Control of Sensing by Navigation on Information Gradients

(Invited Paper)

Sofia Suvorova and Bill Moran
University of Melbourne
Parkville 3010 VIC, Australia

Stephen D. Howard
Defence Science & Technology Organisation
Edinburgh 5111 SA, Australia

Douglas Cochran
Arizona State University
Tempe AZ 85287-5706, USA

***Abstract*—In estimation of parameters residing in a smooth manifold from sensor data, the Fisher information induces a Riemannian metric on the parameter manifold. If the collection of sensors is reconfigured, this metric changes. In this way, sensor configurations are identified with Riemannian metrics on the parameter manifold. The collection of all Riemannian metrics on a manifold forms a (weak) Riemannian manifold, and a smooth trajectory of sensor configurations manifests as a smooth curve in this space. This paper develops the idea of sensor management by following trajectories in the space of sensor configurations that are defined locally by gradients of the metric this space inherits from the abstract space of all Riemannian metrics on the parameter manifold. Theory is developed and computational examples that illustrate sensor configuration trajectories arising from this scheme are presented.**

***Index Terms*—Sensor management, Information geometry**

I. Introduction

Beginning in the 1930s, geometrical aspects of estimation problems started to receive notice in the statistical literature [1], [2]. Nearly simultaneously, in the mid-1940s, C. R. Rao [3] and H. Jeffreys [4] recognized that Fisher information induces a Riemannian metric structure on a smooth manifold of parameters that index a collection of probability distributions. This pioneering work was the genesis of the area now known as information geometry, which owes much of its modern development to S. Amari [5] and his collaborators.

Information-theoretic ideas, together with some information-geometric concepts, have played important roles in the emergence of sensor management over the past fifteen years (see [6, Ch. 3], [7], and the references therein). In particular, various f-divergences [8] have proven useful in defining proxies for actual operational objective functions in this context. Quite recently, some preliminary work has appeared that seeks to capitalize on the geodesic structure of statistical manifolds to inform the design of smooth trajectories in the space of controllable parameters describing the configuration of a sensor system [9], [10].

This paper focuses on the situation where a parameter θ in a smooth parameter manifold is to be estimated using data collected by sensors. The configuration of the sensor suite is assumed to be adjustable within a smooth finite-dimensional manifold of sensor configurations in a smooth fashion, and the objective is to determine what trajectory of the sensor configuration (locally) optimizes information collection in terms of estimating θ. The starting point is the observation that each sensor configuration corresponds to a Riemannian metric on the parameter manifold via the Fisher information corresponding to the problem of estimating θ from data collected in that particular sensor configuration. Established results in Riemannian geometry, outside the context of information geometry, show that the collection of all Riemannian metrics on a Riemannian manifold is itself an (infinite-dimensional) Riemannian manifold. Riemannian metrics arising from Fisher information are a sub-manifold of this geometrical object, and the collection of Riemannian metrics corresponding to Fisher information for actual sensor configurations is typically a finite-dimensional sub-manifold. Subsequent sections develop further this perspective to identify the best sensor management action to take, in terms of gathering the most information relevant to the estimation objective, is locally characterized in terms of geodesic curves in this space.

Section II summarizes the elements of information geometry that are essential to later developments in the paper. A sensor model based that arises from considering the Riemannian manifold of all Riemannian metrics on the parameter manifold is set forth in Section III. The remaining sections describe a scheme for sensor management based on following geodesic paths in the manifold of Riemannian metrics on the parameter manifold and give illustrative numerical results from a simple example involving bearings-only emitter localization using one fixed and one mobile sensor.

II. Information-geometric Preliminaries

Consider a parametric family of conditional densities $p(x|\theta)$ for a random variable $\boldsymbol{x}$ on a probability space $\mathcal{X}$. The parameter θ is assumed to lie in a smooth d-dimensional manifold $\mathcal{M}$. For given x, $p(x|\theta)$ defines a likelihood function for θ on $\mathcal{M}$ for which the log-likelihood function $\ell : \mathcal{M} \to \mathbb{R}$ is defined by $\ell_x(\theta) = \log p(x|\theta)$. Optimal tests for θ versus θ' given data x are of the form

$$\to \theta = \left\{ \log \frac{p(x|\theta)}{p(x|\theta')} > \tau \right\} = \{\ell_x(\theta) - \ell_x(\theta') > \tau\}$$

where τ is a threshold. The Kullback-Leibler (KL) divergence [11], [12]

$$D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')] = \int_{\mathcal{X}} p(x|\theta) \log \frac{p(x|\theta)}{p(x|\theta')}\, dx$$

is a natural measure of discrimination on $\mathbf{M}$. It is assumed throughout this paper that the parametric family of densities is such that $D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')]$ is well-defined for all $\theta,\ \theta' \in \mathbf{M}$; this is a reasonable assumption for most statistical distributions associated with sensing.

A. Divergences

For any choice of connection on $\mathbf{M}$, the second covariant derivative of $D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')]$ at $\theta = \theta'$, defined by that connection, is

$$\nabla^2_\theta D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')]|_{\theta=\theta'} = -\mathsf{E}[\nabla^2 \ell] = \mathsf{E}[d\ell \otimes d\ell]. \quad (1)$$

This object is the *Fisher Information* of the family of distributions $p(x|\theta)$.

The KL divergence exhibits the following three characteristic properties of divergences (see [5]):

1) $D[p(\cdot|\theta) \,\|\, p(\cdot|\theta)] = 0$
2) $D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')] \geq 0$
3) $d_\theta D[p(\cdot|\theta) \,\|\, p(\cdot|\theta')]|_{\theta=\theta'} = \mathsf{E}[d\ell] = 0$

B. The Fisher information metric

A Riemannian metric on a smooth manifold is a (positive definite) inner product on each tangent space that varies smoothly from point to point. Although the KL divergence is not symmetric, it induces a Riemannian metric on $\mathbf{M}$. Specifically, the *Fisher information metric* is defined as in (1) by $F = \mathsf{E}[d\ell \otimes d\ell]$. The corresponding volume form $\mathrm{vol}_F = \sqrt{\det F}\; d\theta_1 \wedge \cdots \wedge d\theta_d$ is the *Jeffreys prior* on the parameter manifold $\mathbf{M}$.

The Fisher information metric may be expressed with respect to local coordinates in the tangent space on $\mathbf{M}$ at θ as follows. Consider a pair of tangent vectors u and v at $\theta \in \mathbf{M}$:

$$u = \sum_{j=1}^{d} u_j \frac{\partial}{\partial \theta_j}\bigg|_\theta$$
$$v = \sum_{j=1}^{d} v_j \frac{\partial}{\partial \theta_j}\bigg|_\theta$$

where the u_j and v_j are respective coordinates of u and v in the tangent space of $\mathbf{M}$ at θ. In terms of these coordinates,

$$\begin{aligned} F_\theta(u,v) &= \mathsf{E}\{d\ell(u) d\ell(v)\} \\ &= \mathsf{E}\{u(\ell)v(\ell)\} = \sum_{j,k=1}^{d} u_j\, [F_\theta]_{jk}\, v_k \end{aligned}$$

where

$$[F_\theta]_{jk} = \int_{\mathcal{X}} \frac{\partial \ell_x(\theta)}{\partial \theta_j} \frac{\partial \ell_x(\theta)}{\partial \theta_k} p(x|\theta) dx$$

III. Riemannian Manifold of Riemannian Metrics

Work in differential geometry [13], [14], outside the context of information geometry, has shown that the collection of all Riemannian metrics on a smooth manifold $\mathbf{M}$ is an infinite-dimensional (weak) Riemannian manifold $\mathcal{M}(\mathbf{M})$. A point in $\mathcal{M}$ is a Riemannian metric on $\mathbf{M}$: it associates a positive definite form g_θ with each point θ in the parameter manifold $\mathbf{M}$. In other words, each point in $\mathcal{M}$ corresponds to a positive definite $(0,2)$-tensor field on $\mathbf{M}$. A tangent vector at a point in $\mathcal{M}$ associates a form h_θ with each point in $\mathbf{M}$; that is, a general symmetric $(0,2)$-tensor field on $\mathbf{M}$. It is assumed that such a tensor field is smooth. The collection of such tensor fields; that is the smooth sections of the symmetric $(0,2)$-tensor bundle is denoted by $\mathbf{\Gamma}(\mathcal{S}T^2)$.

Under assumptions discussed in [14], a metric on $\mathcal{M}$ is given by

$$G_g(h,k) = \int_{\mathbf{M}} \mathbf{Tr}\left(g_\theta^{-1} h_\theta g_\theta^{-1} k_\theta\right) \mathrm{vol}(g), \quad (2)$$

where h and k are smooth $(0,2)$-tensor fields on $\mathbf{M}$. In [14] it is assumed that the tensor fields have compact support. This is inappropriate for "perturbations" of Riemannian metrics in the context of sensing; the effects of changing a sensor will typically affect all of the parameter space, albeit in a diminishing way at infinity. To make (2) valid the tangent space $T_G(\mathcal{M})$ would have to be defined as the set of $h \in \mathbf{\Gamma}(\mathcal{S}T^2)$ satisfying

$$\int_{\mathbf{M}} ||G^{-1}(\theta)h(\theta)||^2\, \mathrm{vol}(G(\theta)) < \infty \quad (3)$$

or a similar constraint. Such a constraint is too severe for many natural sensing problems on manifolds with infinite volume forms; it can be seen from even simple examples that this constraint is rarely satisfied.

If $\mathbf{M}$ is compact, the volume form in this definition is finite and so can be normalized to an uninformative Jeffreys prior probability density on $\mathbf{M}$, and elements of $\mathbf{\Gamma}(\mathcal{S}T^2)$ are bounded so that the integral (3) is finite and the metric (2) on $\mathcal{M}$ is well-defined. If the volume form is infinite, interpretation of the volume form as a probability density on the entire manifold is not possible.

For problems in sensor management it is desirable to accommodate a genuine prior probability distribution on $\mathbf{M}$ to replace the potentially infinite Jeffreys prior. In particular, in multi-stage sensor management applications, results of previous measurements will typically impart a posterior distribution on $\mathbf{M}$ which serves as a prior for the next measurement (or can be updated via a model of the system dynamics to provide a prior for the next measurement). A direct generalization of the metric (2) that incorporates a probability measure P on $\mathbf{M}$ is given by

$$G_g(h,k) = \int_{\mathbf{M}} \mathbf{Tr}\left(g^{-1} h g^{-1} k\right) dP(\theta). \quad (4)$$

Even in this case, the tensor fields need to be constrained to be bounded, but in the context of sensor management this is a a suitable constraint for the tangent space in the sense that the function $\theta \mapsto \|h_\theta\|$ is bounded. This constraint will be imposed in the definition of the tangent space of $\mathcal{M}$ for the remainder of the paper.

IV. Geometric View of Sensor Management

In estimation of $\theta \in \mathbb{M}$, the effect of a particular sensing action s is to produce a log-likelihood

$$\ell_s(\theta) = \log p(x|\theta, s)$$

Thus, choosing a sensing action is tantamount to selecting a Riemannian metric for the parameter manifold $\mathbb{M}$:

$$\begin{aligned} \text{Sensing Action } s &\Rightarrow \text{log-Likelihood on } \ell_s \text{ on } \mathbb{M} \\ &\Rightarrow \text{Fisher Information } F_s \text{ on } \mathbb{M} \\ &\Rightarrow \text{Riemannian metric on } \mathbb{M} \\ &\Rightarrow \text{Point in } \mathcal{M}(\mathbb{M}) \end{aligned}$$

Indeed in the absence of specific measurements, the Kullback-Leibler divergence, and its infinitesimal form the Fisher Information provide an appropriate performance measure for the sensor.

With this perspective, a sensor configuration corresponds to a point in $s \in \mathcal{M}$. It is possible to identify geodesic curves in $\mathcal{M}$ that correspond to smooth trajectories through s and optimize information collection in a sense that will be discussed further in Section IV-C. There is no guarantee, however, that these curves in $\mathcal{M}$ will correspond to viable sensor configurations. Indeed, they may include points in $\mathcal{M}$ that are Riemannian metrics on $\mathbb{M}$ that do not arise from Fisher information as discussed in Section II-B above. The following subsections first introduce the geodesic structure on $\mathcal{M}$ and then show how this can be used to develop geodesics in a sub-manifold of $\mathcal{M}$ corresponding to realizable sensor configurations.

A. Geodesics in $\mathcal{M}$

The geodesic structure of $\mathcal{M}$ has been studied outside the context of information geometry [14]. The "energy" integral on a smooth curve $\gamma : [0, 1] \to \mathcal{M}$ is

$$E_\gamma = \frac{1}{2} \int_0^1 \int_{\mathbb{M}} \mathrm{Tr}\left(\gamma(t)^{-1}\dot\gamma(t)\gamma(t)^{-1}\dot\gamma(t)\right) dP(\theta) dt \quad (5)$$

Geodesics in $\mathcal{M}$ are extremals of E_γ; as described in [14], they satisfy $\ddot\gamma = \dot\gamma\gamma^{-1}\dot\gamma$, which has a solution

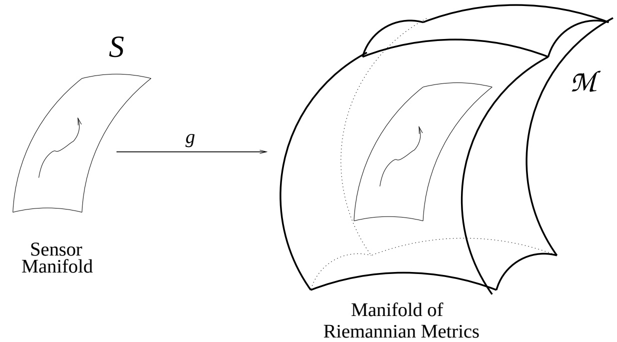
$$\gamma(t) = \gamma(0) \exp\left\{\gamma(0)^{-1}\dot\gamma(0)t\right\},$$

so that, in particular,

$$\det \gamma(t) = \det \gamma(0) e^{\mathrm{Tr}(\gamma(0)^{-1}\dot\gamma(0))t}.$$

B. Sensor Manifold

Assume that the collection of all realizable sensor configurations is parameterized by a smooth, finite-dimensional "sensor manifold" S. A configuration $s \in S$ gives rise to a particular Riemannian metric $g(s)$ on $\mathbb{M}$ and thus a particular point in $g(s) \in \mathcal{M}$. The mapping $g : S \to \mathcal{M}$ taking s to $g(s)$ will be assumed to be smooth embedding.



Although $\mathcal{M}$ is infinite-dimensional, the trajectory planning takes place in the finite-dimensional sub-manifold $g(S)$, and this sub-manifold inherits a metric structure from $\mathcal{M}$. Specifically, the induced metric at a point $s \in S$ is the pullback g^*G:

$$\begin{aligned} G_s(u, v) &= (g^*G)_s(u, v) = G_{g(s)}(g_{*s}u, g_{*s}v) \\ &= \int_{\mathbb{M}} \mathrm{Tr}\left(g_s^{-1} g_{*s}(u) g_s^{-1} g_{*s}(v)\right) dP(\theta) \end{aligned} \quad (6)$$

where u and v are in the tangent space TS_s and g_* is the push-forward of g.

For a smooth curve $\gamma : [0, 1] \to S$ the energy integral (5) restricts to

$$\begin{aligned} E_\gamma = \frac{1}{2} \int_0^1 \int_{\mathbb{M}} \mathrm{Tr}\big(&g(\gamma(t))^{-1} g_*(\dot\gamma(t)) \\ &\cdot g(\gamma(t))^{-1} g_*(\dot\gamma(t))\big) dP(\theta) dt. \end{aligned}$$

The geodesics, which are the extremal curves of E_γ, satisfy

$$\ddot\gamma = -\Gamma_\gamma(\dot\gamma, \dot\gamma)$$

where Γ denotes the Christoffel symbol for the Levi-Civita connection on S that is calculated from the metric (6).

C. A remark on Divergences and Metrics

In [9], the Riemannian metrics on $\mathcal{M}$ obtained from two divergences were derived. The first was the Kullback-Leibler divergence, corresponding to Fisher information on $\mathbb{M}$. The other was a "pseudo-divergence" on $\mathcal{M}$ motivated by symmetrized mutual information, a Shannon information construct. Starting with either of these leads to exactly the same Riemannian metric and geodesic structure on $\mathcal{M}$, the one given by Gil-Medrano and Michor in [14].

V. Example

To provide a more concrete context in which to illustrate some of the concepts set forth in prior sections, it is helpful to examine how the theory manifests in a simple example problem. Two sensors and one stationary target (emitter) are located in the plane $\mathbb{R}^2$. The goal is to successively estimate the position of the target, beginning with a $\mathcal{N}(0, \lambda\mathbb{I})$ prior, from bearings-only measurements taken at the sensors.

One of the sensors is fixed at position $(1, 1)$ while the other is on a mobile platform with constrained maximum speed, as depicted in Figure 1. Since only one sensor is mobile, the sensor management problem is to identify trajectory of that sensor's motion that will yield the best estimate of the target position. For this example, the parameter manifold $\mathcal{M}$ is $\mathbb{R}^2$

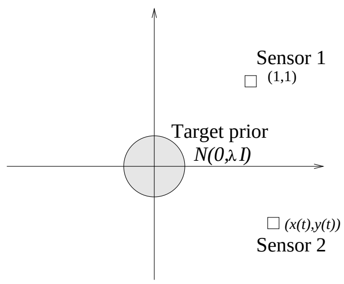


Fig. 1. The example scenario involves estimating the position (x_e, y_e) of a single stationary emitter from bearings-only measurements received at two sensors. One sensor is fixed at $(1, 1)$ and the other is mobile. The prior state of the target is normally discributed with mean $(0, 0)$ and covariance matrix $\lambda\mathbb{I}$.

with $\theta \in \mathcal{M}$ representing the position of the emitter in the plane. The Fisher information for this problem is calculated explicitly in [9]. Figure 2 shows the geodesic vector field for the motion the mobile sensor obtained at each point in the plane as the eigenvector corresponding to the largest eigenvalue of the Fisher information at that point.

Figure 3 shows the evolution of position of the mobile sensor and the target state estimate covariance ellipse as measurements are taken along the sensor trajectory (red points).

VI. Conclusion

Existing work connecting control of sensor configuration according to the geometric structure of a statistical manifold is still embryonic and far removed from practical sensor management algorithms. Even the small example introduced in V leads to extremely cumbersome equations that require burdensome computations which are prone to numerical issues arising from singularities. It is anticipated than much additional research will be necessary to mitigate these issues sufficiently to bring this conceptually promising circle of ideas to practice.

Acknowledgment

This work was supported in part by the U.S. Air Force Office of Scientific Research under grant No. FA9550-09-1-0561.

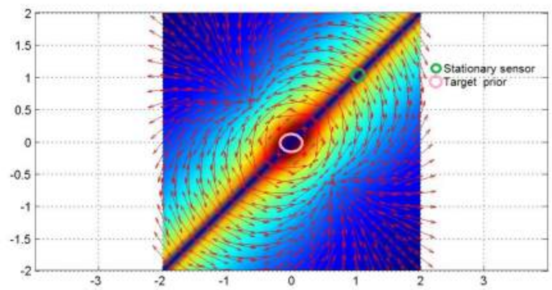


Fig. 2. Geodesic vector field for motion of the mobile sensor.

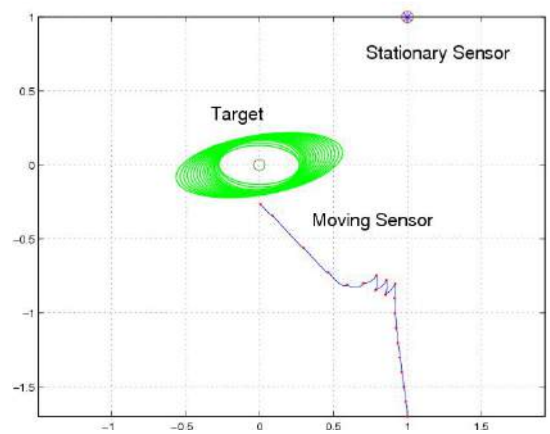


Fig. 3. Evolution of the target state covariance as samples are taken along the information-geometrically determined trajectory.

References

[1] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Science, India*, vol. 2, pp. 49–55, 1936.

[2] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–110, 1943.

[3] C. R. Rao, "Information and accuracy attainable in estimation of statistical parameters" *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[4] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London*. Series A, vol. 186, issue 1007, pp. 453–461, 1946.

[5] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS Translations of Mathematical Monographs, vol. 191, 2000.

[6] A. O. Hero, D. A. Castañón, D. Cochran, and K. Kastella, eds., *Foundations and Applications of Sensor Management*, New York, Springer, 2008.

[7] A. O. Hero III and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, December 2011.

[8] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, October 2006.

[9] W. Moran, S. D. Howard, D. Cochran, and S. Suvorova "Sensor management via Riemannian geometry," *Proceedings of the Allerton Conference on Communications, Control, and Computing*, pp. 358–362, October 2012.

[10] D. Cochran and A. O. Hero III, "Information-driven sensor planning: Navigating a statistical manifold," *Proceedings of the IEEE Global Conference on Signal and Information Processing*, December 2013 (in press).

[11] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[12] S. Kullback, *Information Theory and Statistics*, John Wiley and Sons, 1959.

[13] B. Clarke, "The metric geometry of the manifold of Riemannian metrics over a closed manifold," April 2009 (arXiv:0904.0174).

[14] O. Gil-Medrano and P. W. Michor, "The Riemannian manifold of all Riemannian metrics," *Quarterly Journal of Mathematics (Oxford)*, vol. 42, pp. 183–202, 1991.

[15] H. Jeffreys, *Theory of Probability*, Oxford University Press, 1961.

[16] R. E. Kass and L. Wasserman, "Formal rules of selecting prior distributions: A review and annotated bibliography," *Journal of the American Statistical Association*, vol. 91, pp. 1343–1370, 1996.

[17] W. Kúhnel, *Differential Geometry: Curves - Surfaces - Manifolds*, 2nd Edition, AMS Student Mathematical Library, vol. 16, 2005.