

# A BAYESIAN APPROACH TO THE ESTIMATION OF MAPS BETWEEN RIEMANNIAN MANIFOLDS

LEO T. BUTLER AND BORIS LEVIT

ABSTRACT. Let  $\Theta$  be a smooth compact oriented manifold without boundary, imbedded in a euclidean space  $\mathbf{E}^s$ , and let  $\gamma$  be a smooth map of  $\Theta$  into a Riemannian manifold  $\Lambda$ . An unknown state  $\theta \in \Theta$  is observed via  $X = \theta + \epsilon\xi$  where  $\epsilon > 0$  is a small parameter and  $\xi$  is a white Gaussian noise. For a given smooth prior  $\lambda$  on  $\Theta$  and smooth estimators  $g(X)$  of the map  $\gamma$  we derive a second-order asymptotic expansion for the related Bayesian risk. The calculation involves the geometry of the underlying spaces  $\Theta$  and  $\Lambda$ , in particular, the integration-by-parts formula. Using this result, a second-order minimax estimator of  $\gamma$  is found based on the modern theory of harmonic maps and hypo-elliptic differential operators. This is the final, corrected version as it appears in *Mathematical Methods of Statistics*. 16(4):1–17, 2007.

## 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 to construct asymptotically minimax estimators along with the least favourable Bayesian priors.

The use of differential geometry in optimal statistical estimation has a long history, as documented in a recent article “Information geometry” on Wikipedia, for example. Early applications of differential geometry to the derivation of second-order asymptotic properties of the maximum likelihood estimates are summarized in [1]. However, a rigorous approach to second-order optimality requires a decision-theoretical framework. This approach was developed in [6, 2, 7] and a number of subsequent publications.

In some cases, one is interested in the second-order optimal estimation of a given function of parameters. For an early application of this approach see [5]. As a general rule, such problems require more sophisticated differential-geometric techniques such as the theory of harmonic maps and hypoelliptic differential operators [3, 4].

Consider the following situation:  $\mathbf{E}$  is a real  $s$ -dimensional vector space with inner product  $\sigma$  and  $\Theta$  (resp.  $\Lambda$ ) 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  $\iota$ , and  $\Theta \xrightarrow{\gamma} \Lambda$  is a

---

*Date:* June 4, 2008.

*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.

smooth map. Smooth means infinitely differentiable. These data are summarized by the diagram

$$\begin{array}{ccc} (\mathbf{E}, \sigma) & & \\ \uparrow \iota & \searrow g & \\ (\Theta, \mathbf{g}) & \xrightarrow{\gamma} & (\Lambda, \mathbf{h}), \end{array}$$

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

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

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

$$(1) \quad R_\epsilon(g, \theta) = \int_{x \in \mathbf{E}} \mathbf{dist}(g(x), \gamma(\theta))^2 \psi_\epsilon(x - \iota(\theta)) dx,$$

where  $|\bullet|$  is the norm on  $\mathbf{E}$  induced by  $\sigma$ ,  $\psi_\epsilon(u) = \exp(-|u|^2/2\epsilon^2)/(2\pi\epsilon^2)^{\frac{s}{2}}$  and  $dx$  is the volume form on  $\mathbf{E}$  induced by  $\sigma$ .<sup>2</sup> Define the associated minimax risk

$$(2) \quad r_\epsilon(\Theta) = \inf_g \sup_{\theta \in \Theta} R_\epsilon(g, \theta).$$

**1.1. Results.** The present paper takes a Bayesian approach to the problem of determining the asymptotically minimax estimator  $g$ . In Bayesian statistics, the point  $\theta$  is viewed as a random variable with a prior distribution  $\lambda(\theta)d\theta$  where  $\int_{\theta \in \Theta} \lambda(\theta) d\theta = 1$  ( $d\theta = d\nu_{\mathbf{g}}$  is the riemannian volume of  $(\Theta, \mathbf{g})$ ). The *Bayesian risk* of a map  $g$  is

$$(3) \quad R_\epsilon(g; \lambda) = \int_{\theta \in \Theta} \int_{x \in \mathbf{E}} \mathbf{dist}(g(x), \gamma(\theta))^2 \lambda(\theta) \psi_\epsilon(x - \iota(\theta)) dx d\theta.$$

A *Bayes estimator*  $g : \mathbf{E} \rightarrow \Lambda$  is a map which minimizes the Bayesian risk over all maps. In Theorem 4.1, an expansion in  $\epsilon$  of the Bayes estimator  $\tilde{g}_\epsilon$ , for a fixed Bayesian prior, is computed. The constant term in  $\tilde{g}_\epsilon$  is  $\gamma \circ \pi$ , where  $\pi : N\Theta \rightarrow \Theta$  is the projection map of the normal bundle of  $\Theta \subset \mathbf{E}$ . The order- $\epsilon^2$  term in  $\tilde{g}_\epsilon$  is composed of two parts: the first part is independent of the Bayesian prior and its contribution is to reduce the energy of  $\gamma$ ; the second part is due to the gradient of the prior  $\lambda$  and it tries to move the estimator in the direction which maximizes  $\lambda$ . Theorem 4.1 also computes the Bayesian risk  $R_\epsilon(\tilde{g}_\epsilon; \lambda)$  of  $\tilde{g}_\epsilon$  up to  $O(\epsilon^6)$ .

The results of Theorem 4.1 are used to obtain “the” optimal Bayesian prior. There arises a number of interesting problems of a statistical nature in this regard: foremost is the problem of deciding what should be the *flat* Bayesian prior. Given a flat Bayesian prior, it is shown that the 2nd-order optimal Bayesian prior satisfies an eigenvalue problem. This leads to a second difficulty: in general, the leading term in the Bayesian risk is determined by  $|d\gamma|^2$  and is therefore largely independent

<sup>1</sup>By convention, the covariance operator is the induced inner product on the dual vector space  $\mathbf{E}^*$ . If we regard  $\sigma$  as a linear isomorphism of  $\mathbf{E} \rightarrow \mathbf{E}^*$ , then the covariance operator is the inverse linear isomorphism  $\mathbf{c} = \sigma^{-1} : \mathbf{E}^* \rightarrow \mathbf{E}$ .

<sup>2</sup>One can introduce a  $\sigma$ -orthonormal coordinate system  $x_i$  on  $\mathbf{E}$ . In this case,  $|x|^2 = \sum_i x_i^2$  and  $dx = dx_1 \wedge \cdots \wedge dx_s$ .

of the Bayesian prior. Thus, Theorems 5.1–5.4 give several criteria for second-order optimal Bayesian priors.

**1.2. Note to Reader.** The present paper resulted from the work of B. Levit [6, 2, 7, 5]. This work, and early drafts of the present paper, were done largely in local coordinates using Taylor series. This proved to be both daunting, difficult and unsatisfying because we were forced to assume that  $\Lambda$  was isometrically embedded in some euclidean space and use the ambient distance function. Paradoxically, these computations produced estimators which did not take values in  $\Lambda$ .

The problem with the answer these computations produced was obvious, the reason for the problem was less so. The ultimate reason is that the Taylor series expansion of a function is not a tensorial, or intrinsic, object. Rather, a Taylor series depends on the geometry of the domain and range of the function: it is, in other words, a *geometric* object. It is easy to see why this is: a Taylor series requires the notion of a second derivative to be defined, but it is well-known that a second derivative can be defined only with the aid of an affine connection—a geometric object. Calculations with Taylor series in local coordinates masked this fact and completely mislead us.

This is a roundabout way of explaining the extensive geometric formalism used in the present paper. We hope that the reader will remember that behind this formalism is a simple aim: to define a Taylor series in a rigorous and useful way. As a by-product, the answers that result can be stated in a much more compact way.

This paper proceeds as follows: in section 2, a theory of Taylor-Maclaurin series is developed for riemannian manifolds and several useful curvature and integration-by-parts formulas are developed that are used in subsequent sections; section 3 discusses the existence and uniqueness of a Bayes estimator; section 4 utilizes the theory developed in section 2 to expand the Bayesian risk functional and determined the Bayes estimator up to  $O(\epsilon^6)$ ; section 5 develops criteria for second-order optimal Bayesian priors in terms of the sub-laplacian of a naturally constructed sub-riemannian structure; section 6 computes the examples where  $\gamma$  is a riemannian immersion or submersion, which includes the cases where  $\gamma$  the identity map of  $\Theta$  and the inclusion map  $\iota$  of  $\Theta \subset \mathbf{E}$ .

Throughout, it is assumed that  $\Theta, \Lambda$  are a compact, connected, boundaryless smooth manifolds.

## 2. MACLAURIN SERIES

This section develops a theory of Maclaurin series of a map between riemannian manifolds, then it exposes some useful formulas from riemannian geometry that are used in subsequent sections. First, it is useful recall some constructions.

**2.1. Induced metrics.** Let  $X$  and  $Y$  be real inner-product spaces. The vector space of linear maps  $X \rightarrow Y$  is denoted by  $\text{Hom}(X; Y)$ . Define the inner product of linear maps  $A, B \in \text{Hom}(X; Y)$  by

$$\langle A, B \rangle := \sum_i \langle A.e_i, B.e_i \rangle = \text{Tr}(A'B),$$

where  $e_i$  is an orthonormal base of  $X$ . The Hilbert-Schmidt norm of a linear map is defined in the natural way from this inner product. By construction, if  $x \in X$ , then  $|A.x| \leq |A||x|$ .

We can make  $X^{\otimes n}$  (the  $n$ -fold tensor product of  $X$  with itself) into a real inner-product space by defining

$$\langle a_1 \otimes \cdots \otimes a_n, b_1 \otimes \cdots \otimes b_n \rangle = \langle a_1, b_1 \rangle \cdots \langle a_n, b_n \rangle,$$

for  $a_i, b_i \in X$  and extending by bi-linearity. The previous construction makes  $\text{Hom}(X^{\otimes n}, Y)$  into a real inner-product space. We will use these constructions henceforth without further comment.

**2.2. Maclaurin series.** Let  $(M, g)$  and  $(N, h)$  be riemannian manifolds without boundary and let  $M \xrightarrow{\phi} N$  be a smooth map. For  $x \in M$  and  $y = \phi(x)$ , let  $T_x M$  (resp.  $T_y N$ ) be the tangent space to  $M$  (resp.  $N$ ) at  $x$  (resp.  $y$ ). The exponential map  $\exp_x$  (resp.  $\exp_y$ ) of  $g$  (resp.  $h$ ) is injective on a disk of radius  $a = a(x)$  in  $T_x M$  (resp.  $s = s(y)$  in  $T_y N$ ), while the tangent map of  $\phi$  at  $x$ ,  $d_x \phi$ , maps a disk of radius  $t$  into a disk of radius  $t \times |d_x \phi|$ . If  $r = r(x)$  is defined to be the minimum of  $s(y)/|d_x \phi|$  and  $a(x)$ , then there is a commutative diagram<sup>3</sup>

$$\begin{array}{ccc} (T_x^r M, g_x) & \xrightarrow{\varphi} & (T_y^s N, h_y) \\ \exp_x \downarrow & & \downarrow \exp_y \\ (M, g) & \xrightarrow{\phi} & (N, h), \end{array} \quad \text{i.e. } \varphi = \exp_y^{-1} \circ \phi \circ \exp_x,$$

where  $T_x^r M$  (resp.  $T_y^s N$ ) is the disk radius  $r$  (resp.  $s$ ) in  $T_x M$  (resp.  $T_y N$ ) centred at 0. The map  $\varphi$  is a smooth map between open subsets of euclidean spaces that maps 0 to 0. Its Maclaurin series expansion is well-defined and can be written as

$$(4) \quad \varphi(v) = d\varphi(v) + \frac{1}{2} \nabla d\varphi(v, v) + \frac{1}{6} \nabla^2 d\varphi(v, v, v) + O(|v|^4),$$

for all  $v \in T_x M$ . The hessian  $\nabla d\varphi$  may be understood as the ordinary second derivative of a map between vector spaces, as can  $\nabla^2 d\varphi$ . However, Lemma 2.1 is essential and relates the derivatives of  $d\varphi$  to the covariant derivatives of  $d\phi$  [3].

**Lemma 2.1.** *Let  $v \in T_x M$ . Then  $\nabla^k d\varphi(v, \dots, v)|_0 = \nabla^k d\phi(v, \dots, v)|_x$  for all  $k \geq 0$  and all  $x \in M$ .*

*Proof.* Since  $\exp_y \circ \varphi = \phi \circ \exp_x$  on the open set  $T_x^r M$ , it follows that  $\nabla^k d(\exp_y \circ \varphi)|_0 = \nabla^k d(\phi \circ \exp_x)|_x$ . It suffices to show that the lefthand side equals  $\nabla^k d\varphi|_0$  and the righthand side equals  $\nabla^k d\phi|_x$  when each are evaluated at  $(v, \dots, v)$ . The chain rule, along with  $d_0 \exp_y = \text{id}_{T_y N}$ , shows that

$$\nabla^k d(\exp_y \circ \varphi)|_0 = \nabla^k d\varphi|_0 + T,$$

where  $T$  is a sum of terms which are composition of forms  $\nabla^l d\varphi$ ,  $\nabla^m d\exp_y$  with  $l, m < k$  and  $m \geq 1$ . It therefore suffices to show that

**Claim.**  $\nabla^m d\exp_y|_0 = 0$  for all  $m \geq 1$ .

Let  $v \in T_y N \equiv T_0(T_y N)$ , let  $c(t) = \exp_y(tv)$  be the unique geodesic passing through  $v$ , and let  $m_v(t) = tv$  be the multiplication-by- $v$  map. Since  $c(t) = \exp_x \circ m_v(t)$ ,  $v = dm_v(\partial_t)$  and  $\nabla dm_v = 0$ , we have that  $\nabla_{\dot{c}(t)} \dot{c}(t) = \nabla dc(\partial_t, \partial_t) = \nabla d\exp_x(v, v)$ . Thus  $\nabla d\exp_x(v, v)|_0 = 0$ .

In the general case, for  $m \geq 2$ ,  $\nabla^m d\exp_y(v, \dots, v)|_0 = \nabla_{\dot{c}(t)} (\cdots (\nabla_{\dot{c}(t)} \dot{c}(t)) \cdots)|_{t=0}$ . Since the innermost term vanishes identically in  $t$ , the whole expression vanishes.

<sup>3</sup>The functions  $a$  and  $s$  may be assumed to be smooth.

The proof for  $\phi \circ \exp_x$  is similar.  $\square$

**Lemma 2.2.** *For all  $v \in T_x^r M$ ,*

$$(5) \quad \phi \circ \exp_x(v) = \exp_{\phi(x)} \left( d\phi(v) + \frac{1}{2} \nabla d\phi(v, v) + \frac{1}{6} \nabla^2 d\phi(v, v, v) + O(r^4) \right).$$

**Remarks.** (1) In general, the exponential map of  $(N, h)$  is not a global diffeomorphism. Consequently,  $\varphi$  may not be globally well-defined and its Maclaurin series (5) need not converge globally. The conjugate points of  $\exp_y$  are obstructions to global convergence. If  $(N, h)$  is a simply-connected, non-positively curved manifold, then  $\exp_y$  is a global diffeomorphism and  $\varphi$  is globally defined. (2) If  $v \in T_x M$  is a gaussian with covariance operator  $\epsilon^2 g_x$ , then, since  $\varphi$  is defined on an open neighbourhood of 0 and  $\epsilon$  is small, its expected value is essentially well-defined and equals, by Lemma 2.1 and equation (4),

$$\frac{1}{2} \epsilon^2 \tau(\phi)_x + O(\epsilon^4)$$

where  $\tau(\phi)$  is the trace of the hessian  $\nabla d\phi$ . Riemannian geometers call  $\tau(\phi)$  the *tension field* of  $\phi$ . The tension field has an interesting interpretation: if one views

$$(6) \quad \phi \mapsto \int_M |d\phi|^2 d\nu_g \quad (d\nu_g = \text{riemannian volume form of } g)$$

as the energy of the map  $\phi$ , then  $\phi \mapsto -\tau(\phi)$  is the gradient of this functional. It is known that  $\tau$  is a semilinear elliptic differential operator that is analogous to the laplacian [3]. If one inspects the formula for the bayesian estimator  $\tilde{g}_\epsilon$  in Theorem 4.1, one observes that—neglecting the contribution of the prior  $\lambda$ —the contribution of the  $\frac{1}{2}\tau(\phi)$  is to move the estimate  $\gamma(\hat{\theta})$  in the direction which reduces the energy quickest. Indeed, if one includes both contributions, then their combination can also be viewed in this fashion, but the energy functional depends not on  $\mathbf{g}$  and  $\mathbf{h}$  but  $u \cdot \mathbf{g}$  and  $\mathbf{h}$  where the conformal factor  $u$  is a fractional power of  $\lambda$ .

**2.3. The Ricci Tensor.** This section provides the key inputs to the proofs of Lemma 4.7 and Proposition 4.8 by proving Lemmas 2.4, 2.5 and the integration-by-parts formula in Lemma 2.8. To do this, one must make an excursion into the riemannian geometry of some naturally occurring vector bundles. In this section,  $(M, g)$  and  $(N, h)$  are riemannian manifolds, possibly with boundary, and  $\phi : M \rightarrow N$  is a smooth map.

Let  $\text{Hom}(TM; \phi^*TN)$  be the vector bundle of fibre-linear maps between  $TM$  and  $\phi^*TN$ ; a fibre  $\text{Hom}(TM; \phi^*TN)_x, x \in M$  of  $\text{Hom}(TM; \phi^*TN)$  is the vector space of linear maps from  $T_x M$  to  $T_{\phi(x)} N$ . That is,  $\text{Hom}(TM; \phi^*TN)_x = \text{Hom}(T_x M; T_{\phi(x)} N)$ . One can view  $d\phi$  as a smooth section of  $\text{Hom}(TM; \phi^*TN)$ . There is a natural metric connection on  $\text{Hom}(TM; \phi^*TN)$ , which is denoted by  $\nabla$  or  $\nabla^{\text{Hom}(TM; \phi^*TN)}$ , that is induced by the (Levi-Civita) connections on  $TM$  and  $TN$  respectively. Consequently,  $\nabla d\phi$  is a smooth section of  $T^*M \otimes \text{Hom}(TM; \phi^*TN)$ . This latter vector bundle admits a natural metric connection, in turn, and  $\nabla \nabla d\phi = \nabla^2 d\phi$  is then a smooth section of  $T^*M \otimes T^*M \otimes \text{Hom}(TM; \phi^*TN)$ . In other words,  $\nabla^2 d\phi$  is a 2-form with values in the vector bundle  $\text{Hom}(TM; \phi^*TN)$ . This 2-form has a unique decomposition into a symmetric and anti-symmetric part, *viz.*

$$\nabla^2 d\phi(x, y) = \frac{1}{2} (\nabla_{x,y}^2 + \nabla_{y,x}^2) d\phi + \frac{1}{2} (\nabla_{x,y}^2 - \nabla_{y,x}^2) d\phi,$$

where  $x, y$  are vector fields on  $M$  and  $\nabla_{x,y}^2 d\phi = \nabla_x(\nabla_y d\phi) - \nabla_{\nabla_x y} d\phi$ . Twice the anti-symmetric part of  $\nabla^2 d\phi$  is the curvature tensor of  $(\text{Hom}(TM; \phi^*TN), \nabla)$  and is written as<sup>4</sup>

$$(7) \quad R_{x,y} d\phi = (\nabla_{x,y}^2 - \nabla_{y,x}^2) d\phi.$$

There is a naturally-defined Ricci tensor associated to the curvature  $R$ . Let  $e_j$  be an orthonormal frame. Then for any tangent vector  $x$

$$(8) \quad \text{Ric}_{d\phi}(x) = \sum_j R_{x,e_j} d\phi \cdot e_j$$

which is easily seen to be independent of the choice of orthonormal frame. The Ricci tensor  $\text{Ric}_{d\phi}$  is a section of  $\text{Hom}(TM; \phi^*TN)$ , like  $d\phi$ .

The metric on  $\text{Hom}(TM, \phi^*TN)$  and associated bundles is described in section 2.1.

**Lemma 2.3.** *Let  $v \in T_p M$  be a gaussian with covariance operator (=metric)  $g_p$  and expected value 0. The expected value of*

- (1)  $v \mapsto |d\phi(v)|^2$  equals  $|d\phi|^2$ ;
- (2)  $v \mapsto |\nabla d\phi(v, v)|^2$  equals  $|\tau(\phi)|^2 + 2|\nabla d\phi|^2$ ;
- (3)  $v \mapsto \langle d\phi(v), \nabla^2 d\phi(v, v, v) \rangle$  equals

$$\sum_{i,j} \langle d\phi(e_i), \nabla_{e_j, e_j}^2 d\phi \cdot e_i + (\nabla_{e_i, e_j}^2 + \nabla_{e_j, e_i}^2) d\phi \cdot e_j \rangle,$$

where  $e_i$  is any orthonormal basis of  $T_p M$ .

It is recalled that the tension field  $\tau(\phi)$  equals  $\sum_i \nabla d\phi(e_i, e_i)$  and is a section of  $\phi^*TN$  with its induced norm. The norm of the second fundamental form  $\nabla d\phi$  is the norm of a section of  $T^*M \otimes \text{Hom}(TM; \phi^*TN)$ , so  $|\nabla d\phi|^2 = \sum_{i,j} |\nabla d\phi(e_i, e_j)|^2$ .

*Proof.* A simple calculation.  $\square$

**Lemma 2.4.**

$$\sum_{i,j} \langle d\phi(e_i), \nabla_{e_j, e_j}^2 d\phi \cdot e_i + (\nabla_{e_i, e_j}^2 + \nabla_{e_j, e_i}^2) d\phi \cdot e_j \rangle_p = \langle d\phi, 3\nabla\tau(\phi) - 2\text{Ric}_{d\phi} \rangle_p. \quad (*)$$

*Proof.* Let  $e_i$  be an orthonormal frame at  $p$  and let  $\dagger$  denote the left-hand side of (\*). The subscript  $p$  is dropped in the following. A computation yields

$$\nabla_{e_j, e_i}^2 d\phi \cdot e_j = \nabla_{e_j, e_j}^2 d\phi \cdot e_i \quad \forall i, j.$$

If  $\sum \langle d\phi(e_i), \nabla_{e_j, e_i}^2 d\phi \cdot e_j \rangle$  is added and subtracted to  $\dagger$ , then one obtains

$$\dagger = \sum_{i,j} 3 \langle d\phi(e_i), \nabla_{e_j, e_j}^2 d\phi \cdot e_i \rangle + \langle d\phi(e_i), R_{e_i, e_j} d\phi \cdot e_j \rangle,$$

which simplifies to

$$\dagger = \langle d\phi, 3\text{Trace}(\nabla^2 d\phi) + \text{Ric}_{d\phi} \rangle,$$

where  $\text{Trace}(\nabla^2 d\phi) = \sum_j \nabla_{e_j, e_j}^2 d\phi$ . The identities  $-\Delta d\phi = \text{Trace}(\nabla^2 d\phi) + \text{Ric}_{d\phi}$  and  $-\Delta d\phi = \nabla\tau(\phi)$  yield the lemma [3].  $\square$

---

<sup>4</sup>A nice, concise introduction to the subject of this paragraph is the monograph by Eells and Lemaire [3]. Their curvature tensor is minus that presented here, however. Their Ricci tensor is the same as that here.

The scalar  $\langle d\phi, \text{Ric}_{d\phi} \rangle$  is simplified in the following lemma. Let  $\text{Ric}^M$  be the Ricci tensor of  $(M, g)$ , viewed as a section of  $\text{Hom}(TM, TM)$  and let  $R^N$  be the Riemann curvature tensor of  $(N, h)$ . Let  $e_i$  be an orthonormal frame on  $T_p M$ ,  $u_i = d\phi(e_i)$ . A calculation shows that [3]

$$(9) \quad \langle d\phi, \text{Ric}_{d\phi} \rangle = -\langle d\phi, d\phi(\text{Ric}^M) \rangle + \sum_{i,j} \langle u_i, R^N_{u_i, u_j} u_j \rangle.$$

Since the second term is tensorial in  $u_i$ , this proves that

**Lemma 2.5.** *If  $d\phi|_{T_p M} = d\hat{\phi}|_{T_p M}$ , then  $\langle d\phi, \text{Ric}_{d\phi} \rangle = \langle d\hat{\phi}, \text{Ric}_{d\hat{\phi}} \rangle$  at  $p$ .*

**Lemma 2.6.**  *$\pi$  is harmonic:  $\tau(\pi) = 0$ .*

*Proof.* Since  $\pi \circ \iota = \text{id}_\Theta$ , and the second fundamental form of  $\iota$  is a quadratic form with values in  $N\Theta$ , it follows that  $\nabla d\pi(d\iota, d\iota) = 0$ , i.e.  $\nabla d\pi|_{T\Theta}$  vanishes. On the other hand, if  $\theta \in \Theta$  and  $\eta \in N_\theta\Theta$ , then  $\pi(\theta + s\eta) = \pi(\theta)$  for all  $s$ . Therefore  $\nabla d\pi|_{N\Theta}$  vanishes. These two facts show that the trace of  $\nabla d\pi$ , i.e.  $\tau(\pi)$ , vanishes.  $\square$

**2.4. Integration by Parts.** This section recalls the integration-by-parts formula following the discussion in [3]. Let  $\xi : V \rightarrow M$  be a vector bundle over the riemannian  $m$ -manifold  $(M, g)$  and let  $\mathcal{A}^p$  be the space of smooth sections of  $\Lambda^p M \otimes V$ , i.e.  $\mathcal{A}^p$  is the space of smooth  $p$ -forms on  $M$  with values in  $V$ . Assume that  $V$  is equipped with a metric and a compatible connection. There is a natural metric connection on  $\Lambda^p M \otimes V$ , call it  $\nabla$ , which induces an exterior derivation  $d : \mathcal{A}^p \rightarrow \mathcal{A}^{p+1}$  by skew-symmetrization. Let  $d^* : \mathcal{A}^p \rightarrow \mathcal{A}^{p-1}$  be the adjoint of  $d$  defined by

$$\int_M \langle d\sigma, \rho \rangle d\nu_g = \int_M \langle \sigma, d^*\rho \rangle d\nu_g + \int_{\partial M} (\sigma \wedge * \rho) d\nu_{g|\partial M}$$

for all  $\sigma \in \mathcal{A}^{m-p-1}$ ,  $\rho \in \mathcal{A}^p$ . Here  $*$  :  $\mathcal{A}^p \rightarrow \mathcal{A}^{m-p}$  is the Hodge star operator.

**Lemma 2.7.** *Let  $\lambda : M \rightarrow \mathbf{R}$  be a smooth function. Then*

$$\int_M \lambda \langle d\sigma, \rho \rangle d\nu_g = \int_M \langle \sigma, d^*(\lambda\rho) \rangle d\nu_g + \int_{\partial M} \lambda (\sigma \wedge * \rho) d\nu_{g|\partial M}.$$

*In particular, if  $\lambda|_{\partial M} = 0$ , then*

$$\int_M \lambda \langle d\sigma, \rho \rangle d\nu_g = \int_M \langle \sigma, d^*(\lambda\rho) \rangle d\nu_g.$$

*Proof.* This follows from applying the definition of  $d^*$  with  $\rho' = \lambda\rho$ .  $\square$

We continue with the notation of Lemma 2.7.

**Lemma 2.8.** *Let  $d\phi \in \mathcal{A}^1$  be a 1-form with values in  $V = \phi^*TN$  and assume that  $\lambda|_{\partial M} = 0$ . Then*

$$d^*(\lambda d\phi) = -\lambda \tau(\phi) - d\phi(\nabla\lambda).$$

*Proof.* Let  $e_i$  be an orthonormal frame on  $M$ . If  $\rho \in \mathcal{A}^1$  and  $\rho|_{\partial M} = 0$ , then  $d^*\rho = -\sum_i \nabla_{e_i} \rho \cdot e_i$  [3]. Thus,  $d^*(\lambda d\phi) = -\sum_i \lambda \nabla_{e_i} d\phi \cdot e_i - \sum_i d\phi(e_i) \cdot \nabla_{e_i} \lambda$ . The first term equals  $\lambda \tau(\phi)$ , while the second term equals  $d\phi(\nabla\lambda)$ .  $\square$

## 3. BAYESIAN ESTIMATORS

In Bayesian statistics, the point  $\theta$  is viewed as a random variable with a prior distribution  $\lambda(\theta)d\theta$  where  $\int_{\theta \in \Theta} \lambda(\theta) d\theta = 1$  ( $d\theta = d\nu_{\mathbf{g}}$  is the riemannian volume of  $(\Theta, \mathbf{g})$ ). Recall that the *Bayesian risk* of a map  $g$  is defined to be

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

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

**Existence and Uniqueness of the Bayes Estimator.** Let us sketch a proof of the existence and uniqueness of the Bayes estimator. Define

$$(10) \quad h_\epsilon(x; g, \lambda) = \int_{\theta \in \Theta} \mathbf{dist}(g(x), \gamma(\theta))^2 \lambda(\theta) \psi_\epsilon(x - \iota(\theta)) d\theta.$$

One sees that  $R_\epsilon(g; \lambda) = \int_{x \in \mathbf{E}} h_\epsilon(x; g, \lambda) dx$ . It is clear that a Bayes estimator  $\tilde{g}_\epsilon$  with prior  $\lambda$ , if it exists, will have the property that  $h_\epsilon(x; g, \lambda) \geq h_\epsilon(x; \tilde{g}_\epsilon, \lambda)$  for all  $x$  and all estimators  $g$ . One may assume that the class of estimators is the set of  $L^1$  maps between  $(\mathbf{E}, dx)$  and  $\Lambda$ .

By compactness of  $\Theta$ , there is an  $r_o > 0$  and  $\epsilon_o > 0$ , such that for all  $x \in \mathbf{E}, \theta \in \Theta$ , and  $\epsilon < \epsilon_o$ , the measure  $\psi_\epsilon(x - \iota(\theta))d\theta$  is supported, up to a remainder of  $O(\exp(-1/\epsilon))$ , on the ball of radius  $r_o$  about  $\hat{\theta} = \pi(x)$ . This observation is trivial if  $x$  lies within a distance  $r$  of  $\Theta$ ; and it is trivial if  $x$  lies in the complement of this neighbourhood, since then the measure itself is  $O(\exp(-1/\epsilon^2))$ .

Let  $B_{r_o}(\hat{\theta})$  be the closed ball of radius  $r_o$  centred at  $\hat{\theta}$ . Possibly after shrinking  $r_o$ , the continuity of  $\gamma$  and compactness of  $\Theta$  imply that the image,  $\gamma(B_{r_o}(\hat{\theta}))$ , of  $B_{r_o}(\hat{\theta})$  may be assumed to lie in a closed ball  $D_{s_o}(\gamma(\hat{\theta}))$  of fixed radius  $s_o$  about  $\gamma(\hat{\theta})$ .

Introduce normal coordinates at  $\hat{\theta}$  and  $\gamma(\hat{\theta})$  so the above described balls are isometric to a ball about 0 in a real vector space with an almost euclidean riemannian metric of the form  $\sum_i dx_i \otimes dx_i + O(|x|^2)$ .

We have therefore shown that  $h_\epsilon(x; g, \lambda)$  may be computed, up to a uniform remainder term of  $O(\exp(-1/\epsilon))$ , using a map between two vector spaces that are equipped with metrics that are euclidean up to second order. The techniques used in [7, 5] can be used in this situation to show that the Bayes estimator can be expanded as a formal power series in  $\epsilon^2$  and that this estimator is smooth.

**Remark.** If  $(\Lambda, \mathbf{h})$  is a euclidean vector space, the Bayes estimator exists and has the explicit form

$$(11) \quad \tilde{g}_\epsilon(x) = \frac{\int_{\theta \in \Theta} \gamma(\theta) \lambda(\theta) \psi_\epsilon(x - \iota(\theta)) d\theta}{\int_{\theta \in \Theta} \lambda(\theta) \psi_\epsilon(x - \iota(\theta)) d\theta}.$$

This estimator has some rather curious properties: if  $\gamma = \iota$  is the inclusion map  $\Theta \subset \mathbf{E}$ , then  $\tilde{g}_\epsilon$  is the weighted average of  $\iota(\theta)$ . Because this weighted average need not lie on  $\Theta$ , one finds that this estimator is unsatisfactory. The ultimate reason for this is the poor choice of risk function. somewhat unsatisfactory.



## 4. AN EXPANSION OF THE BAYESIAN RISK

First, introduce a change of variables.

**Lemma 4.1.** *The map  $J_\epsilon : \Theta \times \mathbf{E} \rightarrow \Theta \times \mathbf{E}$  defined by  $\theta = \theta, x = \iota(\theta) + \epsilon z$  is a diffeomorphism such that*

$$(12) \quad R_\epsilon(g; \lambda) = \int_{z \in \mathbf{E}, \theta \in \Theta} \mathbf{dist}(g(\iota(\theta) + \epsilon z), \gamma(\theta))^2 \lambda(\theta) \psi_1(z) \, dz \, d\theta.$$

*Proof.* A straightforward calculation.  $\square$

The expression  $\iota(\theta) + \epsilon z$  equals  $\exp_{\iota(\theta)}(\epsilon z)$  where  $\exp$  is the exponential map of the euclidean space  $\mathbf{E}$ . The Maclaurin series (equation 5) implies that

$$(13) \quad g \circ \exp_{\iota(\theta)}(\epsilon z) = \exp_{g(\iota(\theta))} \left( \epsilon dg(z) + \frac{1}{2} \epsilon^2 \nabla dg(z, z) + \frac{1}{6} \epsilon^3 \nabla^2 dg(z, z, z) + O(\epsilon^4 |z|^4) \right).$$

Since  $\Lambda$  is connected, for each  $a, b \in \Lambda$  there is a geodesic  $c : [0, 1] \rightarrow \Lambda$  such that  $c(0) = a, c(1) = b$  and the length of  $c$  is the distance between  $a$  and  $b$ . That is,  $|w|_a = \mathbf{dist}(a, b)$  where  $w = \dot{c}(0)$ . The tangent vector  $w = w_{a,b}$  is not unique in general, but  $w$  is a measurable function that is smooth off the the cut locus of  $a$ .

For  $a = g(\iota(\theta))$  and  $b = \gamma(\theta)$ , let  $w = w(\theta)$  be the vector  $w_{a,b}$ . The vector  $w(\theta)$  is characterized by the property that  $\exp_{g(\iota(\theta))}(w(\theta)) = \gamma(\theta)$  for all  $\theta$  and  $w(\theta)$  is a shortest vector amongst all such vectors. The Bayesian estimator  $g_\epsilon : \mathbf{E} \rightarrow \Lambda$  is written as

$$(14) \quad g_\epsilon(x) = \exp_{g_o(x)} (\epsilon^2 g_2(x) + O(\epsilon^4)).$$

By definition,  $g_\epsilon$  minimizes the Bayesian risk functional  $g \mapsto R_\epsilon(g; \lambda)$  for each  $\epsilon$ . Since the Bayesian risk functional is an even function of  $\epsilon$ , the Bayesian estimator is, too.

**Lemma 4.2.** *Let the Bayesian risk  $R_\epsilon = A_0 + O(\epsilon^2)$ . Then*

$$(15) \quad A_0(g; \lambda) = \int_{\theta \in \Theta} |w(\theta)|^2 \lambda(\theta) \, d\theta.$$

*Consequently, the Bayes estimator  $g_\epsilon$  satisfies*

$$(16) \quad g_\epsilon \circ \iota(\theta) = \exp_{\gamma(\theta)} (\epsilon^2 g_2(\iota(\theta))) + O(\epsilon^4) \quad \forall \theta.$$

*Proof.* The formula for  $A_0$  is straightforward. Since  $\lambda > 0$  a.e. by hypothesis, and  $A_0 \geq 0$ , it follows that  $A_0 = 0$  only if  $w = 0$  a.e., that is, only if  $g_{\epsilon=0} \circ \iota = \gamma$ . Since  $g_{\epsilon=0} = g_o$ , equation 14 implies equation 16.  $\square$

Lemma 4.2 shows that

**Corollary 4.3.**  $g_o \circ \iota = \gamma$ .

**Lemma 4.4.**  $dg_\epsilon = dg_o + \epsilon^2 dg_2 + O(\epsilon^4)$ .

*Proof.* Let  $x \in \mathbf{E}$  and  $v \in T_x \mathbf{E}$ . It suffices to prove

$$d_x g_\epsilon \cdot v = (d_x g_o + \epsilon^2 d_x g_2 + O(\epsilon^4)) \cdot v. \quad (*)$$

The left-hand side of (\*) is

$$d_x g_\epsilon \cdot v = \left. \frac{d}{dt} \right|_{t=0} \exp_{a(t)} (\epsilon^2 b(t) + O(\epsilon^4)), \quad (**)$$

where  $a(t) = g_o(x + tv)$  is a curve in  $\Lambda$  and  $b(t) = g_2(x + tv)$  is a curve of tangent vectors along  $a(t)$ . The right-hand side of (\*\*\*) is the Jacobi field  $J(s)$  on  $(\Lambda, \mathbf{h})$  with initial conditions  $J(0) = \dot{a}(0)$  and  $\dot{J}(0) = \dot{b}(0) + O(\epsilon^2)$  at the time  $s = \epsilon^2$ . Since  $J(s) = J(0) + s\dot{J}(0) + O(s^2)$  and  $\dot{a}(0) = d_x g_o \cdot v$ ,  $\dot{b}(0) = d_x g_2 \cdot v$  we see that (\*\*\*) implies (\*).  $\square$

**Lemma 4.5.** *Let the Bayesian risk of the Bayesian estimator  $g_\epsilon$  be  $R_\epsilon = \epsilon^2 A_2 + \epsilon^4 A_4 + O(\epsilon^6)$ . If  $A_k = \int a_k \lambda(\theta) d\theta$ , then the integrand  $a_k$  is*

$$(17) \quad a_2 = |dg_o|^2$$

$$(18) \quad a_4 = \begin{cases} \frac{1}{4} |\tau(g_o)|^2 + \frac{1}{2} |\nabla dg_o|^2 + \langle dg_o, \nabla \tau(g_o) \rangle - \frac{2}{3} \text{Ric}_{dg_o} + \\ |g_2|^2 + 2 \langle dg_2, dg_o \rangle + \langle g_2, \tau(g_o) \rangle \end{cases}$$

*Proof.* When one expands  $g_\epsilon(\iota(\theta) + \epsilon z)$  as a Maclaurin series, one obtains

$$(19) \quad \exp_{\gamma(\theta)} \left( \epsilon^2 g_2 + \epsilon dg_\epsilon(z) + \frac{1}{2} \epsilon^2 \nabla dg_\epsilon(z, z) + \frac{1}{6} \epsilon^3 \nabla^2 dg_\epsilon(z, z, z) + O(\epsilon^4 |z|^4) \right).$$

Since  $dg_\epsilon = dg_o + \epsilon^2 dg_2 + O(\epsilon^4)$  by Lemma 4.4, the Maclaurin series equals

$$(20) \quad \exp_{\gamma(\theta)} \left( \epsilon dg_o(z) + \epsilon^2 \left( g_2 + \frac{1}{2} \nabla dg_o(z, z) \right) + \epsilon^3 \left( dg_2(z) + \frac{1}{6} \nabla^2 dg_o(z, z, z) \right) + O(\epsilon^4 |z|^4) \right).$$

The distance between  $g_\epsilon(\iota(\theta) + \epsilon z)$  and  $\gamma(\theta)$  expands to

$$(21) \quad \begin{aligned} \text{dist} &= \epsilon^2 |dg_o(z)|^2 + \epsilon^4 \left( |g_2 + \frac{1}{2} \nabla dg_o(z, z)|^2 + 2 \langle dg_o(z), dg_2(z) \rangle + \frac{1}{6} \nabla^2 dg_o(z, z, z) \right) \\ &+ \epsilon^3 (\cdot) + \epsilon^5 (\cdot) + O(\epsilon^6 |z|^6), \end{aligned}$$

where the coefficients on the odd powers of  $\epsilon$  are odd polynomials in  $z$ . Lemmas 2.3–2.4 now imply this lemma.  $\square$

Recall that  $\pi : N\Theta \rightarrow \Theta$  is the normal bundle of  $\Theta$  in  $\mathbf{E}$ ; the tangent bundle of  $N\Theta$  is isometric to  $T_\Theta \mathbf{E}$  while  $d\pi$  is the orthogonal projection of  $T(N\Theta)$  onto  $T\Theta$ . Corollary 4.3 implies that  $dg_o|_{T\Theta} = d\gamma$ , so on  $\Theta$   $|dg_o|^2 \geq |d\gamma|^2$  with equality iff  $dg_o|_{N\Theta} = 0$  or  $dg_o|_{T_\Theta \mathbf{E}} = d\gamma \circ d\pi$ . By Lemma 4.5, these considerations show that

**Proposition 4.6.** *The Bayesian estimator satisfies*

- (1)  $g_o \circ \iota = \gamma$ ;
- (2)  $dg_o = d(\gamma \circ \pi)$  on  $T_\Theta \mathbf{E}$ .

Define  $\Gamma = \gamma \circ \pi$ . The next step is to show that  $\nabla dg_o = \nabla d\Gamma$  on  $\Theta$ . To do so requires that  $a_4$  (Lemma 4.5) be simplified.

**Lemma 4.7.** *Under the standing hypothesis that  $\lambda > 0$  on  $\Theta$ , we have*

- (1)  $\int_\Theta d\theta \lambda(\theta) \langle dg_o, dg_2 \rangle = \int_\Theta d\theta \lambda(\theta) \langle g_2, \tau(\gamma) + d\gamma(\nabla \log \lambda) \rangle$ ;
- (2)  $g_2 \circ \iota = \tau(\gamma) - \frac{1}{2} \tau(g_o) + d\gamma(\nabla \log \lambda)$ .
- (3)  $a_4 = \frac{1}{2} |\nabla dg_o|^2 - |\tau(\gamma) + d\gamma(\nabla \log \lambda)|^2 - \frac{2}{3} \langle d\Gamma, \text{Ric}_{d\Gamma} \rangle$ ;

*Proof.* The inner product  $\lambda \langle dg_o, dg_2 \rangle$  on  $\Theta$  equals  $\langle \lambda \cdot d(\gamma \circ \pi), (dg_2) \circ \iota \rangle$  which equals  $\langle \lambda \cdot d\gamma, d(g_2 \iota) \rangle$ . The integration-by-parts formula (Lemma 2.8) for sections of  $T^*\Theta \otimes \gamma^* T\Lambda$  yields (1).

(1) along with equation 18 yields

$$a_4 = \begin{cases} |g_2 \circ \iota - \tau(\gamma) - \frac{1}{2}\tau(g_o) + d\gamma(\nabla \log \lambda)|^2 + \\ \langle \tau(g_o), \tau(\gamma) + d\gamma(\nabla \log \lambda) \rangle - |\tau(\gamma) + d\gamma(\nabla \log \lambda)|^2 + \\ \frac{1}{2}|\nabla d g_o|^2 + \langle d g_o, \nabla \tau(g_o) - \frac{2}{3}\text{Ric}_{d g_o} \rangle; \end{cases}$$

It is clear that  $a_4$  is minimized by setting  $g_2$  to that in (2).

Finally, Lemma 2.5 implies that  $\langle d g_o, \text{Ric}_{d g_o} \rangle = \langle d\Gamma, \text{Ric}_{d\Gamma} \rangle$  on  $\Theta$ . A second application of the integration-by-parts formula to  $\lambda \langle d g_o, \nabla \tau(g_o) \rangle$  proves (3).  $\square$

**Proposition 4.8.** *The Bayesian estimator satisfies*

- (1)  $\nabla d g_o = \nabla d\Gamma$  on  $\Theta$ ;
- (2)  $\tau(g_o) = \tau(\gamma)$  on  $\Theta$ ;
- (3)  $g_2 \circ \iota = \frac{1}{2}\tau(\gamma) + d\gamma(\nabla \log \lambda)$ ; and
- (4)  $a_4 = \frac{1}{2}|\nabla d\Gamma|^2 - |\tau(\gamma) + d\gamma(\nabla \log \lambda)|^2 - \frac{2}{3}\langle d\Gamma, \text{Ric}_{d\Gamma} \rangle$ .

*Proof.* By (3) of Lemma 4.7, it is clear that  $a_4$  is minimized iff  $|\nabla d g_o|^2$  is minimized. Let  $\alpha$  (resp.  $\beta$ ) be the orthogonal projection of  $T_\Theta \mathbf{E}$  onto  $T\Theta$  (resp.  $N\Theta$ ). This orthogonal decomposition yields the equality

$$|\nabla d g_o|^2 = |\nabla d g_o(\alpha, \alpha)|^2 + 2|\nabla d g_o(\alpha, \beta)|^2 + |\nabla d g_o(\beta, \beta)|^2.$$

Since  $|\nabla d g_o(\alpha, \alpha)|^2 = |\nabla d g_o(d\iota, d\iota)|^2$  and  $\nabla d(g_o \iota) = \nabla d g_o(d\iota, d\iota) + d g_o \cdot \nabla d\iota$ , Proposition 4.6 yields  $|\nabla d g_o(\alpha, \alpha)|^2 = |\nabla d\gamma|^2 = |\nabla d(\gamma\pi)(\alpha, \alpha)|^2$ . Part (4) follows from this.

A Maclaurin series argument shows that Proposition 4.6 implies that  $\nabla d g_o(\alpha, \beta) = \nabla d(\gamma\pi)(\alpha, \beta)$ , while  $|\nabla d g_o(\beta, \beta)|$  is unconstrained. This is minimized by  $0 = |\nabla d(\gamma\pi)(\beta, \beta)|$ . This proves (1).

The formula  $\tau(g_o) = \tau(\gamma) + d\gamma \cdot \tau(\pi)$  is implied by (1). Since  $\pi$  is harmonic by Lemma 2.6, this implies (2). Lemma 4.7 part (2) implies (3).  $\square$

Let us summarize the results of this section.

**Theorem 4.1.** *Let  $\tilde{g}_\epsilon(x) = \exp_{g_o(x)}(\epsilon^2 g_2(x) + O(\epsilon^4))$  be the Bayesian estimator for the Bayesian risk functional  $R_\epsilon$  (Equation 3) with a fixed Bayesian prior  $\lambda > 0$ . Then*

- (1) for all  $x \in N\Theta$ , where  $\hat{\theta} = \pi(x)$  and  $|x - \hat{\theta}| \leq r$ ,

$$\tilde{g}_\epsilon(x) = \exp_{\gamma(\hat{\theta})} \left( \epsilon^2 \left( \frac{1}{2}\tau(\gamma) + d\gamma(\nabla \log \lambda) \right)_{\hat{\theta}} + O(r\epsilon^4) \right).$$

- (2)

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

where  $\Gamma = \gamma\pi$ .

*Proof.* (1) Let  $x \in \mathbf{E}$  and  $|x - \Theta| \leq r$ . By the hypothesis on the radius  $r$ , the orthogonal projection of  $x$  onto  $\Theta$  is well-defined; this orthogonal projection is denoted by  $\hat{\theta} = \pi(x)$ . Write  $x = \iota(\hat{\theta}) + \epsilon z$ , where by construction,  $z \in N_{\hat{\theta}}\Theta$ . The Maclaurin expansion of  $\tilde{g}_\epsilon$  at  $\iota(\hat{\theta})$  gives

$$(22) \quad \tilde{g}_\epsilon(x) = \tilde{g}_\epsilon \circ \exp_{\iota(\hat{\theta})}(\epsilon z) = \exp_{\tilde{g}_\epsilon \circ \iota(\hat{\theta})} \left( \epsilon d\tilde{g}_\epsilon \cdot z + \frac{\epsilon^2}{2} \nabla d\tilde{g}_\epsilon(z, z) + O(r^3 \epsilon^3) \right)_{\hat{\theta}}.$$

Equation (16) and Proposition 4.8.3 show that  $\tilde{g}_\epsilon \circ \iota(\hat{\theta}) = \exp_{\gamma(\hat{\theta})} w_\epsilon$  where  $w_\epsilon = \epsilon^2 \times \left( \frac{1}{2} \tau(\gamma) + d\gamma(\nabla \log \lambda) \right)_{\hat{\theta}} + O(\epsilon^4)$ . On the other hand, since  $z \in N_{\hat{\theta}}\Theta$ , Proposition 4.6 shows that

$$(23) \quad d\tilde{g}_\epsilon \cdot z = \epsilon^2 \times dg_2 \cdot z + O(r\epsilon^4)$$

while Proposition 4.8.1 shows that

$$(24) \quad \nabla d\tilde{g}_\epsilon(z, z)|_{\hat{\theta}} = \epsilon^2 \times \nabla dg_2(z, z) + O(r^2 \epsilon^4).$$

Equations (22–24) imply that  $\tilde{g}_\epsilon(x) = \exp_{\gamma(\hat{\theta})}(v_\epsilon)$  and that  $v_\epsilon = w_\epsilon + O(r\epsilon^3)$ . Therefore

$$(25) \quad \tilde{g}_\epsilon(x) = \exp_{\gamma(\hat{\theta})} \left( \epsilon^2 \times \left( \frac{1}{2} \tau(\gamma) + d\gamma(\nabla \log \lambda) \right)_{\hat{\theta}} + O(r\epsilon^3) \right).$$

Since the Bayesian estimator  $\tilde{g}_\epsilon$  is an even function of  $\epsilon$ , the error is not  $O(r\epsilon^3)$  but must be  $O(r\epsilon^4)$ .

(2) This is a straightforward application of the preceding work. □

**Remark.** Inspection of the proof above shows that the  $O(r\epsilon^3)$  term in Equation (25) is  $\epsilon^3 \times dg_2 \cdot z$ . Thus, the proof also shows that  $dg_2|_{N_{\Theta}\Theta}$  vanishes.

## 5. OPTIMAL PRIORS

In this section we are interested in the behaviour of the minimax risk which can be defined as

$$r_\epsilon(\Theta) = \inf_{g_\epsilon} \sup_{\theta \in \Theta} (R_\epsilon(g_\epsilon, \theta) - \epsilon^2 |d\gamma(\theta)|^2),$$

where the inf is taken over all possible (sequences of) estimators  $g_\epsilon$ . The problem of finding the asymptotic behaviour of  $r_\epsilon(\Theta)$  can be derived in a relatively straightforward manner from the previous results. Essentially the problem reduces to finding *optimal priors* maximizing the first non-trivial term of the Bayes risk.

Since in the case of smooth functions the minimax risk  $r_\epsilon(\Theta)$  is typically of order  $\epsilon^4$ , we can define the *second-order minimax risk* as

$$(26) \quad r(\Theta) = \lim_{\epsilon \rightarrow 0} \inf_{g_\epsilon} \sup_{\theta \in \Theta} \epsilon^{-4} (R_\epsilon(g_\epsilon, \theta) - \epsilon^2 |d\gamma(\theta)|^2).$$

Sometimes, a more general minimax risk may be of interest. Let  $p$  and  $q > 0$  be given function defined on  $\Theta$ . Then equation (26) can be modified as

$$r^*(\Theta) = \lim_{\epsilon \rightarrow 0} \inf_{g_\epsilon} \sup_{\theta \in \Theta} \frac{R_\epsilon(g_\epsilon, \theta) - \epsilon^2 |d\gamma(\theta)|^2 - \epsilon^4 p(\theta)}{\epsilon^4 q(\theta)}.$$

Even more useful is the following equivalent definition of the second order minimax risk

$$(27) \quad r^*(\Theta) = \inf\{r|\exists g : R_\epsilon(g_\epsilon, \theta) \leq \epsilon^2|d\gamma(\theta)|^2 + \epsilon^4(p(\theta) + rq(\theta))\}.$$

The advantage of the last formula is that, unlike the previous one, it allows consideration of smooth functions  $q(\theta) \geq 0$ ,  $q(\theta) \neq \text{const}$ . It is thus in this form that the second-order minimax risk will be considered below.

Theorem 4.1, part 2, gives a formula for the Bayesian risk expanded up to  $O(\epsilon^6)$  of the Bayesian estimator  $\tilde{g}_\epsilon$ . One would like to determine the Bayesian prior distribution that minimizes the Bayesian risk. There are a couple interesting twists that arise at this point. First, the implicit *flat* prior is a constant multiple of the riemannian volume form  $d\theta$ . However, there is no reason to single out the riemannian volume form as the flat prior. Rather, one can introduce the flat prior  $d\nu = a(\theta) d\theta$  ( $a > 0$  a.e.) and the Bayesian prior  $\eta(\theta) d\nu = \lambda(\theta) d\theta$ , where  $\eta = \lambda/a$ . Let us stress that the change from  $d\theta$  to  $d\nu$ , and  $\lambda$  to  $\eta$ , does not change the foregoing calculations and results. Second, the minimizers of the Bayesian risk functional are also the minimizers of the functional

$$(28) \quad \tilde{R}_\epsilon(g; \lambda) = R_\epsilon(g; \lambda) - \epsilon^2 \int_{\Theta} d\theta \lambda |d\gamma|^2,$$

since the second term is independent of  $g$ . One may, therefore, elect to minimize the functional  $\tilde{R}_\epsilon$ , to obtain the Bayesian estimator  $\tilde{g}_\epsilon$ —which is implicitly a function of the Bayesian prior  $\lambda d\theta$ —and proceed to determine the second-order optimal prior by minimizing  $\lambda \mapsto \tilde{R}_\epsilon(\tilde{g}_\epsilon; \lambda)$ . Finally, inspection of part (2) of Theorem 4.1 shows that one needs tools to understand how to simplify the term  $|\tau(\gamma) + d\gamma(\nabla \log \lambda)|^2$ . The requisite tool is known as *sub-riemannian* geometry.

**5.1. Sub-riemannian geometry.** Let us describe a particular construction of a sub-riemannian geometry. Let  $(M, g) \xrightarrow{\phi} (N, h)$  be a smooth map, and let  $D_p = \ker d_p\phi^\perp$  for  $p \in M$ . The collection  $D = \cup_p D_p$  is a singular distribution on  $M$ . It is equipped with an inner product  $s$  – a sub-riemannian metric – by declaring that  $d_p\phi|_{D_p} \rightarrow \text{im } d_p\phi \subset T_pN$  is an isometry. That is,  $s = \phi^*h|_D$ .

One may think of the subriemannian structure  $(D, s)$  as a singular distribution of directions in which one may travel, along with a metric which allows one to measure speed (and angles). Subriemannian structures arise in optimal control problems quite frequently [8, 9].

One may equivalently characterize the sub-riemannian structure  $(D, s)$  by a bundle map  $\mu : T^*M \rightarrow TM$  such that (i)  $\mu$  is self-adjoint; and (ii) the image of  $\mu$  equals  $D$ . In the present context, the map  $\mu$  is characterized by the identity

$$\mu(du, dv) = s(\nabla u, \nabla v) = \langle d\phi(\nabla u), d\phi(\nabla v) \rangle,$$

for all smooth functions  $u, v : M \rightarrow \mathbf{R}$ . Equivalently,  $\mu \cdot du = d\phi' d\phi(\nabla u)$ .

An augmented sub-riemannian structure  $\mathfrak{D} = (D, s, d\nu)$  is a sub-riemannian structure  $(D, s)$  plus a volume form  $d\nu$ . The augmented sub-riemannian structure permits one to define a *sub-laplacian*  $\Delta_{\mathfrak{D}}$ , which is a second-order, self-adjoint

differential operator.<sup>5</sup> In local coordinates

$$(29) \quad \Delta_{\mathfrak{D}} = - \sum_{ij} \frac{1}{f} \frac{\partial}{\partial x^i} \left( f \cdot \mu^{ij} \frac{\partial}{\partial x^j} \right),$$

where  $d\nu = f dx^1 \wedge \cdots \wedge dx^m$ . The sub-laplacian is defined invariantly by

$$(30) \quad \int u \cdot \Delta_{\mathfrak{D}} v \, d\nu = \int \mu(du, dv) \, d\nu,$$

for all smooth functions that vanish on  $\partial M$ . The self-adjointness of  $\mu$  implies  $\Delta_{\mathfrak{D}}$  is self-adjoint.

If  $a$  is a positive function, then let the augmented sub-riemannian structure  $(D, s, a \cdot d\nu)$  be denoted by  $a \cdot \mathfrak{D}$ . Equation (30) shows that the sub-laplacian of the augmented sub-riemannian structures differ by a differential operator of first order

$$(31) \quad \Delta_{a \cdot \mathfrak{D}} = \Delta_{\mathfrak{D}} - \mu \cdot d \log a.$$

**5.2. Optimal priors, I.** The discussion of sub-riemannian geometry allows the expansion of the Bayesian risk (Theorem 4.1). The term  $\int d\theta \lambda |\tau(\gamma) + d\gamma(\nabla \log \lambda)|^2$  expands to

$$(32) \quad \int d\theta \{ \omega^2 |\tau(\gamma)|^2 + 4\omega \langle \tau(\gamma), d\gamma(\nabla \omega) \rangle + 4\omega \Delta_{\mathfrak{E}} \omega \},$$

where  $\lambda = \omega^2$  and  $\mathfrak{E} = (E, s, d\theta)$  where  $E = \ker d\gamma^\perp$  and  $s = \gamma^* \mathbf{h}|E$ . Define

$$(33) \quad \kappa = \frac{1}{2} |\nabla d\Gamma|^2 - \frac{2}{3} \langle d\Gamma, \text{Ric}_{d\Gamma} \rangle + |\tau(\gamma)|^2 + 2 \langle d\gamma, \nabla \tau(\gamma) \rangle,$$

$$(34) \quad L = 4\Delta_{\mathfrak{E}} + \kappa.$$

From this discussion, and an application of the integration-by-parts formula to  $\int d\theta \lambda \langle \tau(\gamma), d\gamma(\nabla \log \lambda) \rangle$ , the following is clear.

**Theorem 5.1.** *The Bayesian risk functional at  $\tilde{g}_\epsilon$  with prior  $\lambda = \omega^2$  equals*

$$(35) \quad R_\epsilon(\tilde{g}_\epsilon; \omega^2) = \epsilon^2 \int d\theta \omega^2 |d\gamma|^2 + \epsilon^4 \int d\theta \omega \cdot L\omega + O(\epsilon^6),$$

while

$$(36) \quad \tilde{R}_\epsilon(\tilde{g}_\epsilon; \omega^2) = \epsilon^4 \int d\theta \omega \cdot L\omega + O(\epsilon^6).$$

Define a differential operator  $\mathcal{H}_\epsilon$  on  $\Theta$  by

$$(37) \quad \mathcal{H}_\epsilon := \epsilon^2 L + |d\gamma|^2.$$

The operator  $\mathcal{H}_\epsilon$  is the Schrödinger operator for a unit-mass particle on  $\Theta$  in a potential field  $V = |d\gamma|^2 + \epsilon^2 \kappa$  with kinetic energy  $T = \frac{1}{2} \langle \mu(p), p \rangle$  induced by the sub-riemannian metric and Planck constant  $\hbar = 8\epsilon^2$ . From Theorem 5.1 it is apparent that  $R_\epsilon(\tilde{g}_\epsilon | \omega^2) = \epsilon^2 \int d\theta \omega \cdot \mathcal{H}_\epsilon \omega + O(\epsilon^6)$ .

---

<sup>5</sup>Warning: the sign of  $\Delta_{\mathfrak{D}}$  conflicts with the sign in Montgomery's exposition [8], but it accords with the sign convention in riemannian geometry.

**Theorem 5.2.** *Let  $\alpha_\epsilon$  be the largest eigenvalue of  $\mathcal{H}_\epsilon$  with eigenfunction  $\omega = \omega_\epsilon$  which has  $\int d\theta \omega_\epsilon^2 = 1$ . Then*

$$(38) \quad R_\epsilon(\tilde{g}_\epsilon; \omega_\epsilon^2) = \epsilon^2 \alpha_\epsilon + O(\epsilon^6).$$

*Let  $\alpha$  be the largest eigenvalue of  $L$  with eigenfunction  $\omega$  normalized so that  $\int d\theta \omega^2 = 1$ . Then*

$$(39) \quad \tilde{R}_\epsilon(\tilde{g}_\epsilon; \omega^2) = \epsilon^4 \alpha + O(\epsilon^6),$$

and

$$(40) \quad r(\Theta) = \alpha.$$

**Remarks.** 1/ In Theorem 5.2, it is assumed that  $\mathcal{H}_\epsilon$  (resp.  $L$ ) does possess a largest eigenvalue. Non-compactness of  $\Theta$  may negate this assumption; it may also be negated by properties of the singular distribution  $E$ . A reformulation of the theorem in the event that  $\mathcal{H}_\epsilon$  (resp.  $L$ ) has no largest eigenvalue is clear. 2/ When are the eigenvalues of  $\mathcal{H}_\epsilon$  (resp.  $L$ ) constant? This depends on the *accessibility* property of the singular distribution  $E$ . If sections of  $E$  generate  $T\Theta$  under repeated Lie brackets, then Hörmander has shown that  $\mathcal{H}_\epsilon$  (resp.  $L$ ) is hypoelliptic. At the opposite extreme, the distribution  $E$  might be integrable, in which case the eigenvalues of  $\mathcal{H}_\epsilon$  (resp.  $L$ ) will vary from leaf to leaf. The optimal prior in this latter case is a singular function (a distribution, in the functional-analytic sense) concentrated on the leaf with the largest eigenvalue. 3/ The operator  $\mathcal{H}_\epsilon$  is a singular perturbation of a multiplication operator, so one generally cannot naïvely expand  $\alpha_\epsilon$  in a power series. However, when  $|\mathrm{d}\gamma| = \alpha_0$  is constant, the naïve idea is correct. In this case, one sees that  $\alpha_\epsilon = \alpha_0 + \epsilon^2 \alpha$  where  $\alpha$  is the largest eigenvalue of  $L$  (modulo the remarks in 1/). 4/ Important special cases include  $\gamma$  being a riemannian submersion or immersion.

**5.3. Optimal priors, II.** As noted in the beginning of this section, there is no natural reason why one should choose  $d\theta$  as the flat prior. Let us investigate the effect of choosing the flat prior to be  $d\nu = a^2 d\theta$ . With  $a\eta = \omega$ , one computes from equations (30,31) that

$$(41) \quad \begin{aligned} 4 \int d\theta \omega \Delta_{\mathfrak{E}} \omega &= 4 \int d\nu \{ \eta^2 |\mathrm{d}\log a|^2 + 2\eta\mu(\mathrm{d}\log a, \mathrm{d}\eta) + \eta \Delta_{a^2 \cdot \mathfrak{E}} \eta \}, \\ &= 4 \int d\theta \times a^2 \times \{ \eta^2 |\mathrm{d}\log a|^2 + \eta \Delta_{\mathfrak{E}} \eta \}, \end{aligned}$$

where  $|\mathrm{d}\log a|^2 = \mu(\mathrm{d}\log a, \mathrm{d}\log a)$ . Define

$$(42) \quad \kappa_a = \kappa + |\mathrm{d}\log a|^2,$$

$$(43) \quad L_a \eta = (4\Delta_{\mathfrak{E}} + \kappa_a) \eta.$$

From this discussion and the results of the previous section, the following is clear.

**Theorem 5.3.** *The Bayesian risk of  $\tilde{g}_\epsilon$  with prior  $\lambda d\theta = \eta^2 d\nu$  ( $d\nu = a^2 d\theta$ ) equals*

$$(44) \quad R_\epsilon(\tilde{g}_\epsilon; \lambda) = \epsilon^2 \int d\nu \eta^2 |\mathrm{d}\gamma|^2 + \epsilon^4 \int d\nu \eta \cdot L_a \eta + O(\epsilon^6),$$

while

$$(45) \quad \tilde{R}_\epsilon(\tilde{g}_\epsilon; \lambda) = \epsilon^4 \int d\nu \eta \cdot L_a \eta + O(\epsilon^6).$$

Since the operator  $L_a$  is self-adjoint with respect to the inner product determined by  $d\theta$ , one knows that the prior that maximizes  $\tilde{R}_\epsilon(\tilde{g}_\epsilon; \lambda)$  occurs at a solution to the eigenvalue problem

$$L_a \eta = \alpha a^2 \eta. \quad (\$)$$

**Theorem 5.4.** *Let  $\alpha$  be the largest eigenvalue of the eigenvalue problem (§) with eigenfunction  $\eta$  normalized so that  $\int d\nu \eta^2 = 1$ . Then, with  $\lambda = a^2 \eta^2$ ,*

$$(46) \quad \tilde{R}_\epsilon(\tilde{g}_\epsilon; \lambda) = \epsilon^4 \alpha + O(\epsilon^6),$$

and

$$(47) \quad r^*(\Theta) = \alpha.$$

[ $r^*(\Theta)$  is defined in Eq. (27)].

## 6. APPLICATIONS

There are several cases in which the formulas of Theorem 5.2 yield especially nice results.

**6.1. Riemannian immersions.** Recall that  $\phi : (M, g) \rightarrow (N, h)$  is a riemannian immersion if  $\phi^* h = g$ . If  $\gamma : (\Theta, \mathbf{g}) \rightarrow (\Lambda, \mathbf{h})$  is a riemannian immersion, then the riemannian structure and the induced sub-riemannian structure coincide, while  $|d\gamma|^2 = \dim \Theta$  is constant. Remark 3/ following Theorem 5.2 shows that

**Corollary 6.1.** *Let  $L = 4\Delta + \kappa$  where  $\Delta$  is the laplacian of  $(\Theta, \mathbf{g})$  and  $\kappa$  is defined in Eq. (33). Let  $\alpha$  be the largest eigenvalue of  $L$  with eigenfunction  $\omega$  of unit  $L^2$ -norm. Then  $\alpha_\epsilon = \dim \Theta + \epsilon^2 \alpha$ ,  $\omega_\epsilon = \omega$  and*

$$\begin{aligned} R_\epsilon(\tilde{g}_\epsilon; \omega_\epsilon^2) &= \epsilon^2 (\dim \Theta + \epsilon^2 \alpha) + O(\epsilon^6), \\ &= \tilde{R}_\epsilon(\tilde{g}_\epsilon; \omega_\epsilon^2). \end{aligned}$$

There are two interesting special cases of this corollary: when  $\gamma = \text{id}_\Theta$  and when  $\gamma = \iota$  (the inclusion map of  $\Theta$  into  $\mathbf{E}$ ). By corollary 6.1, the sub-laplacian is the same in each case. However, the curvatures of the identity map differ substantially from those of the inclusion map. One sees that for  $x$  in neighbourhood of  $\Theta$

$$(48) \quad g_\epsilon(x) = \begin{cases} \exp_{\pi(x)} (2\epsilon^2 \nabla \log |\omega|) + O(\epsilon^4) & \text{if } \gamma = \text{id}_\Theta, \\ \pi(x) + \epsilon^2 (\tau(\iota) + 2\nabla \log |\omega|) + O(\epsilon^4) & \text{if } \gamma = \iota. \end{cases}$$

The tension field of the inclusion map  $\tau(\iota)$  is  $\dim \Theta$  times the mean curvature vector field – in particular, it is normal to  $\Theta$  – so  $g_\epsilon(x) \notin \Theta$  in the second case. It should be noted that  $\omega$  is not the same function in each line. The curvature term  $\kappa$  equals

$$(49) \quad \kappa = \begin{cases} \frac{5}{3} |\nabla d\iota|^2 - \frac{2}{3} |\tau(\iota)|^2 & \text{if } \gamma = \text{id}_\Theta, \\ \frac{3}{2} |\nabla d\iota|^2 - |\tau(\iota)|^2 & \text{if } \gamma = \iota. \end{cases}$$

While the two estimation problems are incomparable, strictly speaking, it is interesting to observe that  $\kappa$  – and consequently, the dominant eigenvalue of  $L$  and bayesian risk – is least for the estimator of the inclusion map. This comes with an expense: the estimator of  $\iota$  does not take values that are on  $\Theta$ , while the estimator of the identity is forced to do so.

B. Levit, in his unpublished *Habilitation* thesis, computes  $\kappa$  in the case where  $\gamma = \iota$ . His calculations are carried out in a system of local coordinates, which masks the difference between the inclusion and the identity map. The present



paper's formalism, based on the Maclaurin series, clarifies these differences and explains why the earlier estimator takes values off the manifold  $\Theta$ . The formulas in equation 49 are proven in the next section.

6.1.1. *Calculations for riemannian immersions.* Let  $(M, g) \xrightarrow{\phi} (N, h)$  be a riemannian immersion. For the present calculations, it may be assumed that  $N$  is a riemannian vector bundle over  $M$  and  $\phi$  is the inclusion of the zero section. The projection map  $N \rightarrow M$  is denoted by  $\pi$ .  $N$  is naturally identified with the normal bundle of  $M$ . The tangent bundle to  $N$  along  $M$  is denoted by  $T_M N = TM \oplus N$ .

**Lemma 6.2.** *Let  $p \in M$  and  $x + y, u + v \in T_p M \oplus N_p$ . Then*

$$\nabla d\pi(x + y, u + v) = B'_x v + B'_u y,$$

where  $B_\bullet : T_p M \rightarrow N_p$  is defined by  $B_\bullet = \nabla d\phi(\bullet, \cdot)$  and  $B'_\bullet : N_p \rightarrow T_p M$  is the transposed map.

*Proof.* For a smooth vector field  $x$  on  $M$ , let  $\tilde{x}$  be a smooth vector field on  $N$  that equals  $x$  at  $M$ . The Levi-Civita connection on  $M$  (resp.  $N$ ) is  $\nabla$  (resp.  $\tilde{\nabla}$ ). Since  $\phi$  is a riemannian immersion,  $\nabla_x y = d\pi(\tilde{\nabla}_{\tilde{x}} \tilde{y})$ . From this fact it follows that  $\nabla d\pi|_{T_p M}$  vanishes. On the other hand, since  $\pi \circ \exp_p|_{N_p} = p$ ,  $\nabla d\pi|_{N_p}$  vanishes. Finally, if  $x \in T_p M$  and  $v \in N_p$ , then  $\nabla d\pi(x, v) = -d\pi(\tilde{\nabla}_x v)$  since  $d\pi(v) = 0$ . Therefore, if  $z \in T_p M$  then

$$\begin{aligned} \langle z, \nabla d\pi(x, v) \rangle &= -\langle z, d\pi(\tilde{\nabla}_x v) \rangle = -\langle z, \tilde{\nabla}_x v \rangle = \langle \tilde{\nabla}_{\tilde{x}} \tilde{z}, v \rangle, \\ &= \langle v, \tilde{\nabla}_{\tilde{x}} \tilde{z} - \nabla_x z \rangle = \langle v, \nabla d\phi(x, z) \rangle, \\ &= \langle B'_x v, z \rangle. \end{aligned}$$

This completes the proof, since  $\nabla d\pi$  is bilinear.  $\square$

Let us compute the riemannian curvature tensor of  $M$  in terms of that of  $N$  and the curvature of the immersion  $\phi$ . From the fact that the Levi-Civita connection on  $M$  is obtained by orthogonally projecting the connection of  $N$ , we have that for all vector fields  $x, y, z$  on  $M$

$$\begin{aligned} R_{x,y}^M z &= \nabla_x (\nabla_y z) - \nabla_y (\nabla_x z) - \nabla_{[x,y]} z, \\ &= d\pi \left( \nabla_{\tilde{x}} d\pi \cdot \tilde{\nabla}_{\tilde{y}} \tilde{z} + \tilde{\nabla}_{\tilde{x}} \tilde{\nabla}_{\tilde{y}} \tilde{z} - \nabla_{\tilde{y}} d\pi \cdot \tilde{\nabla}_{\tilde{x}} \tilde{z} - \tilde{\nabla}_{\tilde{y}} \tilde{\nabla}_{\tilde{x}} \tilde{z} - \tilde{\nabla}_{[\tilde{x}, \tilde{y}]} \tilde{z} \right), \\ &= d\pi \left( R_{\tilde{x}, \tilde{y}}^N \tilde{z} \right) + (B'_x B_y - B'_y B_x) z, \end{aligned}$$

where we have used the identity  $\nabla_{\tilde{x}} d\pi \cdot \tilde{\nabla}_{\tilde{y}} \tilde{z} = B'_x \circ (1 - d\pi) \cdot \tilde{\nabla}_{\tilde{y}} \tilde{z}$ , since  $\nabla d\pi$  vanishes on the horizontal part. Since  $(1 - d\pi) \cdot \tilde{\nabla}_{\tilde{y}} \tilde{z} = \nabla d\phi(y, z)$ , this demonstrates the final line.

From this equation, it follows that

$$(50) \quad \text{Ric}^M(x) = d\pi \left( \sum_i R_{\tilde{x}, \tilde{e}_i}^N \tilde{e}_i \right) + \sum_i (B'_x B_{e_i} - B'_{e_i} B_x) e_i,$$

where  $x \in T_p M$  and  $e_i$  is an orthonormal basis of  $T_p M$ . Application of equation 50 to equation 9 yields

$$(51) \quad \langle d\phi, \text{Ric}_{d\phi} \rangle = |\nabla d\phi|^2 - |\tau(\phi)|^2.$$

The scalar curvature of  $M$  is the trace of the Ricci tensor, which equals  $\sum_{i,j} \langle e_i, \mathbf{R}_{e_i, e_j}^N e_j \rangle + |\tau(\phi)|^2 - |\nabla d\phi|^2$ , where we omit the  $\tilde{\cdot}$ . When  $\phi$  is the inclusion  $\iota$  of  $M$  into  $\mathbf{E}$ , we see that

$$(52) \quad \text{scal}_M = |\tau(\iota)|^2 - |\nabla d\iota|^2.$$

To compute  $\langle d\phi, \nabla \tau(\phi) \rangle$ , note that since  $\tau(\phi)$  is orthogonal to  $T_p M$ ,  $\langle d\phi \cdot e_i, \nabla_{e_i} \tau(\phi) \rangle + \langle \nabla_{e_i} d\phi \cdot e_i, \tau(\phi) \rangle$  vanishes for all  $i$ . Therefore

$$(53) \quad \langle d\phi, \nabla \tau(\phi) \rangle = -|\tau(\phi)|^2.$$

**Proposition 6.3.** *[Following the notation of section 4.] Let  $\gamma$  be a riemannian immersion. Then  $\Delta_{\mathfrak{E}}$  is the laplacian of  $(\Theta, \mathfrak{g})$ , and*

$$(54) \quad \kappa = \frac{5}{3} |\nabla d\iota|^2 - \frac{2}{3} |\tau(\iota)|^2 - \frac{1}{6} |\nabla d\gamma|^2 - \frac{1}{3} |\tau(\gamma)|^2.$$

If  $\gamma$  is totally geodesic (iff  $\nabla d\gamma = 0$  iff  $\gamma(\Theta)$  is totally geodesic), then

$$(55) \quad \kappa = \frac{5}{3} |\nabla d\iota|^2 - \frac{2}{3} |\tau(\iota)|^2.$$

*Proof.* From equation 53 and Theorem 5.1,  $\kappa = \frac{1}{2} |d\Gamma|^2 - \frac{2}{3} \langle d\Gamma, \text{Ric}_{d\Gamma} \rangle - |\tau(\gamma)|^2$ . It remains to compute the first two terms.

- (1) Since  $\Gamma = \gamma \circ \pi$ , one sees that  $\nabla d\Gamma = \nabla d\gamma(d\pi, d\pi) + d\gamma \cdot \nabla d\pi$ , which is an orthogonal decomposition. Since  $\gamma$  is a riemannian immersion,

$$|\nabla d\Gamma|^2 = |\nabla d\gamma|^2 + 2|\nabla d\iota|^2.$$

- (2) From equation 9, and the fact that  $\text{Ric}^{\mathbf{E}}$  vanishes,  $\langle d\Gamma, \text{Ric}_{d\Gamma} \rangle$  equals  $\sum_{i,j} \langle u_i, \mathbf{R}_{u_i, u_j}^{\Lambda} u_j \rangle$  where  $u_i = d\gamma(e_i)$ . Equation 9 and the hypothesis that  $\gamma$  is a riemannian immersion implies that this equals  $\langle d\gamma, \text{Ric}_{d\gamma} \rangle + \text{scal}_{\Theta}$ . Equations 51 and 52 imply that

$$\langle d\Gamma, \text{Ric}_{d\Gamma} \rangle = |\nabla d\gamma|^2 - |\tau(\gamma)|^2 - |\nabla d\iota|^2 + |\tau(\iota)|^2.$$

The two equations prove the formula for  $\kappa$ . □

**Corollary 6.4.** *The formulas in equation 49 are correct.*

*Proof.* 1/ If  $\gamma = \text{id}_{\Theta}$ , then  $\nabla d\gamma = 0$  so  $\tau(\gamma) = 0$ , also. 2/ When  $\gamma = \iota$  is the inclusion map, the equation is clearly correct. □

**6.2. Riemannian submersions.** Recall that  $(M, g) \xrightarrow{\phi} (N, h)$  is a riemannian submersion if  $d\phi|_D : (D, g|_D) \rightarrow (TN, h)$  is an isometry where  $D = (\ker d\phi)^{\perp}$ . In this case  $|\text{d}\phi|^2 = \dim N$ . Equation 9 implies that

$$(56) \quad \langle d\phi, \text{Ric}_{d\phi} \rangle = -\text{scal}_D + \text{scal}_N \circ \phi,$$

where  $\text{scal}_D$  is defined to be the trace of  $\text{Ric}^M|_D$ . Since  $|\text{d}\phi|$  is constant, the identity  $\frac{1}{2} \Delta |\text{d}\phi|^2 = \langle d\phi, \text{Ric}_{d\phi} \rangle - \nabla \tau(\phi) - |\nabla d\phi|^2$  implies that

$$(57) \quad |\nabla d\phi|^2 = \text{scal}_N \circ \phi - \text{scal}_D - \langle d\phi, \nabla \tau(\phi) \rangle.$$

**Proposition 6.5.** *[Following the notation of section 4.] Let  $\gamma$  be a riemannian submersion. Then*

$$(58) \quad \kappa = -\frac{1}{6} \text{scal}_\Lambda \circ \gamma + |\tau(\gamma)|^2 + \frac{3}{2} \langle d\gamma, \nabla \tau(\gamma) \rangle.$$

*If  $\gamma$  is harmonic (iff  $\tau(\gamma) = 0$ ) then*

$$(59) \quad \kappa = -\frac{1}{6} \text{scal}_\Lambda \circ \gamma.$$

*Let  $e_i$  be an orthonormal frame of  $E_\theta$ . The sublaplacian  $\Delta_{\mathbf{e}}$ , when applied to a smooth function  $f$ , equals*

$$(60) \quad \Delta_{\mathbf{e}} f = \sum_{i=1}^{\dim E} \nabla_{e_i, e_i}^2 f \quad \text{at } \theta.$$

*Proof.* Equation 56, when applied to the submersion  $\Gamma = \gamma \circ \pi$ , yields  $\langle d\Gamma, \text{Ric}_{d\Gamma} \rangle = \text{scal}_\Lambda \circ \Gamma$  since  $\mathbf{E}$  is flat. In addition, since  $|d\Gamma|$  is constant, equation 57 and lemma 2.6 implies that  $|\nabla d\Gamma|^2 = \text{scal}_\Lambda \circ \Gamma - \langle d\gamma, \nabla \tau(\gamma) \rangle$ . Equation 33, along with  $\pi \circ \iota = \text{id}_\Theta$ , implies equation 58.

For the proof of equation 60, let  $x^j$  be a system of normal coordinates centred at  $\theta$  and let  $f_j = \frac{\partial}{\partial x^j} + O(|x|^2)$  be an orthonormal frame. Assume that  $e_i = f_i$  for  $i = 1, \dots, \dim E$ . The bundle map  $\mu : T^*\Theta \rightarrow T\Theta$  that characterizes the subriemannian structure is written at  $\theta$  as

$$\mu = \sum_i^{\dim E} f_i \otimes f_i = \sum_i^{\dim E} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^i} + O(|x|^2).$$

Since the riemannian metric  $\mathbf{g} = \sum_{i=1}^n dx^i \otimes dx^i + O(|x|^2)$ , where  $n = \dim \Theta$ , the riemannian volume form  $d\theta = (1 + O(|x|^2)) dx^1 \wedge \dots \wedge dx^n$ , and so the sub-laplacian at  $\theta$  is

$$\Delta_{\mathbf{e}} = \sum_{i=1}^{\dim E} \frac{\partial^2}{(\partial x^i)^2},$$

which equals  $\sum_{i=1}^{\dim E} \nabla_{e_i, e_i}^2$  in invariant notation. □

**Remark.** If one compares the formulas in Propositions 6.3 and 6.5, one sees that both  $\Delta_{\mathbf{e}}$  and  $\kappa$  depend on the inclusion map  $\iota$  when  $\gamma$  is a riemannian immersion; when  $\gamma$  is a riemannian submersion  $\Delta_{\mathbf{e}}$  and  $\kappa$  do not depend on the inclusion map  $\iota$ . In the latter case, the geometric information carried by the riemannian submersion  $\gamma$  subsumes that carried by  $\iota$ .

#### REFERENCES

- [1] S. Amari. Differential-geometrical methods in statistics. Lecture notes in statistics, Springer-Verlag, Berlin, 1985
- [2] P. Berkhin and B. Levit. Second order asymptotically minimax estimates of the mean of a normal population, Problems of Information Transmission 16 (1980), 212-229.
- [3] J. Eells and L. Lemaire. Selected topics in harmonic maps. C.B.M.S. Regional Conference Series. AMS. 1983.
- [4] J. Jost. Riemannian geometry and geometric analysis. Springer-Verlag. Fourth edition. 2005.
- [5] Z. Landsman and B. Levit. Second order asymptotically minimax estimation in the presence of nuisance parameters, Problems of Information Transmission 26 (1990), 50-66.

- [6] B. Levit. On second order asymptotically minimax estimates, *Theor. Probab. Appl.* 25 (1980), 552-568.
- [7] B. Levit. Minimax estimation and the positive solutions of elliptic equations, *Theor. Probab. Appl.* 27 (1982), 525-546.
- [8] R. Montgomery. A tour of subriemannian geometries, their geodesics and applications. *Mathematical Surveys and Monographs*, 91. American Mathematical Society, Providence, RI, 2002.
- [9] A. Bloch. Nonholonomic mechanics and control. With the collaboration of J. Baillieul, P. Crouch and J. Marsden. With scientific input from P. S. Krishnaprasad, R. M. Murray and D. Zenkov. *Interdisciplinary Applied Mathematics*, 24. Systems and Control. Springer-Verlag, New York, 2003.

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`