

$$S'_j = S_j^2 + R_j$$

defined for $0 < t < t_j$ with values in symmetric matrices such that:

(1) $U := S_2 - S_1$ has a continuous extension to $t = 0$ with $U(0) \geq 0$;
(2) $R_j(t)$ is a smooth function with $R_2(t) \geq R_1(t)$ for all $t \geq 0$.

Then $t_1 \leq t_2$, and $S_1 \leq S_2$ on $(0, t_1)$.

We recall that an inequality of the form $P \leq Q$ between symmetric matrices means that the difference $Q - P$ is positive semidefinite. It implies that the maximal eigenvalue of P is less than that of Q. As a consequence, the inequality $-cI \leq P \leq cI$ for a positive number c implies the bound $\|P\| \leq c$ on the operator norm of P.

We will use Theorem 3.10 to estimate the hessian of the potential function for a gentle pair, by comparison with manifolds of constant curvature ± 1. Our first step will be to examine and estimate the solutions of the Riccati equation when $R(t)$ is a constant multiple of the identity matrix.

It suffices to consider the scalar case, i.e. the equations

$$s' = s^2 \pm 1$$

for a real-valued function $s(t)$ defined for $t > 0$. We must consider two kinds of boundary conditions, corresponding to the vertical and horizontal cases in the geometric situation.

For the vertical case (directions normal to N and their parallel translates along normal geodesics), we choose the solution with a pole at 0, namely $-\cot t$ when the sign of the curvature ± 1 is positive and $-\coth t$ when it is negative. For convenience, we will sometimes use the notation $\cot_+$ for $\cot$ and $\cot_-$ for $\coth$, as well as $\cot_\pm$ to denote whichever is appropriate; thus the general solution can be written as

$$s_{\mathrm{ver}}(t) = \cot_\pm t.$$

For the horizontal case (directions tangent to N and their parallel translates along normal geodesics), we take the solution with initial value b (which can be positive or negative), which is

$$s_{\mathrm{hor}}(t) = \frac{\tan_\pm t + b}{1 \mp b \tan_\pm t}.$$

We will now estimate the function $h(t) = ts(t)$ on the interval $[0, 1]$, since it is this product rather than $s(t)$ alone which gives the hessian of the potential function.

For the vertical direction, we have $h_{\mathrm{ver}}(t) = t \cot_\pm t$ (notice that the pole has disappeared, as it should). With positive curvature, $h(t)$ decreases from 1 at $t = 0$ to $\cot 1 \sim .642\ldots$ at $t = 1$. With negative curvature, $h(t)$ increases from 1 at $t = 0$ to $\coth 1 \sim 1.313\ldots$ at $t = 1$. Thus, we have the estimate

$$.64 < h_{\mathrm{ver}}(t) < 1.32 \ \text{ for } \ 0 \leq t \leq 1.$$

Notice that this estimate also applies to the hessian in the radial direction, which is identically 1, even though the relation between the hessian and the Riccati solution is more complicated in that case.

For the horizontal direction, we will need only an upper bound, and there is no pole to worry about, so we can estimate s_{hor} first and multiply by t afterward. For positive curvature,

$$s_{\text{hor}}(t) = \frac{\tan t + b}{1 - b \tan t}$$

is monotonically increasing in b, so we can set $b = \frac{3}{2}$, by our estimate on the second fundamental form under the assumption of bounded geometry. This gives us

$$|s_{\text{hor}}(t)| < \frac{\tan t + \frac{3}{2}}{1 - \frac{3}{2}\tan t}.$$

To keep the denominator under control, we will require that $\frac{3}{2}\tan t < \frac{1}{2}$, i.e. $\tan t < \frac{1}{3}$, which is satisfied when $t < .321\ldots$. To keep the numbers simple, we will require that $t < \frac{1}{4}$, which leads to the upper bound of $\frac{13}{6}$, which for simplicity we replace by 3. This gives us the estimate

$$|h_{\text{hor}}(t)| \leq 3t \text{ for } 0 \leq t \leq \frac{1}{4}.$$

We are left with the negative case. But since tanh grows more slowly than tan, the estimate from the positive case still applies.

Summarizing the estimates above and applying Theorem 3.10, with the initial conditions given by Proposition 3.9 gives the following:

Proposition 3.11 *Let (M, N) be a gentle pair. At a point of M with distance $r \leq \frac{1}{4}$ from N, the hessian H_N when expressed in block matrix form with respect to the decomposition into vertical and horizontal parts satisfies the inequality:*

$$\begin{bmatrix} .64I & 0 \\ \hline 0 & -3rI \end{bmatrix} < H_N < \begin{bmatrix} 1.32I & 0 \\ \hline 0 & 3rI \end{bmatrix}. \tag{2}$$

Corollary 3.12 *On the $\frac{1}{4}$-tube around N, $\|H_N\| \leq 1.32$.*

Proof. The statement follows from the bound $-.75I \leq H_N \leq 1.32I$.

$\square$

Corollary 3.13 *At a point of M with distance $r \leq \frac{1}{4}$ from N, for a horizontal vector v and a vertical vector w, $|H_N(v, w)| \leq 3\sqrt{r}\,\|v\|\,\|w\|$.*

Proof. We may assume that v and w are unit vectors. By Proposition 3.11, $|H_N(v, v)| < 3r$ and $.64 < H_N(w, w) < 1.32$. In addition, for any real number t, $H_N(tv + w, tv + w) < 3rt^2 + 1.32$. Thus the quadratic function $t \mapsto t^2(H_N(v, v) - 3r) + 2tH_N(v, w) + H_N(w, w) - 1.32$ is negative for all t and hence its discriminant is negative. So $4H_N(v, w)^2 < 4(H_N(v, v) - 3r)(H_N(w, w) - 1.32) \leq 4 \cdot 6r \cdot 0.68$. Therefore, $H_N^2(v, w) < 4.08r$, and so $|H_N(v, w)| < 3\sqrt{r}$.

<div align="right">□</div>

Remark 3.14 The square root in Corollary 3.13 is responsible for the presence of $\sqrt{\epsilon}$ rather than ϵ in the estimates in the main theorems. It seems plausible that the square root could be removed here, since the off-diagonal block of H_N is zero along N, and so it must grow linearly with r on any particular manifold. Because of the singular nature of the relation between S_N and H_N, though, we have not been able to get a linear bound which is uniform for all gentle pairs.

3.6. Geometry relative to nearby submanifolds

In this section, we estimate the difference between geometric objects in the tubular neighborhoods of two nearby submanifolds. To begin, we show that the retractions to nearby submanifolds are close to one another.

Lemma 3.15 *Let (M, N) and (M, N') be gentle pairs such that $d(N, N') < \frac{1}{2}$, and let $x \in M$ with $d(x, N) < \frac{1}{4}$. Then $d(\pi_N(x), \pi_N\pi_N'(x)) < 14d(N, N')$ and $d(\pi_N(x), \pi_N'(x)) < 15d(N, N')$.*

Proof. Let $w = \pi_N(x)$, $y = \pi_{N'}(x)$, and $z = \pi_N(y)$. Let ξ be the unit tangent vector at y to the geodesic segment yx, let η be the parallel translate of ξ to z along the geodesic segment yw, and let ζ be the nearest unit vector to η in the normal space $\nu_z N$. Let r and s be the endpoints of the geodesic segments zr and zs with initial directions η and ζ respectively, each with the same length as yx. We will show below that $d(x, r)$ and $d(r, s)$ are small. The resulting estimate on $d(x, s)$, combined with Lemma 3.3, will give us an estimate on $d(w, z)$, which will lead to the result.

   To estimate $d(x, r)$, we observe that x and r are the endpoints of geodesic segments departing from y and z with initial tangent vectors related by parallel translation along the geodesic yz. It follows that $d(x, r) \leq \lambda d(y, z)$, where λ is a bound on the growth of a Jacobi field with initial derivative 0 along a geodesic segment of length $d(x, y)$, under the assumption that sectional curvatures are ≥ -1.

   We write α for $d(N, N')$ and β for $d(x, N)$. Since $d(x, y) \leq d(x, N) + d(N, N') = \beta + \alpha$, we can take $\lambda = \cosh(\beta + \alpha)$, and since $d(y, z) \leq$

$d(N, N') = \alpha$ we have

$$d(x, r) < \alpha \cosh(\beta + \alpha).$$

For $d(r, s)$, we use the fact that r and s are the endpoints of geodesic segments with the same initial point, length at most $\alpha + \beta$, and initial (unit) tangent vectors making an angle of at most α with one another (since $d(\Gamma_N(z), \Gamma_{N'}(z)) \leq d(N, N') = \alpha$). The usual Jacobi field estimate, using again the lower bound of -1 on curvature, gives

$$d(r, s) < \alpha \sinh(\beta + \alpha),$$

and hence

$$d(x, s) < \alpha(\cosh(\beta + \alpha) + \sinh(\beta + \alpha)) = \alpha e^{\beta + \alpha} < \alpha e^{\frac{1}{3}} < 2\alpha.$$

By Corollary 3.4, $d(w, z) < 14\alpha$. Since $d(y, z) \leq \alpha$, the triangle inequality gives the required estimate for $d(\pi_N(x), \pi'_N(x))$. $\square$

Next we estimate the difference between the tangent vectors from a point x to the submanifolds N and N'.

Lemma 3.16 *Let (M, N) and (M, N') be gentle pairs such that $d(N, N') < \frac{1}{50}$, and let $x \in \mathcal{T}N$ be such that $d(x, N) < \frac{1}{4}$. Then the difference between the initial tangent vectors $\xi_N(x)$ and $\xi_{N'}(x)$ of the minimizing geodesics from x to $\pi_N(x)$ and $\pi_{N'}(x)$ respectively is less than $\frac{45}{2} \, d(N, N')$.*

Proof. As in Lemma 3.15, we let $\alpha = d(N, N')$. According to that lemma, we can connect $\pi_N(x)$ to $\pi_{N'}(x)$ by a path σ of length less than $15\alpha < \frac{3}{10}$. Since the distance from x to N is less than $\frac{1}{4}$, all the points of σ are within a distance 1 of x, so that we can lift σ by the exponential map at x to a path $\tilde{\sigma}$ in the tangent space $T_x M$ and thereby produce a smooth family $\{\gamma_s\}$ of geodesic segments joining x to the points of σ. Note that $\tilde{\sigma}(0)$ and $\tilde{\sigma}(1)$ equal $\xi_N(x)$ and $\xi_{N'}(x)$ respectively.

We estimate the length ratio $|\sigma|/|\tilde{\sigma}|$ by using the Jacobi fields along the segments γ_s. Since the sectional curvature in M is bounded above by 1, the segments have length less than 1, and $\sin x/x > \frac{2}{3}$ for $0 \leq x \leq 1$, we have the estimate

$$|\sigma|/|\tilde{\sigma}| > \tfrac{2}{3}.$$

$\frac{3}{2} \cdot 15\alpha = \frac{45}{2}\alpha$ is therefore an upper bound for the length of $\tilde{\sigma}$ and hence for the difference between $\xi_N(x)$ and $\xi_{N'}(x)$. $\square$

Finally, we estimate the distance between the extended Gauss maps of nearby submanifolds.

Proposition 3.17 *Let (M, N) and (M, N') be gentle pairs such that $d(N, N') \leq \frac{1}{2}$. If $x \in \mathcal{T}_{\frac{1}{4}} N$, then $d(\Gamma_N(x), \Gamma_{N'}(x)) < 250d(N, N')$.*

Proof. Let $y = \pi_{N'}(x)$, $w = \pi_N(x)$, and $z = \pi_N(y) = \pi_N \pi_{N'}(x)$. The polygon $xyzwx$ is defined to consist of segments which are geodesics in M, except for zw, which is taken to be a geodesic in N. We lift the polygon $xyzwx$ to GM as follows. On the segments xy and yz, we take the parallel section $\Gamma_{N'}$; note that

$$d(\Gamma_{N'}(z), \Gamma_N(z)) \leq d(N, N') < \alpha.$$

At this point, we jump to $\Gamma_N(z)$, which we parallel translate to w along the geodesic zw in N. Arriving at w, we have a space whose distance from $\Gamma_N(w)$ is less than $\frac{3}{2}$ (our bound for the second fundamental form of N') times the length $|zw|$. Finally, we jump to $\Gamma_N(w)$ and parallel translate it to x along the normal geodesic wx to arrive at $\Gamma_N(x)$. The distance between the starting and ending values of our section is bounded above by the sum of the sizes of the jumps at z and w and $2^{-\frac{1}{2}}$ times the distance from the identity to the operator $\mathcal{H}$ of holonomy around $xyzwx$.

Let $\alpha = d(N, N')$. By Lemma 3.15, $d(z, w) < 14\alpha$; hence, $|zw| < 7 \cdot 14\alpha < 100\alpha$ by Corollary 3.5, and so the jump at w has size at most 150α. The distance $d(I, \mathcal{H})$ from the identity to $\mathcal{H}$ is bounded by the product of the curvature (≤ 1) and the area of a surface spanning the polygon. We build such a surface[7] with the minimizing geodesics from x to the points on the segments yz and zw. All these points are within a distance $d(x, N) + 2\alpha < \frac{1}{4} + 2\alpha < \frac{3}{4}$ of x, and Jacobi fields along these geodesic segments with the value zero at x are always increasing in length (no focal points), so the area of the surface is at most

$$\tfrac{1}{2}(d(x, N) + 2\alpha)(|yz| + |zw|) < (\tfrac{1}{4} + 1)(\alpha + 100\alpha) < 130\alpha.$$

Adding the estimates for the jumps and the holonomy, we conclude that

$$d(\Gamma_N(x), \Gamma_{N'}(x)) < \alpha + 150\alpha + 2^{-\frac{1}{2}} \cdot 130\alpha < 250\alpha.$$

$\square$

Remark 3.18 We can derive from the results of this section a crude "quasisymmetry" estimate for the C^1 distance: if $d(N, N') < \frac{1}{4}$, then $d(N', N) < 250d(N, N')$.

---

[7] This is one of the places where we use Condition (iii) in the definition of bounded geometry; see Remark 2.1.

4. The main theorem

In this section, we will prove Theorem 2.3, of which Theorem 2.2 is a direct consequence. For convenience, we repeat the statement of the former:

Theorem 2.3 *Let M be a riemannian manifold and $\{N_g\}$ a family of compact submanifolds of M parametrized in a measurable way by elements g of a probability space G, such that all the pairs (M, N_g) are gentle. If $d(N_g, N_h) < \epsilon < \frac{1}{20000}$ for all g and h in G, there is a well defined* **center of mass** *submanifold $\overline{N}$ with $d(N_g, \overline{N}) < 136\sqrt{\epsilon}$ for all g in G. The center of mass construction is equivariant with respect to isometries of M and measure preserving automorphisms of G.*

Proof. We begin with an outline of our proof and leave details to a series of lemmas which follow.

For each g in the parameter space G, the potential function $P_g = P_{N_g}$ is defined on the tubular neighborhood $\mathcal{T}N_g$. On the intersection $\mathcal{T}$ of these neighborhoods, the average over g of the functions P_g is a function $\overline{P}$ whose gradient is the average of the ∇P_g. The Gauss maps $\Gamma_g = \Gamma_{N_g}$ are all defined on $\mathcal{T}$ as well, and since the manifolds N_g are C^1 close to one another, the average of the Γ_g (by the construction in the Appendix) is a well-defined section $\overline{\Gamma}$ of the Grassmann bundle. The zero set $\overline{N}$ of the projection $\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ of $\nabla \overline{P}$ into the averaged vertical bundle $\overline{\Gamma}(\mathcal{T})$ is obviously an invariant subset of M.

To see that $\overline{N}$ is a smooth manifold which is C^1 close to all the N_g, we choose a base point e in G and look at the restriction of $\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ to each of the normal slices for N_e. To study these restrictions, we make $\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ into a vector field along the normal slices by applying the projection $\mathbb{P}_{Q_e}$ into the quasi-vertical bundle (tangents to the normal slices) for N_e. Since the vertical and quasi-vertical bundles are close, by Proposition 3.7, $\mathbb{P}_{Q_e}$ gives an isomorphism between the two bundles, and $\overline{N}$ is again the zero set of the vector field $\xi = \mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ along the normal slices. (We note that ξ is neither invariant nor a gradient, but this is not a problem.)

Combining several estimates from Sect. 3, we will show that all the zeros of $\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ are nonsingular with index $+1$ along the normal slices, and that $\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}$ points outward along the sphere of radius 100ϵ in each slice. As a result, there is a unique zero inside that sphere, and the collection of all these zeros forms a manifold which is transverse to the normal slices and, in fact, is C^1 close to N_e. Since e was arbitrary, we get C^1 closeness of $\overline{N}$ to all the N_g, and we are done.

Remark 4.1 The arguments in the preceding paragraph require M to be finite dimensional. To extend the result to the infinite-dimensional case will

require further analytic estimates on the projected gradient, so that some infinite-dimensional degree theory can be applied.

We will divide the detailed proof into a series of steps, using the assumption that the (M, N_g) are gentle pairs with $d(N_g, N_h) < \epsilon$ for all g and h. We will add restrictions on ϵ as necessary. Lemmas will be embedded in the proof; the end of the proof of a lemma will be indicated by the symbol $\triangle$.

Step 1. To begin, we establish a region on which all the potential functions are defined.

Lemma 4.2 *For any r such that $\epsilon < r < 1$, $\mathcal{T}^r = \cap_{g \in G} \mathcal{T}^r N_g$ contains $\mathcal{T}^{r-\epsilon} N_h$ for every h in G.*

Proof. If $x \in \mathcal{T}^{r-\epsilon} N_h$, then there is a $y \in N_h$ with $d(x, y) < r - \epsilon$. For any g, since $d(N_g, N_h) < \epsilon$, there is a z in N_g with $d(z, y) < \epsilon$. Then $d(x, z) < r$, and so $x \in \mathcal{T}^r N_g$. Thus $\mathcal{T}^{r-\epsilon} N_h$ is contained in each $\mathcal{T}^r N_g$.

$\triangle$

We will leave r undetermined for the moment and study geometry in the G-invariant subset $\mathcal{T}^r$.

Step 2. The Gauss maps $\Gamma_g = \Gamma N_g$ are all defined on $\mathcal{T}^r$. We will show that, when r and ϵ are sufficiently small, they are close enough to one another so that their values can be averaged by the method of the Appendix to produce an equivariant Gauss map $\overline{\Gamma}$. This averaged Gauss map is close to both the vertical and quasi-vertical bundles of the N_g.

Lemma 4.3 *If $r < \frac{1}{4}$ and $\epsilon < \frac{1}{1000}$, then the average $\overline{\Gamma}(x)$ of the subspaces $\Gamma_g(x) \subset T_x M$ can be constructed by Proposition A.8 for each $x \in \mathcal{T}^r$, with the Finsler distance $d(\overline{\Gamma}(x), \Gamma_g(x)) < 1000\epsilon$ for all g in G. The mapping $\overline{\Gamma} : \mathcal{T}^r \to TM$ is smooth and G-equivariant, and the construction of $\overline{\Gamma}$ from the family $\{N_g\}$ is equivariant with respect to isometries of M and measure preserving automorphisms of G.*

Proof. Since x has distance less than $\frac{1}{4}$ from each N_g, and $d(N_g, N_h) < \epsilon$, Proposition 3.17 gives the estimate $d(\Gamma_g(x), \Gamma_h(x)) < 250\epsilon < \frac{1}{4}$. Proposition A.8 then produces in each tangent space $T_x M$ an averaged subspace $\overline{\Gamma}(x)$ with $d(\overline{\Gamma}(x), \Gamma_g(x)) < \sin^{-1}(3 \cdot 250\epsilon)$ for all g. Since $\sin^{-1} x/x < \frac{4}{3}$ when $0 < x < \frac{3}{4}$, we conclude that $d(\overline{\Gamma}(x), \Gamma_g(x)) < 1000\epsilon$.

The equivariance properties of $\overline{\Gamma}$ and its construction follow from the absence of arbitrary choices in the construction. The smoothness of $\overline{\Gamma}$ follows from the corresponding assertion in Proposition A.8 (using local trivializations).

$\triangle$

Lemma 4.4 *For each x in $\mathcal{T}^r$ and g in G, the distance from $\overline{\Gamma}(x)$ to the quasi-vertical space $Q_g = \ker T_x \pi_{N_g}$ is less than $1000\epsilon + \frac{1}{4}r^2$.*

Proof. Add the estimates in Propositions 3.7 and A.8.

$\triangle$

Step 3. The averaged potential function

$$\overline{P} = \int_G P_g \, dg$$

is defined on $\mathcal{T}^r$. Since the metric is invariant,

$$\nabla \overline{P} = \int_G \nabla P_g \, dg.$$

Lemma 4.5 *If $r < \frac{7}{10}$, $\epsilon < \frac{1}{50}$, and $x \in \mathcal{T}^r$, then*

$$\|\nabla \overline{P}(x) - \nabla P_g(x)\| < \frac{45}{2}\epsilon.$$

Proof. By the usual averaging argument, it suffices to check that

$$\|\nabla P_g - \nabla P_h\| < \frac{45}{2}\epsilon$$

for all g and h in g. Observing that $-\nabla P_g(x)$ is just the initial tangent vector of the minimizing geodesic from x to N_g, we get the result immediately from Lemma 3.16.

$\triangle$

Step 4. As indicated in our outline of the proof, we define $\overline{N}$ to be the set of zeros in $\mathcal{T}^r$ of the projected gradient vector field $\mathbb{P}_{\overline{\Gamma}} \nabla \overline{P}$. It is obviously a G-invariant subset of M. To study this zero set, we fix a reference element e of G and replace $\mathbb{P}_{\overline{\Gamma}} \nabla \overline{P}$ by its projection $\mathbb{P}_{Q_e} \mathbb{P}_{\overline{\Gamma}} \nabla \overline{P}$ into the quasi-vertical bundle Q_e for N_e. Since this projected field is tangent to the normal slices for N_e, it will have at least one zero in each normal slice if we can show that it points outward along some sphere, i.e. if it has a positive inner product with ∇P_e. As long as the hypotheses of Lemma 4.3 are satisfied, Lemma 4.4 implies that the distance $\|\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{Q_e}\|$ is less than $\sin(1 + \frac{1}{16}) < 1$, so the projection from the averaged normal bundle into the normal slices is an isomorphism, and the zero set is not changed. To simply formulas, we will denote the "doubly projected gradient" vector field $\mathbb{P}_{Q_e} \mathbb{P}_{\overline{\Gamma}} \nabla \overline{P}(x)$ by the symbol $\mathcal{V}$.

Lemma 4.6 *If $\epsilon < \frac{1}{2250}$ and $99\epsilon < d(x, N_e) < \frac{1}{4}$, then the inner product $\langle \mathcal{V}, \nabla P_e(x) \rangle$ is positive.*

Proof. Let $\rho = d(x, N_e)$. Under the assumptions of the Lemma, it follows from Lemma 4.4 and Proposition A.4 that $\|\mathbb{P}_{Q_e} - \mathbb{P}_{\overline{\Gamma}}\| < \sin(\frac{1000}{2250} + \frac{1}{16}) < \frac{1}{2}$, and from Lemma 3.16 that $\|\nabla \overline{P} - \nabla P_e\| < \frac{45}{2}\epsilon < \frac{45}{198}\rho$.

Then

$$
\begin{aligned}
\langle \mathcal{V}, \nabla P_e(x) \rangle &= \langle \mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\nabla \overline{P}(x), \nabla P_e(x) \rangle = \langle \nabla \overline{P}(x), \mathbb{P}_{\overline{\Gamma}}\nabla P_e(x) \rangle \\
&= \langle \nabla P_e(x) + (\nabla \overline{P}(x) - \nabla P_e(x)), \left[\mathbb{P}_{Q_e} + (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{Q_e}) \right] \nabla P_e(x) \rangle \\
&= \langle \nabla P_e(x), \nabla P_e(x) \rangle + \langle \nabla P_e(x), (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{Q_e})\nabla P_e(x) \rangle \\
&\quad + \langle \nabla \overline{P}(x) - \nabla P_e(x)), \nabla P_e(x) \rangle \\
&\quad + \langle \nabla \overline{P}(x) - \nabla P_e(x)), (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{Q_e})\nabla P_e(x) \rangle \\
&\geq \rho^2 - \left[\rho \cdot \tfrac{1}{2}\rho + \tfrac{45}{198}\rho \cdot \rho + \tfrac{45}{198}\rho \cdot \tfrac{1}{2} \cdot \rho \right] > \tfrac{1}{10}\rho^2 > 0.
\end{aligned}
$$

$\triangle$

Following Lemma 4.6, we require $\epsilon < \frac{1}{2250}$ and set $r = 100\epsilon$. Then the set $\overline{N}$ intersects each normal slice at least once in $\mathcal{T}^r$.

Step 5. We will show that each zero of $\mathcal{V}$ is nondegenerate with degree 1 when restricted to a normal slice. It will follow that $\overline{N}$ is a smooth section of the retraction π_{N_e}, with the C^0 distance from N_e to $\overline{N}$ less than 100ϵ. The estimate used to establish nondegeneracy will also give us the desired bound on the C^1 distance $d(N, N_e)$. For our considerations in this section, we require $\epsilon < \frac{1}{20000}$.

At each point x of $\overline{N}$, the projected gradient $\mathcal{V}$, being zero at that point, has an **intrinsic derivative** $\mathcal{DV}$ which is a linear map from $T_x M$ to the quasi-vertical space Q_e at x. We will show that $\mathcal{DV}$ is surjective with positive determinant when restricted to Q_e. From this, it will follow from the implicit function theorem that $\overline{N}$ is a smooth manifold which is the graph of a smooth section of the retraction π_{N_e}. By our construction in Step 5, the C^0 distance from N_e to $\overline{N}$ less than 100ϵ. In fact, since any element of G could have been chosen as e, and the definition of $\overline{N}$ is independent of this choice, the C^0 distance from each N_g to $\overline{N}$ less than 100ϵ.

Since the tangent space to $\overline{N}$ at x is the kernel of $\mathcal{DV}$, we will be able to estimate the C^1 distance by comparing the sizes of the restrictions of $\mathcal{DV}$ to Q_e and its orthogonal complement. Although $\mathcal{DV}$ is defined only at the zeros of $\mathcal{V}$, it agrees there with an object defined throughout $\mathcal{T}^r$, namely the covariant derivative $\mathbf{D}\mathcal{V}$, considered as the operator $v \mapsto \mathbf{D}_v\mathcal{V}$. (Although the values of $\mathbf{D}\mathcal{V}$ do not generally lie in Q_e, they do lie in there at the zeros of $\mathcal{V}$, as we shall see.)

To simplify notation, we will denote the operation of averaging over G by $\mathcal{A}$. Then

$$
\begin{aligned}
\mathbf{D}\mathcal{V} &= \mathbf{D}\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\nabla \mathcal{A}P_g \\
&= (\mathbf{D}\mathbb{P}_{Q_e})\mathbb{P}_{\overline{\Gamma}}\mathcal{A}\nabla P_g + \mathbb{P}_{Q_e}(\mathbf{D}\mathbb{P}_{\overline{\Gamma}})\mathcal{A}\nabla P_g + \mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\mathbf{D}\mathcal{A}\nabla P_g \,.
\end{aligned}
$$

We analyze the last expression above term by term.

The first term, $(\mathbf{D}\mathbb{P}_{Q_e})\mathbb{P}_{\overline{\Gamma}}\nabla\mathcal{A}P_g$, is zero along the zero set $\overline{N}$ of the projected gradient $\mathbb{P}_{\overline{\Gamma}}\nabla\mathcal{A}P_g = \mathbb{P}_{\overline{\Gamma}}\nabla\overline{P}$.

For the second term, we have $\|\mathbb{P}_{Q_e}(\mathbf{D}\mathbb{P}_{\overline{\Gamma}})\mathcal{A}\nabla P_g\| \leq \|\mathbf{D}\mathbb{P}_{\overline{\Gamma}}\| \sup_g \|\nabla P_g\|$. By Proposition 3.8, this is bounded above by 11 times the distance from N_g, i.e. by $11 \cdot 100\epsilon = 1100\epsilon$.

In the last term, $\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\mathbf{D}\mathcal{A}\nabla P_g = \mathcal{A}\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\mathbf{D}\nabla P_g$, the factor $\mathbf{D}\nabla P_g$ is the hessian H_{N_g}, which we will denote by H_g. Using Proposition 3.11, we will estimate the expression $\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\mathbf{D}_v\nabla P_g$ by looking at the inner product $\langle\mathbb{P}_{Q_e}\mathbb{P}_{\overline{\Gamma}}\mathbf{D}_v\nabla P_g, w\rangle = H_g(v, \mathbb{P}_{\overline{\Gamma}}w)$, where w is an arbitrary quasi-vertical vector. We analyze separately the cases where v is quasi-vertical and quasi-horizontal. Since Proposition 3.11 uses the splitting into vertical and horizontal spaces, we have to estimate several "correction terms."

Lemma 4.7 *For any quasi-vertical vector v at any point of $\overline{N}$, $\langle\mathcal{D}\mathcal{V}(v), v\rangle \geq \frac{89}{200}\|v\|^2$.*

Proof. For v quasi-vertical,

$$
\begin{aligned}
H_g(v, \mathbb{P}_{\overline{\Gamma}}w) &= H_g(\mathbb{P}_{Q_e}v, \mathbb{P}_{\overline{\Gamma}}w) \qquad\qquad\qquad\qquad\qquad (3)\\
&= H_g(\mathbb{P}_{\Gamma_g}v, \mathbb{P}_{\Gamma_g}w) + H_g((\mathbb{P}_{Q_e} - \mathbb{P}_{\Gamma_g})v, \mathbb{P}_{\Gamma_g}w)\\
&\quad + H_g(\mathbb{P}_{\Gamma_g}v, (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{\Gamma_g})w)\\
&\quad + H_g((\mathbb{P}_{Q_e} - \mathbb{P}_{\Gamma_g})v, (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{\Gamma_g})w).
\end{aligned}
$$

We estimate first the differences between projections. By Proposition 3.7 and Proposition 3.17,

$$
\begin{aligned}
\|\mathbb{P}_{Q_e} - \mathbb{P}_{\Gamma_g}\| &\leq \|\mathbb{P}_{Q_e} - \mathbb{P}_{\Gamma_e}\| + \|\mathbb{P}_{\Gamma_e} - \mathbb{P}_{\Gamma_g}\| < \tfrac{1}{5}d(x, N_e)^2 + 250\epsilon\\
&< \tfrac{1}{5}\left(\tfrac{100}{20000}\right)^2 + \tfrac{250}{20000} < \tfrac{1}{50}\,.
\end{aligned}
$$

By Lemma 4.3, $\|\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{\Gamma_g}\| < 1000\epsilon < \frac{1000}{20000} = \frac{1}{20}$. From this we have in particular that

$$
\|\mathbb{P}_{\Gamma_g}v\| \geq (1 - \|\mathbb{P}_{\Gamma_g} - \mathbb{P}_{\Gamma_e}\|)\|v\| > \left(1 - \tfrac{250}{20000}\right)\|v\| > \tfrac{49}{50}\|v\|\,.
$$

Applying Proposition 3.11 now gives the estimate $H_g(\mathbb{P}_{\Gamma_g}v, \mathbb{P}_{\Gamma_g}v) > .64(\frac{49}{50})^2\|v\|^2 > \frac{3}{5}\|v\|^2$. Since, by 3.11, $\|H_g\| < 1.32$, the remaining three terms in the last expression in Eq. 3 are bounded above in norm by $1.32\|v\|^2$ times $\frac{1}{50}$, $\frac{1}{20}$, and $\frac{1}{50}\frac{1}{20}$ respectively, so that $H_g(v, \mathbb{P}_{\overline{\Gamma}}v) \geq \frac{1}{2}\|v\|^2$ for each quasi-vertical vector v.

We return now to Eq. 3 to estimate the derivative of $\mathcal{V}$ in the quasi-vertical direction. From that equation, we know that, for quasi-vertical v, $\langle\mathcal{D}\mathcal{V}(v), v\rangle = \langle\mathbf{D}_v\mathcal{V}\rangle$ is the sum of three terms, the first of which vanishes and second of which is bounded in absolute value by $1100\epsilon\|v\|^2$

along $\overline{N}$. The third term is the average over G of quantities which we have just shown to be bounded below by $\frac{1}{2}\|v\|^2$. Since $\epsilon < \frac{1}{20000}$, we have $\langle \mathcal{D}\mathcal{V}(v), v \rangle \geq \frac{89}{200}\|v\|^2$ for all quasi-vertical v.

$\triangle$

It follows from Lemma 4.7 that the derivative $\mathcal{D}\mathcal{V}$ is invertible on the quasi-vertical space, so that all the zeros of $\mathcal{V}$ are nondegenerate along the normal slices. Moreover the positivity of $\langle \mathcal{D}\mathcal{V}(v), v \rangle$ shows that the degree of each zero is 1, so there is exactly one zero along each such slice. Thus, $\overline{N}$ is the image of a smooth section of the normal bundle to N_e, i.e. of the tubular neighborhood retraction π_{N_e}. To estimate the derivative of this section, and hence the C^1 distance from N_e to $\overline{N}$, we need to estimate $\mathcal{D}\mathcal{V}$ when applied to quasi-horizontal vectors.

Lemma 4.8 *For any quasi-horizontal vector v at any point of the zero set $\overline{N}$ of $\mathcal{V}$, $\|\mathcal{D}\mathcal{V}(v)\| \leq 60\sqrt{\epsilon}\,\|v\|$.*

Proof. We will estimate the inner product $\langle \mathcal{D}\mathcal{V}(v), w \rangle =$ for quasi-horizontal v and quasi-vertical w. Since we already have the bound of $1100\epsilon\|v\|\,\|w\|$ for what corresponds to the first two terms of Eq. 3, what remains to be estimated is $H_g(v, \mathbb{P}_{\overline{\Gamma}}w)$. For this we use Corollary 3.13, making the usual slight adjustment between horizontal and quasi-horizontal vectors, analogous to Eq. 3.

$$
\begin{aligned}
H_g(v, \mathbb{P}_{\overline{\Gamma}}w) &= H_g(\mathbb{P}_{Q_e^\perp}v, \mathbb{P}_{\overline{\Gamma}}w) \qquad\qquad\qquad\qquad (4)\\
&= H_g(\mathbb{P}_{\Gamma_g^\perp}v, \mathbb{P}_{\Gamma_g}w) + H_g((\mathbb{P}_{Q_e^\perp} - \mathbb{P}_{\Gamma_g^\perp})v, \mathbb{P}_{\Gamma_g}w)\\
&\quad + H_g(\mathbb{P}_{\Gamma_g^\perp}v, (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{\Gamma_g})w)\\
&\quad + H_g((\mathbb{P}_{Q_e^\perp} - \mathbb{P}_{\Gamma_g^\perp})v, (\mathbb{P}_{\overline{\Gamma}} - \mathbb{P}_{\Gamma_g})w).
\end{aligned}
$$

We estimate this expression term by term, as in the proof of Lemma 4.7, using Corollary 3.13 (and the duality of Corollary A.6): $|H_g(v, \mathbb{P}_{\overline{\Gamma}}w)| \leq [3\sqrt{r} + 1.32(\frac{1}{5}r^2 + 250\epsilon + 1000\epsilon + (\frac{1}{5}r^2 + 250\epsilon)1000\epsilon)]\|v\|\,\|w\|$. Using the assumptions $r = 100\epsilon$ and $\epsilon < \frac{1}{20000}$ and some arithmetic we get $H_g(v, \mathbb{P}_{\overline{\Gamma}}w) \leq (30\sqrt{\epsilon} + 1670\epsilon)\|v\|\,\|w\|$.

As with the previous lemma, we return to Eq. 3 which expresses $\mathcal{D}\mathcal{V}(v)(v) = \mathbf{D}_v(v)$ as a sum of three terms, the first of which vanishes and second of which is bounded in norm by $1100\epsilon\|v\|$ along $\overline{N}$. The third term is the average over G of operators which we have just shown to be bounded above in norm by $(30\sqrt{\epsilon} + 1670\epsilon)\|v\|$. As a result, we find $\|\mathcal{D}\mathcal{V}(v)\| \leq (30\sqrt{\epsilon} + 2770\epsilon)\|v\| \leq 60\sqrt{\epsilon}\|v\|$.

$\triangle$

Combining the results of Lemmas 4.7 and 4.8, we conclude that the operator from $Q_e^\perp$ to Q_e whose graph is the tangent space $T_x\overline{N}$ has norm at most

$60\frac{200}{89}\sqrt{\epsilon} < 135\sqrt{\epsilon}$. By Corollary A.5, $d(T_x\overline{N}, Q_e^{\perp}) < \tan^{-1}(135\sqrt{\epsilon}) < 135\sqrt{\epsilon}$. Combining this estimate with Proposition 3.7, we find that the Finsler distance from $\overline{N}$ to the parallel translate $\Gamma_e^{\perp}$ of the tangent space to N_e at the foot of the normal from x is less than $136\sqrt{\epsilon}$. This Finsler distance is one of the two quantities entering in the definition of the C^1 distance from N_e to $\overline{N}$. The other is simply $d(x, N_e)$, which by Step 4 is bounded above by $r = 100\epsilon < \sqrt{\epsilon}$. Since e could have been any element of G, we conclude that $d(N_g, \overline{N}) < 136\sqrt{\epsilon}$ for all g in G, and Theorem 2.3 is proven.

$\square$

5. Other ways of averaging submanifolds

Before (and while) obtaining the present proof of Theorem 2.3, we tried several other ways to average submanifolds. None of them have succeeded yet, but there may be means of circumventing the difficulties which we found. In each subsection below, we will describe a proposed procedure for averaging a pair of submanifolds; we leave it to the reader to general each one to arbitrary weighted families.

5.1. Pursuit

Given submanifolds N_1 and N_2, we may move each point on N_i half-way along the geodesic segment to the nearest point on the other submanifold. The resulting two submanifolds should be closer together than the original pair, and we may then iterate the process and hope that the pair converges to a single submanifold. Variants on this method would involve going only a small fraction of the way from each submanifold to the other at each step, or letting each submanifold evolve continuously by the (time dependent) vector field pointing toward the nearest points on the other one.

We did not succeed in obtaining the a priori bound on derivatives which would insure convergence for any version of this procedure. It might be interesting to test the idea with numerical experiments.

5.2. The space of submanifolds

One could consider N_1 and N_2 as points in the space of submanifolds, and then apply Grove-Karcher averaging there. (For a pair of submanifolds, this would just mean going half-way along the geodesic from one manifold to the other.) The problem here is that there is no natural differentiable structure on a Banach manifold of unparametrized submanifolds. In fact, if one uses the natural coordinate systems given by sections of normal bundles,

the transition maps fail to be differentiable at points which do not have extra smoothness. This difficulty is related to the fact that multiplication and inversion are not differentiable maps for any Banach manifold of diffeomorphisms. Perhaps there is still a tractable variational problem for paths in the space of submanifolds for which two nearby points are connected by a unique minimizer. (See Sect. 6.4 for a related remark.)

5.3. Distribution theory

The average of N_1 and N_2 is well defined as a de Rham current on M, i.e. as a continuous linear functional on the differential forms of degree equal to the dimension of the N_i (at least if the submanifolds are orientable). This current is closed and is "approximately a manifold" in some sense. Perhaps there is a theorem which allows one to "condense" such a current to an honest manifold. We don't even know how to begin proving such a theorem.

Alternatively, since the submanifolds inherit riemannian metrics from M, one could consider their average as a measure on N and, once again, try to turn this measure into one supported on a submanifold. This idea seems even harder to implement than the one using currents, since one does not easily read the differentiable structure of the submanifolds from the measures.

6. Final remarks

6.1. Is the average gentle?

Unfortunately, our method gives no information on the extrinsic curvature of $\overline{N}$, so cannot establish that $(M, \overline{N})$ is a gentle pair. In particular, we have no way to estimate the opposite distances $d(\overline{N}, N_g)$. It would be interesting to know whether the averaging of gentle pairs (even gentle curves in euclidean 3-space) could produce a highly "wrinkled" submanifold. Perhaps one of the alternative averaging methods suggested in Sect. 5 would be preferable in this respect.

We note that, for a pair of hypersurfaces with equal weights, our averaging method just produces the hypersurface of points which are equidistant from the original two. In this special case, there is no "loss of regularity," because the average hypersurface can be found directly from the potential functions of the original two, rather than from their gradients.

6.2. Possible extensions of the theorem

Is there a more global version of Theorem 2.2, which would subsume Cartan's fixed point theorem (see the Introduction). In addition to the as-

sumption of nonpositive curvature, one would presumably need to impose conditions on the submanifold N, for instance that its normal exponential map be a diffeomorphism.

It should be possible to extend our theorem to the case where the ambient manifold M is infinite-dimensional. This extension would be essential for applications to almost-actions of non-finite compact Lie groups. (See the Introduction.) If M is a Hilbert manifold, the only place where more work needs to be done is in Steps 4 and 5 of the proof of the averaging theorem, where degree theory is used. (See Remark 4.1.) If M is a general Banach manifold, we would also need to extend our study of the Grassmann manifolds to this case. This would be complicated by the fact that orthogonal projections are no longer available.

There should also be versions of our theorem in symplectic geometry, e.g. for isotropic, lagrangian, or coisotropic submanifolds. If there are not, the obstructions to solutions should be interesting.

6.3. Morphing

Jamie Sethian pointed out that a solution to our averaging problem for a pair N_1 and N_2 of submanifolds would, on repeated application, give a family N_t of submanifolds defined for all binary fractions $t \in [1, 2]$. Presumably, this could be extended by continuity to all real $t \in [1, 2]$, thus giving a deformation, or "morphing", between the two submanifolds. (Alternatively, we could vary the weight on the parameter space $\{1, 2\}$ to move from one submanifold to the other.) Conversely, any "reversible" procedure for morphing one manifold into another would give a solution to the averaging problem.

6.4. Mass transport

If the two submanifolds N_i are parametrized by the same manifold N, the midpoint problem becomes trivial, since we can simply take the midpoints of the geodesic segments joining correspondingly parametrized points. (Of course, some estimates would still be needed to show that the resulting set of midpoints is a manifold.) Conversely, a solution of the midpoint problem gives a diffeomorphism between N_1 and N_2, obtained by projecting the average manifold onto both of them. David Aldous pointed out to me that this problem of finding a "distinguished" diffeomorphism between two submanifolds may be thought of as a special case of the Monge-Kantorovich mass-transport problem (see, for example, [12]), in which an optimal correspondence is sought between the supports of two measures on the same space. Unfortunately, the correspondences found in most known solutions of this problem are multiple-valued.

6.5. Evolution of submanifolds

The representation of submanifolds as zeros of projected gradient functions may be useful for describing the evolution of manifolds under flows, generalizing the level set method for hypersurfaces. Ruuth et al. [27] have studied such evolutions numerically. On the other hand, Ambrosio and Soner [2] study these evolutions by representing a submanifold as the zero set of its distance or potential function. Although they use a single function and not a section of a bundle, they must analyze the derivatives of the potential function and the curvature of tube boundaries just as we do.

A. Appendix: Finsler distance and averaging in Grassmann manifolds

We review in this section some geometry of Grassmann manifolds. Specifically, we will describe a center-of-mass construction derived from the averaging of projections, and we will analyze this construction in terms of a natural Finsler metric on the grassmannian. Many of the results here are probably not new, but there seems to be no convenient reference for them in the form which we need.

   We will work with finite dimensional subspaces of (possibly infinite dimensional) Hilbert spaces. By duality, we will extend our results to subspaces of finite codimension (such as the normal spaces to finite dimensional submanifolds), but this is as far as we can go. The study of grassmannians of general subspaces is much more subtle. (See, for example, Chap. 7 of [25].)

   For our purposes, a Finsler manifold will be a Banach manifold with a continuous field of norms on the tangent spaces. The norms should define the topology but need not be smooth or strictly convex. These conditions are enough to define length of curves and an inner metric (see [26], §16) compatible with the topology, though nearby points may be connected by many length-minimizing paths, such as in the case of the L^∞ ("taxicab") metric on the plane.

A.1. The Grassmannian and its metric

For a real Hilbert space E and $k = 0, 1, 2, \ldots$ we denote by $G_k(E)$ the Grassmann manifold of k-dimensional subspaces of E, and by $G(E)$ the manifold having the $G_k(E)$'s as its connected components. Each $G_k(E)$ is a homogeneous space under the natural action of the orthogonal group $O(E)$; the isotropy of an element $F \subset E$ is the direct product $O(F) \times O(F^\perp)$. Each tangent space $T_F G(E)$ is naturally isomorphic to the space $\mathrm{Hom}(F, F^\perp)$ of continuous linear maps from F to $F^\perp$.

   Endowing each tangent space $\mathrm{Hom}(F, F^\perp)$ with the operator norm gives an $O(E)$-invariant Finsler metric on $G(E)$. We will use this metric rather

than the riemannian metric (also $O(E)$-invariant) coming from the inner products $\langle Q_1, Q_2 \rangle = \mathrm{Tr}(Q_1 Q_2^*)$ on the tangent spaces. The Finsler distance $d(F, F')$ between two subspaces is then defined as the infimum of the lengths of paths joining F to F'.

We embed $G(E)$ into the space $S(E)$ of self-adjoint endomorphisms of E by mapping each F to the operator $P(F) = \mathbb{P}_F$ of orthogonal projection onto F. The image of $G_k(E)$ under this embedding consists of the self-adjoint projections with trace k. The tangent space to $P(G(E))$ at $P(F)$ consists of those q in $S(E)$ such that $\mathbb{P}_F q = q(I - \mathbb{P}_F)$. The image of $Q \in T_F G(E)$ under the derivative $T_F P$ has the block diagonal form

$$q = \begin{bmatrix} 0 & Q^* \\ Q & 0 \end{bmatrix}$$

with respect to the decomposition $E = F \oplus F^\perp$.

The Hilbert norm $\sqrt{\mathrm{Tr}(q^* q)}$ of q is $\sqrt{2}$ times that of Q, but the operator norms of q and Q are exactly the same. Thus P is an isometric embedding of the Finsler manifold $G(E)$ into the Banach space $S(E)$.[8]

We will also make use of the antisymmetric operator

$$h(Q) = \begin{bmatrix} 0 & -Q^* \\ Q & 0 \end{bmatrix}$$

in the Lie algebra $\mathfrak{o}(E)$ of $O(E)$. We call it the **horizontal lift** of the tangent vector Q, because the horizontal lifts of all $Q \in T_F(G(E))$ form a complement to the isotropy subalgebra $\mathfrak{o}(F) \times \mathfrak{o}(F^\perp)$ of F. By right translating the horizontal space around $O(E)$, we obtain a connection in the principal bundle $O(E) \to G_k(E)$. Thus, each path $\{F_t\}$ in $G(E)$ has a horizontal lift $\{g_t\}$ in $O(E)$ such that $F_t = g_t(F_0)$, and dF_t/dt and dg_t/dt have the same norm.

It will be useful to compare the Finsler distance with some other notions of distance between subspaces. We have the following results.

Proposition A.1 *If F and F' are subspaces of the same (finite) dimension in E, then the Finsler distance $d(F, F')$ is equal to the C^0 distance $d_0(F, F')$ between the unit spheres of F and F', considered as submanifolds of the unit sphere in E (with the standard riemannian metric).*

Corollary A.2 *The Finsler distance between any two elements of $G_k(E)$ is at most $\pi/2$.*

Corollary A.3 *The C^0 distance $d_0(F, F')$ is symmetric in F and F'.*

---

[8] The map P keeps the length of curves fixed, but it generally decreases distances; see Proposition A.4 below.

Proposition A.4 *If F and F' are subspaces of the same (finite) dimension in E, then $\|\mathbb{P}_F - \mathbb{P}_{F'}\| = \sin d(F, F')$.*

Proof of Proposition A.1. To begin, we note that, by an arbitrarily small perturbation of F, we can insure that

$$F \cap F'^{\perp} = \{0\}. \qquad (*)$$

It will suffice, then, to prove our equality for pairs having property $(*)$. (When it is not satisfied, $d_0(F, F') = \pi/2$.)

We will show first that $d(F, F') \leq d_0(F, F')$. By our assumption $(*)$, with respect to the decomposition $E = F \oplus F^{\perp}$, F' is the graph of a linear map $\phi : F \to F^{\perp}$. The non-negative symmetric operator $\phi^*\phi$ on F has k eigenvalues which we take to be the squares of

$$\lambda_1 \geq \ldots \geq \lambda_k \geq 0.$$

We denote an orthonormal basis of eigenvectors with these eigenvalues by $e_1, \ldots, e_k$, and we call the angles $\theta_j = \tan^{-1} \lambda_j$ the **canonical angles**[9] between the subspaces F and F'.

Since

$$\langle \phi e_j, \phi e_k \rangle = \langle \phi^* \phi e_j, e_k \rangle = \lambda_j^2 \delta_{jk},$$

we get an orthonormal basis for F' by taking the vectors

$$\frac{e_j + \phi e_j}{\sqrt{1 + \lambda_j^2}} = \frac{e_j + \phi e_j}{\sqrt{1 + \tan^2 \theta_j}} = (\cos \theta_j)(e_j + \phi e_j).$$

We get an orthonormal set in $F^{\perp}$ by taking $f_j = (1/\lambda_j)\phi e_j$ whenever $\lambda_j \neq 0$. For the remaining values of j, we set $f_j = 0$.

Our basis elements for F' can now be written as $\cos \theta_j e_j + \sin \theta_j f_j$, and we can now connect F and F' by the path

$$F_t = \cos(t\theta_j)e_j + \sin(t\theta_j)f_j, \quad 0 \leq t \leq 1.$$

The velocity of this path at $t = 0$ is the map $e_j \mapsto \theta_j f_j$, which has norm θ_1, the largest canonical angle. Since the F_t form a trajectory starting at F_1 for a 1-parameter group of isometries (generated by the skew-symmetric transformation which takes e_j to $\theta_j f_j$ and is zero on $(F \oplus \phi F)^{\perp}$), the speed of the path is constant, so its length is θ_1. Hence $d(F, F') \leq \theta_1$. But

---

[9]   These angles have a long history, going back to Jordan [19]. We refer to [22] for an extensive discussion of this history, and add as well the references [1] and [23]. Finally, we note that Porta and Recht [24] have quite explicitly studied some aspects of the Finsler geometry of grassmannians. They represent subspaces by the involutions $2\mathbb{P}_F - I$, so their Finsler metric is twice ours.

$\theta_1 \leq d_0(F, F')$ because θ_1 is the distance from $\cos\theta_1 e_1 + \sin\theta_1 f_1 \in F'$ to the unit sphere in F, so we have proven that $d(F, F') \leq d_0(F, F')$.

For the reverse inequality, we will show that $d_0(F, F')$ is bounded above by the length of any path from F to F'. Let $\{F_t\}$ be such a path, $\{g_t\}$ its horizontal lift to the orthogonal group. Any unit vector in F' is $g_1 v$ for some unit vector v in F; the distance from $g_1 v$ to the unit sphere in F is estimated by using the path $\{g_t v\}$:

$$d(v, g_1 v) \leq \int_0^1 \left\| \frac{d}{dt}(g_t v) \right\| \, dt \leq \int_0^1 \left\| \frac{d}{dt} g_t \right\| \, dt = \int_0^1 \left\| \frac{d}{dt} F_t \, dt \right\|,$$

which is the length of the path $\{F_t\}$.

$\square$

Proof of Proposition A.4. Let $\theta_1 \geq \ldots \geq \theta_k$ be the canonical angles between F and F'. According to the preceding proposition, $d(F, F') = \theta_1$. In an orthonormal basis beginning with the orthonormal set $e_1, f_1, e_2, f_2, \ldots$ constructed in the proof above, the projection $\mathbb{P}_{F'}$ is block diagonal with 2×2 blocks of the form

$$\begin{bmatrix} \cos^2\theta_j & \cos\theta_j \sin\theta_j \\ \cos\theta_j \sin\theta_j & \sin^2\theta_j \end{bmatrix}$$

followed by zero blocks, while $\mathbb{P}_F$ has the same form with the θ_j's replaced by 0. The difference $\mathbb{P}_F - \mathbb{P}_{F'}$ is built of diagonal blocks

$$h(Q) = \begin{bmatrix} 1 - \cos^2\theta_j & -\cos\theta_j \sin\theta_j \\ -\cos\theta_j \sin\theta_j & -\sin^2\theta_j \end{bmatrix} = \sin\theta_j \begin{bmatrix} \sin\theta_j & -\cos\theta_j \\ -\cos\theta_j & -\sin\theta_j \end{bmatrix},$$

each of which has norm $\sin\theta_j$, so

$$\|\mathbb{P}_F - \mathbb{P}_{F'}\| = \sin\theta_1 = \sin d(F, F').$$

$\square$

The identification of $d(F, F')$ with the largest canonical angle also leads immediately to the following result.

Corollary A.5 *If F' is the graph of a linear map ϕ from F to $F^\perp$, then $d(F, F') = \tan^{-1}\|\phi\|$.*

It is easy to go back and forth between subspaces and their orthogonal complements.

Corollary A.6 *If F and F' are elements of $G_k(E)$, and E is finite dimensional, then $d(F, F') = d(F^\perp, F'^\perp)$.*

Proof. Note that the projections on a subspace and its orthogonal complement sum to the identity operator. Now apply Proposition A.4. $\square$

Remark A.7 Thanks to Corollary A.6 (or simply using Proposition A.4 as a definition of the distance), we can define the distance between subspaces of finite *codimension* in a Hilbert space to be the distance between their orthogonal complements.

We also note that our distance functions are unaffected by rescaling of the inner product on E.

A.2. Center of mass in the Grassmannian

We turn now to the averaging of subspaces. Using Grove-Karcher averaging in the grassmannian makes it more difficult to obtain estimates of the kind we need, since it is hard to control the behavior of the Finsler distance along geodesics. Instead, we average the orthogonal projections associated to the subspaces. This approach was used already by de la Harpe and Karoubi in [9], but their work is more complicated and their estimates less sharp than ours because their projections are not necessarily orthogonal.

Suppose, then, that we are given a family F_g of elements of $G_k(E)$ parametrized measurably by elements g of a probability space G. If we average the projections $\mathbb{P}_{F_g}$, we obtain a self-adjoint operator which is not, in general, a projection. However, if the F_g are sufficiently close together, this operator is an "almost projection," and we will see that it can be approximated in a canonical way by a self-adjoint projection having rank k. Here is the precise result.

Proposition A.8 *Let F_g be a family of elements of $G_k(E)$ parametrized measurably by elements g of a probability space G such that $d(F_g, F_h) < \epsilon$ for all g and h in G. As long as $\epsilon < \frac{1}{2}$, we can define an "average" element $\overline{F} \in G_k(E)$ such that $d(F_g, \overline{F}) < \sin^{-1} 2\epsilon$ for all g in G. This averaging construction is equivariant with respect to isometries of E and measure preserving automorphisms of G. If the F_g also depend smoothly on other parameters μ, the average $\overline{F}$ also depends smoothly on these parameters. Furthermore, if $\epsilon < \frac{1}{4}$, the magnitude of the derivative $\frac{\partial \overline{F}}{\partial \mu}$ is at most 8 times the supremum over g of the magnitudes of the derivatives $\frac{\partial F_g}{\partial \mu}$.*

Proof. We note first of all that, by Proposition A.4, $\|\mathbb{P}_{F_g} - \mathbb{P}_{F_h}\| < \epsilon$. For the average

$$\overline{P} = \int_G \mathbb{P}_{F_g} \, dg$$

in the space of self-adjoint operators on E, a simple computation shows that $\|\overline{P} - \mathbb{P}_{F_g}\| < \epsilon$ for all g in G.

We will now modify the approximate projection $\overline{P}$ to make it into an orthogonal projection. The estimate $\|\overline{P} - \mathbb{P}_{F_g}\| < \epsilon$ tells us that the spectrum of $\overline{P}$ is contained in intervals of radius ϵ about 0 and 1. Since $\epsilon < \frac{1}{2}$, we can apply to $\overline{P}$ the holomorphic function h, defined on the complex λ plane with the line Re $\lambda = \frac{1}{2}$ removed, having the value 0 to the left of the line and 1 to the right. The operator $h(\overline{P})$ is a spectral projection of $\overline{P}$ which can be written as an integral around a circle γ of radius $\frac{1}{2} - \delta$ around $\lambda = 1$, where δ can be an arbitrarily small positive number.

$$h(\overline{P}) = \frac{1}{2\pi i} \int_\gamma (\lambda I - \overline{P})^{-1} \, d\lambda.$$

The application of h to $\overline{P}$ amounts to moving the spectrum to 0 and 1 while keeping the invariant subspaces the same, so that $\|h(\overline{P}) - \overline{P}\| < \epsilon$.

Denoting by $\overline{F}$ the image of $h(\overline{P})$, we have $h(\overline{P}) = \mathbb{P}_{\overline{F}}$, and $\|\mathbb{P}_{\overline{F}} - \mathbb{P}_{F_g}\| \leq \|h(\overline{P}) - \overline{P}\| + \|\overline{P} - \mathbb{P}_{F_g}\| < \epsilon + \epsilon = 2\epsilon < 1$ for all g in G. Since $\|\mathbb{P}_{\overline{F}} - \mathbb{P}_{F_g}\| < 1$, the spaces $\overline{F}$ and F_g have the same dimension (see [20], Theorem 6.32), so $\overline{F} \in G_k(E)$. By Proposition A.4, $d(\overline{F}, F_g) < \sin^{-1} 2\epsilon$.

The equivariance of our construction is obvious, since all the steps were canonical. To establish smooth dependence and estimate derivatives with respect to parameters, we continue our analysis of the operator function h by observing first that differentiation under the integral sign gives smoothness (in fact analyticity) of h. To estimate the derivative, we assume that the spectrum of an operator family $P(\mu)$ lies in the union of discs of radius ϵ around 0 and 1 and compute:

$$
\begin{aligned}
\frac{\partial h(P(\mu))}{\partial \mu} &= \frac{1}{2\pi i} \int_\gamma \frac{\partial}{\partial \mu}(\lambda I - \overline{P})^{-1} \, d\lambda \\
&= -\frac{1}{2\pi i} \int_\gamma (\lambda I - P(\mu))^{-1} \frac{\partial P(\mu)}{\partial \mu} (\lambda I - P(\mu))^{-1} \, d\lambda.
\end{aligned}
$$

Hence,

$$\left\| \frac{\partial h(P(\mu))}{\partial \mu} \right\| \leq \frac{1}{2\pi} (\text{length } \gamma) \sup_\lambda \left\| (\lambda \mathrm{I} - P(\mu))^{-1} \right\|^2 \left\| \frac{\partial P(\mu)}{\partial \mu} \right\|.$$

As the radius of γ approaches $\frac{1}{2}$, the length of γ approaches π, and so

$$\left\| \frac{\partial h(P(\mu))}{\partial \mu} \right\| \leq \frac{1}{2\pi} \pi \frac{1}{(\frac{1}{2} - \epsilon)^2} \left\| \frac{\partial P(\mu)}{\partial \mu} \right\| \leq \frac{2}{(1 - 2\epsilon)^2} \left\| \frac{\partial P(\mu)}{\partial \mu} \right\|.$$

As long as $\epsilon < \frac{1}{4}$, we have

$$\left\| \frac{\partial h(P(\mu))}{\partial \mu} \right\| \leq 8 \left\| \frac{\partial P(\mu)}{\partial \mu} \right\|,$$

and we shall be content with this crude estimate (though the factor 8 can probably be brought close to unity as ϵ becomes small).

We can now finish the proof of our Proposition. Let the spaces F_g depend smoothly on μ. Since the map $F \mapsto \mathbb{P}_F$ is a Finsler isometry, $\|\frac{\partial F_g(\mu)}{\partial \mu}\| = \|\frac{\partial \mathbb{P}_{F_g}}{\partial \mu}\|$. Averaging does not increase norms, so $\|\frac{\partial \overline{P}(\mu)}{\partial \mu}\| \leq \sup_g \|\frac{\partial F_g}{\partial \mu}\|$. By our previous estimate, $\|\frac{\partial \overline{F}}{\partial \mu}\| = \|\frac{\partial \mathbb{P}_{\overline{F}}}{\partial \mu}\| \leq 8 \sup_g \|\frac{\partial F_g}{\partial \mu}\|$.

$\square$

References

1. Afriat, S.N.: Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. Proc. Camb. Philos. Soc. **53**, 800–816 (1957)
2. Ambrosio, L., Soner, H.M.: Level set approach to mean curvature flow in arbitrary codimension. J. Diff. Geom. **43**, 693–737 (1996)
3. Anderson, R.M.: "Almost" implies "near". Trans. Amer. Math. Soc. **296**, 229–237 (1986)
4. Bishop, R.L., Crittenden, R.J.: Geometry of Manifolds. New York: Academic Press 1964
5. Buser, P., Karcher, H.: Gromov's almost flat manifolds, Astérisque **81**, Société Mathématique de France, Paris 1981
6. Cannas da Silva, A., Weinstein, A.: Lectures on Geometric Models for Noncommutative Algebras. Berkeley Math. Lecture Notes, Amer. Math. Soc., Providence, 1999; electronic version at `http://math.berkeley.edu/~alanw`
7. Cartan, É.: Groupes simples clos et ouverts et géométrie riemannienne. J. Math. Pures Appl. **8**, 1–33 (1929)
8. Cartan, É.: Leçons sur la Géométrie des Espaces de Riemann. Paris: Gauthier-Villars 1928
9. de la Harpe, P., Karoubi, M.: Représentations approchées d'un groupe dans une algèbre de Banach. Manuscripta Math. **22**, 293–310 (1977)
10. Eschenburg, J.: Comparison theorems and hypersurfaces. Manuscripta Math. **59**, 295–323 (1987)
11. Eschenburg, J.-H., Heintze, E.: Comparison theory for Riccati Equations. Manuscripta Math. **68**, 209–214 (1990)
12. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. **177**, 113–161 (1996)
13. Gray, A.: Tubes. Redwood City: Addison–Wesley 1990
14. Grove, K., Karcher, H.: How to conjugate C^1-close group actions. Math. Z. **132**, 11–20 (1973)
15. Grove, K., Karcher, H., Ruh, E.A.: Group actions and curvature. Invent. math. **23**, 31–48 (1974)
16. Helgason, S.: Differential geometry and symmetric spaces. New York: Academic Press 1962
17. Hyers, D.H.: On the stability of the linear functional equation. Proc. Nat. Acad. Sci. U.S.A. **27**, 222–224 (1941)

18. Jost, J., Karcher, H.: Geometrische Methoden zur Gewinnung von a-priori-Schranken für harmonische Abbildungen. Manuscripta Math. **40**, 27–77 (1982)
19. Jordan, C.: Essai sur la géométrie à n dimensions. Bull. Soc. Math. France **3**, 103–174 (1875)
20. Kato, T.: Perturbation Theory for Linear Operators, 2d edn. Berlin, New York: Springer 1976
21. Lang, S.: Fundamentals of Differential Geometry. New York: Springer 1999
22. Paige, C.C., Wei, M.: History and generality of the CS decomposition. Lin. Alg. Appl. **208/209**, 303–326 (1994)
23. Petkanschin, B.: Integralgeometrie 6. Zusammenhänge zwishen den Dichten der lineare Unterräume in n-dimensionalen Raum. Abh. Math. Sem. Hamburg. Univ. **11**, 249–310 (1936)
24. Porta, H., Recht, L.: Minimality of geodesics in Grassmann manifolds. Proc. Amer. Math. Soc. **100**, 464–466 (1987)
25. Pressley, A., Segal, G.: Loop Groups. New York: Oxford University Press 1986
26. Rinow, W.: Die Innere Geometrie der Metrischen Räume. Berlin: Springer 1961
27. Ruuth, S.J., Merriman, B., Xin, J., Osher, S.: Diffusion-generated motion by mean curvature for filaments. Preprint, UCLA, 1998
28. Ulam, S.: Sets, Numbers, and Universes. Cambridge: MIT Press 1974
29. Warner, F.W.: Extensions of the Rauch comparison theorem to submanifolds. Trans. Amer. Math. Soc. **122**, 341–356 (1966)
30. Weinstein, A.: From Riemann geometry to Poisson geometry and back again. Lecture at Chern Symposium, MSRI, 1998, available at `http://msri.org/publications/video/contents.html`
31. Whitney, H.: Differentiable manifolds. Ann. Math. **37**, 645–680 (1936)