**Fiftieth Annual Allerton Conference**
**Allerton House, UIUC, Illinois, USA**
**October 1 - 5, 2012**

# Sensor Management via Riemannian Geometry

William Moran[1], Stephen D. Howard[2], Douglas Cochran[3], and Sofia Suvorova[1]

***Abstract*— In estimation of scenario parameters from sensor data, the Fisher information induces a Riemannian metric on the manifold of parameters. 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 smooth changes in configuration of the sensor suite manifests as a smooth curve in this space. This paper examines the idea of sensor management via navigation along geodesics in a sub-manifold of this space corresponding to physically viable sensor configurations; i.e., curves that optimize an energy integral in the sub-manifold. In addition to development of the theory, computational examples that illustrate sensor configuration trajectories arising from this scheme in small-scale scenarios are presented.**

## I. Introduction

Information geometry [1] brings a differential-geometric perspective to statistical estimation theory. In particular, it lends a rigorous geometric framework to the problem of measuring the power of measured data to discriminate values of parameters. Although the foundational ideas of information geometry date back to Rao [13], who showed that the Fisher information of a likelihood used in an estimation problem can be seen as a Riemannian metric on the parameter manifold, their application in the context of quantifying the value of potential sensor measurements for resource allocation in sensor management applications has only started to receive attention in the past few years (see [4, Ch. 3], [5], and the references therein). In particular, this recent work has demonstrated the use of mutual information and divergences as surrogates for actual operational objective functions in sensor management applications.

This paper focuses on the situation where a parameter $\theta$ 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 $\theta$. 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 $\theta$ 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 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.

The following section introduces the elements of information geometry that are essential to later developments in the paper. Subsequent sections proceed to describe the nature of a sensor model from an information-geometric viewpoint, to define the parameter manifold, the sensor manifold, its metric structure, and to derive a differential equation that characterizes geodesic curves on the sensor manifold. The introduction of informative prior distributions on the parameter manifold is one aspect that distinguishes this information-geometric development from previous differential-geometric work on the manifold of Riemannian metrics, in which attention is restricted to a volume form corresponding to the Jeffreys prior. Additionally, it is noted that the Riemannian metric with respect to which geodesics maximize "energy integrals" on the sensor manifold arises from both the Kullback-Leibler divergence and a "pseudo-divergence" corresponding to symmetrized mutual information. Finally, an example involving emitter position estimation using noisy bearings-only measurements from a pair of mobile sensor platforms is presented to illustrate the mathematical concepts developed in a more concrete setting.

## 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 $\theta$ is assumed to lie in a smooth $d$-dimensional manifold $M$. For given $x$, $p(x|\theta)$ defines a likelihood function for $\theta$ on $M$ for which the log-likelihood function $\ell : M \to \mathbb{R}$ is defined by $\ell_x(\theta) = \log p(x|\theta)$. Optimal tests for $\theta$ versus $\theta'$ given data $x$ are of the form

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

This work was supported in part by the Defence Science and Technology Organisation of Australia, the University of Michigan and the U.S. Army Research Office under MURI award No. W911NF-11-1-0391, and by the U.S. Air Force Office of Scientific Research under Grant No. FA9550-09-1-0561.

[1]W. Moran and S. Suvorova are with the University of Melbourne, Parkville VIC, Australia

[2]S. D. Howard is with the Defence Science and Technology Organisation, PO Box 1500, Edinburgh, 5111, Australia

[3]D. Cochran is with Arizona State University, Tempe AZ, USA

where $\tau$ is a threshold. The Kullback-Leibler (KL) divergence [9], [10]

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

is a natural measure of discrimination on $M$. 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 $\theta,\ \theta' \in M$; this is a reasonable assumption for most statistical distributions associated with sensing.

### A. Divergences

For any choice of connection on $M$, 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 [1]):

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 $M$. Specifically, the *Fisher information metric* is defined as in (1) by

$$F = \mathsf{E}[d\ell \otimes d\ell].$$

The corresponding volume form on the manifold

$$\mathrm{vol}_F = \sqrt{\det F}\, d\theta_1 \wedge \cdots \wedge d\theta_d$$

is the *Jeffreys prior* on the parameter manifold $M$.

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

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

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

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

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 [3], [2], outside the context of information geometry, has shown that the collection of all Riemannian metrics on a smooth manifold $M$ is an infinite-dimensional (weak) Riemannian manifold $\mathcal{M}(M)$. A point in $\mathcal{M}$ is a Riemannian metric on $M$: it associates a positive definite form $g_\theta$ with each point $\theta$ in the parameter manifold $M$. In other words, each point in $\mathcal{M}$ corresponds to a positive definite $(0,2)$-tensor field on $M$. A tangent vector at a point in $\mathcal{M}$ associates a form $h_\theta$ with each point in $M$; that is, a general symmetric $(0,2)$-tensor field on $M$. 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 $\Gamma(\mathcal{S}T^2)$.

Under assumptions discussed in [3], a metric on $\mathcal{M}$ is given by

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

where $h$ and $k$ are smooth $(0,2)$-tensor fields on $M$. In [3] 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 $h \in \Gamma(\mathcal{S}T^2)$ satisfying

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

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 $M$ is compact, the volume form in this definition is finite and so can be normalized to an uninformative Jeffreys prior probability density on $M$, and elements of $\Gamma(\mathcal{S}T^2)$ 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 $M$ 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 $M$ 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 $M$ is given by

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

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 $\theta \in M$, 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 $M$:

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

Indeed in the absence of specific measurements, the Kullback-Leibler divergence, and its infinitessimal 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 V. 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 $M$ 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 [3]. The "energy" integral on a smooth curve $\gamma : [0, 1] \to \mathcal{M}$ is

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

Geodesics in $\mathcal{M}$ are extremals of $E_\gamma$; as described in [3], they satisfy $\ddot\gamma = \dot\gamma\gamma^{-1}\dot\gamma$, which has a solution

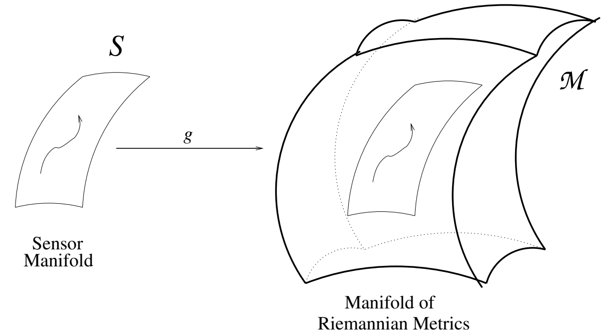
$$\gamma(t) = \gamma(0)\exp\left\{\gamma(0)^{-1}\dot\gamma(0)t\right\},$$

so that, in particular,

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

### 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 $M$ 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 $g^*G$:

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

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

$$I_\gamma = \frac{1}{2}\int_0^1 \int_M \operatorname{Tr}\left(g(\gamma(t))^{-1}g_*(\dot\gamma(t)) \cdot g(\gamma(t))^{-1}g_*(\dot\gamma(t))\right) dP(\theta)dt.$$

The geodesics, which are the extremal curves of $I_\gamma$, satisfy

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

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

## V. Divergences and Metrics

A divergence on a smooth manifold $R$ is a function $D[\cdot\|\cdot] : R \times R \to \mathbb{R}$ such that, for $p, q \in R$, $D[p \,\|\, q] \geq 0$ and $D[p \,\|\, q] = 0$ if and only if $p = q$. A divergence has the properties [1]

1) $\partial_p D[p \,\|\, q]|_{p=q} = 0$, and
2) $g = \nabla_p^2 D[p \,\|\, q]|_{p=q}$ defines a Riemannian metric on $R$.

To get a Riemannian metric per the second of these properties, it is not necessary that $D[p\|q] = 0$ if and only if $p = q$. Rather it is sufficient of this holds locally.

### A. Metrics on $\mathcal{M}$ from specific divergences

The Kullback-Leibler divergence between zero-mean normal distributions $\mathcal{N}(0, g)$ and $\mathcal{N}(0, h)$ is

$$\frac{1}{2}\operatorname{Tr}\left(gh^{-1} - \mathbb{I}\right) + \frac{1}{2}\log|hg^{-1}|$$

This leads to a divergence on $\mathcal{M}$

$$D_{\mathrm{KL}}(g \,\|\, h) = \int_M \frac{1}{2}\operatorname{Tr}\left(gh^{-1} - \mathbb{I}\right) + \frac{1}{2}\log|hg^{-1}|\, dP(\theta)$$

The corresponding Riemannian metric on $\mathcal{M}$ is

$$\nabla_g^2 D_{\mathrm{KL}}|_{g=h}(u,v) = \int_M \mathrm{Tr}\left(g^{-1}ug^{-1}v\right) dP(\theta),$$

which is exactly the metric defined in (4).

Similarly, a "pseudo-divergence" on $\mathcal{M}$ motivated by symmetrized mutual information can be defined by

$$D_{\mathrm{MI}}(g \,\|\, h) = \int_M \left\{ \log\left(\left|\frac{1}{2}(\mathbb{I}+g^{-1}h)\right|\right) + \log\left(\left|\frac{1}{2}(\mathbb{I}+h^{-1}g)\right|\right)\right\} dP(\theta)$$

$D_{\mathrm{MI}}$ is not a true divergence because it does not satisfy the condition $D_{\mathrm{MI}}[g \,\|\, h] = 0$ if and only if $g = h$. Nevertheless, it satisfies the local condition mentioned above so as to yield a metric on $\mathcal{M}$. Specifically, the corresponding Riemannian metric on $\mathcal{M}$ is

$$\nabla_g^2 D_{\mathrm{MI}}|_{g=h}(u,v) = \int_M \mathrm{Tr}\left(g^{-1}ug^{-1}v\right) dP(\theta)$$

The point of elaborating these two cases is that starting from either KL divergence (a Fisher information construct) or mutual information (a Shannon information construct) leads to exactly the same Riemannian metric and geodesic structure on $\mathcal{M}$. This suggests the possibility that this metric is a fundamental descriptor of information accumulation for estimation problems.

## VI. Example

To provide a more concrete context in which to illustrate some of the concepts that arise in later sections, it is helpful to examine how the theory manifests in an example problem. Suppose two mobile sensor platforms and one stationary target (emitter) are located in the plane $\mathbb{R}^2$, as depicted in Figure 1. The goal is to estimate the position of the target from bearings-only measurements taken at the sensors. Since the sensors are mobile, the sensor management problem is to identify the trajectories of sensor motion that will yield the best estimate of the target position. More specifically, the target position is $(x_e, y_e)$ and the sensor positions are $(x_j, y_j)$ for $j = 1, 2$. Denoting $\tilde{x}_j = x_j - x_e$ and $\tilde{y}_j = y_j - y_e$, the bearing of the target from sensor $j$ is $\varphi_j = \arctan(\tilde{y}_j/\tilde{x}_j)$. The sensor measurements are independent and von Mises distributed, each with common concentration parameter $\kappa$ and with the measurement at sensor $j$ having circular mean $\varphi_j$.

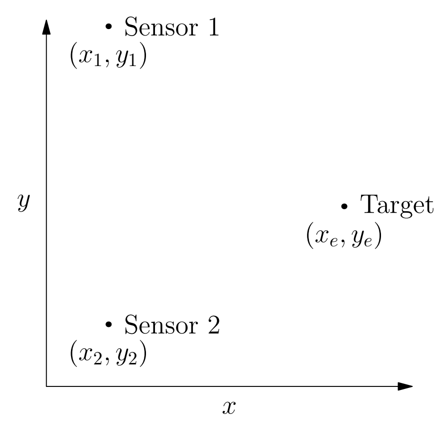



Fig. 1. An illustrative scenario involves estimating the position $(x_e, y_e)$ of a single stationary emitter from bearings-only measurements received at two mobile sensing platforms located at respective coordinates $(x_1, y_1)$ and $(x_2, y_2)$ in the plane.

For this example, the parameter manifold $M$ is $\mathbb{R}^2$ with $\theta \in M$ representing the position of the emitter in the plane. The Fisher information is

$$F = \kappa \frac{I_1(\kappa)}{I_0(\kappa)} \sum_{j=1}^{2} \frac{1}{R_j^4} (\tilde{y}_j - \tilde{x}_j) \otimes (\tilde{y}_j - \tilde{x}_j)$$

where $R_j^2 = \tilde{x}_j^2 + \tilde{y}_j^2$.

For computational tractability, it is convenient to further constrain the problem by assuming:

1) The sensor positions are known, and
2) The sensors remain at right angles relative to the target.

This situation is depicted in Figure 2. The Fisher information is, for some $C(\kappa)$,

$$F = \frac{C(\kappa)}{2} \begin{pmatrix} 1/r_1^2 & 0 \\ 0 & 1/r_2^2 \end{pmatrix}.$$

So

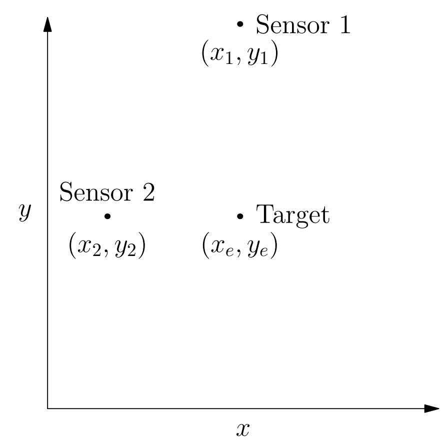



Fig. 2. A constrained scenario with the sensors at right angles relative to the emitter.

$$I_r = 4 \int \left(\frac{\dot{r}_1}{r_1}\right)^2 dt + 4 \int \left(\frac{\dot{r}_2}{r_2}\right)^2 dt.$$

The extremals take the form

$$r_i(t) = r_i(0) \exp\left(\frac{\dot{r}_i(0)}{r_i(0)} t\right)$$

Figure 3 shows numerical results for the evolution of the positions of a pair of sensors in a more complex situation. In this case the sensors start at different distances from the emitter and are not constrained to remain at right angles relative to the emitter. The sensor trajectories lead to the sensors being at the same distance from the emitter and to them moving toward the emitter. The zig-zag pattern arises because no penalty is imposed on trajectories that are not smooth, allowing either geodesic direction to be chosen in each processing epoch.

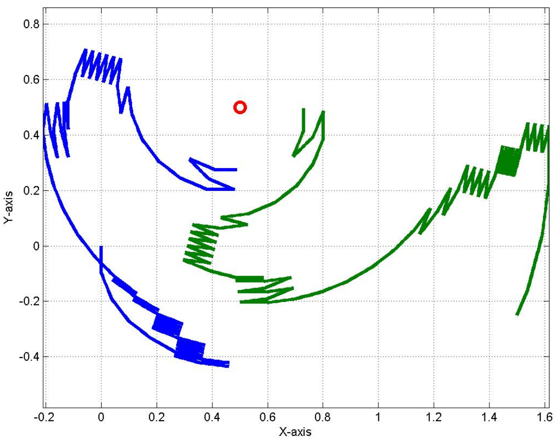



Fig. 3. Evolution of the positions of a pair of sensors that start at different distances from the emitter.

## VII. Conclusion

The preliminary work presented here is still mostly conceptual; much is needed to move from this point to practical sensor management algorithms. Even the small example introduced above leads to extremely cumbersome equations and high computational complexity. Approaches to mitigate this complexity are being sought in ongoing research.

## References


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

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

[3] 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.

[4] A. O. Hero III, D. Castañón, D. Cochran, and K. Kastella, eds., *Foundations and Applications of Sensor Management*, Springer, 2007.

[5] 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.

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

[7] 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.

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

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

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

[11] E. L. Lehmann and G. Casella, G., *Theory of Point Estimation*, 2nd edition, 1998.

[12] B. Moran, S. D. Howard, and D. Cochran, "An information-geometric approach to sensor management," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5261–5264, March 2012.

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