INTRINSIC AND EXTRINSIC MEANS ON THE CIRCLE – A MAXIMUM LIKELIHOOD INTERPRETATION

Anders Brun[1], *Carl-Fredrik Westin*[2], *Magnus Herberthson*[3], *Hans Knutsson*[1]

[1] Department of Biomedical Engineering, Linköpings Universitet, Linköping, Sweden
[2] Laboratory of Mathematics in Imaging, Harvard Medical School, Boston, MA, USA
[3] Department of Mathematics, Linköpings universitet, Linköping, Sweden
email: andbr@imt.liu.se

ABSTRACT

For data samples in $\mathbb{R}^n$, the mean is a well known estimator. When the data set belongs to an embedded manifold M in $\mathbb{R}^n$, e.g. the unit circle in $\mathbb{R}^2$, the definition of a mean can be extended and constrained to M by choosing either the intrinsic Riemannian metric of the manifold or the extrinsic metric of the embedding space. A common view has been that extrinsic means are approximate solutions to the intrinsic mean problem. This paper study both means on the unit circle and reveal how they are related to the ML estimate of independent samples generated from a Brownian distribution. The conclusion is that on the circle, intrinsic and extrinsic means are maximum likelihood estimators in the limits of high SNR and low SNR respectively.

Index Terms— Signal Processing, Maximum likelihood estimation, Signal representations, Diffusion equations

1. INTRODUCTION

The mean of a set of scalar- or vector-valued data points is a well known quantity, often used to estimate a parameter in presence of noise. Manifold-valued data is gaining importance in applications and for this kind of data several extensions of the mean have been proposed[1, 2, 3]. While the mean for scalar- and vector-valued data can be defined as a point in the data space minimizing the sum of squared distances to all the other points, the natural extension to manifold-valued data is to replace the metric and restrict the search to a minimizer on the manifold.

1.1. The Intrinsic Mean

The intrinsic mean for a set of N data points x_i in a compact manifold M is defined using the Riemannian metric $d_M(x, y)$, i.e. the geodesic distance between two points x and y in the manifold [1]:

$$\overline{x}_{\text{int}} = \arg\min_{q \in M} \sum_{i=1}^{N} d_M^2(x_i, q).$$

While the (set of) global minimizer(s) might be difficult to compute, one may look for local minimizers, which can be guaranteed to be unique if the distributions of points x_i are enough localized in M [1]. The intrinsic mean is often seen as the natural generalization of means to manifold-valued data. The drawback is that it is relatively complicated to compute, when implemented as a (local) minimization over a non-linear manifold. The procedure is illustrated in Fig. 1.

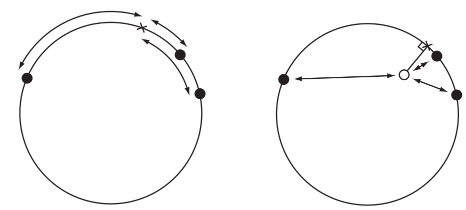

Fig. 1. A schematic view of how the intrinsic mean (**left**) and extrinsic mean (**right**) are calculated on $\mathbb{S}^1$. Black dots are data points and the means are marked by crosses. The intrinsic mean is a point on $\mathbb{S}^1$ minimizing the sum of squared intrinsic distances (curved arrows), while the extrinsic mean is a point on the circle minimizing the sum of squared extrinsic distances (straight arrows). The white dot is an intermediate result in the calculation of the extrinsic mean, i.e. the mean of the data points in the extrinsic space $\mathbb{R}^2$, which is followed by an orthogonal projection back to $\mathbb{S}^1$. This procedure is equivalent to the minimization in (1), which explains the popularity of the extrinsic mean [3].

1.2. The Extrinsic Mean

When the manifold M is embedded in a Euclidean space, $\mathbb{R}^n$, it is sometimes faster to calculate the so called extrinsic mean. This involves two steps: 1) Calculation of the mean of the data points seen as vectors in the Euclidean space. 2) A shortest distance projection back to the manifold. This is illustrated in Fig. 1 and is equivalent [3] to solving the following mini-

We acknowledge the support of the Manifold Valued Signal Processing project by the Swedish Research Council and the SIMILAR network of excellence (http://www.similar.cc). Also thanks to Thomas Schön for valuable comments on the manuscript and to Steven Haker for initial discussions about the diffusion equation.

mization problem

$$\overline{x}_{\text{ext}} = \arg\min_{q\in M} \sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{q}|^2. \tag{1}$$

It is essentially the same expression as for the intrinsic mean, except that the Riemannian metric is replaced by the Euclidean metric. Note that boldface, e.g. $\mathbf{q}$, is used when we may interpret the point as a vector in a vector space $\mathbb{R}^n$, while q is used for a point in a general manifold M and sometimes refer to its coordinate (angle). If there exists a natural embedding of a manifold for which the shortest projection back to the manifold is easy to compute, then the main advantage of the extrinsic mean is that iterative optimization over a non-linear manifold can essentially be replaced by a two-step procedure. This is the case for the unit circle, $\mathbb{S}^1$, but also for other compact symmetric manifolds such as n-dimensional spheres, $\mathbb{S}^n$, and n-dimensional projective spaces $\mathbb{RP}^n$.

2. MODELING NOISE BY BROWNIAN MOTION

It is well known that the mean for a set of data points in $\mathbb{R}^n$ is also the maximum likelihood (ML) estimate of $\mathbf{x}$ for the model $\mathbf{x}_i = \mathbf{x} + \mathbf{n}_i$ where the noise is modeled by a Gaussian distribution, $\mathbf{n}_i \in N(\mathbf{0}, \sigma\mathbf{I})$, generating a set of independent and identically distributed (i.i.d.) data points. In $\mathbb{R}^n$ the Gaussian distribution is also a model for Brownian motion, i.e. the resulting distribution of a random walk or diffusion process. The concept of diffusion is easy to extend to manifolds in general and for this reason we choose to model noise by a Brownian distribution. We will now start with an interpretation of the mean value as the ML estimate for a model where noise in $\mathbb{R}^n$ is modeled using Brownian motion and then proceed to the case of Brownian noise on $\mathbb{S}^1$.

2.1. Means as ML estimates in $\mathbb{R}^n$

The isotropic Gaussian distribution in $\mathbb{R}^n$ is related to Brownian motion and the diffusion equation, which is also equivalent to the heat equation. Given a distribution $I(\mathbf{p}, 0)$, describing the amount of particles at position $\mathbf{p}$ and time $t = 0$, the diffusion equation states

$$I_t(\mathbf{p}, t) = D\Delta_p I(\mathbf{p}, t) \tag{2}$$

where D is the diffusion coefficient, I_t is the derivative of I w.r.t. time and Δ_p is the Laplacian operator acting in the spatial domain. Since D is not important in this paper, we let $D = 1/4$ for simplicity. The solution to the diffusion equation at a time t is obtained by convolution in the spatial domain,

$$I(\mathbf{p}, t) = \int_{\mathbb{R}^n} K(\mathbf{p}, \mathbf{q}, t) I(\mathbf{q}, 0) dq.$$

$K(\mathbf{p}, \mathbf{q}, t)$ is the so called diffusion kernel,

$$K(\mathbf{p}, \mathbf{q}, t) = \frac{1}{(\pi t)^{n/2}} \exp\left[-\frac{|\mathbf{p} - \mathbf{q}|^2}{t}\right].$$

To study the behavior of a single particle moving according to a Brownian motion diffusion process, one may choose $I(\mathbf{p}, 0)$ to be a Dirac function $\delta(\mathbf{p} - \mathbf{x})$.

Modeling noise using a Brownian (Gaussian) distribution in $\mathbb{R}^n$ now yields the following likelihood function for a set of i.i.d. data points:

$$\begin{aligned} L(\mathbf{x}) &= P(\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_N | \mathbf{x}) \\ &= P(\mathbf{x}_1|\mathbf{x}) P(\mathbf{x}_2|\mathbf{x}) \ldots P(\mathbf{x}_N|\mathbf{x}) \\ &= C_1 \prod_{i=1}^{N} \exp\left[-\frac{1}{t}(\mathbf{x}_i - \mathbf{x})^T(\mathbf{x}_i - \mathbf{x})\right] \\ &= C_1 \exp\left[-\frac{1}{t}\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{x})^T(\mathbf{x}_i - \mathbf{x})\right] \\ &= C_2 \exp\left[-\frac{1}{t}N(\overline{\mathbf{x}} - \mathbf{x})^T(\overline{\mathbf{x}} - \mathbf{x})\right], \end{aligned}$$

for some constants C_1 and C_2. From this we see that regardless of t, the ML estimate of $\mathbf{x}$ is the mean $\overline{\mathbf{x}}$. We also note that both the intrinsic and extrinsic mean in $\mathbb{R}^n$ is $\overline{\mathbf{x}}$, since $\mathbb{R}^n$ is flat.

2.2. Intrinsic Means as ML estimates in $\mathbb{S}^1$

Given the results for $\mathbb{R}^n$ it is a reasonable approach to investigate the ML estimate of i.i.d. Brownian distributions on $M = \mathbb{S}^1$, the unit circle. The diffusion kernel on $\mathbb{S}^1$ can be modeled using a wrapped Gaussian distribution [4],

$$K(p, q, t) = \frac{1}{\sqrt{\pi t}} \sum_{k=-\infty}^{+\infty} \exp\left[-\frac{(d_M(p, q) + 2\pi k)^2}{t}\right]. \tag{3}$$

Modeling noise using $P(x_i|x) = K(x_i, x, t)$ gives an expression for the likelihood, similar to the case for $\mathbb{R}^n$, which we seek to maximize,

$$\begin{aligned} \arg\max_{x\in M} L(x) &= \arg\max_{x\in M} P(x_1, x_2 \ldots x_N|x) \\ &= \arg\max_{x\in M} P(x_1|x)P(x_2|x)\ldots P(x_N|x) \\ &= \arg\max_{x\in M} \sum_{i=1}^{N} \log(P(x_i|x)). \end{aligned}$$

Finding the ML estimate in the general case is difficult and for this reason we first study what happens in the limit when $t \to 0^+$. Due to a formula by Varadhan [5, 4], it is known that

$$\lim_{t\to 0^+} t\log(K(p, q, t)) = -\frac{d_M^2(p, q)}{2}$$

uniformly in $(p, q) \in \mathbb{S}^1 \times \mathbb{S}^1$. For any fix $t > 0$ we have

$$\arg\max_{x\in M} \log(L(x)) = \arg\max_{x\in M} t\log(L(x)),$$

and for this reason

$$\begin{aligned}\lim_{t\to 0^+}\arg\max_{x\in M} L(x) &= \lim_{t\to 0^+}\arg\max_{x\in M} t\log(L(x)) \\ &= \arg\max_{x\in M}\sum_{i=1}^{N} -\frac{d_M^2(x,x_i)}{2} \\ &= \arg\min_{x\in M}\sum_{i=1}^{N} d_M^2(x,x_i) \\ &= \overline{x}_{int}.\end{aligned}$$

This means that the above ML estimate converges to $\overline{x}_{int}$ when $t \to 0^+$.

2.3. Extrinsic Means as ML estimates in $\mathbb{S}^1$

Since $L(x)$ approached $\overline{x}_{int}$ in the limit $t \to 0^+$, it is now interesting to also investigate the behavior when $t \to \infty$. Instead of direct use of (3), Fourier series are applied to solve (2) to obtain the diffusion kernel on $\mathbb{S}^1$ [6]. At $t = 0$,

$$\begin{aligned}K(p,q,0) &= \delta(d_M(p-q)) \\ &= \frac{1}{2}A_0 + \sum_{n=1}^{\infty}\left(A_n\cos(np) + B_n\sin(np)\right), \\ A_n &= \frac{1}{\pi}\cos(nq) \quad (n = 0,1,2,\ldots) \\ B_n &= \frac{1}{\pi}\sin(nq) \quad (n = 1,2,3,\ldots),\end{aligned}$$

where p and q are either points on $\mathbb{S}^1$ or angles in the interval $[-\pi, \pi[$. This kernel evolves according to

$$K(p,q,t) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty} e^{-n^2t/4}[A_n\cos(np) + B_n\sin(np)].$$

Once again, the data is modeled by $P(x_i|x) = K(x_i, x, t)$. We observe that

$$P(x_i|x) = \frac{1}{2\pi} + \varepsilon[A_1\cos(x_i) + B_1\sin(x_i)] + O(\varepsilon^2)$$

where $\varepsilon \to 0$ when $t \to \infty$. Thus when $t \to \infty$, the likelihood function is

$$\begin{aligned}L(x) &= \prod_{i=1}^{N} P(x_i|x) \\ &= \frac{1}{(2\pi)^N} + \frac{\varepsilon}{2\pi}\sum_{i=1}^{N}[A_1\cos(x_i) + B_1\sin(x_i)] \\ &\quad +O(\varepsilon^2).\end{aligned}$$

Any such likelihood function will converge towards a constant value, $L(x) \to 1/(2\pi)^N$, when $t \to \infty$. The dominant terms however, important for finding the maximum of $L(x)$, are generically A_1 and B_1 and

$$\begin{aligned}\arg\max_{x\in M} L(x) &= \arg\max_{x\in M}\sum_{i=1}^{N}\cos x\cos x_i + \sin x\sin x_i \\ &= \arg\max_{x\in M}\sum_{i=1}^{N}\mathbf{x}^T\mathbf{x}_i = \\ &= \overline{\mathbf{x}}/|\overline{\mathbf{x}}| = \overline{x}_{ext}.\end{aligned}$$

Strange as it might seem, searching for the maximizer of a function which converges towards a constant value, it will in fact always exist a unique maximum for every $0 < t < \infty$, and generically also a unique maximizer.

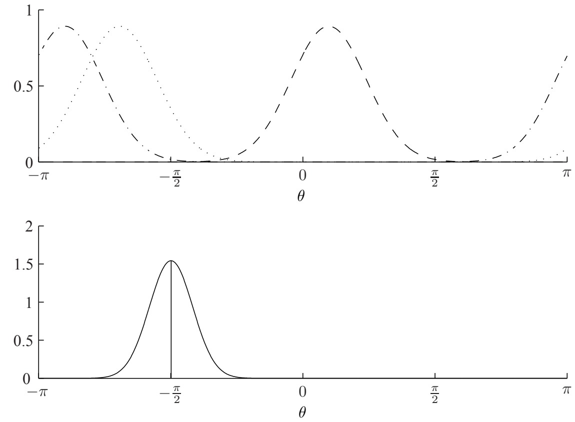

Fig. 2. **Top:** Three samples x_i have been collected on $\mathbb{S}^1$, -2.80, -2.11 and 0.34. For $t = 0.1$ their individual likelihood functions look like in the plot. **Bottom:** The total normalized likelihood function $L(x)$ peaks around -1.52, which is close to the intrinsic mean: $\overline{x}_{\text{int}} = (-2.80 - 2.11 + 0.34)/3 \approx -1.52$.

3. EXPERIMENTS

To verify the results we implemented the diffusion equation on the unit circle in MATLAB and calculated the likelihood as a function of t. The results on a small data set x_i are shown for three choices of t in Fig. 2–4.

4. DISCUSSION

In this paper, we let a Brownian distribution replace the traditional Gaussian distribution. By varying the parameter t we model the variance of the noise in the i.i.d. samples (measurements) $x_i \in \mathbb{S}^1$. The signal model is a constant manifold-valued function with the value $x \in \mathbb{S}^1$. Both the theoretical analysis and the experiments in this paper show that the intrinsic and extrinsic means on $\mathbb{S}^1$ can be regarded as ML estimates in the limits of high and low SNR respectively for this particular choice of models.

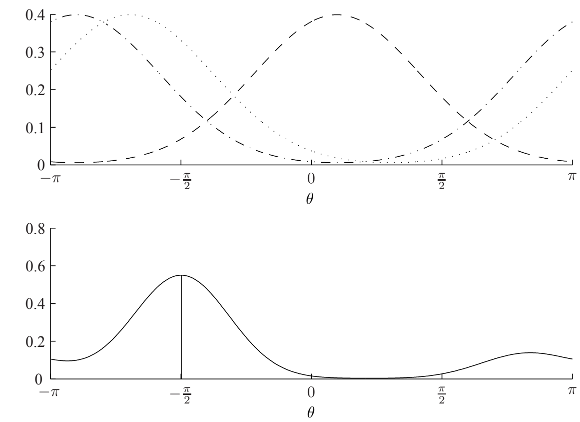

Fig. 3. Same as in Fig. 2, but $t = 0.5$. **Top:** Individual likelihood functions. **Bottom:** The total normalized likelihood.

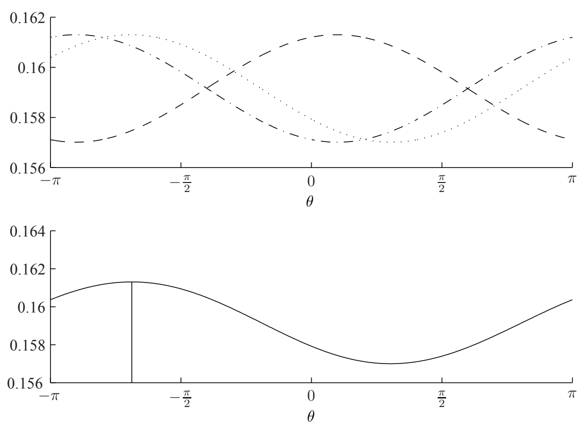

Fig. 4. Same as in Fig. 2, but $t = 1.0$. **Top:** Individual likelihood functions. **Bottom:** The total normalized likelihood peaks around -2.11, which is close to the extrinsic mean:
$\overline{x}_{\text{ext}} = \tan^{-1} \frac{(\sin(-2.80)+\sin(-2.11)+\sin(0.34))}{(\cos(-2.80)+\cos(-2.11)+\cos(0.34))} - \pi \approx -2.11.$

A close inspection of the experiment shown in Fig. 2–4, for a wider range of t than shown in the figures, revealed convergence to both the intrinsic and extrinsic mean when $t \to 0^+$ and $t \to \infty$. The only reason for not including figures of experiments with very large or small t in this paper was the difficulty in obtaining a reasonable scaling of the plots. In Fig. 3 we observe the possibility of several local maxima for certain choices of t, while Fig. 2 and 4 demonstrate the typical behavior in the limits.

The result of this paper points towards a more balanced view of the intrinsic and extrinsic means, since they are both extreme cases for our model on $\mathbb{S}^1$. Other researchers, see for instance [2], have regarded the intrinsic mean for e.g. rotation matrices as the "natural" mean, while the extrinsic mean has been regarded as an approximation. The question is if a more balanced view, advocated in this paper for $\mathbb{S}^1$, is valid for a general compact manifold M.

Due to the generality of Varadhan's formula [5, 4], it is in fact possible to extend the results for the ML estimate when $t \to 0^+$, from $\mathbb{S}^1$ to any connected and compact manifold. This gives a probabilistic motivation for intrinsic means on such manifolds in general. Indirectly it also motivates the use of the squared geodesic distance, $d_M^2(x, y)$, as a building block in other estimates on manifolds, for instance estimates facilitating basic interpolation and filtering. While this paper show the essence of the idea on $\mathbb{S}^1$, the details for the general case will be investigated in future research.

Despite the apparent symmetry of intrinsic and extrinsic means on $\mathbb{S}^1$ presented in the paper, extending the results for the extrinsic mean and the ML estimate when $t \to \infty$ to general manifolds will not be as easy as for the case $t \to 0^+$ hinted above. In particular, the extrinsic mean depends on how the manifold M is embedded in $\mathbb{R}^n$. For "natural" embeddings of certain symmetric and compact manifolds, such as $\mathbb{S}^n$ and $\mathbb{RP}^n$, which also include important special cases such as the sphere $\mathbb{S}^2$ and the group of rotations in $\mathbb{R}^3$, we do expect that the ML estimate will converge towards the extrinsic mean when $t \to \infty$. Thus we expect that future research will give a probabilistic motivation, based on the Brownian model of noise, for extrinsic means on e.g. the unit spheres and rotation matrices in $\mathbb{R}^n$.

In summary, this paper has revealed a more balanced view on intrinsic and extrinsic means on $\mathbb{S}^1$, which shows the essence of an idea which we believe is useful for the understanding of a wider class of algorithms performing signal processing and estimation on manifold-valued signals and data.

5. REFERENCES

[1] X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," in *Proc. of Nonlinear Signal and Image Processing*, Antalya, Turkey, 1999, IEEE-EURASIP, pp. 194–198.

[2] C. Gramkow, "On averaging rotations," *Journal of Mathematical Imaging and Vision*, vol. 15, pp. 7–16, 2001.

[3] A. Srivastava and E. Klassen, "Monte carlo extrinsic estimators for manifold-valued parameters," *Special issue of IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 299–308, February 2002.

[4] D. W. Strook and J. Turetsky, "Short time behavior of logarithmic derivatives of the heat kernel," *Asian J. Math.*, vol. 1, no. 1, pp. 17–33, March 1997.

[5] S. R. Varadhan, "Diffusion processes in a small time interval," *Comm. Pure Appl. Math.*, vol. 20, pp. 659–685, 1967.

[6] Walter A. Strauss, *Partial differential equations: An introduction*, John Wiley & Sons Inc., New York, 1992.