A BAYESIAN APPROACH TO THE ESTIMATION OF MAPS BETWEEN RIEMANNIAN MANIFOLDS, II: EXAMPLES

LEO T. BUTLER AND BORIS LEVIT

Abstract. Let Θ be a smooth compact oriented manifold without boundary, imbedded in a euclidean space $\mathbf{E}^s$, and let γ be a smooth map of Θ into a Riemannian manifold Λ. An unknown state $\theta \in \Theta$ is observed via $X = \theta + \epsilon\xi$ where $\epsilon > 0$ is a small parameter and ξ is a white Gaussian noise. For a given smooth prior λ on Θ and smooth estimators $g(X)$ of the map γ we have derived a second-order asymptotic expansion for the related Bayesian risk [3]. In this paper, we apply this technique to a variety of examples.

The second part examines the first-order conditions for equality-constrained regression problems. The geometric tools that are utilised in [3] are naturally applicable to these regression problems.

1. Introduction

In many estimation problems, one has a state which lies on a manifold but one observes this state plus some error in a euclidean space. It is desirable to utilise the underlying geometry to construct an estimator of the state. The present paper uses a Bayesian approach and the Bayesian estimator derived in [3] and computes the estimator in a variety of examples.

In many cases, the geometric framework of [3] naturally extends to regression problems. In an estimation problem, the map is known while the state is observed with noise and one attempts to infer the 'true' state; in a regression problem, the map is unknown and one observes the input-output states with some noise and attempts to infer the map. In this paper, we will assume that the regression map belongs to a given compact finite-dimensional manifold. In such a situation, one may formally transpose the regression problem in the sense that one may regard the map as the state that one observes with noise and the input-output states may be regarded as (evaluation) maps. This transposition is commonly used in topology and differential geometry. In the second part of this note, we derive first-order conditions for regression problems on manifolds. It is shown in several cases that this duality between estimation and regression is exact: the two viewpoints lead to the same estimator.

Consider the following situation: $\mathbf{E}$ is a real s-dimensional vector space with inner product σ and Θ (resp. Λ) is a smooth manifold with riemannian metric $\mathbf{g}$ (resp. $\mathbf{h}$). Assume that the smooth riemannian manifold $(\Theta, \mathbf{g})$ is isometrically embedded in a euclidean space $(\mathbf{E}, \sigma)$ via the inclusion map ι, and $\Theta \xrightarrow{\gamma} \Lambda$ is a smooth map. Smooth means infinitely differentiable. These data are summarized

Date: August 18, 2009.

2000 *Mathematics Subject Classification.* Primary 62C10; Secondary 62C20, 62F12, 53B20, 53C17, 70G45.

Key words and phrases. Bayesian problems, Bayes estimators, Minimax estimators, riemannian geometry, sub-riemannian geometry, sub-laplacian, harmonic maps.

The first author thanks the Carnegie Trust for the Universities of Scotland for supporting this research.

by the diagram

$$\begin{tikzcd}
\mathrm{N}(\Theta) \arrow[r, "\text{incl.}"] \arrow[dr, bend right, "\pi"'] & (\mathbf{E},\sigma) \arrow[dr, dotted, "g"] & \\
 & (\Theta,\mathbf{g}) \arrow[ul, "\text{incl.}"'] \arrow[u, "\iota"] \arrow[r, "\gamma"] & (\Lambda,\mathbf{h}),
\end{tikzcd}$$

where $\mathrm{N}(\Theta)$ is an open neighbourhood of Θ in $\mathbf{E}$ and π is the orthogonal projection onto Θ. A basic result of differential geometry is that if Θ is compact, then there is an $r > 0$ such that π is a smooth map on the set of all vectors within a distance r of Θ [13].

Suppose that $X \in \mathbf{E}$ is a gaussian random variable with mean $\theta \in \Theta$ and covariance operator[1] $\epsilon^2\mathfrak{c}$, *i.e.*

$$X \sim \mathcal{N}(\theta, \epsilon^2\mathfrak{c}), \qquad \theta \in \Theta.$$

A basic statistical problem is to determine an estimator "$\gamma(X)$," by which we mean an optimal extension of γ off Θ, in the minimax sense. To make this precise, let $g : \mathbf{E} \to \Lambda$ be an estimator (map), and let **dist** be the riemannian distance function of $(\Lambda, \mathbf{h})$. Define a loss function by

$$R_\epsilon(g,\theta) = \int_{x\in\mathbf{E}} \mathbf{dist}(g(x),\gamma(\theta))^2\, \psi_\epsilon(x-\iota(\theta))\,\mathrm{d}x,$$

where $\psi_\epsilon(u) = \exp(-|u|^2/2\epsilon^2)/(2\pi\epsilon^2)^{\frac{s}{2}}$, $|\bullet|$ is the norm on $\mathbf{E}$ induced by σ, and $\mathrm{d}x$ is the volume form on $\mathbf{E}$ induced by σ.[2] Define the associated minimax risk

$$r_\epsilon(\Theta) = \inf_g \sup_{\theta\in\Theta} R_\epsilon(g,\theta).$$

1.1. **Results: Bayesian estimation.** One may use a Bayesian approach to determine the asymptotically minimax estimator g. Here one views θ is viewed as a random variable with a prior distribution $\lambda(\theta)\mathrm{d}\theta$ where $\int_{\theta\in\Theta}\lambda(\theta)\,\mathrm{d}\theta = 1$ ($\mathrm{d}\theta = \mathrm{d}\nu_{\mathbf{g}}$ is the riemannian volume of $(\Theta,\mathbf{g})$). The *Bayesian risk* of a map g is

$$R_\epsilon(g;\lambda) = \int_{\theta\in\Theta}\int_{x\in\mathbf{E}} \mathbf{dist}(g(x),\gamma(\theta))^2\,\lambda(\theta)\,\psi_\epsilon(x-\iota(\theta))\;\mathrm{d}x\,\mathrm{d}\theta.$$

A *Bayes estimator* $g : \mathbf{E} \to \Lambda$ is a map which minimizes the Bayesian risk over all maps.

Before stating the main result of [3], recall that a riemannian connection permits one to define higher-order derivatives. In particular, $\nabla\mathrm{d}$ is used to denote the hessian (second derivative) and $\tau = \mathrm{Tr}\,(\nabla\mathrm{d})$ denotes the tension field (laplacian), while Ric denotes the Ricci curvature [6, 3].

In [3], the present authors proved

Theorem 1.1. *Let $\tilde{g}_\epsilon(x) = \exp_{g_o(x)}\left(\epsilon^2 g_2(x) + O(\epsilon^4)\right)$ be the Bayesian estimator for the Bayesian risk functional R_ϵ (Equation 1.1) with a fixed Bayesian prior $\lambda > 0$, where g_o, g_2 are the lowest order terms in the expansion. Then for all ϵ sufficiently small*

---

[1] By convention, the covariance operator is the induced inner product on the dual vector space $\mathbf{E}^*$. If we regard σ as a linear isomorphism of $\mathbf{E} \to \mathbf{E}^*$, then the covariance operator is the inverse linear isomorphism $\mathfrak{c} = \sigma^{-1} : \mathbf{E}^* \to \mathbf{E}$. It is common to think of $\mathbf{E}$ as a space of column vectors, and the dual as a space of row vectors, in which case $\mathfrak{c}$ is the transpose map $x \mapsto x'$ from row to column vectors.

[2] One can introduce a σ-orthonormal coordinate system x_i on $\mathbf{E}$. In this case, $|x|^2 = \sum_i x_i^2$ and $\mathrm{d}x = \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_s$.

(1) *for all* $x \in \mathrm{N}(\Theta)$ *with* $|x - \pi(x)| \le r$

$$\tilde{g}_\epsilon(x) = \exp_{\Gamma(x)}\left(\epsilon^2 \left[\frac{1}{2}\tau(\gamma) + \mathrm{d}\gamma(\nabla \log \lambda)\right]_{\pi(x)} + O(\epsilon^4)\right),$$

where $\Gamma = \gamma\pi$, *and* exp *is the exponential map of* $(\Lambda, \mathbf{h})$*; and*

(2)

$$R_\epsilon(\tilde{g}_\epsilon; \lambda) = \epsilon^2 \int \mathrm{d}\theta\, \lambda\, |\mathrm{d}\gamma|^2 + \epsilon^4 \int \mathrm{d}\theta\, \lambda \left\{\frac{1}{2}|\nabla \mathrm{d}\Gamma|^2 - |\tau(\gamma) + \mathrm{d}\gamma(\nabla \log \lambda)|^2 - \frac{2}{3}\langle \mathrm{d}\Gamma, \mathrm{Ric}_{\mathrm{d}\Gamma}\rangle\right\} + O(\epsilon^6).$$

In section 2, this notes applies Theorem 1.1 to compute the bayesian estimator and risk of the identity map for a wide class of compact group orbits and a 'linear' prior (see Theorem 2.3).

1.2. **Results: Bayesian regression.** Let Θ, Λ be smooth manifolds. Let $\theta_1, \ldots, \theta_k$ be a collection of design points on a manifold Θ and let $y_1, \ldots, y_k$ be a random sample of points on Λ. Assume that the conditional probability density of y_l given θ_l is $f(y_l|\gamma(\theta_l))$, where $\gamma : \Theta \to \Lambda$ is an unknown map and θ is a point on Θ. One is interested in estimating the unknown map γ by minimising a discrepancy function

$$\mathsf{s}(\gamma) = \frac{1}{k}\sum_{l=1}^{k} \ell(y_l, \gamma(\theta_l))$$

for a given loss function $\ell : \Lambda \times \Lambda \to [0, \infty)$ (see section 3.2). If one assumes that the space of admissible maps γ is parameterized by a finite-dimensional manifold Γ, a solution to this regression problem is

$$\hat{\gamma} = \mathrm{argmin}\, \{\mathsf{s}(\gamma) \;:\; \gamma \in \Gamma\}.$$

One may also consider the regression problem from a bayesian perspective. In this case, one assumes there is a smooth volume form $\mathrm{d}\gamma$ on Γ and a prior distribution $\lambda(\gamma)\,\mathrm{d}\gamma$. The bayesian regression problem is to derive the regressor by minimising the risk functional

$$\mathcal{R}(\hat{\gamma}) = \int_{\mathbf{y} \in \Lambda^k} \int_{\gamma \in \Gamma} \ell(\hat{\gamma}, \gamma)\, f(\mathbf{y}|\gamma)\, \lambda(\gamma)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\gamma.$$

over regressors $\hat{\gamma} : \Lambda^k \to \Gamma$.

In section 3, these two regression problems are examined. First-order conditions that determine the regressors are proven. In addition, we examine the special cases where $\Theta, \Lambda \subset \mathbf{E}$, Γ is a manifold of linear maps and

(1) ℓ is determined by the ambient euclidean structure;
(2) ℓ is determined by the riemannian distance function on Λ induced by the euclidean structure.

The special case where Θ and Λ are both the 2-dimensional unit sphere $S^2 \subset \mathbf{E}^3$ and Γ is the group of orientation-preserving linear isometries of $\mathbf{E}^3$, SO(3) is examined in detail in each case.

The results of section 3 are formulated in Propositions 3.1, 3.7 and 3.11.

2. Estimation of states on group orbits

Let $(\mathbf{E}^s, \sigma)$ be an s-dimensional euclidean space: that is, $\mathbf{E}^s$ is an s-dimensional real vector space and σ is a symmetric, positive-definite quadratic form on $\mathbf{E}^s$. The group of linear isometries of $\mathbf{E}^s$ is denoted by $\mathrm{O}(\mathbf{E}^s, \sigma)$ and called the *orthogonal* *group* of $(\mathbf{E}^s, \sigma)$. This group is denoted by $\mathrm{O}(\mathbf{E})$ when the euclidean structure σ and dimension s are understood. By choice of an orthonormal basis, $(\mathbf{E}^s, \sigma)$ is linearly isometric to $\mathbf{R}^s$ with its standard orthonormal basis; the orthogonal group of this latter model euclidean space is denoted by O_s, while SO_s is the subgroup of O_s with unit determinant.

The set of linear transformations $\mathbf{E}^s \to \mathbf{E}^r$ is denoted by $\mathrm{Hom}(\mathbf{E}^s, \mathbf{E}^r)$. It is naturally a euclidean space with the trace inner product $(x, y) \mapsto \mathrm{Tr}\,(x'y)$. There is an orthogonal decomposition of $\mathrm{Hom}(\mathbf{E}^s, \mathbf{E}^s)$ into the sets of skew-symmetric transformations (denoted so_s) and symmetric transformations (denoted $\mathrm{sym}(\mathbf{E}^s)$).

Let $G \subset \mathrm{O}(\mathbf{E}^s, \sigma)$ be a compact group of linear isometries of $(\mathbf{E}^s, \sigma)$. A tangent vector $\xi \in T_1 G$ in the tangent space to the identity of G can be identified with a matrix. The matrix exponential map restricts naturally to give a map $\exp : T_1 G \to G$. For each $g \in G$, the curve $t \to \exp(t\xi) \cdot g$ is a curve in G passing through g at $t = 0$. Its derivative $\xi \cdot g$ is therefore a tangent vector in $T_g G$. Thus, each tangent space is canonically isomorphic to $T_1 G$ via right translations.[3] One typically writes $T_1 G = \mathfrak{g}$, and calls $\mathfrak{g}$ the Lie algebra of the Lie group G. As a set of matrices, $\mathfrak{g}$ is equipped with the Lie bracket denoted by $[\xi, \eta] = \xi \cdot \eta - \eta \cdot \xi$. One can easily verify that $\xi, \eta \in \mathfrak{g}$ implies that $[\xi, \eta] \in \mathfrak{g}$. In addition, for each $g \in G$, $\xi \in \mathfrak{g}$, the element $g \cdot \xi \cdot g^{-1} \in \mathfrak{g}$. It is conventional to write $\mathrm{Ad}_g \xi = g \cdot \xi \cdot g^{-1}$ and observe that $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ is a representation, called the adjoint representation. One knows that $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} \mathrm{Ad}_{\exp(t\xi)} \eta = [\xi, \eta]$, so the derivative $\mathrm{d}_1 \mathrm{Ad} =: \mathrm{ad} : \mathfrak{g} \to \mathrm{gl}(\mathfrak{g})$ is a linear representation of $\mathfrak{g}$.

The trace form $(\xi, \eta) \mapsto \mathrm{Tr}\,(\xi'\eta)$ is positive definite on $\mathfrak{g}$. Moreover, the trace form is invariant under the adjoint representation of G, *i.e.* G acts as a group of isometries of this euclidean structure on $\mathfrak{g}$. For a subspace $V \subset \mathfrak{g}$, let $V^\perp$ denote its orthogonal complement with respect to the trace form.

For each $\vartheta \in \mathbf{E}$, let the set $G\vartheta = \{\phi \in \mathbf{E} \ : \ \exists g \in G \text{ and } \phi = g \cdot \vartheta\}$ be the G-orbit of ϑ and let $G_\vartheta = \{g \in G \ : \ g \cdot \vartheta = \vartheta\}$ be the G-stabilizer of ϑ. It is a well-known theorem that $G{\cdot}\vartheta$ is a smooth submanifold of $\mathbf{E}$. The tangent space to $G{\cdot}\vartheta$ at ϕ can be identified with $\mathfrak{g}_\phi^\perp$, where $\mathfrak{g}_\phi \subset \mathfrak{g}$ is the Lie algebra of G_ϕ. Indeed, since G acts transitively, the map $\mathfrak{g} \to T_\phi(G{\cdot}\vartheta) : \xi \mapsto \xi \cdot \phi$ is onto and its kernel is $\mathfrak{g}_\phi$. If $\phi = g \cdot \vartheta$, then one sees that $G_\phi = g \cdot G_\vartheta \cdot g^{-1}$ and similarly for the Lie algebras.

The normal bundle $\mathrm{N}(G{\cdot}\vartheta)$ of $G{\cdot}\vartheta$ is isomorphic to the vector bundle

$$\mathrm{N}(G{\cdot}\vartheta) = G \times_{G_\vartheta} (T_\vartheta G{\cdot}\vartheta)^\perp = G \times_{G_\vartheta} \mathrm{N}_\vartheta(G{\cdot}\vartheta).$$

Here, $G \times \mathrm{N}_\vartheta(G\vartheta)$ is the cartesian product of the group G with the orthogonal complement $\mathrm{N}_\vartheta(G\vartheta)$ to the tangent space to G's orbit through ϑ. The stabiliser G_ϑ acts linearly on $\mathrm{N}_\vartheta(G\vartheta)$ and by right translation on G. The set $G \times_{G_\vartheta} \mathrm{N}_\vartheta(G\vartheta)$ is the quotient space whose points are the sets (G_ϑ-orbits) $[g, v] = \{(g \cdot h, h \cdot v) \ : \ h \in G_\vartheta\}$ for each $(g, v) \in G \times \mathrm{N}_\vartheta(G{\cdot}\vartheta)$.

It is also a well-known fact that there is an open neighbourhood of $G{\cdot}\vartheta$ which is G-equivariantly diffeomorphic to an open neighbourhood T of $G{\cdot}\vartheta$ in $\mathrm{N}(G{\cdot}\vartheta)$. See [13] for generalities and [7, 9] for specifics on linear Lie groups.

To simplify notation, Θ is used to denote $G{\cdot}\vartheta$ in some cases.

2.1. **The projection map onto $G{\cdot}\vartheta$.** Let us now derive the projection map $\pi : \mathrm{T} \to G{\cdot}\vartheta$. Given $x \in \mathbf{E}$, assume that there is a $g \in G$ such that

$$g^{-1}x \in \vartheta + \mathrm{N}_\vartheta(G{\cdot}\vartheta) \qquad \text{which implies } x \in g\vartheta + \mathrm{N}_{g\vartheta}(G{\cdot}\vartheta), \tag{1}$$

[3] One can equally use left translations.

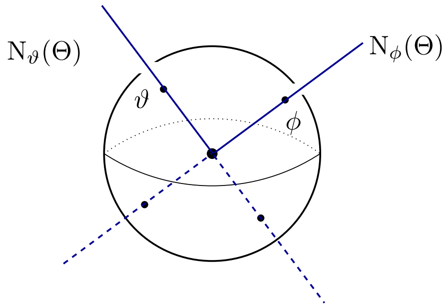


FIGURE 1. The tubular neighbourhood T and the normal bundle $\mathrm{N}(\Theta)$.

by G-equivariance. In this case, we can define

$$\pi(x) = g\vartheta. \tag{2}$$

Lemma 2.1. *There is an open neighbourhood* T *of* $\Theta = G{\cdot}\vartheta$ *such that the map* π *defined in* (2) *is independent of* ϑ *in* Θ*. In addition,* π *is a real-analytic submersion whose fibres are open neighbourhoods of* $0 \in \mathrm{N}_\phi(\Theta)$ *for each* $\phi \in \Theta$*.*

Proof. It suffices to observe that if $x \in \mathrm{N}_\vartheta(G{\cdot}\vartheta)$, then one can take $g = 1 \bmod G_\vartheta$ in (1) and $\pi(x) = \vartheta$, and that (1–2) defines π as a G-equivariant map from $\mathrm{N}(G{\cdot}\vartheta)$ to $G{\cdot}\vartheta$. The lemma then follows from the tubular neighbourhood theorem [13]. □

Remark 2.2. 1/ In general, the affine planes $\vartheta + \mathrm{N}_\vartheta(\Theta)$ and $\phi + \mathrm{N}_\phi(\Theta)$ intersect each other, as in figure 1. At such a point of intersection π is not single-valued; hence these points obstruct the extension of π from a tubular neighbourhood of Θ to a globally-defined map on $\mathbf{E}$. 2/ Lemma 2.1 is a consequence of the tubular neighbourhood theorem for group orbits. Moreover, many linear-algebraic decompositions are, in fact, an application of this tubular neighbourhood result.

2.2. **Linear priors on $G{\cdot}\vartheta$.** As noted in the introduction, the euclidean structure σ induces a linear isomorphism $\mathbf{E} \to \mathbf{E}^* : v \mapsto \hat{v}(\bullet) = \sigma(v, \bullet)$.[4] For each $v \in \mathbf{E}$, let $\hat{v} \in \mathbf{E}^*$ be the dual vector induced by the euclidean structure σ and let $f_v = \hat{v}|G{\cdot}\vartheta$ be the restriction of $\hat{v}$ to the group orbit. In terms of the inclusion map $\iota : G\vartheta \to \mathbf{E}$, one can write $f_v = \hat{v} \circ \iota$. Let $\underline{f}_v$ be the minimum value of f_v and $\bar{f}_v = \int f_v(\phi)\, \mathrm{d}\phi$ be the mean value of f_v with respect to $\mathrm{d}\phi$, the unique G-invariant probability measure on $G{\cdot}\vartheta$. (One can define $\bar{\phi} = \int \iota(\phi)\, \mathrm{d}\phi$ to be the mean element of $G{\cdot}\vartheta$, in which case $\bar{f}_v = \langle v, \bar{\phi} \rangle$. Since $\bar{\phi}$ is a fixed point of G, $\bar{\phi} = 0$ unless $\mathbf{E}$ contains a trivial representation of G.) Define a bayesian prior density $\lambda = \lambda_v$ by

$$\lambda_v = \alpha f_v + \beta \tag{3}$$

where the real numbers $\alpha > 0$ and β satisfy $\alpha \bar{f}_v + \beta = 1$ and $\alpha \underline{f}_v + \beta = c > 0$.

The chain rule shows that $\mathrm{d}f_v = \mathrm{d}\hat{v} \circ \mathrm{d}\iota$, whence

$$\begin{aligned} \nabla f_v(\phi) &= \mathrm{d}_\phi \pi(v) \\ \nabla \log \lambda_v(\phi) &= \frac{\alpha}{\lambda_v} \times \mathrm{d}_\phi \pi(v) \qquad\qquad \text{for all } \phi \in G{\cdot}\vartheta. \end{aligned}$$

The gradient vanishes at ϕ iff $v \in \mathrm{N}_\phi(G{\cdot}\vartheta)$. The chain rule for second derivatives shows that $\nabla \mathrm{d} f_v = \nabla \mathrm{d}\hat{v}(\mathrm{d}\iota, \mathrm{d}\iota) + \mathrm{d}\hat{v} \circ \nabla \mathrm{d}\iota = \mathrm{d}\hat{v} \circ \nabla \mathrm{d}\iota$ since $\nabla \mathrm{d}\hat{v} = 0$ because $\hat{v}$ is linear. The tensor field $\nabla \mathrm{d}\iota$ is the second fundamental form of $G{\cdot}\vartheta$ in $\mathbf{E}$ and it is a measure of the curvature of $G{\cdot}\vartheta$. Application of the definition of $\nabla \mathrm{d}\iota$ [6] shows

[4] One often thinks of this map as $v \mapsto v'$, v maps to v-transpose. This notation is also used below.

that $\nabla \mathrm{d}\iota(\xi_\phi, \eta_\phi) = (1 - \mathrm{d}_\phi \pi)\xi \cdot \eta \cdot \phi$ for all $\xi, \eta \in \mathfrak{g}$. Thus $\nabla \mathrm{d} f_\vartheta(\xi, \xi)|_\vartheta = -|\xi \cdot \vartheta|^2$ for all $\xi \in \mathfrak{g}_\vartheta^\perp \simeq T_\vartheta G{\cdot}\vartheta$, so the maximum is a non-degenerate critical point of f_ϑ. A well-known theorem in Morse theory states that for almost all $v \in \mathbf{E}$, f_v is a Morse function on $G{\cdot}\vartheta$ [13].

For the purposes of imposing a strong prior, a natural choice is $v = \vartheta$. The Cauchy-Schwarz inequality plus the fact that G acts by isometries implies that f_ϑ (hence λ_ϑ) attains its unique maximum value at ϑ.

Let $\mathrm{d}g$ be the Haar probability measure on G. The Haar measure factors as $\mathrm{d}h \cdot \mathrm{d}\vartheta$, where $\mathrm{d}h$ is the Haar probability measure on H, the stabiliser of ϑ, and $\mathrm{d}\vartheta$ is the unique G-invariant probability measure on $\Theta = G{\cdot}\vartheta$. Define

$$\tilde{v}_\vartheta = \int_{g \in G} \mathrm{d}g \, \log \lambda_{g\cdot v}(\vartheta) \, g \cdot v$$

to be the mean of v over $G{\cdot}v$ taken with respect to a peculiar measure.

Theorem 2.3. *Let $\Theta = G{\cdot}\vartheta$ and γ be the identity map of Θ. Let the bayesian prior density $\lambda = \lambda_v$ be defined by (3). Let $x \in \mathrm{T}$, $\hat{\vartheta} = \pi(x)$ and $\xi \in \mathfrak{g}_{\hat{\vartheta}}^\perp$ be the unique vector such that $\xi \cdot \hat{\vartheta} = \mathrm{d}_{\hat{\vartheta}}\pi(v)$. The bayesian estimator $\tilde{g}_\epsilon$ and its risk equal*

$$\tilde{g}_\epsilon(x) = \exp(s\xi + O(\epsilon^4)) \cdot \hat{\vartheta} \qquad s = \frac{\alpha\epsilon^2}{\lambda_v(\hat{\vartheta})} \tag{4}$$

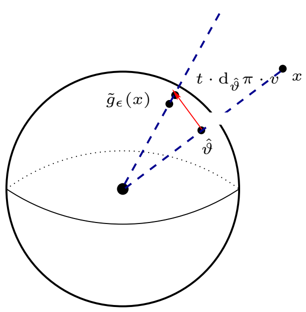
$$R_\epsilon(\tilde{g}_\epsilon; \lambda_v) = \epsilon^2 \, dim\, \Theta + \epsilon^4 \left(\frac{1}{3} \, \mathrm{scal}_{\Theta,\lambda} + \left\langle \tilde{v}_{\hat{\vartheta}}, \tau(\iota)_{\hat{\vartheta}} \right\rangle \right) + O(\epsilon^6) \tag{5}$$

where exp *is the exponential map of the Lie group G,* $\mathrm{scal}_{\Theta,\lambda}$ *is the average of the scalar curvature of Θ with respect to $\mathrm{d}\vartheta\lambda$ and $\tau(\iota)$ is the normal vector field of Θ.*

The proof of (4) applies Theorem 1.1 and the fact that G acts as a transitive isometry group of Θ. It should be noted that, although $\tilde{v}_\vartheta$ is not independent of ϑ, the inner product $\langle \tilde{v}_\vartheta, \tau(\iota)_\vartheta \rangle$ is independent. In addition, the integration-by-parts formula is needed to demonstrate (5).

In the particular case of $v = 0$, λ is a flat prior density, the G-invariant measure $\mathrm{d}\theta$ is the flat distribution and the estimator is $\tilde{g}_\epsilon(x) = \hat{\vartheta} + O(\epsilon^4)$ with risk $R_\epsilon(\tilde{g}_\epsilon; 1) = \epsilon^2 \dim \Theta + \epsilon^4 \mathrm{scal}_\Theta / 3 + O(\epsilon^6)$ where scal_Θ is the mean scalar curvature of Θ. The flat prior produces the minimax estimator in this case.

2.2.1. *A sample application: S^2.* Let us apply theorem 2.3 to the case where $\mathbf{E} = \mathbf{E}^3$, Θ is the 2-dimensional unit sphere S^2 in $\mathbf{E}^3$ and $G = \mathrm{SO}(3)$ is the group of linear, orientation-preserving isometries of $\mathbf{E}^3$. In this case, the projection map is $\pi(x) = x/|x|$ and the projection of v onto $T_{\hat{\vartheta}}\Theta$ is the orthogonal projection $\bar{v} = \mathrm{d}_{\hat{\vartheta}}\pi(v) = v - \langle v, \hat{\vartheta} \rangle \, \hat{\vartheta}$. The bayesian estimator in this case is



$$\tilde{g}_\epsilon(x) = \exp(t\hat{v}) \cdot \hat{\vartheta} + O(\epsilon^4), \qquad t = \frac{\epsilon^2 |\bar{v}|}{\langle v, \hat{\vartheta} \rangle + \beta}$$

where $\hat{v} \in \mathrm{so}(3)$ is the rotation by $\pi/2$ radians counterclockwise in the plane orthogonal to $\bar{v}$.

If one supposes that $v \in S^2$, then $\beta = 1$ and $0 \leq \alpha < 1$. Since $\tau(\iota)_\vartheta = \vartheta$, one computes that

$$\langle \tilde{v}_v, \tau(\iota)_v \rangle = (\alpha^{-2} - 1) \log \sqrt{\frac{1-\alpha}{1+\alpha}} + \frac{1}{\alpha}. \tag{6}$$

On the other hand, the scalar curvature of the 2-dimensional unit sphere is twice the Gaussian curvature, hence is 2, and the mean of λ is 1, so $\mathrm{scal}_{\Theta,\lambda} = 2$. The bayesian risk is therefore

$$R(\tilde{g}_\epsilon; \lambda_v) = 2\epsilon^2 + \left(\frac{2}{3} + \langle \tilde{v}_v, \tau(\iota)_v \rangle\right) \epsilon^4 + O(\epsilon^6).$$

Inspection of (6) shows that the right-hand side is α at $\alpha = 0, 1$ and it is monotone increasing on $[0, 1]$. This verifies that the flat prior ($\alpha = 0$) yields the second-order minimax estimator.

2.3. **Derivation and application of the projection map.** This section applies Lemma 2.1 to a wide range of orbit spaces. Lemma 2.1 says that to construct the bayesian estimator of theorem 1.1 it is necessary to give a concrete description of a tubular neighbourhood T and the projection map $\pi : \mathrm{T} \to \Theta$ from the tubular neighbourhood to the group orbit Θ.

It is also necessary to give a concrete description of the exponential map of the riemannian manifold $(\Theta, \mathbf{g}) \subset (\mathbf{E}^s, \sigma)$. This problem is solved as in Theorem 2.3, where one uses the linear isomorphism between $T_\vartheta\Theta$ and $\mathfrak{g}_\vartheta^\perp$, which pulls back the riemannian exponential map to the Lie group's exponential map.

[4, Chapters 1–3] provide a nice background, aimed at statisticians, for many applications of several of the orbit spaces considered below.

2.3.1. *The sphere S^{n-1}.* Let $G = \mathrm{O}(\mathbf{E}^n)$, and $\vartheta \in \mathbf{E}^n$ be non-zero. The group orbit $G{\cdot}\vartheta$ is the sphere of radius $r = |\vartheta|$. Without loss of generality, one can suppose that $r = 1$ and $\mathbf{E}^n$ has a basis e_i where $e_1 = \vartheta$. In this case, $\mathrm{N}_\vartheta(G{\cdot}\vartheta) = \mathbf{R}\vartheta$ and $g^{-1}x \in \vartheta + \mathrm{N}_\vartheta(G{\cdot}\vartheta)$ iff $g^{-1}x = \lambda e_1$ iff $x = \lambda\, g e_1$ and $\lambda = \pm|x|$. Because π must be the identity on $G{\cdot}\vartheta$, one sees that $\lambda > 0$ and therefore $\forall x \neq 0$

$$\pi(x) = g\vartheta = x/|x|.$$

In this case $\mathrm{T} = \mathbf{E} - \{0\}$.

2.3.2. *The Stiefel manifold.* There is a natural generalisation of the unit sphere introduced by Steifel [4, 8]. Let $v = [v_1 \cdots v_k]$ be a k-tuple of unit vectors $v_i \in \mathbf{E}^n$ for $k \leq n$ which are mutually orthogonal. The set of all such orthonormal k-frames v is called a Stiefel manifold and denoted by $V_k(\mathbf{E}^n)$. One can naturally identify $V_k(\mathbf{E}^n)$ as a subset of $\mathbf{E} = \mathrm{Hom}(\mathbf{E}^k, \mathbf{E}^n)$ (the $n \times k$ real matrices). The euclidean structure $\sigma = \langle \cdot, \cdot \rangle$ on $\mathbf{E}$ is defined by

$$\langle x, y \rangle = \mathrm{Tr}\,(x'y) \tag{7}$$

for all $x, y \in \mathbf{E}$ where x' is the transpose of x. The group $G = \mathrm{O}(\mathbf{E}^n)$ acts on $\mathbf{E}$ by left multiplication and with the frame $\vartheta = [e_1 \cdots e_k]$

$$V_k(\mathbf{E}^n) = G{\cdot}\vartheta.$$

Given $x \in \mathbf{E}$, the map

$$\kappa(x) = x'x \qquad\qquad \kappa : \mathrm{Hom}(\mathbf{E}^k, \mathbf{E}^n) \to \mathrm{sym}(\mathbf{E}^k),$$

from the $k \times k$ matrices to the symmetric $k \times k$ matrices, defines a submersion when x is of maximal rank k and G acts transitively on the fibre of κ. Thus, if $x \in \mathbf{E}$ is of maximal rank, then the normal space $x + \mathrm{N}_x(G{\cdot}x)$ can be identified with $x'x + \mathrm{sym}(\mathbf{E}^k)$ via the linearized map $\mathrm{d}_x\kappa$.

To compute the projection map $\pi : \mathrm{T} \to V_k(\mathbf{E}^n)$: let T be the connected component containing ϑ of the set of $x \in \mathbf{E}$ of maximal rank. For each $x \in \mathrm{T}$, $x'x$ is a symmetric, positive-definite matrix and therefore $x'x$ has a unique symmetric positive-definite square root $\tau =: (x'x)^{\frac{1}{2}}$. Let us define

$$\pi(x) = x(x'x)^{-\frac{1}{2}} \qquad\qquad \pi : \mathrm{T} \to V_k(\mathbf{E}^n). \tag{8}$$

It is clear that π is a G-equivariant map, $\pi(\vartheta) = \vartheta$ and since $\kappa \circ \pi$ maps T to $1 \in \mathrm{sym}(\mathbf{E}^k)$, the image is $V_k(\mathbf{E}^n)$ and $\pi|_{V_k(\mathbf{E}^n)} = id$. These facts suffice to show that the map π is indeed the projection map of the normal bundle. (If one had taken another square root of $x'x$ to define π, then $\pi(\vartheta) \neq \vartheta$, so that map could not be the projection map of a tubular neighbourhood).

In the general case, let $\vartheta \in \mathbf{E}$ be of maximal rank and let τ be the unique positive-definite symmetric square root of $\theta'\vartheta$. The projection map $\pi : \mathrm{T} \to G{\cdot}\vartheta$ is then

$$\pi(x) = x(x'x)^{-\frac{1}{2}}\tau \qquad\qquad \pi : \mathrm{T} \to G{\cdot}\vartheta,$$

where T is the set of maximal rank elements in $\mathbf{E}$.

One can specialize the above construction to obtain:

$k = 1$: In this case, $V_1(\mathbf{E}^n) = S^{n-1}$ and $(x'x)^{\frac{1}{2}} = |x|$, so (8) specializes to yield the projection map onto S^{n-1};

$k = 2$: In this case, $V_2(\mathbf{E}^n)$ is the unit sphere bundle of S^{n-1}, so (8) specializes to yield the projection map from the set of non-collinear vectors in $\mathbf{E}^n \times \mathbf{E}^n$ to the unit sphere bundle $S(S^{n-1})$;

$k = n$: In this case, $V_n(\mathbf{E}^n) = \mathrm{O}(\mathbf{E}^n)$ and $p = (x'x)^{\frac{1}{2}}$ is the polar factor in the polar decomposition $x = gp$ where $g \in \mathrm{O}(\mathbf{E}^n)$ and $p \in \mathrm{sym}(\mathbf{E}^n)$. Thus (8) specializes to yield the projection map of an arbitary invertible matrix onto its orthogonal part. See example 2.3.9 below for a general construction.

2.3.3. *The real Grassmannian manifold.* Another group orbit space that is closely related to the Steifel manifold is the manifold of unoriented k-dimensional planes in $\mathbf{E}^n$, called the Grassmannian manifold [4, 8]. Let $G_k(\mathbf{E}^n)$ denote the Grassmannian manifold of unoriented k-planes in $\mathbf{E}^n$. A k-plane Π in $\mathbf{E}^n$ is uniquely characterized by an orthogonal projection $\mathbf{p}_\Pi \in \mathrm{Hom}(\mathbf{E}^n, \mathbf{E}^n)$ which is symmetric, has an image equal to Π and kernel equal to $\Pi^\perp$. Since each plane Π and its orthogonal complement admit an orthonormal basis, we have the following natural description of the Grassmannian manifold as an orbit space $(k + l = n)$

$$G_k(\mathbf{E}^n) = G{\cdot}\vartheta \qquad\qquad G = \mathrm{O}(\mathbf{E}^n), \quad \vartheta = \begin{bmatrix} 1_k & 0 \\ 0 & 0_l \end{bmatrix}$$

where the action of G on the symmetric matrices $\mathrm{sym}(\mathbf{E}^n) \subset \mathrm{Hom}(\mathbf{E}^n, \mathbf{E}^n)$ is by conjugation/congruence

$$g \cdot x = gxg' \qquad\qquad \forall g \in G, x \in \mathrm{sym}(\mathbf{E}^n).$$

The Grassmannian manifold is equivariantly diffeomorphic to

$$G_k(\mathbf{E}^n) = \mathrm{O}(\mathbf{E}^n)/\mathrm{O}(\mathbf{E}^k) \times \mathrm{O}(\mathbf{E}^l).$$

To identify the normal space $\mathrm{N}_\vartheta(G_k(\mathbf{E}^n))$, a computation shows that

$$T_\vartheta G_k(\mathbf{E}^n) = \left\{ \begin{bmatrix} 0_k & \alpha \\ \alpha' & 0_l \end{bmatrix} \ : \ \alpha \in \mathrm{Hom}(\mathbf{E}^l, \mathbf{E}^k) \right\}$$

whence the normal space is

$$\mathrm{N}_\vartheta(G_k(\mathbf{E}^n)) = \left\{ \begin{bmatrix} \beta & 0 \\ 0 & \gamma \end{bmatrix} \ : \ \beta \in \mathrm{sym}(\mathbf{E}^k), \gamma \in \mathrm{sym}(\mathbf{E}^l) \right\}.$$

Recall that every $x \in \mathrm{sym}(\mathbf{E}^n)$ is congruent via some $g \in \mathrm{O}(\mathbf{E}^n)$ to a diagonal matrix $\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ where the eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

One may then define, for all those x with $\lambda_k > \lambda_{k+1}$, the projection of x onto the Grassmannian manifold via

$$x = g\lambda g' \qquad\qquad \Longrightarrow \qquad\qquad \pi(x) = g\vartheta g'.$$

Equivalently, x is congruent via a $g_1 \in \mathrm{O}(\mathbf{E}^n)$ to a matrix y in $\mathrm{N}_\vartheta(G_k(\mathbf{E}^n))$ where the eigenvalues of β dominate those of γ. In this case, one can define $\pi(x) = g_1\vartheta g_1'$. The two definitions of π coincide since $g = g_1 \bmod$ the stabilizer of ϑ. In this case, the tubular neighbourhood T is the connected component containing ϑ of the set of $x \in \mathrm{sym}(\mathbf{E}^n)$ which have eigenvalues such that $\lambda_k > \lambda_{k+1}$.

Remark 2.4. In the above construction, one may replace $(\mathbf{E}^n, \mathrm{sym}(\mathbf{E}^n), \mathrm{O}(\mathbf{E}^n))$ and the real transpose by $(\mathbf{C}^n, \mathrm{sym}(\mathbf{C}^n), \mathrm{U}_n \subset \mathrm{O}(\mathbf{C}^n))$ and the conjugate transpose (resp. $(\mathbf{H}^n, \mathrm{sym}(\mathbf{H}^n), \mathrm{Sp}(\mathbf{H}^n) \subset \mathrm{O}(\mathbf{H}^n))$ and the quaternionic conjugate transpose) to obtain the grassmannian of complex k-planes in $\mathbf{C}^n$ (resp. the grassmannian of quaternionic k-planes in quaternionic n-space $\mathbf{H}^n$). In these cases, one views $\mathbf{C}^n$ (resp. $\mathbf{H}^n$) as a real euclidean vector space, where the euclidean structure is provided by the real part of the hermitian (resp. quaternionic) structure, and the isometries preserve both the euclidean structure and the complex (resp. quaternionic) structure. The construction of the projection map of the tubular neighbourhood is essentially the same. Since the conclusions of Theorem 1.1 rely only on the real euclidean structure, the conclusions remain valid.

2.3.4. *The singular-value decomposition.* Let $\mathbf{E} = \mathrm{Hom}(\mathbf{E}^k, \mathbf{E}^n)$ (the $n \times k$ real matrices) with the euclidean structure defined as in 2.3.2 and let $G = \mathrm{O}(\mathbf{E}^n) \times \mathrm{O}(\mathbf{E}^k)$ act on $\mathbf{E}$ by

$$g \cdot x = g_1 x g_2^{-1} \qquad\qquad \forall g = (g_1, g_2) \in G, x \in \mathbf{E}. \tag{9}$$

Without loss of generality, one may assume that $k \geq n$. In this case, the well-known singular-value decomposition says that there is a $g \in G$ such that $g^{-1} \cdot x = \vartheta$ where ϑ is in the "diagonal" form

$$\vartheta = \left[\begin{array}{cccc|c} \vartheta_1 & 0 & \cdots & 0 & 0 \\ 0 & \vartheta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \vartheta_n & 0 \end{array}\right] \qquad \text{and } \vartheta_1 \geq \vartheta_2 \geq \cdots \geq \vartheta_n \geq 0. \tag{10}$$

To compute the normal space of $G{\cdot}\vartheta$ is somewhat involved, but one can simplify the computation in the following way.

A non-degenerate symmetric bilinear form η is of index (n, k) if it is positive definite (resp. negative definite) on a subspace of dimension n (resp. k). Let $\mathbf{E}^{n,k}$ be a real vector space with indefinite inner product η of index (n, k) [14]. The orthogonal group, $H = \mathrm{O}(\mathbf{E}^{n,k})$, of this pseudo-euclidean space is non-compact but its maximal compact subgroup is $G = \mathrm{O}(\mathbf{E}^n) \times \mathrm{O}(\mathbf{E}^k)$.[5] The Lie algebra, $\mathfrak{h} = \mathrm{o}_{n,k}$, of $\mathrm{O}(\mathbf{E}^{n,k})$ contains the subalgebra $\mathfrak{g} = \mathrm{o}_n \oplus \mathrm{o}_k$ and its orthogonal complement relative to the trace form, the subspace

$$\mathfrak{p} = \left\{ \begin{bmatrix} 0_n & x \\ x' & 0_k \end{bmatrix} \;:\; x \in \mathrm{Hom}(\mathbf{E}^k, \mathbf{E}^n) \right\}.$$

The action by conjugation of $G = \mathrm{O}(\mathbf{E}^n) \times \mathrm{O}(\mathbf{E}^k)$ on $\mathfrak{p}$ is naturally identified with the action defined in (9).

For $\boldsymbol{\vartheta} \in \mathfrak{p}$, the orbit of G has

$$T_{\boldsymbol{\vartheta}}(G{\cdot}\boldsymbol{\vartheta}) = \mathrm{ad}_{\mathfrak{g}}\boldsymbol{\vartheta} \qquad\qquad \mathrm{N}_{\boldsymbol{\vartheta}}(G{\cdot}\boldsymbol{\vartheta}) = \mathfrak{p} \cap (\mathrm{ad}_{\mathfrak{g}}\boldsymbol{\vartheta})^{\perp} = \mathfrak{p}_{\boldsymbol{\vartheta}} \tag{11}$$

[5] In the special case of $k = 1$, one has Lorentzian geometry, which is important in special relativity [14].

where $\mathfrak{p}_{\boldsymbol{\vartheta}}$ is the intersection of the centralizer of $\boldsymbol{\vartheta}$ in $\mathfrak{h}$ with $\mathfrak{p}$. If one supposes that

$$\boldsymbol{\vartheta} = \begin{bmatrix} 0_n & \vartheta \\ \theta' & 0_k \end{bmatrix}, \qquad \text{and } \vartheta \text{ is in the diagonal form of}$$

then $\mathfrak{p}_{\boldsymbol{\vartheta}}$ contains all elements of the same form as $\boldsymbol{\vartheta}$. Therefore, for

$$\mathbf{x} = \begin{bmatrix} 0_n & x \\ x' & 0_k \end{bmatrix} \in \mathfrak{p}$$

one knows from the singular-value decomposition of x, that $\mathbf{x}$ is conjugate via a $g \in G$ to an element in $\mathfrak{p}_{\boldsymbol{\vartheta}}$. Thus, one has that

$$\begin{aligned} \pi(\mathbf{x}) &= g\boldsymbol{\vartheta}g^{-1} \qquad \text{where } g = (g_1, g_2) \text{ and } g_1^{-1}xg_2 \text{ is diagonal} \\ \pi &: \mathbf{T} \to G\cdot\boldsymbol{\vartheta} \end{aligned}$$

or, equivalently

$$\pi(x) = g_1\vartheta g_2^{-1} \qquad \pi : \mathrm{T} \to G\cdot\vartheta$$

where T is the set of x whose singular values have collisions nowhere except possibly where those of ϑ collide and $\mathbf{T}$ is defined similarly.

Remark 2.5. From an applied point of view, one may wish to approximate the matrix $x \in \mathrm{Hom}(\mathbf{E}^k, \mathbf{E}^n)$ by a low rank matrix. To do this, one could specify ϑ in (10) with $\vartheta_i = 1$ for $i = 1, \ldots, l$ and $\vartheta_i = 0$ for $i > l$. One can then use (2.3.4–2.3.4) to project x onto the low-rank matrix orbit $G\cdot\vartheta$. In applied mathematics, one approximates x by a low rank matrix in a slightly different manner. One computes the singular-value decomposition $\lambda = g_1^{-1}xg_2$ with λ in the diagonal form (10), one truncates λ to a diagonal matrix λ_0 by zeroing out the singular values $\lambda_{l+1}, \ldots, \lambda_n$ and then one defines $x_0 = g_1\lambda_0 g_2^{-1}$ to be the low-rank approximation (in practice, $l \ll n, k$ so one saves only the l right and left singular vectors not g_1 and g_2). In this case, one knows that the set of rank l matrices is the union over all rank l orbits, and one uses x to determine the particular orbit onto which x is projected. By construction, this determines the rank l orbit that is closest to x.

2.3.5. *The lagrangian grassmannian.* A totally real subspace Π of $\mathbf{C}^n$ is a real subspace which has the distinguished property that $\Pi \cap i\Pi = 0$; a totally real n-plane in $\mathbf{C}^n$ is also called a lagrangian plane. Let $\Lambda_n = \mathrm{U}_n/\mathrm{O}_n$ be the manifold of lagrangian planes in $\mathbf{C}^n$. This manifold arose in Maslov's work on quantisation [11]. It is well-known in hamiltonian mechanics that the stable and unstable subspaces of a hyperbolic linear hamiltonian system are both lagrangian planes [1].

One can embed Λ_n into $\mathbf{E} = \mathrm{sym}(\mathbf{C}^n)$, the subspace of complex $n \times n$ matrices which are symmetric under the transpose (not conjugate transpose). The euclidean structure $\sigma = \langle \cdot, \cdot \rangle$ on $\mathbf{E}$ is defined by

$$\langle x, y \rangle = \mathrm{Tr}\,(x^* y) \qquad \forall x, y \in \mathrm{sym}(\mathbf{C}^n),$$

where x^* is the conjugate transpose of x. The unitary group $G = \mathrm{U}_n$ acts by isometries of $(\mathbf{E}, \sigma)$ by

$$g \cdot x = gxg' \qquad \forall g \in \mathrm{U}_n, x \in \mathbf{E}. \tag{12}$$

The stabilizer of $\vartheta = 1 \in \mathbf{E}$ under this action is the real orthogonal subgroup O_n, so $\Lambda_n = G\cdot\vartheta$.

To compute the projection map, one observes that $T_\vartheta(G\cdot\vartheta) = i \cdot \mathrm{sym}(\mathbf{R}^n)$ and $\mathrm{N}_\vartheta(G\cdot\vartheta) = \mathrm{sym}(\mathbf{R}^n)$. To define the projection of $x \in \mathbf{E}$ onto $G\cdot\vartheta$, it is necessary that there exists a unitary $g \in \mathrm{U}_n$ such that $g^{-1} \cdot x \in \vartheta + \mathrm{N}_\vartheta(G\cdot\vartheta) = \mathrm{sym}(\mathbf{R}^n)$. Thus,

$$\exists g \in \mathrm{U}_n, p \in \mathrm{sym}(\mathbf{R}^n) \text{ such that } x = gpg' \qquad \Longrightarrow \qquad \pi(x) = gg'. \tag{13}$$

To see the connection with the above description of Λ_n, note that if $x \in \mathrm{sym}(\mathbf{C}^n)$ admits a factorization as in (13), then the symmetric quadratic from $q_x(u,v) = u'xv$ is totally real on the real subspace $w \cdot \mathbf{R}^n \subset \mathbf{C}^n$ where $w = (g')^{-1}$; and conversely, if q_x is totally real on $w \cdot \mathbf{R}^n$, then x admits a factorization as in (13). Provided that x is non-degenerate, $w \in \mathrm{U}_n$ and hence g, is uniquely defined up to an element in O_n.

Remark 2.6. As in the singular-value decomposition, there is a natural Cartan decomposition that is associated with this example [7]. Let $H = \mathrm{Sp}(\mathbf{R}^{2n})$ be the group of symplectic automorphisms of $\mathbf{R}^{2n} = \mathbf{C}^n$ where the symplectic form is the skew-symmetric bilinear form

$$\omega(x,y) = \mathrm{Im}\, x^* y \qquad\qquad \forall x, y \in \mathbf{C}^n.$$

The Lie algebra $\mathfrak{h}$ of H admits a Cartan decomposition $\mathfrak{h} = \mathfrak{k} + \mathfrak{p}$ where

$$\mathfrak{k} = \left\{ \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \;:\; \alpha = -\alpha', \beta = \beta' \right\}, \qquad \mathfrak{p} = \left\{ \begin{bmatrix} a & b \\ b & -a \end{bmatrix} \;:\; a' = a, b' = b \right\}$$

where all matrices are real. The maps $(\alpha, \beta) \mapsto \alpha + i\beta$ and $(a,b) \mapsto a + ib$ shows that $\mathfrak{k} \simeq \mathrm{u}_n$ and $\mathfrak{p} \simeq \mathrm{sym}(\mathbf{C}^n)$. The action of $K \simeq \mathrm{U}_n$ on $\mathfrak{p}$ by conjugation is identified with the action on $\mathrm{sym}(\mathbf{C}^n)$ by congruences (12). We note that, by the theory of Cartan subalgebras, an $\mathbf{x} = (a,b) \in \mathfrak{p}$ may be diagonalized over $\mathbf{R}$ on a basis that is simultaneously symplectic and orthogonal, that is, $\mathbf{x} \in \mathfrak{p}$ is conjugate to a real diagonal matrix via a unitary transformation [7]. This implies the validity of the decomposition (13) on the set of $\mathbf{x} \in \mathfrak{p}$ ($x \in \mathrm{sym}(\mathbf{C}^n)$) which are non-singular.

To compute the projection map $\pi : \mathrm{T} \to G{\cdot}\vartheta$ when ϑ is in general position, it is most easy to apply (11).

Remark 2.7. In the above construction, one may replace $(\mathbf{E}^n, \mathbf{C}^n, \mathbf{E} = \mathrm{sym}(\mathbf{C}^n), \mathrm{U}_n = \mathrm{O}(\mathbf{C}^n))$ and the real transpose by $(\mathbf{C}^n, \mathbf{H}^n, \mathbf{E} = \mathrm{sym}(\mathbf{H}^n), \mathrm{Sp}(\mathbf{H}^n) = \mathrm{O}(\mathbf{H}^n))$ and the complex conjugate transpose. The orbit of $\vartheta = 1$ is the homogeneous space $\mathrm{Sp}(\mathbf{H}^n)/\mathrm{U}(\mathbf{C}^n)$, which is the grassmannian of totally complex n-planes in $\mathbf{H}^n$.

2.3.6. *The isotropic grassmannians.* There are several distinguished orbits in $\mathrm{sym}(\mathbf{C}^n)$ in addition to the lagrangian grassmannian. From the natural embedding $\mathbf{R}^k \subset \mathbf{R}^n \subset \mathbf{C}^n$, one obtains the grassmannian manifold of isotropic (or totally real) k-planes in $\mathbf{C}^n$

$$\Lambda_{k,n} = \mathrm{U}_n{\cdot}\mathbf{R}^k = \mathrm{U}_n/\mathrm{O}_k \times \mathrm{U}_l \qquad\qquad (k + l = n).$$

If one defines

$$\vartheta = \begin{bmatrix} 1_k & 0 \\ 0 & 0_l \end{bmatrix} \qquad\qquad \Lambda_{k,n} = \mathrm{U}_n{\cdot}\vartheta \subset \mathrm{sym}(\mathbf{C}^n).$$

The grassmannian of totally real k-planes in $\mathbf{C}^n$ arise naturally in hamiltonian mechanics. For example, the tangent spaces to an orbit of the Keplerian 2-body problem trace a closed curve in $\Lambda_{k,n}$.

To compute the tubular neighbourhood of $\Lambda_{k,n}$ and its projection map, remark (2.6) implies that each $x \in \mathrm{sym}(\mathbf{C}^n)$ is congruent via a $g \in \mathrm{U}_n$ to a real diagonal $\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. If $\lambda_k > \lambda_{k+1}$, then the first k largest eigenvalues and consequently the sum of the eigenspaces is uniquely determined. One can then define $\pi(x) = g\vartheta g'$, which is well-defined.

Geometrically, this condition amounts to the following. Remark (2.6) implies that there is some lagrangian plane ℓ on which q_x is totally real. If the condition on the eigenvalues of x is satisfied, then there is a unique isotropic k-plane $\ell_k \subset \ell$ such that $\ell = \ell_k \oplus \ell_k^\perp$. The lagrangian ℓ and $\ell_k^\perp$ are not uniquely defined (if q_x is degenerate), but ℓ_k itself is. The map $x \mapsto \ell_k$ is the projection map of the tubular neighbourhood.

2.3.7. *The manifold of orthogonal complex structures on* $\mathbf{R}^{2n}$. A complex structure J on $\mathbf{R}^{2n}$ is a linear map such that $J^2 = -1$; it is orthogonal if $J' = J^{-1}$. A complex structure has eigenvalues $\pm i$ (repeated n times), so it is necessarily orientation preserving. The conjugate of an orthogonal complex structure is also an orthogonal complex structure, and conversely, any complex structure is conjugate to the standard complex structure by an orthogonal conjugacy.

Let us construct the manifold of orthogonal complex structures as a homogeneous space. Embed $\mathrm{U}(\mathbf{C}^n)$ into $G = \mathrm{SO}(\mathbf{E}^{2n})$ by

$$x = \alpha + i\beta \mapsto \mathbf{x} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \qquad \forall x \in \mathrm{Hom}(\mathbf{C}^n, \mathbf{C}^n)$$

where α (resp. β) is the real (resp. imaginary) part of x. Let $\mathrm{SO}(\mathbf{E}^{2n})$ act on its Lie algebra $\mathbf{E} = \mathrm{so}_{2n}$ by conjugation. The standard complex structure on $\mathbf{R}^{2n}$ is the element

$$\vartheta = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \text{which has the orbit } G{\cdot}\vartheta = \mathrm{SO}_{2n}/\mathrm{U}_n$$

since U_n is the stabilizer group of ϑ.

The tangent and normal space to $G{\cdot}\vartheta$ at ϑ are equal to

$$T_\vartheta(G{\cdot}\vartheta) = \mathrm{ad}_{\mathfrak{p}}\vartheta, \qquad \mathrm{N}_\vartheta(G{\cdot}\vartheta) = \mathrm{u}_n, \qquad \mathfrak{p} = \left\{ \begin{bmatrix} \delta & \gamma \\ \gamma & -\delta \end{bmatrix} \ : \ \delta, \gamma \in \mathrm{so}_n \right\}.$$

where $\mathfrak{p}$ is the orthogonal complement of u_n in so_{2n}. It follows from the fact that every $x \in \mathrm{so}_{2n}$ is contained in a Cartan subalgebra that

$$\exists g \in \mathrm{SO}_{2n} \text{ such that } g'xg = \left[\begin{array}{c|c} \mathbf{0}_n & \begin{matrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{matrix} \\ \hline \begin{matrix} -\alpha_1 & & \\ & \ddots & \\ & & -\alpha_n \end{matrix} & \mathbf{0}_n \end{array}\right] = \alpha,$$

and the stabilizer of α is contained in U_n provided that $\det x \neq 0$. Thus, the tubular neighbourhood T of $G{\cdot}\vartheta$ is the connected component containing ϑ of the set of $x \in \mathrm{so}_{2n}$ such that $\det x \neq 0$. The projection map $\pi(x) = g\vartheta g'$ is therefore defined on T.

Remark 2.8. One can also define the manifold of unitary quaternionic structures J on $\mathbf{C}^{2n}$. In this case, the homogeneous space is $\mathrm{SU}(\mathbf{C}^{2n})/\mathrm{Sp}(\mathbf{H}^n)$ and the construction is essentially the same as above.

2.3.8. *Adjoint orbits.* Let $G \subset H$ be compact Lie groups and let $\mathfrak{g} \subset \mathfrak{h}$ be their Lie algebras. The negative Cartan-Killing form on $\mathfrak{h}$, $(x, y) \mapsto -\mathrm{Tr}\,(\mathrm{ad}_x \cdot \mathrm{ad}_y)$, defines a G-invariant euclidean structure, where G acts on $\mathfrak{h}$ by the adjoint action (conjugation) [7]. Let $\vartheta \in \mathfrak{h}$ and let G_ϑ be the stabilizer of ϑ in G, and $\mathfrak{g}_\vartheta$ be its Lie algebra. One has

$$T_\vartheta(G{\cdot}\vartheta) = \mathrm{ad}_{\mathfrak{g}}\vartheta, \qquad \mathrm{N}_\vartheta(G{\cdot}\vartheta) = \mathrm{ad}_\vartheta^{-1}\left(\mathfrak{g}^\perp\right)$$

and

$$G{\cdot}\vartheta = G/G_\vartheta$$

One knows that there is an equivariant tubular neighbourhood T of $G{\cdot}\vartheta$ such that the projection map $\pi : \mathrm{T} \to G{\cdot}\vartheta$ is defined. The examples above may be formulated in these terms.

2.3.9. *The group G itself.* If $G \subset \mathrm{O}(\mathbf{E}^n)$, then $G \subset \mathrm{Hom}(\mathbf{E}^n, \mathbf{E}^n) = \mathbf{E}$. With the euclidean structure on $\mathbf{E}$ defined as in (7), one obtains a decomposition

$$\mathbf{E} = \mathfrak{g} + \mathfrak{p} = T_1 G + \mathrm{N}_1(G).$$

The projection map of the tubular neighbourhood T can be defined as: if, given $x \in \mathbf{E}$, there is a unique $g \in G$ such that $g^{-1}x \in \mathfrak{p}$, then $\pi(x) := g$. We see that the projection is a generalization of the polar decomposition encountered above in the $k = n$ case of example 2.3.2.

3. Regression problems

This section deals primarily with the first-order conditions for a class of non-linear regression problems. Despite the fact that section 2 showed the construction of second-order minimax estimators, the geometry that underlay those constructions is very similar to that required here. We also give a numerical example of the derived regressor in the specific setting.

Let $\theta_1, \ldots, \theta_k$ be a collection of design points on a manifold Θ and let $y_1, \ldots, y_k$ be a random sample of points on a manifold Λ embedded in a euclidean space as in diagram (14). Let the conditional probability density of y given $\theta \in \Theta$ be $f(y|\gamma(\theta))$, where γ is an unknown map.

$$\begin{tikzcd} & (\mathbf{E}, \sigma) \\ (\Theta, \mathbf{g}) \arrow[ur, "\jmath \cdot \gamma"] \arrow[r, "\gamma"] & (\Lambda, \mathbf{h}), \arrow[u, "\jmath"'] \end{tikzcd} \tag{14}$$

One is interested in estimating the unknown map γ by, say, minimizing the discrepancy function

$$\mathsf{s}(\gamma) = \frac{1}{k} \sum_{l=1}^{k} \ell(y_l, \gamma(\theta_l)). \tag{15}$$

The loss function $\ell : \Lambda \times \Lambda \to \mathbf{R}$ is assumed to satisfy

(1) ℓ is continuous everywhere and smooth a.e.;
(2) $\ell(x, y) = \ell(y, x)$ for all $x, y \in \Lambda$;
(3) $\ell(x, y) \geq 0$ for all $x, y \in \Lambda$ and equals 0 iff $x = y$;
(4) the hessian $\nabla \mathrm{d}\ell|_{\mathrm{N}(\Delta\Lambda)}$ is non-degenerate, where $\Delta\Lambda = \{(x, x) \ : x \in \Lambda\}$ is the diagonal.

Natural examples of loss functions include: 1) that induced by the euclidean structure, $\ell(x, y) = |\jmath x - \jmath y|^2$ for all $x, y \in \Lambda$; and 2) that induced by the intrinsic riemannian distance on Λ, $\ell(x, y) = \mathbf{dist}(x, , y)^2$ for all $x, y \in \Lambda$.

If one assumes that the space of admissible maps γ is parameterized by a compact finite-dimensional manifold Γ, a solution to this estimation problem is

$$\hat{\gamma} = \mathrm{argmin} \, \{\mathsf{s}(\gamma) \ : \ \gamma \in \Gamma\}. \tag{16}$$

To highlight the geometry and minimise the analysis, it is assumed throughout that the space of admissible maps Γ is a compact finite-dimensional submanifold of $C^\infty(\Theta, \Lambda)$.

3.1. **A first-order condition for the least-squares solution.** To state the first-order condition for a solution to (16), one needs some results from differential topology. The space $C^\infty(\Theta, \Lambda)$ may be equipped with the structure of a Fréchet manifold and one may consider Γ as a smooth submanifold.[6] There are canonical smooth maps $\mathrm{ev}_i : \Gamma \to \Lambda$ defined by

$$\begin{tikzcd}
\Theta \times C^\infty(\Theta, \Lambda) \arrow[r, "\mathrm{ev}"] & \Lambda \\
\Theta \times \Gamma \arrow[u, "1\times\mathrm{incl.}"] \arrow[ur, "\mathrm{ev}"] & \\
\Gamma \arrow[u, "\theta_i \times 1"] \arrow[uur, "\mathrm{ev}_i"'] &
\end{tikzcd} \qquad \text{where } \mathrm{ev}(\theta, \gamma) = \gamma(\theta).$$

Proposition 3.1. *The first-order condition for $\gamma \in \Gamma$ to be a minimizer of* s (15) *is that* $\mathrm{d}_\gamma \mathsf{s} \in \mathrm{N}_\gamma(\Gamma) \subset T^*_\gamma C^\infty(\Theta, \Lambda)$. *That is,*

$$\sum_{i=1}^{k} \left(\mathrm{d}_{\lambda_i}\jmath \cdot \mathrm{d}_\gamma \mathrm{ev}_i\right)^* \cdot \left(\jmath(y_i) - \jmath\mathrm{ev}_i(\gamma)\right)\Big|_{T_\gamma\Gamma} = 0 \qquad \textit{where } \lambda_i = \gamma(\theta_i). \tag{17}$$

Remark 3.2. We have seen above that compact group orbits are important examples of smooth manifolds; and each of these lie within a sphere of constant radius. Thus, if Λ is contained in a sphere of constant radius, proposition 3.1 yields the first-order condition

$$\sum_{i=1}^{k} \left(\mathrm{d}_{\lambda_i}\jmath \cdot \mathrm{d}_\gamma \mathrm{ev}_i\right)^* \cdot \jmath(y_i)\Big|_{T_\gamma\Gamma} = 0.$$

Proof of Proposition 3.1. Let us recall the definition of a tangent vector $v \in T_\gamma C^\infty(\Theta, \Lambda)$. One may view v as the derivative at $t = 0$ of an equivalence class of smooth curves γ_t with $\gamma_{t=0} = \gamma$. For each $\theta \in \Theta$, $\gamma_t(\theta)$ is a smooth curve on Λ through $\gamma(\theta)$. Thus, a tangent vector $v \in T_\gamma C^\infty(\Theta, \Lambda)$ is a smooth map $v : \Theta \to T\Lambda$ such that $v(\theta) \in T_{\gamma(\theta)}\Lambda$ for all θ (differential geometers say that v is a smooth section of $\gamma^* T\Lambda$). It follows that if $v \in T_\gamma\Gamma$, then $\mathrm{d}_\gamma \mathrm{ev}_i \cdot v$ is a tangent vector in $T_{\gamma(\theta_i)}\Lambda$.

If $v \in T_\gamma\Gamma$ is a tangent vector, then the chain rule shows that

$$\begin{aligned}
\mathrm{d}_\gamma \mathsf{s} \cdot v &= \frac{2}{k} \sum_{i=1}^{k} \langle \jmath(y_i), \mathrm{d}_{\lambda_i}\jmath \cdot \mathrm{d}_\gamma \mathrm{ev}_i \cdot v \rangle - \langle \jmath\mathrm{ev}_i(\gamma), \mathrm{d}_{\lambda_i}\jmath \cdot \mathrm{d}_\gamma \mathrm{ev}_i \cdot v \rangle \\
&= \frac{2}{k} \sum_{i=1}^{k} \left\langle \left(\mathrm{d}_{\lambda_i}\jmath \cdot \mathrm{d}_\gamma \mathrm{ev}_i\right)^* \cdot \left(\jmath(y_i) - \jmath\mathrm{ev}_i(\gamma)\right), v \right\rangle,
\end{aligned}$$

where one uses the fact that the euclidean structure on $\mathbf{E}$ allows one to identify $T_\bullet\mathbf{E}$ and $T^*_\bullet\mathbf{E}$. This yields (17). □

3.1.1. *Least-squares for linear maps.* Assume that Θ, Λ are isometrically embedded in euclidean spaces $\mathbf{E}_0, \mathbf{E}_1$ with inclusion maps $\iota, \jmath$ respectively.

Definition 3.3. *Let $\Gamma \subset C^\infty(\Theta, \Lambda)$. One says that Γ is a set of* linear maps *if there is a subset $\Delta \subset \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$ that map Θ into Λ such that $\Delta\iota = \jmath\Gamma$.*

[6] A Fréchet space is a Hausdorff, locally convex vector space, with a complete translationally invariant metric [15]. A Fréchet manifold is a Hausdorff topological space with an atlas of smooth coordinate charts into a Fréchet space.

In other words, Γ is a set of linear maps if, for each $\alpha \in \Gamma$, there is an $A \in \Delta$ such that the following commutes:

$$\begin{tikzcd} \mathbf{E}_0 \arrow[r, "A"] & \mathbf{E}_1 \\ \Theta \arrow[u, "\iota"] \arrow[r, "\alpha"] & \Lambda. \arrow[u, "\jmath"'] \end{tikzcd}$$

Since $\jmath$ is a smooth embedding (resp. ι is a smooth immersion when Θ spans $\mathbf{E}_0$), it is permissible to abuse notation and identify Γ and Δ as smooth manifolds.

Corollary 3.4. *Let $\Theta \subset \mathbf{E}_0$ be a spanning set and let Γ be a submanifold of linear maps. If γ is a least-squares solution to* (16)*, then the matrices*

$$\nu = \sum_{i=1}^{k} y_i \otimes \theta_i', \qquad \tau = \sum_{i=1}^{k} \theta_i \otimes \theta_i' \tag{18}$$

satisfy

$$\nu \equiv \gamma\tau \qquad \mod \mathrm{N}_\gamma(\Gamma), \tag{19}$$

where $T_\gamma \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1) = T_\gamma\Gamma \oplus \mathrm{N}_\gamma(\Gamma)$ as in (2)*, and $\nu \in \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1), \tau \in \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_0)$.*

The condition (19) specialises to the least-squares regression formula when $\Lambda = \mathbf{E}^1$, $\Theta = \mathbf{E}^s$ and Γ is the space of linear functions $\mathrm{Hom}(\mathbf{E}^s, \mathbf{E}^1)$ so that the normal space is trivial. In the usual least-squares regression formula, the coefficient vector is viewed as a column vector, whereas here one views the coefficient vector as a linear function and hence a row vector. The standard formula is recovered by tranposing the normal equations (19).

An especially useful application of corollary 3.4 is when $\Theta = \Lambda$ and $\Gamma \subset \mathrm{Hom}(\mathbf{E})$ is a group. In this case the first-order condition simplifies to

$$\nu\gamma' \equiv \gamma\tau\gamma' \qquad \mod \mathrm{N}_1(\Gamma), \tag{20}$$

and when $\Gamma \subset \mathrm{O}(\mathbf{E})$, since τ is symmetric,

$$\gamma'\nu \equiv 0 \qquad \mod \mathrm{N}_1(\Gamma). \tag{21}$$

In other words, γ is the orthogonal projection onto Γ of the matrix ν.

Remark 3.5. Kim [10] looks at the spherical regression problem where one has n known design points x_i on S^2 and there are n observations y_i on S^2 which are distributed about αx_i where $\alpha \in \mathrm{SO}(\mathbf{E}^3)$ is unknown. For a uniform bayesian prior on S^2 and the discrepancy function $\mathsf{s}(a) = \frac{1}{n}\sum_{i=1}^{n} |y_i - ax_i|^2$, Kim shows that the bayesian estimator is the "least-squares" estimator obtained as follows. Let

$$\begin{aligned} x &= \frac{1}{n} Y X' && \text{where } Y = [y_1 \cdots y_n], X = [x_1 \cdots x_n] \\ &= u\sigma v' && u, v \in \mathrm{SO}(\mathbf{E}^3), \sigma = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3) \\ & && \text{with } \sigma_1 \geq \sigma_2 \geq \sigma_3. \end{aligned} \tag{22}$$

The least squares estimator is then

$$\hat{\alpha} = uv'.$$

If one observes that the singular-value-like decomposition of x in (22) can be rewritten to obtain a polar-like decomposition of x,

$$x = uv'(v\sigma v') = gp \qquad g = uv' \in \mathrm{SO}(\mathbf{E}^3), p = v\sigma v' \in \mathrm{sym}(\mathbf{E}^3)$$

whence

$$\hat{\alpha} = \pi(x) \qquad\qquad \text{provided } \sigma_3 > 0.$$

One can see that (21) generalises Kim's [10] formula for the spherical regression problem.

Proof of Corollary 3.4. Since Γ acts linearly on $\mathbf{E}_0$, one sees that the normal equations (17) are linear in γ and simplify to: for all $v \in T_\gamma\Gamma$

$$0 = \sum_{i=1}^{k} \langle y_i - \gamma \cdot \theta_i, v \cdot \theta_i \rangle = \sum_{i=1}^{k} \langle y_i \otimes \theta_i' - \gamma \cdot \theta_i \otimes \theta_i', v \rangle \tag{23}$$

where the second inner product is the trace inner product as in (7) and the inclusion map $\jmath$ is dropped to simplify notation. Rearranging (23) yields (19).

To arrive at (21), one notes that when $\Gamma \subset \mathrm{Hom}(\mathbf{E}, \mathbf{E})$ is a group, then each tangent vector $v \in T_\gamma\Gamma$ is of the form $v = \xi \cdot \gamma$ where $\xi \in T_1\Gamma$. The normal equations (23) are then transformed to

$$0 = \sum_{i=1}^{k} \langle y_i \otimes \theta_i' \cdot \gamma' - \gamma \cdot \theta_i \otimes \theta_i' \cdot \gamma', \xi \rangle \tag{24}$$

for all $\xi \in T_1\Gamma$. Rearranging (24) yields (20). $\square$

3.1.2. *Regression with the intrinsic distance.* Let $\mathbf{dist} : \Lambda \times \Lambda \to \mathbf{R}$ be the riemannian distance of the riemannian manifold $(\Lambda, \mathbf{h})$. One can define the discrepancy functional (15) using $\mathbf{dist}$ to be

$$\mathsf{s}(\gamma) = \frac{1}{k} \sum_{l=1}^{k} \mathbf{dist}(y_l, \gamma(\theta_l))^2, \tag{25}$$

for $\gamma \in C^\infty(\Theta, \Lambda)$.

For each $y \in \Lambda$, the function $x \mapsto \mathbf{dist}(y, x)^2$ is smooth on the open set of x such that there is a unique minimising geodesic from y to x. The set of x on which this function is not differentiable is the cut locus of y — a closed, nowhere dense subset of Λ. If x is not in the cut locus of y, then there exists a unique shortest tangent vector $w = w_y(x) \in T_x\Lambda$ such that $\exp_x w = y$ and $|w| = \mathbf{dist}(y, x)$. One may write $w = \log_x y$; one knows that $w = w_y(x)$ is a smooth vector field off the cut locus of y.

Proposition 3.6. *For x in the complement of the cut locus of y,*

$$\frac{\partial}{\partial x} \mathbf{dist}(y, x)^2 = -2 \log_x y,$$

where $T_x^*\Lambda$ and $T_x\Lambda$ are identified via the metric $\mathbf{h}$.

Proof. Let x_t be a smooth curve such that $x_{t=0} = x$ and $v = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} x_t$. Let $c_t(s)$ be the unique minimal geodesic from x_t to y. It is clear from figure 2 that the derivative of $\frac{1}{2}\,\mathbf{dist}(y, x_t)^2$ is $\langle c'(1), v \rangle$ where $c = c_0$. Since there is a unique shortest geodesic joining y to x, reversibility shows that $c'(1) = -w_y(x) = -\log_x y$. $\square$

Proposition 3.7. *If $\gamma \in \Gamma$ is a minimiser of* s (25), *then* either s *is differentiable at γ and*

$$\mathrm{d}_\gamma \mathsf{s} = -\frac{2}{k} \sum_{l=1}^{k} (\mathrm{d}_\gamma \mathrm{ev}_l)^* \left. \log_{\gamma(\theta_l)} y_l \right|_{T_\gamma\Gamma} = 0;$$

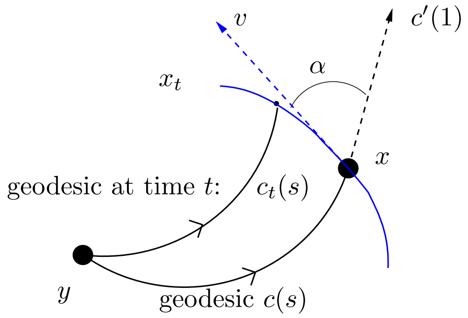


FIGURE 2. The derivative of $\mathbf{dist}(y,x)^2$.

or *there is an l such that $\gamma(\theta_l)$ lies in the cut locus of y_l.*

In the first case where s is defined via the extrinsic distance (15), proposition 3.1 results in a closed form solution for the minimising estimator in many interesting cases. The intrinsic distance leads to a system of normal equations which, even in simple cases, appear opaque. However, there is additional information which one may obtain from these equations. In the first case, since Γ is a finite-dimensional manifold, let us equip it with some riemannian metric. It is well-known that the hessian of a smooth function may be defined using riemannian structures, but that this hessian at a critical point is independent of those structures. Thus, if one lets $\phi_l(x) = \mathbf{dist}(y_l, x)^2$, then the calculus of second derivatives gives

$$\nabla \mathrm{d}\mathsf{s}|_\gamma = \frac{1}{k}\sum_{l=1}^{k} \nabla \mathrm{d}\phi_l(\mathrm{d}_\gamma \mathrm{ev}_l, \mathrm{d}_\gamma \mathrm{ev}_l) + \mathrm{d}_{\gamma(\theta_l)}\phi_l \cdot \nabla \mathrm{d}\,\mathrm{ev}_l$$

where $\nabla \mathrm{d}\phi_l$ is the hessian of ϕ_l, etc.. One knows that $\nabla \mathrm{d}\phi_l(v,w)$ is the second variation of the energy functional $E[c] = \int_0^1 |c'(s)|^2\,\mathrm{d}s$ along the Jacobi fields determined by $v, w \in T_{\gamma(\theta_l)}\Lambda$ and the minimising geodesic c from y_l to $\gamma(\theta_l)$. If it is assumed that these do not lie in the cut locus of the other, then this second variation is necessarily positive. Thus, the only way for $\nabla \mathrm{d}\mathsf{s}$ to not be positive definite is for one or more of the forms $\mathrm{d}\phi_l \cdot \nabla \mathrm{d}\,\mathrm{ev}_l$ to be negative definite along some subspace. This cannot happen if $\nabla \mathrm{d}\,\mathrm{ev}_l$ vanishes for all l.

Proposition 3.8. *If there is a riemannian structure on Γ such that $\nabla \mathrm{d}\,\mathrm{ev}_l = 0$ for all l, and $\gamma \in \Gamma$ is a smooth critical point of s, then γ is a local minimum.*

Example 3.9. Let $\Phi = \Lambda$ be the unit sphere S^2 in $\mathbf{E}^3$ and let $\Gamma = \mathrm{SO}(3)$ be the group of orientation-preserving isometries of S^2. In this case, the distance function is the angle between vectors

$$\mathbf{dist}(y,x) = \arccos(\alpha)$$

while the inverse to the exponential function is

$$\log_x y = \frac{\alpha}{\sin\alpha}\,(x\wedge y)\wedge x \qquad \text{where} \qquad \cos\alpha = \langle y, x\rangle\,, \tag{26}$$

and $y \neq -x$ in (26). One computes that for each $\gamma \in \mathrm{SO}(3)$ and $\xi \in \mathrm{so}(3)$

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathsf{s}(\gamma e^{t\xi}) = -\frac{2}{k}\sum_{l=1}^{k} \mathrm{Tr}\left(\log_{\theta_l}(\gamma' y_l)\otimes \theta_l' \cdot \xi\right)$$

so the pullback of $\mathrm{d}_\gamma \mathsf{s}$ to $T_1\mathrm{SO}(3)$ is

$$2\tau = \frac{2}{k}\sum_{l=1}^{k} \log_{\theta_l}(\gamma' y_l) \otimes \theta_l' \Big|_{\mathrm{so}(3)}.$$

Thus, $\mathrm{d}_\gamma \mathsf{s}$ vanishes iff τ is a symmetric matrix. But modulo $\mathrm{N}_1(\mathrm{SO}(3)) = \mathrm{sym}(\mathbf{E}^3)$ one has

$$\tau = \gamma'\nu, \qquad \text{where } \nu = \frac{1}{k}\sum_{l=1}^{k} \frac{\alpha_l}{\sin\alpha_l} y_l \otimes \theta_l' \tag{27}$$

so one concludes that the first-order condition is that γ is the orthogonal projection of ν onto SO(3). This is similar to the least-squares condition (19) — except that in (27) the matrix ν is a function of γ through the angles α_l. However, if one writes $y_l = \gamma'\theta_l + \epsilon r_l$, and expands the matrix ν in the small parameter ϵ, then one has $\nu = \epsilon\nu_1 + O(\epsilon^2)$ and ν_1 formally is the same as ν in (18). In other words, the intrinsic-distance regressor is a perturbation of the least-squares regressor.

In this example, $\mathrm{ev}_l(\gamma) = \gamma(\theta_l)$ so the map $\mathrm{ev}_l : \mathrm{SO}(3) \to S^2$ is the canonical projection map. In particular, this map has vanishing hessian – $\nabla \mathrm{d}\, \mathrm{ev}_l = 0$ – so proposition 3.8 implies that a smooth solution to the first-order condition $\tau \equiv 0 \bmod \mathrm{sym}(\mathbf{E}^3)$ is a local minimiser of s. Moreover, one knows

Lemma 3.10. *Let $\mathfrak{N}$ be the set of γ at which s is not differentiable. Then $\mathfrak{N}$ is a union of translates of subgroups isomorphic to* SO(2).

If γ is a local minimum point of s, then s is differentiable at γ. In particular, the regression estimator γ satisfies the property that

$$\gamma'\nu \equiv 0 \qquad \mod \mathrm{N}_1(\mathrm{SO}(3))$$

where ν is defined in (27).

Proof. Since $\mathbf{dist}(y, x)$ is differentiable in x on the set $S^2 - \{-y\}$, s is differentiable at γ iff $\gamma(\theta_l) \neq -y_l$ for all l. Thus, if s is not differentiable at γ, then there is a y_{l_0} such that $\gamma(\theta_{l_0}) = -\theta_{l_0} = y_{l_0}$. If γ_l is some solution to $\gamma(\theta_l) = -y_l$, then the set of all solutions to the latter is $\gamma_l \cdot \mathrm{stab}(\theta_l)$, which is a translate of a group isomorphic to SO(2). Thus $\mathfrak{N} = \cup_{l=1}^{k} \gamma_l \cdot \mathrm{stab}(\theta_l)$.

Let γ be a local minimum point of s. Assume that $\gamma \in \mathfrak{N}$. Without loss of generality, it can be assumed that there is an $l_0 > 0$ such that $\gamma(\theta_l) = -y_l$ (resp. $\gamma(\theta_l) \neq -y_l$) for $l \leq l_0$ (resp. $l > l_0$).

Let s_0 (resp. s_1) be the part of s contributed for $l \leq l_0$ (resp. $l > l_0$). Then,

$$\mathsf{s}_0(\gamma) = l_0\pi^2, \qquad d_\gamma \mathsf{s}_1 = 0.$$

Moreover, if $y, x \in S^2$, then since $y, -y, x$ is a degenerate triangle, $\mathbf{dist}(y, x) = \mathbf{dist}(y, -y) - \mathbf{dist}(-y, x) = \pi - \mathbf{dist}(-y, x)$. Therefore, one knows that

$$\begin{aligned}
\mathsf{s}_0(\hat\gamma) &= \sum_{l \leq l_0} (\pi - \mathbf{dist}(-y_l, \hat\gamma(\theta_l)))^2 \\
&= l_0\pi^2 - 2\pi \sum_{l \leq l_0} \mathbf{dist}(\gamma(\theta_l), \hat\gamma(\theta_l)) + O(|\gamma - \hat\gamma|^2), \\
\mathsf{s}_1(\hat\gamma) &= \mathsf{s}_1(\gamma) + O(|\gamma - \hat\gamma|^2).
\end{aligned}$$

Since the orbit map $\mathrm{ev}_l : \mathrm{SO}(3) \to S^2$ is a riemannian submersion, there are $\hat\gamma$ such that, for a fixed l, $\mathbf{dist}(\gamma(\theta_l), \hat\gamma(\theta_l)) = \mathbf{dist}(\gamma, \hat\gamma)$. This implies that s_0 decreases along $\hat\gamma$ more than s_1 increases. But γ is a local minimum. Absurd. Therefore, if γ is a local minimum, then s is differentiable at γ. □

Let $R_j(s)$ be the counterclockwise rotation of $\mathbf{E}^3$ by s radians in the plane orthogonal to the j-th standard basis vector. Elements of SO(3) may be parameterised in terms of '3-1-3' Euler angles: $\gamma = R_3(a)R_1(b)R_3(c)$ where $a, c \in [0, 2\pi]$ and $b \in [0, \pi]$ [2]. In figure 3, one has an empirical distribution of the regressor $\hat{\gamma} = \hat{\gamma}(\mathbf{y})$ in Euler angles. For $k = 100$ design points θ_l, drawn from the uniform distribution on S^2, $y_l = u_l/|u_l|$ where $u_l = \gamma\theta_l + \sigma \cdot \epsilon_l$ and ϵ_l is an i.i.d. gaussian in $\mathbf{E}^3$. $N = 1000$ draws are made and the first-order condition (27) is numerically solved for $\sigma = 0.1$ to 0.9 in increments of 0.1. All computations are performed in Octave [5]. The starting point for the numerical solution of (27) is provided by the orthogonal projection of $\sum_{l=1}^{k} y_l \otimes \theta_l'$ onto SO(3).

Figure 4 shows the histograms of the normalised empirical distributions of the Euler angles of the regressor $\hat{\gamma} = \hat{\gamma}(\mathbf{y})$ and reports the Kolmogorov-Smirnov p-value for normality. The normalised Euler angles are of the form $\xi = C^{-1}\mathbf{x}$, where $\mathbf{x}$ is the regressor's Euler angle, and the sample covariance matrix of $\mathbf{x}$ is CC'.

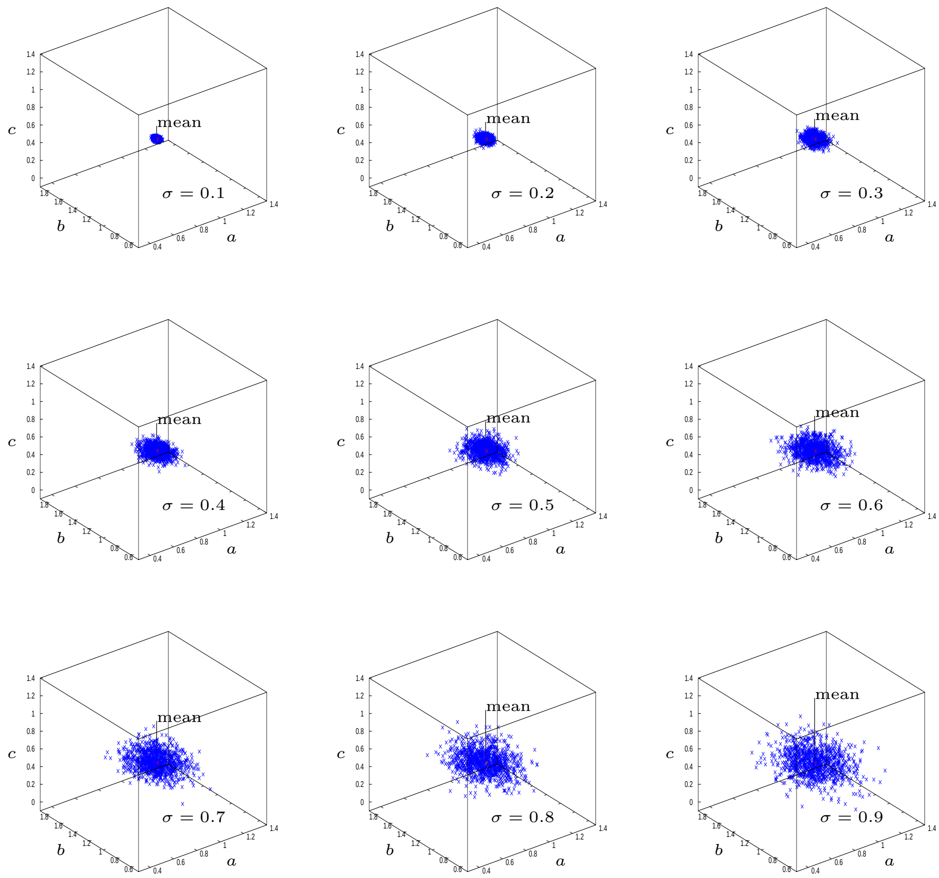


Figure 3. The empirical distribution of the regressor $\hat{\gamma} = \hat{\gamma}(\mathbf{y})$. See text for further information.

3.2. **A bayesian approach.** Let $\mathbf{y} = (y_1, \ldots, y_k) \in \Phi = \Lambda^k$ and let $\lambda(\gamma)\,\mathrm{d}\gamma$ be a bayesian prior on Γ (Γ is only assumed to be a smooth submanifold of $C^\infty(\Theta, \Lambda)$

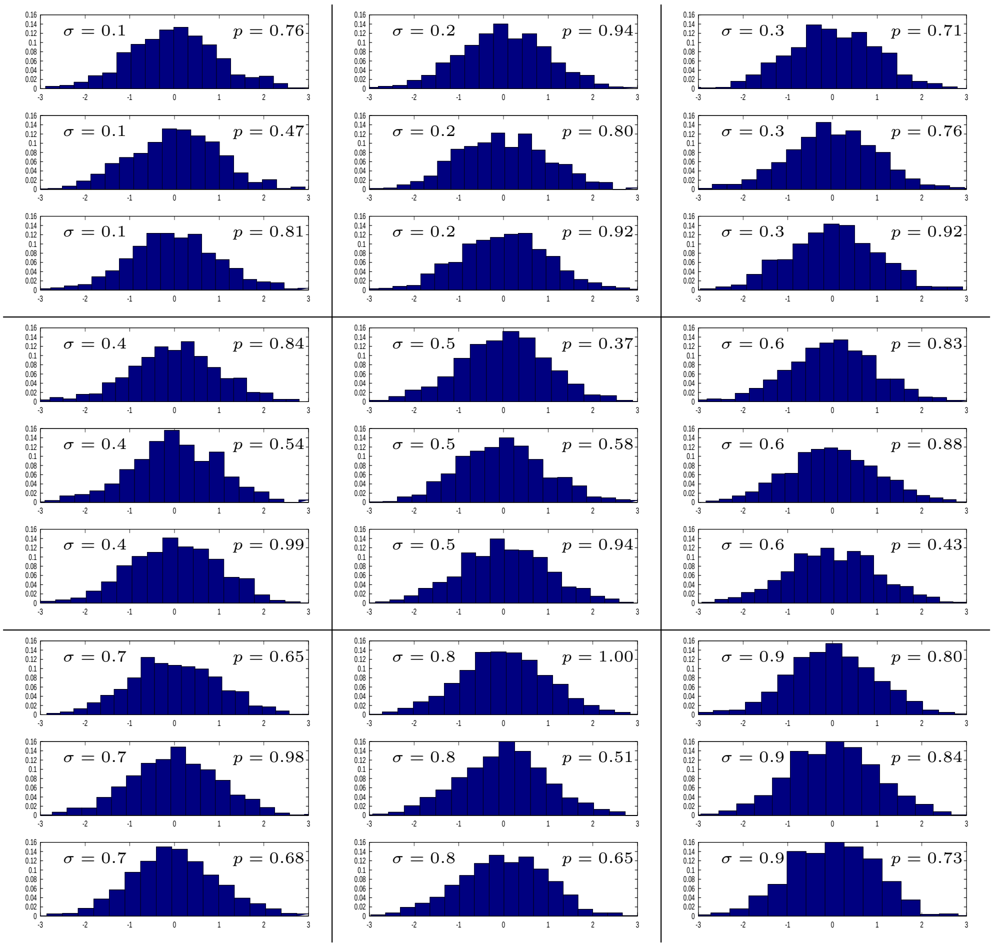


FIGURE 4. The histogram of the normalised deviations from the mean of the regressor $\hat{\gamma} = \hat{\gamma}(\mathbf{y})$. The p-value for the Kolmogorov-Smirnov test of normality is reported.

at this point). Let $\ell : \Gamma \times \Gamma \to \mathbf{R}$ be a loss function as defined in the introduction to section 3 and assume that $f(\mathbf{y}|\gamma) = \prod_i f(y_i|\gamma(\theta_i))$ is the conditional density of $\mathbf{y}$. The bayesian risk of $\hat{\gamma} \in \Gamma$ is then

$$\mathcal{R}(\hat{\gamma}) = \int_{\mathbf{y}\in\Phi} \int_{\gamma\in\Gamma} \ell(\hat{\gamma}, \gamma)\, f(\mathbf{y}|\gamma)\, \lambda(\gamma)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\gamma.$$

One can define quantities

$$\mu(\mathbf{y}) = \int_{\gamma\in\Gamma} f(\mathbf{y}|\gamma)\, \lambda(\gamma)\, \mathrm{d}\gamma, \qquad \lambda(\gamma|\mathbf{y}) = \frac{f(\mathbf{y}|\gamma)\, \lambda(\gamma)}{\mu(\mathbf{y})},$$
$$\mathcal{R}(\hat{\gamma}|\mathbf{y}) = \int_{\gamma\in\Gamma} \ell(\hat{\gamma}, \gamma)\, \lambda(\gamma|\mathbf{y})\, \mathrm{d}\gamma$$

to arrive at

$$\mathcal{R}(\hat{\gamma}) = \int_{\mathbf{y}\in\Phi} \mathcal{R}(\hat{\gamma}|\mathbf{y})\, \mu(\mathbf{y})\, \mathrm{d}\mathbf{y},$$

where dependence on the design points θ_i has been omitted for notational compactness. Therefore, one can choose a bayesian estimator $\hat{\gamma}$ by minimising the posterior risk

$$g(\mathbf{y}) = \operatorname{argmin}\left\{\mathcal{R}(\hat\gamma|\mathbf{y}) \ :\ \hat\gamma \in \Gamma\right\}, \qquad\qquad g : \Phi \to \Gamma. \tag{28}$$

Since Γ is assumed to be compact, the bayesian estimator g is defined for all $\mathbf{y}$ and measurable. If, in addition, ℓ is smooth, then $\mathcal{R}(\hat\gamma|\mathbf{y})$ is smooth in $\hat\gamma$.

The following notation is useful in formulating the first-order necessary condition to determine $g(\mathbf{y})$. Let $\ell = \ell(\hat\gamma, \gamma)$ be a smooth function that is defined for all pairs of maps $\hat\gamma, \gamma$ in $C^\infty(\Theta, \Lambda)$. One may view ℓ as a function of $\hat\gamma$ depending on the parameter γ. Let

$$\frac{\partial \ell}{\partial \hat\gamma} \in T^*_{\hat\gamma} C^\infty(\Theta, \Lambda) \tag{29}$$

be the 1-form defined by fixing γ and taking the derivative with respect to $\hat\gamma$. In this case, the map $\gamma \mapsto \frac{\partial \ell}{\partial \hat\gamma}$ is a smooth map from $C^\infty(\Theta, \Lambda)$ to the vector space $T^*_{\hat\gamma} C^\infty(\Theta, \Lambda)$.

Proposition 3.11. *Assume that the loss function ℓ is a smooth function on $C^\infty(\Theta, \Lambda) \times C^\infty(\Theta, \Lambda)$. Then,*

$$\frac{\partial}{\partial \hat\gamma} \mathcal{R}(\hat\gamma|\mathbf{y}) = \int_{\gamma \in \Gamma} \frac{\partial \ell}{\partial \hat\gamma}\, \lambda(\gamma|\mathbf{y})\, \mathrm{d}\gamma$$

If $\hat\gamma = g(\mathbf{y})$ is a bayesian estimator satisfying (28)*, then*

$$\frac{\partial}{\partial \hat\gamma} \mathcal{R}(\hat\gamma|\mathbf{y}) \text{ lies in } \mathrm{N}_{\hat\gamma}(\Gamma) \subset T^*_{\hat\gamma} C^\infty(\Theta, \Lambda).$$

The proof of this proposition is straightforward. One observes that the integral on the right-hand side is well defined since, by (29), one is integrating a smooth function which takes values in a single vector space.

3.2.1. *The squared-norm loss function.* As $(\Lambda, \mathbf{h})$ is assumed to be isometrically embedded in $(\mathbf{E}, \sigma)$ as in (14), one may define an L^2 metric on $C^\infty(\Theta, \Lambda)$ by means of the ambient euclidean structure

$$|\gamma|^2 = \int_{\theta \in \Theta} |\gamma(\theta)|^2\, \mathrm{d}\theta \qquad\qquad \forall \gamma \in C^\infty(\Theta, \Lambda). \tag{30}$$

A natural squared-norm loss function is then

$$\ell(\hat\gamma, \gamma) = |\hat\gamma - \gamma|^2 \qquad\qquad \forall \hat\gamma, \gamma \in C^\infty(\Theta, \Lambda). \tag{31}$$

(The requisite '$\jmath$'s in (30–31) are suppressed for simplicity).

Proposition 3.12. *Let $\Gamma \subset C^\infty(\Theta, \Lambda)$ be a smooth submanifold and the loss function ℓ be defined as in* (31)*. If $\hat\gamma = g(\mathbf{y})$ is a bayesian estimator as in* (28)*, then*

$$\jmath\bar\gamma := \int_{\gamma \in \Gamma} \jmath\gamma\, \lambda(\gamma|\mathbf{y})\, \mathrm{d}\gamma \qquad \textit{satisfies } \jmath\bar\gamma \in \mathrm{N}_{\hat\gamma}(\Gamma) \subset T^*_{\hat\gamma} C^\infty(\Theta, \mathbf{E}). \tag{32}$$

Remark 3.13. One considers $\jmath\bar\gamma$ to be a form in $T^*_{\hat\gamma} C^\infty(\Theta, \mathbf{E})$ and not in $T^*_{\hat\gamma} C^\infty(\Theta, \Lambda)$ in equation (32) due to the natural embedding $\Lambda \subset \mathbf{E}$.

Proof. In this case, the smoothness of $\mathcal{R}(\hat\gamma|\mathbf{y})$ in $\hat\gamma$ is immediate from the loss function. One computes that

$$\frac{\partial \ell}{\partial \hat\gamma} = \int_{\theta \in \Theta} \mathrm{d}_{\hat\gamma(\theta)}\jmath^*\ (\jmath\hat\gamma(\theta) - \jmath\gamma(\theta))\ \mathrm{d}\theta \qquad\qquad \in T^*_{\hat\gamma} C^\infty(\Theta, \mathbf{E}),$$

whence

$$\frac{\partial}{\partial \hat\gamma} \mathcal{R}(\hat\gamma|\mathbf{y}) = \int_{\theta \in \Theta} \mathrm{d}\theta\, \mathrm{d}_{\hat\gamma(\theta)}\jmath^* \left\{ \int_{\gamma \in \Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y})\ (\jmath\hat\gamma(\theta) - \jmath\gamma(\theta)) \right\}. \tag{33}$$

Proposition 3.11 shows that the left-hand side of (33) lies in $\mathrm{N}_{\hat\gamma}(\Gamma)$ if $\hat\gamma$ is a bayesian estimator. Define $\xi \in T^*_{\hat\gamma} C^\infty(\Theta, \mathbf{E})$ by

$$\xi(\theta) = \mathrm{d}_{\hat\gamma(\theta)}\jmath^* \left\{ \int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y})\ (\jmath\hat\gamma(\theta) - \jmath\gamma(\theta)) \right\} \qquad \forall\theta\in\Theta.$$

One observes that the right-hand side of (33) vanishes on $T_{\hat\gamma}\Gamma$ if ξ vanishes on $T_{\hat\gamma}\Gamma$, and ξ vanishes on $T_{\hat\gamma}\Gamma$ if ξ vanishes, *i.e.*, if

$$\jmath\hat\gamma(\theta) \equiv \int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y})\, \jmath\gamma(\theta) \qquad \mathrm{mod}\, \mathrm{N}_{\hat\gamma}(\Gamma)_\theta,$$

where $\mathrm{N}_{\hat\gamma}(\Gamma)_\theta$ is the subspace of $T^*_{\hat\gamma(\theta)}\mathbf{E}$ generated by elements of $\mathrm{N}_{\hat\gamma}(\Gamma)$ (which are sections of $\hat\gamma^* T^*\mathbf{E}$) evaluated at θ. Therefore, one obtains

$$\jmath\hat\gamma \equiv \int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y})\, \jmath\gamma \qquad \mathrm{mod}\, \mathrm{N}_{\hat\gamma}(\Gamma),$$

which proves the proposition. □

3.2.2. *Estimation of Linear Maps.* Assume that both Θ and Λ are isometrically embedded in euclidean spaces $\mathbf{E}_0$ and $\mathbf{E}_1$ respectively. Let $\Gamma \subset \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$ be a manifold of linear maps that maps Θ to Λ. Inspection of the right-hand side of (32) shows that $\bar\gamma$ is itself the restriction of a linear map to Θ, so the bayesian estimator $\hat\gamma$ can be described using only the geometry of $\mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$.

Define a positive semi-definite quadratic form on $\mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$ by

$$\langle\langle \alpha, \beta \rangle\rangle = \mathrm{Tr}\,(\alpha' \cdot \beta \cdot \tau) \qquad \forall \alpha, \beta \in \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1),$$

where

$$\tau = \int_{\theta\in\Theta} \jmath(\theta) \otimes \jmath(\theta)'\, \mathrm{d}\theta \in \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_0).$$

The first-order condition (32) implies that the bayesian estimator $\hat\gamma$ satisfies

$$\hat\gamma \cdot \tau \equiv \bar\gamma \cdot \tau \qquad \mathrm{mod}\, \mathrm{N}_{\hat\gamma}(\Gamma),$$

where $\mathrm{N}_{\hat\gamma}(\Gamma)$ is the normal space to $T_{\hat\gamma}\Gamma$ in $\mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$.

Proposition 3.14. *Let* $\Gamma \subset \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$ *be a submanifold and the loss function* ℓ *be defined as in* (30). *Suppose that* Θ *spans* $\mathbf{E}_0$. *If* $\hat\gamma = g(\mathbf{y})$ *is a bayesian estimator as in* (28), *then the linear transformation*

$$\bar\gamma := \int_{\gamma\in\Gamma} \gamma\, \lambda(\gamma|\mathbf{y})\, \mathrm{d}\gamma \qquad \textit{satisfies } \bar\gamma \equiv \hat\gamma \bmod \mathrm{N}_{\hat\gamma}(\Gamma) \cdot \tau^{-1}. \tag{34}$$

Proof. The only thing that remains to prove is that τ is non-degenerate if Θ spans $\mathbf{E}_0$. If $v \in \mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$ and

$$0 = \langle\langle v, v \rangle\rangle = \int_{\theta\in\Theta} |v\cdot\theta|^2\, \mathrm{d}\theta, \qquad \text{then } \Theta \subseteq \ker v.$$

Therefore, $\mathbf{E}_0 = \mathrm{span}\,\Theta \subseteq \ker v$, so $v = 0$. □

Remark 3.15. Let Θ be the unit sphere in $\mathbf{E}_0$. One computes that τ is a scalar multiple of the identity matrix, whence condition (34) is simply that $\hat\gamma$ is the orthogonal projection onto Γ of $\bar\gamma$.

Let ℓ be the loss function on Γ induced by the inner product on $\mathrm{Hom}(\mathbf{E}_0, \mathbf{E}_1)$:

$$\ell(\hat\gamma, \gamma) = |\hat\gamma - \gamma|^2 = \mathrm{Tr}\,((\hat\gamma - \gamma)'(\hat\gamma - \gamma)) \qquad \forall \hat\gamma, \gamma \in \Gamma.$$

When $\mathbf{E}_0 = \mathbf{E}_1$ and $\Gamma \subset \mathrm{O}(\mathbf{E})$, the loss function simplifies to $2s - 2\,\mathrm{Tr}\,(\hat\gamma'\gamma)$ where $s = \dim \mathbf{E}$.

3.2.3. *The intrinsic distance loss function.* Let $\hat\gamma, \gamma \in C^\infty(\Theta, \Lambda)$ be smooth maps between the riemannian manifolds $(\Theta, \mathbf{g})$ and $(\Lambda, \mathbf{h})$. For each $\theta \in \Theta$, let $w(\theta) \in T_{\hat\gamma(\theta)}\Lambda$ be a tangent vector to a shortest geodesic $c(s) = \exp_{\hat\gamma(\theta)}(s \cdot w(\theta))$ such that $c(1) = \gamma(\theta)$.

If $\gamma(\theta)$ does not lie in the cut locus of $\hat\gamma(\theta)$, the tangent vector $w(\theta)$ is uniquely defined and one may unambiguously write $w(\theta) = \log_{\hat\gamma(\theta)}(\gamma(\theta))$. It is apparent that there are measurable maps $\theta \mapsto w(\theta)$, and this map is smooth off the above-mentioned set of "bad" points. In particular, if the graph of γ lies in a tubular neighbourhood of the graph of $\hat\gamma$, then the map w is a uniquely defined, smooth map.

Let $\mathfrak{C} = \mathfrak{C}_{\gamma,\hat\gamma} \subset \Theta$ be the set of points θ such that $\gamma(\theta)$ lies in the cut locus of $\hat\gamma(\theta)$. If the measure of $\mathfrak{C}$ is zero, then compactness of Λ implies that w is square-integrable. Therefore, one may define a one-form $\omega = \omega_{\gamma,\hat\gamma} \in T^*_{\hat\gamma}C^\infty(\Theta, \Lambda)$ by

$$\langle \omega, v\rangle = \int_{\theta\in\Theta} \mathrm{d}\theta \cdot \mathbf{h}(w(\theta), v(\theta))_{\hat\gamma(\theta)}.$$

for each $v \in T_{\hat\gamma}C^\infty(\Theta, \Lambda)$.

Proposition 3.16. *Let*

$$\ell(\hat\gamma, \gamma) = \frac{1}{2}\int_{\theta\in\Theta} \mathrm{d}\theta \cdot \mathbf{dist}(\hat\gamma(\theta), \gamma(\theta))^2,$$

where **dist** *is the riemannian distance function of* $(\Lambda, \mathbf{h})$. *If* $\mathfrak{C}_{\gamma,\hat\gamma}$ *has measure zero and* Λ *is compact, then* $\frac{\partial \ell}{\partial\hat\gamma}$ *exists at* $(\gamma, \hat\gamma)$ *and equals*

$$\frac{\partial \ell}{\partial\hat\gamma} = -\omega_{\gamma,\hat\gamma}.$$

Proof. Let $\hat\gamma_t$ be a curve of smooth maps such that $\hat\gamma_{t=0} = \hat\gamma$ and $v = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \hat\gamma_t$. Let $\theta \in \Theta - \mathfrak{C}$ be fixed, and let $c_t(s)$ be the minimal geodesic from $\gamma(\theta)$ to $\hat\gamma_t(\theta)$. It is clear from figure 2 that the derivative of $\frac{1}{2}\,\mathbf{dist}(\hat\gamma_t(\theta), \gamma(\theta))^2$ is $\langle c'(1), v(\theta)\rangle$ where $c = c_0$. From the above discussion, it is clear that $c'(1) = -w(\theta)$.

If $\mathfrak{C}_{\gamma,\hat\gamma}$ has zero measure, then the discussion above shows that $\frac{\partial \ell}{\partial\hat\gamma}$ exists at $(\gamma, \hat\gamma)$ and equals $-\omega_{\gamma,\hat\gamma}$. ∎

Let $\mathfrak{C}_\Gamma \subset \Gamma$ be the set of maps $\hat\gamma$ such that $\int_{\gamma\in\Gamma}\int_{\mathfrak{C}_{\gamma,\hat\gamma}} \mathrm{d}\theta\,\mathrm{d}\gamma = 0$. By proposition 3.16, if $\hat\gamma \in \mathfrak{C}_\Gamma$, then $\frac{\partial \ell}{\partial\hat\gamma}$ exists for almost all $(\gamma, \hat\gamma) \in \Gamma \times \{\hat\gamma\}$. The following theorem is a consequence of Proposition 3.16 and Fubini's theorem.

Theorem 3.1. (1) *If* $\hat\gamma \in \mathfrak{C}_\Gamma$ *and* $\hat\gamma = g(\mathbf{y})$ *is a bayesian estimator as in Proposition 3.11, then* $\hat\gamma$ *satisfies*

$$\int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y}) \left\langle \log_{\hat\gamma(\theta)}(\gamma(\theta)), v(\theta)\right\rangle = 0 \qquad\qquad \forall v \in T_{\hat\gamma}\Gamma.$$

(2) *In particular, if* $\hat\gamma \in \mathfrak{C}_\Gamma$ *satisfies the equation*

$$\int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\gamma|\mathbf{y})\, \log_{\hat\gamma(\theta)}(\gamma(\theta)) = 0 \qquad\qquad \bmod \mathrm{N}_{\hat\gamma}(\Gamma)_\theta$$

for a.a. $\theta \in \Theta$, *then* $\hat\gamma$ *is a candidate for a bayesian estimator as in Proposition 3.11.*

Example 3.17. Let us examine an application of both parts of Theorem 3.1. Let $\Theta = \Lambda = S^2 \subset \mathbf{E}^3$ be the unit sphere and let $\Gamma = \mathrm{SO}(3)$ be the group of orientation-preserving isometries of S^2 with normalised Haar measure $\mathrm{d}\gamma$.

(1) Since Γ is a transitive group of isometries, the logarithm function is Γ-equivariant, so part (1) of 3.1 implies that

$$w_{\hat{\gamma}}(\theta|\mathbf{y}) = \int_{\gamma\in\Gamma} \mathrm{d}\gamma\, \lambda(\hat{\gamma}\gamma|\mathbf{y})\, \log_\theta(\gamma\theta)$$

must integrate to zero on S^2 against any vector field of the form $v(\theta) = \xi\cdot\theta$, $\xi \in \mathrm{so}(3)$. One uses the fact that $\log_\theta(\gamma\theta) = \frac{\alpha}{\sin\alpha} \times (\gamma\theta - \langle\gamma\theta,\theta\rangle\cdot\theta)$ (*c.f.* 26) to compute that

$$\begin{aligned}\langle w_{\hat{\gamma}}, v\rangle &= \int_{\theta\in S^2}\int_{\gamma\in\mathrm{SO}(3)} \mathrm{d}\theta\,\mathrm{d}\gamma\,\lambda(\hat{\gamma}\gamma|\mathbf{y})\,\frac{\alpha}{\sin\alpha}\,\langle\gamma\theta,\xi\theta\rangle\,, \qquad \cos\alpha = \langle\gamma\theta,\theta\rangle \\ &= -\frac{1}{3}\times\,\mathrm{Tr}\left(\tau(\hat{\gamma})\cdot\xi\right),\end{aligned}$$

where

$$\tau(\hat{\gamma}) = 3\int_{\theta\in S^2}\int_{\gamma\in\mathrm{SO}(3)} \mathrm{d}\theta\,\mathrm{d}\gamma\,\lambda(\hat{\gamma}\gamma|\mathbf{y})\,\frac{\alpha}{\sin\alpha}\,\gamma\theta\theta' \qquad (\theta' = \text{transpose of } \theta)$$

is defined analogous to (32). Since $\int_{\theta\in S^2}\mathrm{d}\theta\;\theta\otimes\theta' = \frac{1}{3}I$, if the weight $\alpha/\sin\alpha$ were identically 1, then $\bar{\gamma} = \hat{\gamma}\tau(\hat{\gamma})$ would coincide with that defined in (32). It follows that if $\hat{\gamma}$ equals the bayesian estimator $g(\mathbf{y})$, then $\tau(\hat{\gamma})$ must be symmetric. In other words, $\hat{\gamma}$ is the orthogonal projection of $\bar{\gamma}(\hat{\gamma}) = \hat{\gamma}\cdot\tau(\hat{\gamma})$ onto Γ, similar to (27).

(2) On the other hand, let us investigate condition (2) of Theorem 3.1. Let $\mathbf{x} : (S^2)^k \to \mathrm{SO}(3)$ be an equivariant map and let the joint conditional density of $\mathbf{y}$ be $f(\mathbf{y}|\gamma) = 1 + c\,\mathrm{Tr}\,(\gamma'\cdot\mathbf{x}(\mathbf{y}))$. Assume that the mean of γ' with respect to the bayesian prior $\lambda(\gamma)$ is zero. The posterior density $\lambda(\gamma|\mathbf{y})$ is therefore equal to $f(\mathbf{y}|\gamma)$.

To fix ideas, one may take $\mathbf{x}(\mathbf{y}) = \pi(\sum_{l=1}^k y_l\otimes\theta_l')$, where $\pi : \mathrm{gl}(3)\to\mathrm{SO}(3)$ is the orthogonal projection, and $\lambda(\gamma) = 1$ for all γ.

Let $\mathbf{e}\in S^2$ be a given point. Since Γ acts transitively, one can write $\theta = \alpha\cdot\mathbf{e}$ for some $\alpha\in\Gamma$. One therefore finds that the vanishing of $w_{\hat{\gamma}}(\theta|\mathbf{y})$ is equivalent to the vanishing of

$$\int_{\alpha\in\Gamma}\mathrm{d}\alpha\,\left|\int_{\gamma\in\Gamma}\mathrm{d}\gamma\,\lambda(\hat{\gamma}\alpha\gamma\alpha^{-1}|\mathbf{y})\,\log_{\mathbf{e}}(\gamma\cdot\mathbf{e})\right|^2. \tag{35}$$

If one introduces Euler angles on SO(3) relative to an orthonormal frame $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3=\mathbf{e}\}$, then one can write $\gamma_j = R_3(a_j)R_1(b_j)R_3(c_j)$ where $R_i(s)$ is a rotation in the plane orthogonal to $\mathbf{e}_i$ counterclockwise by angle s. The vanishing of (35) is equivalent to the vanishing of the multi-integral

$$\begin{aligned}\int_{[0,2\pi]^4\times[0,\pi]^2} &\frac{1}{64\pi^4}\,\mathrm{d}a_1\,\mathrm{d}a_2\,\mathrm{d}c_1\,\mathrm{d}c_2\,\mathrm{d}b_1\,\mathrm{d}b_2 \\ &\times\sin(b_1)\sin(b_2)\,b_1b_2\cos(a_1-a_2)\,\lambda(\hat{\gamma}\alpha\gamma_1\alpha^{-1}|\mathbf{y})\,\lambda(\hat{\gamma}\alpha\gamma_2\alpha^{-1}|\mathbf{y})\end{aligned} \tag{36}$$

for every $\alpha\in\mathrm{SO}(3)$.

Let the special orthogonal matrix $\alpha^{-1}\hat{\gamma}'\mathbf{x}\alpha$ be factorised as $R_3(x)R_1(y)R_3(z)$ in terms of Euler angles. Maxima computes the integral (36) to be $\pi^4c^2\sin(y)^2/256$ [12]. Therefore, the integral vanishes for all α iff $\hat{\gamma}'\mathbf{x} = I$, *i.e.* $\hat{\gamma} = \mathbf{x}$.

References

[1] R. Abraham and J. E. Marsden. *Foundations of Mechanics.* Addison-Wesley. 1978.

[2] V. I. Arnold. *Mathematical methods of classical mechanics.* 2nd ed. Springer-Verlag. 1989.

[3] L. Butler and B. Levit. A Bayesian approach to the estimation of maps between riemannian manifolds. *Mathematical Methods of Statistics.* 16(4):1–17. 2007.

[4] Y. Chikuse. *Statistics on Special Manifolds.* Lecture Notes in Statistics, 174. 2003.

[5] J. W. Eaton *GNU Octave Manual.* 2nd edition. Network Theory Limited. 2008.

[6] J. Eells and L. Lemaire. Selected topics in harmonic maps. C.B.M.S. Regional Conference Series. AMS. 1983.

[7] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, 1978.

[8] D. Husemoller. *Fiber Bundles.* 3rd edition. Springer-Verlag, New York. 1994.

[9] R. A. Horn and C. R. Johnson Topics in matrix analysis. Corrected reprint of the 1991 original. Cambridge University Press, Cambridge, 1994.

[10] P.T. Kim Decision theoretic analysis of spherical regression, *Journal of Multivariate Analysis*, 38:233–240. 1991.

[11] V. P. Maslov. *Théorie des perturbations et méthodes asymptotiques.* 1972.

[12] Maxima.sourceforge.net *Maxima CAS version 5.18.1.* `http://maxima.sourceforge.net/`. April 2009.

[13] J. Milnor *Morse Theory.* Princeton University Press. 1963.

[14] B. O'Neill. *Semi-riemannian geometry.* Academic Press. 1983.

[15] W. Rudin. *Functional Analysis*, McGraw-Hill. 1991.

LB: School of Mathematics, 6214 James Clerk Maxwell Building, The University of Edinburgh, Edinburgh, UK, EH9 3JZ BL: Department of Mathematics and Statistics Queen's University, Kingston, ON, Canada, K7L 3N6

E-mail address: `l.butler@ed.ac.uk, blevit@mast.queensu.ca`