

Geometric remarks on Kalman filtering with intermittent observations

Andrea Censi

Abstract

Sinopoli *et al.* (TAC, 2004) considered the problem of optimal estimation for linear systems with Gaussian noise and intermittent observations, available according to a Bernoulli arrival process. They showed that there is a “critical” arrival probability of the observations, such that under that threshold the expected value of the covariance matrix (i.e., the quadratic error) of the estimate is unbounded. Sinopoli *et al.*, and successive authors, interpreted this result implying that the behavior of the system is qualitatively different above and below the threshold. This paper shows that this is not necessarily the only interpretation. In fact, the critical probability is different if one considers the average error instead of the average quadratic error. More generally, finding a meaningful “average” covariance is not as simple as taking the algebraic expected value. A rigorous way to frame the problem is in a differential geometric framework, by recognizing that the set of covariance matrices (or better, the manifold of Gaussian distributions) is not a flat space, and then studying the intrinsic Riemannian mean. Several metrics on this manifold are considered that lead to different critical probabilities, or no critical probability at all.

I. INTRODUCTION

The Kalman filter was conceived in the 1960s [1] and found immediate use at the forefront of engineering [2]. For the successive decades, the state-space approach of the Kalman filter was the tool of choice for many filtering and tracking problems, both in its algebraically equivalent formulations (e.g., Information filter [3], square root and “array” algorithms [4])

The author is with the Control & Dynamical Systems department, Division of Engineering and Applied Sciences, California Institute of Technology. Address: MC 107-81, 1200 E. California Blvd., 91125, Pasadena, CA. E-mail: andrea@cds.caltech.edu

and its extensions to nonlinear problems (e.g., extended Kalman filter (EKF), unscented Kalman filter), only recently giving way to Monte Carlo methods (particle filters).

In recent years, in many engineering fields, estimation problems have been considered where the availability of observations, or their structure, is subject to random phenomena, and one is interested in characterizing the “average accuracy”. For example, in robotics, the EKF is used in problems such as Simultaneous Localization and Mapping (SLAM); the observations structure depends on the landmark configuration, which is unknown a priori, yet it is of interest to study average accuracy results [5].

In the control literature, random observations can model packet drops, which is one of the important phenomena in network-based estimation and control. Sinopoli *et al.* [6] considered the problem of Kalman filtering when the observations are available intermittently with Bernoulli probability. They showed that there exists a critical value of the arrival probability such that, under that threshold, the expected value of the error covariance matrix is unbounded. Other successive papers improved on the same results by better characterizing the critical probability or considering non-independent packet drops [7]–[10].

The way the result of Sinopoli *et al.* is often interpreted is that the system has a qualitatively different behavior above and below the critical probability. The purpose of this note is to show that this is not necessarily the only interpretation. A motivating example is given in Section II. Considering the expected value of the covariance is equivalent to considering the expected value of the squared error norm $\mathbb{E}\{\|e\|_2^2\}$. If one instead considers the error norm $\mathbb{E}\{\|e\|_2\}$, which is equivalent to considering the expected value of the standard deviation, a different — and lower — critical probability is obtained. This raises doubts about the significance of Sinopoli *et al.*’s critical probability. More generally, what is critical is the way one defines the “average” uncertainty. Because the operation of expected value is not invariant to change of coordinates, the result is different if one averages the covariances, the standard deviations, or the information matrices: in general, $\mathbb{E}\{\mathbf{P}\} \neq \mathbb{E}\{\sqrt{\mathbf{P}}\}^2 \neq \mathbb{E}\{\mathbf{P}^{-1}\}^{-1}$. This paper advocates a geometric point of view. The basic assumption is that covariance matrices are only a particular choice of coordinates to represent Gaussian distributions, which is a Riemannian manifold with a very rich structure.

Section III deals with how to extend the idea of “mean” to Riemannian manifolds, and how that depends on the choice of a metric. Section IV discusses several metrics one can use for the manifold of Gaussian distributions. After the obvious metrics are discussed (which lead to averaging covariances, information matrices, etc.), a non-trivial Riemannian metric is introduced that is shown to be the most natural when dealing with Gaussian distributions, or, in general, when considering the intrinsic properties of the set of positive definite matrices. These different metrics lead to different critical probabilities, or no critical probability at all.

Notation: All matrices are assumed to be real. Let \mathbf{A}^* be the transpose of the matrix \mathbf{A} , let $\text{Tr}(\mathbf{A})$ be its trace, and $\{\lambda_i(\mathbf{A})\}$ its eigenvalues. Let $\text{GL}(n)$ be the set of $n \times n$ invertible matrices; let $\text{O}(n)$ be the set of orthogonal matrices; let $\mathcal{S}(n)$ be the set of symmetric $n \times n$ matrices; and let $\mathcal{P}(n) \subset \mathcal{S}(n)$ be the set of positive definite matrices. Let $\mathcal{G}(n)$ be the manifold of Gaussian distributions on \mathbb{R}^n , and $\mathcal{G}_0(n) \subset \mathcal{G}(n)$ the submanifold of Gaussian distributions with mean $\mathbf{0}$. An element of $\mathcal{G}(n)$ is denoted as $\mathcal{G}(\boldsymbol{\mu}, \mathbf{P})$, where the mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and the covariance $\mathbf{P} \in \mathcal{P}(n)$ serve as coordinates on $\mathcal{G}(n)$. Let $\|\cdot\|$ be the operator norm ($\|\mathbf{A}\|^2 = \lambda_{\max}(\mathbf{A}\mathbf{A}^*)$), and let $\|\cdot\|_F$ be the Frobenius norm ($\|\mathbf{A}\|_F^2 = \text{Trace}(\mathbf{A}\mathbf{A}^*)$). For $\mathbf{P} \in \mathcal{P}(n)$, let $\sqrt{\mathbf{P}}$ be the unique matrix in $\mathcal{P}(n)$ such that $(\sqrt{\mathbf{P}})^2 = \mathbf{P}$. All inequalities between matrices are to be interpreted in the Löwner partial order: $\mathbf{P}_1 \geq \mathbf{P}_2$ iff $\mathbf{P}_1 - \mathbf{P}_2$ is semidefinite positive.

II. MOTIVATING EXAMPLE

Consider the discrete-time linear dynamical system

$$\begin{aligned}\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\boldsymbol{\omega}(k), \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \boldsymbol{\epsilon}(k),\end{aligned}$$

with $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\omega} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$, and \mathbf{A} , \mathbf{B} , \mathbf{C} real matrices of appropriate sizes. Assume $\boldsymbol{\omega}(k)$ and $\boldsymbol{\epsilon}(k)$ are white Gaussian sequences with zero mean and covariance matrix equal to the identity, and that the initial prior for $\mathbf{x}(0)$ is Gaussian with mean $\hat{\mathbf{x}}(0)$ and covariance $\mathbf{P}(0)$. Moreover, assume that the observations are available randomly, i.e., one has available the observations $\mathbf{y}'(k) = \gamma(k)\mathbf{y}(k)$, where $\gamma(k)$ is a sequence of independent Bernoulli random variables, such

that $\mathbb{P}(\{\gamma(k) = 1\}) = \bar{\gamma}$ and $\mathbb{P}(\{\gamma(k) = 0\}) = 1 - \bar{\gamma}$. The conditional estimate of $\mathbf{x}(k)$, given the available observations until time k is still Gaussian [6], and is indicated by the mean $\hat{\mathbf{x}}(k)$ and the covariance $\mathbf{P}(k)$. Define the error estimate $\mathbf{e}(k) \triangleq \hat{\mathbf{x}}(k) - \mathbf{x}(k)$. Then $\mathbf{e}(k)$ has a Gaussian distribution with mean $\mathbf{0}$ and covariance $\mathbf{P}(k)$. This is the setup considered in [6] and is henceforth called Linear/Gaussian/Bernoulli (LGB); the name ‘‘Kalman filtering’’ is not used because the results are independent of the particular representation of the optimal filter.

Let $\mathbf{Q} \triangleq \mathbf{B}\mathbf{B}^*$ and $\mathcal{I} \triangleq \mathbf{C}^*\mathbf{C}$. If the observations are always available ($\bar{\gamma} = 1$), the evolution of $\mathbf{P}(k)$ is deterministic and obeys the recursion¹

$$g : \mathbf{P} \mapsto ((\mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q})^{-1} + \mathcal{I})^{-1}. \quad (1)$$

If (\mathbf{A}, \mathbf{B}) is stabilizable and (\mathbf{A}, \mathbf{C}) is detectable, then g has a fixed point \mathbf{P}_∞ to which $\mathbf{P}(k)$ tends regardless of the initial value $\mathbf{P}(0)$ [3]. The Kalman filter and analogous variants implement the recursion with different representations for \mathbf{P} , and faster and more numerically stable algorithms than (1), which is used in the present analysis for convenience and compactness (apply one of the matrix inversion lemmas to obtain the usual Riccati recursion).

If $\bar{\gamma} \in (0, 1)$, the evolution of $\mathbf{P}(k)$ is not deterministic anymore. A convenient way to represent the evolution of $\mathbf{P}(k)$ is in the form of an Iterated Function System [12]

$$\mathcal{S} = \begin{cases} g : \mathbf{P} \mapsto ((\mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q})^{-1} + \mathcal{I})^{-1}, & p_g = \bar{\gamma}, \\ h : \mathbf{P} \mapsto \mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q}, & p_h = 1 - \bar{\gamma}. \end{cases} \quad (2)$$

A stationary distribution for \mathbf{P} exists for all values of $\bar{\gamma}$ (under much more general conditions than Bernoulli observations) [13]. In the following, the stationary distribution is referred to as the Linear/Gaussian/Bernoulli (LGB) distribution, and \mathbf{P} refers to the stationary variable.

Sinopoli *et al.* [6] showed that there is a threshold $\bar{\gamma}_c$ such that, for $\bar{\gamma} < \bar{\gamma}_c$, $\mathbb{E}\{\mathbf{P}\}$ is unbounded.

¹Note that this paper uses the posterior covariance matrix $(\mathbf{P}(k) = \mathbf{P}_{k|k} = \text{cov}(\hat{\mathbf{x}}(k) - \mathbf{x}(k)|\mathbf{y}'(1), \dots, \mathbf{y}'(k)))$. The a-priori covariance $\mathbf{P}_{k|k-1}$ and a-posteriori $\mathbf{P}_{k|k}$ are linked by the simple relation $\mathbf{P}_{k|k} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^* + \mathbf{Q}$ hence there is no loss of generality for investigating the boundedness of the stationary distribution: $\mathbb{E}\{\mathbf{P}_{k|k-1}\}$ is bounded if and only if $\mathbb{E}\{\mathbf{P}_{k|k}\}$ is. Using the posterior covariance matrix seems a better choice for LGB filtering because, when written with information matrices ($\mathbf{Y} \mapsto (\mathbf{A}\mathbf{Y}^{-1}\mathbf{A}^* + \mathbf{Q})^{-1} + \mathcal{I}$) the difference between the maps g and h is the constant term \mathcal{I} [11].

This is equivalent to say that the square of the Euclidean norm of the estimation error, $\mathbb{E}\{\|e\|_2^2\}$, is unbounded. The threshold $\bar{\gamma}_c$ depends non trivially on the parameters of the system, and a precise characterization is object of current research [6]–[9].

The way this result is often interpreted is that for $\bar{\gamma} \geq \bar{\gamma}_c$ the behavior of the system is qualitatively different than for $\bar{\gamma} < \bar{\gamma}_c$, and therefore $\bar{\gamma}_c$ is called “critical” probability. However, if one considers another measure of performance, the critical probability changes, as described by the following result.

Proposition 1: Consider the problem of LGB filtering in the scalar case, with $\mathbf{A} = a > 1$, $\mathcal{I} = \mathcal{I} > 0$, $\mathbf{Q} = Q \geq 0$. Then the expected squared error $\mathbb{E}\{e^2\}$ and the expected error $\mathbb{E}\{|e|\}$ have two different critical probabilities:

- 1) $\mathbb{E}\{e^2\}$ is bounded if and only if $\bar{\gamma} > \bar{\gamma}_c = 1 - 1/a^2$.
- 2) $\mathbb{E}\{|e|\}$ is bounded if and only if $\bar{\gamma} > \bar{\gamma}'_c = 1 - 1/|a|$.

Proof: It is convenient to define two others systems, which are respectively an “optimist” and a “pessimist” approximations to the iteration defined by (2). The stationary distribution has support in the set $\{P|P \geq P_\infty\}$. The optimist approximation simplifies the map g to a constant by considering the best scenario $P = P_\infty$:

$$\mathcal{S}_{\text{opt}} = \begin{cases} g_{\text{opt}} : & P_{\text{opt}} \mapsto W_{\text{opt}}, & W_{\text{opt}} \triangleq ((a^2 P_\infty + Q)^{-1} + \mathcal{I})^{-1}, \\ h_{\text{opt}} : & P_{\text{opt}} \mapsto a^2 P_{\text{opt}} + Q. \end{cases} \quad (3)$$

The pessimist approximation considers the worst case ($P \rightarrow \infty$):

$$\mathcal{S}_{\text{pes}} = \begin{cases} g_{\text{pes}} : & P_{\text{pes}} \mapsto W_{\text{pes}}, & W_{\text{pes}} \triangleq \mathcal{I}^{-1}, \\ h_{\text{pes}} : & P_{\text{pes}} \mapsto a^2 P_{\text{pes}} + Q. \end{cases} \quad (4)$$

The two systems are identical except for the “reset” values W_{opt} and W_{pes} after an observation is received. It is straightforward to check that, if $P_{\text{opt}}(0) = P(0) = P_{\text{pes}}(0)$ and the three systems see the same sequences of observations, $P_{\text{opt}}(k) \leq P(k) \leq P_{\text{pes}}(k)$. Therefore, for the stationary variables, $\mathbb{E}\{P_{\text{opt}}\} \leq \mathbb{E}\{P\} \leq \mathbb{E}\{P_{\text{pes}}\}$ and $\mathbb{E}\{\sqrt{P_{\text{opt}}}\} \leq \mathbb{E}\{\sqrt{P}\} \leq \mathbb{E}\{\sqrt{P_{\text{pes}}}\}$.

Obviously $\mathbb{E}_{\omega, \epsilon}\{e^2(k)\} = P(k)$, by the definition of covariance and the fact that $\mathbb{E}_{\omega, \epsilon}\{e(k)\} = 0$. Moreover, in the case of a Gaussian distribution, one can show that $\mathbb{E}_{\omega, \epsilon}\{|e(k)|\} = \sqrt{2/\pi}\sqrt{P(k)}$. Therefore, upper and lower bounds for $\mathbb{E}\{e^2\}$ and $\mathbb{E}\{|e|\}$ can be found as

$$\mathbb{E}\{P_{\text{opt}}\} \leq \mathbb{E}\{e^2\} \leq \mathbb{E}\{P_{\text{pes}}\}, \quad (5)$$

$$\sqrt{2/\pi}\mathbb{E}\{\sqrt{P_{\text{opt}}}\} \leq \mathbb{E}\{|e|\} \leq \sqrt{2/\pi}\mathbb{E}\{\sqrt{P_{\text{pes}}}\}. \quad (6)$$

The rest of the proof estimates the terms in these expressions and is inspired by some ideas in [14]. The pdf for the stationary distribution for the two IFSs (3)–(4) can be computed in closed form. Consider, for example, the IFS in (3). The value of P_{opt} at time k can be written in closed form as a function of $\tau(k)$, the number of steps that passed without receiving an observation ($\tau(k) = 0$ if the last observation was received):

$$P_{\text{opt}}(k) = (a^2)^{\tau(k)}W_{\text{opt}} + \sum_{i=0}^{\tau(k)-1} (a^2)^i Q.$$

Assuming independent arrivals, $\tau(k)$ has the probability distribution $\mathbb{P}(\{\tau(k) = j\}) = (1 - \bar{\gamma})^j \bar{\gamma}$. The expected value $\mathbb{E}\{P_{\text{opt}}\}$ can be computed as $\sum_{j=0}^{\infty} \mathbb{P}(\{\tau(k) = j\}) P_{\text{opt}}(\tau(k))$, giving

$$\mathbb{E}\{P_{\text{opt}}\} = \bar{\gamma} \left(W_{\text{opt}} + \frac{Q}{a^2 - 1} \right) \sum_{j=0}^{\infty} [a^2(1 - \bar{\gamma})]^j - \frac{Q}{a^2 - 1}.$$

The series converges, and $\mathbb{E}\{P_{\text{opt}}\}$ is bounded, if and only if $a^2(1 - \bar{\gamma}) < 1$ (as already proved in [6]). Analogously, the expected value $\mathbb{E}\{\sqrt{P_{\text{opt}}}\}$ can be computed as

$$\mathbb{E}\{\sqrt{P_{\text{opt}}}\} = \bar{\gamma} \sum_{j=0}^{\infty} ((1 - \bar{\gamma})|a|)^j \sqrt{\left(W_{\text{opt}} + \frac{Q}{a^2 - 1} \right) - \frac{Q}{(a^2)^j(a^2 - 1)}}.$$

The series converges, and $\mathbb{E}\{\sqrt{P_{\text{opt}}}\}$ is bounded, if and only if $|a|(1 - \bar{\gamma}) < 1$.

Because the proof did not rely on the value of W_{opt} , the same convergence critical values are valid for the pessimist approximations $\mathbb{E}\{P_{\text{pes}}\}$ and $\mathbb{E}\{\sqrt{P_{\text{pes}}}\}$ as well. By taking into account (5) and (6), we see that $a^2(1 - \bar{\gamma}) < 1$ is a necessary and sufficient condition for boundedness of $\mathbb{E}\{e^2\}$, and likewise $|a|(1 - \bar{\gamma}) < 1$ for boundedness of $\mathbb{E}\{|e|\}$. ■

Because $\bar{\gamma}_c > \bar{\gamma}'_c$, there is a range of values $(\bar{\gamma}'_c, \bar{\gamma}_c]$ such that $\mathbb{E}\{|e|\}$ is bounded, but $\mathbb{E}\{e^2\}$ is not. The value $\bar{\gamma}'_c$ is at least as “critical” as $\bar{\gamma}_c$. The goal of this paper is not to advocate the use of the boundedness of $\mathbb{E}\{\|e\|\}$ rather than $\mathbb{E}\{\|e\|^2\}$ as a criterion of stability; rather, it is of more interest to discuss what are the assumptions behind using one or the other. Is one “intrinsically” more correct? Other than in Kalman filtering with intermittent observations, similar questions arise in other problems where one must compute an “average” accuracy [5]. In general, the expected value is not invariant to change of coordinates, so a different average accuracy is obtained if one considers the average of covariances, of standard deviations, or of information matrices.

Some answers to these questions can be found by setting the problem in a geometric framework. In particular, instead of considering the set $\mathcal{P}(n)$ of covariance matrices as a subset of $\mathbb{R}^{n \times n}$, one can consider, more abstractly, the manifold $\mathcal{G}(n)$ of Gaussian distributions. The next section shows how the operation of expected value $\mathbb{E}\{\cdot\}$ can be generalized to Riemannian manifolds, such that one can define a “Riemannian mean” $\mathbb{M}\{\cdot\}$ independently of the choice of coordinates.

III. MEANS ON RIEMANNIAN MANIFOLDS

Classical mathematical statistics [15] developed in the first decades of last century in the context of Euclidean spaces. Subsequently, it became clear that many applications would benefit from rigorous coordinate-free approaches to statistics on manifolds. Examples of such applications and corresponding manifolds include robotics [16] (motion groups), shape analysis [17] (size-and-shape spaces), radar imaging [18] (Grassman manifold), diffusion tensor magnetic resonance imaging [19], [20] (positive definite tensors), and Lie groups in general [21]. This section recalls the definition of Riemannian mean on manifolds; the reader is assumed to be familiar with basic differential geometry (e.g., [22]).

Let X be a random variable taking values in \mathbb{R}^n with joint cumulative distribution function μ . The expected value of X (or Euclidean mean, or simply *mean*) is defined, in the most general

terms, as the Lebesgue-Stieltjes integral

$$\mathbb{E}\{X\} \triangleq \int_{\mathbb{R}^n} \mathbf{x} d\mu(\mathbf{x}). \quad (7)$$

This definition is not directly generalizable to manifolds because it assumes that the set has a vector space structure. However, the mean satisfies a variational property, being the point that minimizes the quadratic risk

$$\mathbb{E}\{X\} = \arg \min_y \mathbb{E} \{ \|X - y\|_2^2 \}. \quad (8)$$

Definitions (7) and (8) are easily seen to be equivalent in the case of vector spaces. The second definition has the benefit that it can be generalized to any metric space.

In particular, it can be generalized to Riemannian manifolds. Recall that a Riemannian manifold (\mathcal{M}, g) is a differentiable manifold \mathcal{M} equipped with a smooth metric g on the tangent space. The “length” $\ell(\gamma)$ of a differentiable curve $\gamma : [0, 1] \rightarrow \mathcal{M}$ is defined by

$$\ell(\gamma) = \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))} dt.$$

Given the notion of length, one defines the distance between two points $m_1, m_2 \in \mathcal{M}$ as

$$d(x, y) = \inf \{ \ell(\gamma) \mid \gamma \text{ is a differentiable curve joining } x \text{ and } y \}.$$

Consider a Riemannian manifold (\mathcal{M}, g) with corresponding distance d . Generalizing (8) for a random variable X taking values in \mathcal{M} , define the Riemannian mean (also called: Riemannian barycenter, Riemannian center of mass, Frechét mean or Karcher mean) as the point that minimizes the average quadratic distance:

$$\mathbb{M}\{X\} \triangleq \arg \inf_{y \in \mathcal{M}} \mathbb{E} \{ d^2(X, y) \}. \quad (9)$$

The Riemannian mean is unique for a simply connected manifold of non positive sectional curvature [23] — as counterexamples, the reader may consider the distribution consisting of a pair of antipodal points on the unit circle \mathbb{S}^1 (a non-simply connected, zero curvature manifold)

and on the unit sphere \mathbb{S}^2 (a simply connected, positive curvature manifold). See [24] for an alternative characterization of the Riemannian mean using the inverse of the exponential map, and, more in general, see [18] for a short introduction to modern intrinsic estimation on manifolds.

IV. DIFFERENT METRICS FOR THE MANIFOLD OF GAUSSIAN DISTRIBUTIONS

The random availability of the observations in the LGB filtering setup induces a stationary distribution on $\mathcal{G}(n)$, the manifold of Gaussian distributions on \mathbb{R}^n , for the estimate $(\hat{\mathbf{x}}, \mathbf{P})$. In particular, the estimation error $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x}$ has a Gaussian distribution with zero mean, therefore we focus on the submanifold $\mathcal{G}_0(n)$ of Gaussian distributions with mean $\mathbf{0}$. The Riemannian mean depends on the choice of a metric, and this section considers several such options. In the following, for compactness of notation, sometimes we confound $\mathcal{G}_0(n)$ with $\mathcal{P}(n)$, for example by writing the distance between Gaussian distributions $d(\mathcal{G}(\mathbf{0}, \mathbf{P}_1), \mathcal{G}(\mathbf{0}, \mathbf{P}_2))$ directly as $d(\mathbf{P}_1, \mathbf{P}_2)$.

A. Flat metric for covariances

The traditional way to represent a Gaussian distribution is by using its mean and covariance, by identifying $\mathcal{G}(n)$ with $\mathbb{R}^n \times \mathcal{P}(n)$. This is what most consider to be the “natural” representation. Considering $\mathcal{P}(n)$ as a convex cone of $\mathbb{R}^{n \times n}$ seems also to fit very well with the operation of expected value, because the expectation is nothing other than a glorified convex linear combination; the convexity is also useful in optimization, for example in semidefinite programming [25].

From this point of view, $\mathcal{P}(n)$ inherits the Euclidean metric of $\mathbb{R}^{n \times n}$. This is the metric implicitly used by Sinopoli *et al.* [6]. The distance between two Gaussian distributions reduces to the Frobenius distance between the two covariance matrices:

$$d(\mathcal{G}(\mathbf{0}, \mathbf{P}_1), \mathcal{G}(\mathbf{0}, \mathbf{P}_2)) = \|\mathbf{P}_1 - \mathbf{P}_2\|_F. \quad (10)$$

For the Riemannian mean, one obtains that the covariance of the mean distribution is the expected value of the covariances: $\mathbb{M}\{\mathcal{G}(\mathbf{0}, \mathbf{P})\} = \mathcal{G}(\mathbf{0}, \mathbb{E}\{\mathbf{P}\})$. Note that this mean is affine-invariant, i.e. invariant to a change of coordinate $\mathbf{P} \mapsto \mathbf{A}\mathbf{P}\mathbf{A}^*$ for $\mathbf{A} \in \text{GL}(n)$, but the distance (10) is not.

B. Flat metric for information matrices

The Information filter [3] utilizes a different parametrization $(\boldsymbol{\eta}, \mathbf{Y})$ to represent a Gaussian distribution ($\boldsymbol{\eta} = \mathbf{P}^{-1}\hat{\mathbf{x}}$ and $\mathbf{Y} = \mathbf{P}^{-1}$); this gives a *different* embedding of $\mathcal{G}(n)$ in $\mathbb{R}^n \times \mathcal{P}(n)$. One could make the argument that information matrices are a more natural parametrization for Gaussian distributions: in the canonical representation of Gaussian distributions as an exponential family [15], the information matrix is the natural parameter; in fact, one writes the probability density function using \mathbf{P}^{-1} . With this choice, the distance between distributions is given by

$$d(\mathcal{G}(0, \mathbf{P}_1), \mathcal{G}(0, \mathbf{P}_2)) = \|\mathbf{P}_1^{-1} - \mathbf{P}_2^{-1}\|_F.$$

The Riemannian mean is $\mathbb{M}\{\mathcal{G}(0, \mathbf{P})\} = \mathcal{G}(0, \mathbb{E}\{\mathbf{P}^{-1}\}^{-1})$. It is easy to see that, in the Linear/Gaussian/Bernoulli case, $\mathbb{M}\{\mathcal{G}(0, \mathbf{P})\}$ exists for all values of $\bar{\gamma} \in (0, 1]$, because \mathbf{P}^{-1} is bounded by \mathbf{P}_∞^{-1} . As in the previous case, the mean is affine-invariant, but the distance is not.

C. Flat metric for square root of covariances

The matrix equivalent of the scalar standard deviation is the square root of the covariance matrix; the eigenvalues of $\sqrt{\mathbf{P}}$ are the standard deviations. The distance between distributions can be defined as

$$d(\mathcal{G}(0, \mathbf{P}_1), \mathcal{G}(0, \mathbf{P}_2)) = \|\sqrt{\mathbf{P}_1} - \sqrt{\mathbf{P}_2}\|_F,$$

and consequently the Riemannian mean is $\mathbb{M}\{\mathcal{G}(0, \mathbf{P})\} = \mathcal{G}(0, \mathbb{E}\{\sqrt{\mathbf{P}}\}^2)$. By Jensen's inequality and the fact that $\mathbf{P} \mapsto \sqrt{\mathbf{P}}$ is operator-concave, it follows that $\mathbb{E}\{\sqrt{\mathbf{P}}\}^2 \leq \mathbb{E}\{\mathbf{P}\}$. Thus the Riemannian mean for this distance exists in all cases when $\mathbb{E}\{\mathbf{P}\}$ exists; moreover, as shown by Proposition 1, in some cases, the critical probability for boundedness of $\mathbb{E}\{\sqrt{\mathbf{P}}\}$ is strictly less than the critical probability for $\mathbb{E}\{\mathbf{P}\}$.

D. Fisher Information Metric

The problem we are analysing is special in two regards: 1) we are concerned with doing statistics on a certain manifold $\mathcal{G}(n)$; and 2) the elements of the manifold $\mathcal{G}(n)$ represents

probability distribution themselves. The branch of statistics that studies the properties of the families of probability distributions considered as a manifold is called *information geometry* and is a relatively recent development with respect to classical mathematical statistics [26], [27].

From this point of view, the manifold has a natural metric given by the generalization of the Fisher Information Matrix (FIM) as a Riemannian metric. We recall the definition of the FIM in the Gaussian case [15]. Assume that the available observations $\mathbf{z} \in \mathbb{R}^q$ have a Gaussian distribution whose mean and covariance are parametrized by an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^n$: $\mathbf{z} \sim \mathcal{G}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. The FIM for $\boldsymbol{\theta}$ is the $n \times n$ semidefinite positive matrix $\mathcal{I}[\boldsymbol{\theta}]$ defined as

$$\mathcal{I}[\boldsymbol{\theta}]_{a,b} = \frac{\partial \boldsymbol{\mu}^*}{\partial \theta_a} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_b} + \frac{1}{2} \text{Tr} \left\{ \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_a} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_b} \right\}. \quad (11)$$

The FIM gives the information contained in the samples about the value of $\boldsymbol{\theta}$; for example, using the FIM one defines the Cramér-Rao Bound for unbiased estimators as $\text{cov}(\hat{\boldsymbol{\theta}}) \geq \mathcal{I}[\boldsymbol{\theta}]^{-1}$. The FIM can be generalized to be a Riemannian metric for the manifold $\mathcal{G}(n)$. If we restrict to the submanifold $\mathcal{G}_0(n)$, given two elements $\mathbf{X}, \mathbf{Y} \in \mathcal{S}(n)$ in the tangent space at $\mathcal{G}(0, \mathbf{P})$, the Fisher Information Metric is

$$g(\mathbf{X}, \mathbf{Y}) = \frac{1}{2} \text{Tr} \{ \mathbf{P}^{-1} \mathbf{X} \mathbf{P}^{-1} \mathbf{Y} \}. \quad (12)$$

Compare (12) with the second term in (11). The distance induced by this metric is (e.g., [19]):

$$d(\mathcal{G}(\mathbf{0}, \mathbf{P}_1), \mathcal{G}(\mathbf{0}, \mathbf{P}_2)) = \left[\sum_{i=1}^n \log^2(\lambda_i(\mathbf{P}_1 \mathbf{P}_2^{-1})) \right]^{1/2} \quad (13)$$

This distance is “natural” in the sense that it is linked to the probability of distinguishing the two distributions $\mathcal{G}(\mathbf{0}, \mathbf{P}_1)$ and $\mathcal{G}(\mathbf{0}, \mathbf{P}_2)$ by observing their samples, in a sense which is made precise in [26]. Unfortunately, a closed form expression for writing $\mathbb{M}\{\mathcal{G}(\mathbf{0}, \mathbf{P})\}$ is not known.

The use of this natural distance on $\mathcal{G}_0(n) \simeq \mathcal{P}(n)$ allows to show that some naive results obtained using the flat metric on covariance are incorrect [18]. For example, in basic mathematical statistics courses, one teaches that, given a set of samples $\{\mathbf{x}_i\}_{i=1}^n$ from a distribution with covariance \mathbf{P} , the bias-corrected sample covariance matrix $\hat{\mathbf{P}}_{sc} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i)^*$

is an *unbiased* and *efficient* estimator of \mathbf{P} , i.e. $\mathbb{E}\{\hat{\mathbf{P}}_{sc}\} = \mathbf{P}$ and $\hat{\mathbf{P}}_{sc}$ reaches the Cramér-Rao bound. However, it is also well known that $\hat{\mathbf{P}}_{sc}$ performs poorly at low sample support. Smith’s [18] explanation to this conundrum is that $\hat{\mathbf{P}}_{sc}$ is not unbiased according to the natural metric: $\mathbb{M}\{\hat{\mathbf{P}}_{sc}\} \neq \mathbf{P}$.

Ignoring the Fisher Information Metric interpretation, the distance defined by (13) is also natural for $\mathcal{P}(n)$ when it is considered either as a symmetric space, or as the quotient space $GL(n)/O(n)$ [18]. The distance has several other useful properties in the context of LGB filtering. $(\mathcal{P}(n), d)$ is a complete metric space [13] and a geodesically complete manifold with nonpositive curvature [20]. The distance d induces the usual topology [13] on $\mathcal{P}(n)$. The distance is invariant to affine transformations, and also to inversion $\mathbf{P} \mapsto \mathbf{P}^{-1}$; this last property is useful because one can use either covariance matrices or information matrices: $\mathbb{M}\{\mathbf{P}\} = \mathbb{M}\{\mathbf{P}^{-1}\}^{-1}$, which is not true if one uses the expected value. Using this distance it is also easy to show contraction properties for the Riccati iterations g, h that guarantee the existence of the stationary distribution [13].

It is possible to show that there is no “critical probability” if one uses this metric. To this end, one should first prove that the system has a stationary distribution for all values of $\bar{\gamma} > 0$. This is done in the next section. Then, in Section IV-F, it is proved that the Riemannian mean of this distribution exists.

E. Existence of the stationary distribution

In this section, we prove the existence of the stationary distribution, for every value of $\bar{\gamma} > 0$. This can be done by using some results from Bougerol [13] regarding the contraction properties of the maps h and g , and some results from Barnsley *et al.* [28] about the convergence of Iterated Function Systems². Once these results are recalled, the conclusion will be immediate.

We need some preliminaries from [28]. Let (X, d) be a complete metric space. Let f_i , $i = 1, \dots, n$ be Lipschitz functions from X to X , that is, there exists $s_i > 0$ such that $d(f_i(x), f_i(y)) \leq s_i d(x, y)$ for all x, y in X . We say that f_i is “nonexpansive” if $s_i \leq 1$, and

²Because this paper is not available electronically yet, the results are stated here extensively.

we say that it is a “strict contraction mapping” if $s_i < 1$. Assign a set of probabilities p_i , to these functions, such that $p_i > 0$ and $\sum_{i=1}^n p_i = 1$. Consider now the Iterated Function System $\{(f_i, p_i)\}$ and the corresponding Markov chain, which we denote $\{Z_i\}, i \geq 0$. We say that a measure μ is “attractive” if, for every initial distribution of Z_0 , the process Z_i converges in distribution to μ , that is, $\lim_{i \rightarrow \infty} \mathbb{E}\{f(Z_i)\} = \int f d\mu$ for every bounded continuous function f on X . It is an intuitive result that, if all the f_i are strict contractions, then the process “tends to forget” the initial conditions, and a stationary distribution exists. What is not trivial is that IFSs converge in distribution with much weaker hypotheses, as shown by the following result.

Theorem 1: (Barnsley *et al.* [28]) Suppose that the f_i satisfies an *average contractivity condition* as follows: for all $x, y \in X$,

$$\sum_{i=1}^n p_i \log \frac{d(f_i(x), f_i(y))}{d(x, y)} < 0. \quad (14)$$

Then there exists a unique, attractive invariant probability measure for the IFS.

Remark 1: Note that it is not assumed that the single f_i are strict contractions ($s_i < 1$) or even contractions ($s_i \leq 1$). This theorem can also be generalized to the case in which the transition probabilities depend on the state ($p_i = p_i(x)$), although some additional hypotheses are required [29].

We now recall the following from [13]:

Lemma 1: (Bougerol [13]) In the metric d defined by (13),

- 1) The maps h and g are nonexpansive mappings: $d(h(\mathbf{P}_1), h(\mathbf{P}_2)) \leq d(\mathbf{P}_1, \mathbf{P}_2)$, and equivalently for g .
- 2) If \mathbf{A} is non-singular, (\mathbf{A}, \mathbf{B}) is controllable, (\mathbf{A}, \mathbf{C}) observable, the composition $g^n = g \circ \dots \circ g$ of n copies of g is a strict contraction mapping; that is, there exists $\rho = \rho(\mathbf{A}, \mathbf{B}, \mathbf{C}) < 1$ such that $d(g(\mathbf{P}_1), g(\mathbf{P}_2)) \leq \rho d(\mathbf{P}_1, \mathbf{P}_2)$.

From these results, the following result is easily proved.

Proposition 2: If \mathbf{A} is non-singular, (\mathbf{A}, \mathbf{B}) controllable, (\mathbf{A}, \mathbf{C}) observable, then the stationary distribution for \mathbf{P} exists for all $\bar{\gamma} > 0$.

Proof: Consider the behavior of the system at intervals of n steps. This corresponds to considering the “power” IFS $\mathcal{S}^n = \{(g^n, \bar{\gamma}^n), (g \circ h^{n-1}, \bar{\gamma}(1 - \bar{\gamma})^{n-1}), \dots, (h^n, (1 - \bar{\gamma})^n)\}$, which is created by all 2^n possible combinations of length n of the functions g, h , with corresponding probabilities. By Lemma 1, g^n is a strict contraction, and all the other combinations are non-expansive mappings. Therefore, assuming $\bar{\gamma} > 0$, the system satisfies the average contractivity condition (14), and by Theorem 1 the stationary distribution exists. ■

F. Existence of the Riemannian mean for Fisher Information Metric

After having ascertained that the stationary distribution exists (Proposition 2), we now prove existence of the Riemannian mean.

Proposition 3: The Riemannian mean of the LGB distribution for the distance (13) exists for all $\bar{\gamma} \in (0, 1]$.

Proof: The stationary distribution of the IFS $\mathcal{S} = \{(g, \bar{\gamma}), (h, 1 - \bar{\gamma})\}$ is equivalent to that of the power IFS $\mathcal{S}^n = \{(g^n, \bar{\gamma}^n), (g \circ h^{n-1}, \bar{\gamma}(1 - \bar{\gamma})^{n-1}), \dots, (h^n, (1 - \bar{\gamma})^n)\}$, obtained by considering compositions of length n of the functions (g, h) with corresponding probabilities. We now build the IFS $\mathcal{S}_{\text{pes}}^n$, a “pessimist” approximation to \mathcal{S}^n . By recalling that g and h are order-preserving, and $g(\mathbf{M}) \leq h(\mathbf{M})$ for all \mathbf{M} [6], one can bound all mixed terms in g, h in \mathcal{S}^n by h^n . Furthermore, one can also find an upper bound for g^n : because the system is observable, the uncertainty is bounded over all the state space after n consecutive observation are received. Therefore, $\mathbf{W}_{\text{pes}} \triangleq \sup_{\mathbf{P} \geq \mathbf{P}_\infty} g^n(\mathbf{P})$ exists and is bounded: $\mathbf{P}_\infty \leq \mathbf{W}_{\text{pes}} < \infty$. Thus the pessimist approximation to \mathcal{S}^n is $\mathcal{S}_{\text{pes}}^n = \{(\mathbf{W}_{\text{pes}}, \bar{\gamma}^n), (h^n, 1 - \bar{\gamma}^n)\}$. Seeing the stationary variables \mathbf{P} (for \mathcal{S}^n) and \mathbf{P}_{pes} (for $\mathcal{S}_{\text{pes}}^n$) as functions of the past infinite sequence of arrivals $\omega = \{\gamma(0), \gamma(-1), \gamma(-2), \dots\} \in \{0, 1\}^{\mathbb{N}}$, one has that

$$\mathbf{P}_\infty \leq \mathbf{P}(\omega) \leq \mathbf{P}_{\text{pes}}(\omega). \quad (15)$$

To prove that $\mathbb{M}\{\mathbf{P}\}$ is bounded for all $\bar{\gamma}$, it is sufficient to show that the minimization problem (9) is feasible for all $\bar{\gamma}$. To prove this, it is sufficient to show that $\mathbb{E}\{d^2(\mathbf{X}, \mathbf{P})\}$ is bounded for some matrix \mathbf{X} ; it is convenient to choose $\mathbf{X} = \mathbf{P}_\infty$. By (15) and Lemma 2 below,

we obtain that $d(\mathbf{P}_\infty, \mathbf{P}(\omega)) \leq d(\mathbf{P}_\infty, \mathbf{P}_{\text{pes}}(\omega))$ and thus $\mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P})\} \leq \mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P}_{\text{pes}})\}$. It follows that $\mathbb{M}\{\mathbf{P}\}$ is bounded if $\mathbb{M}\{\mathbf{P}_{\text{pes}}\}$ is.

We now investigate boundedness of $\mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P}_{\text{pes}})\}$. Choose an \mathbf{M} such that $\mathbf{M}\mathbf{P}_\infty\mathbf{M}^* = \mathbf{I}$ and do a change of coordinates $\mathbf{P} \mapsto \mathbf{M}\mathbf{P}\mathbf{M}^*$. One finds that $\mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P}_{\text{pes}})\} = \mathbb{E}\{d^2(\mathbf{I}, \mathbf{P}'_{\text{pes}})\}$, where $\mathbf{P}'_{\text{pes}} \triangleq \mathbf{M}\mathbf{P}_{\text{pes}}\mathbf{M}^* \geq \mathbf{I}$. Note that because $\lambda_i(\mathbf{P}'_{\text{pes}}) \geq 1$, we can find the bound $d^2(\mathbf{I}, \mathbf{P}'_{\text{pes}}) = \sum_{i=1}^n \log^2(\lambda_i(\mathbf{P}'_{\text{pes}})) \leq n \log^2(\|\mathbf{P}'_{\text{pes}}\|)$. An expression for $\mathbf{P}_{\text{pes}}(k)$ can be written explicitly as in the proof of Proposition 1 as a function of $\tau(k)$, the number of steps passed without receiving an observation (recall that one time step in the IFS \mathcal{S}^n corresponds to n steps of \mathcal{S}):

$$\mathbf{P}_{\text{pes}}(k) = \mathbf{A}^{n\tau(k)} \mathbf{W}_{\text{pes}} (\mathbf{A}^*)^{n\tau(k)} + \sum_{i=0}^{n\tau(k)-1} \mathbf{A}^i \mathbf{Q} (\mathbf{A}^*)^i.$$

From this one finds the bound $\|\mathbf{P}_{\text{pes}}(k)\| \leq c_1 \|\mathbf{A}\|^{2n\tau(k)}$ for some $c_1 > 0$. Thus $n \log^2(\|\mathbf{P}'_{\text{pes}}\|) \leq \tau(k)^2 c_2 + \tau(k) c_3 + c_4$ for some $c_2, c_3, c_4 > 0$. As in the proof of Proposition 1, the expectation can be computed with respect to $\tau(k)$, and one obtains

$$\mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P})\} \leq \bar{\gamma}^n \sum_{i=0}^{\infty} (1 - \bar{\gamma}^n)^i [i^2 c_2 + i c_3 + c_4].$$

Series of the kind $\sum_{i=0}^{\infty} i^k x^i$ with $k \geq 0$ are convergent if $|x| < 1$, hence $\mathbb{E}\{d^2(\mathbf{P}_\infty, \mathbf{P})\}$ is always bounded if $\bar{\gamma} \in (0, 1]$. Therefore, the Riemannian mean is always bounded. ■

Lemma 2: For the distance defined in (13), $\mathbf{P}_1 \leq \mathbf{P}_2 \leq \mathbf{P}_3 \Rightarrow d(\mathbf{P}_1, \mathbf{P}_2) \leq d(\mathbf{P}_1, \mathbf{P}_3)$.

Proof: (sketch) First, reduce to the case $\mathbf{P}_1 = \mathbf{I}$ by letting $\mathbf{P}'_i = \mathbf{M}\mathbf{P}_i\mathbf{M}^*$, with \mathbf{M} chosen such that $\mathbf{M}\mathbf{P}_1\mathbf{M}^* = \mathbf{I}$. Then verify $d(\mathbf{I}, \mathbf{P}'_2) \leq d(\mathbf{I}, \mathbf{P}'_3)$ by direct computation using (13). ■

After one has proved the existence of $\mathbb{M}\{\mathbf{P}\}$, uniqueness follows from the fact that the manifold has nonpositive curvature [23].

V. CONCLUSIONS

Algebra is the offer made by the devil to the mathematician. The devil says: ‘I will give you this powerful machine, it will answer any question you like. All you need to do is giving me your soul: *give up geometry* and you will have this marvellous machine.’ [30]

— Sir Michael Atiyah (1929–)

The righteous engineer must refuse the devil’s offer. Reframing problems in a geometric framework usually allows to spot the hidden assumptions, and therefore to check whether the results have a physical meaning, or they are just figments of the mathematical formalization.

The hidden assumption in the work of Sinopoli *et al.* is that positive definite matrices are treated as a convex cone of $\mathbb{R}^{n \times n}$. This is perhaps the most intuitive interpretation, and has good consequences in certain contexts, such as semidefinite programming [25]. However, this could lead to incorrect conclusions when doing rigorous intrinsic estimation. This is well shown by Smith’s example [18], that even though the sample covariance matrix is unbiased in the naive sense ($\mathbb{E}\{\hat{\mathbf{P}}_{sc}\} = \mathbf{P}$), it is biased in the intrinsic sense ($\mathbb{M}\{\hat{\mathbf{P}}_{sc}\} \neq \mathbf{P}$), thereby contradicting what is taught in elementary statistics courses.

In the case of the Linear/Gaussian/Bernoulli filtering problem, if one uses the average standard deviations, instead of the average covariances, one obtains a different critical probability (Proposition 1). It is pointless to discuss which critical probability is more critical than the other, but surely considering the average error is more natural than considering the average error *squared*. The point is that the boundedness of the expected value cannot be considered as a criterion for “stability”; there are plenty of well-behaved probability distributions which have infinite moments³.

Sinopoli’s critical probability is critical only in the sense that it is the threshold under which $\mathbb{E}\{\mathbf{P}\}$ ceases to be meaningful as a performance measure. Under the threshold, the qualitative behavior of the system does not change, as one can see from considering the Riemannian mean derived from the intrinsic Fisher Information Metric (Proposition 3) — that is possibly a less intuitive, but more natural way to represent the concept of “average uncertainty”. Thus the intense research effort in trying to characterize $\bar{\gamma}_c$ seems misplaced; of more interest is studying the entire LGB distribution [11], [14].

³ Consider as an example the Pareto distribution, defined as $\mathbb{P}(\{X > x\}) = x^{-k}$ for $x \geq 1, k > 0$: for this distribution, $\mathbb{E}\{X\}$ is bounded only if $k > 1$, and $\mathbb{E}\{X^2\}$ only if $k > 2$; yet it has a regular power law. All the statistics that remain finite change smoothly when k goes through the “critical” values 1 and 2.

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME – Journal of Basic Engineering*, no. 82 (Series D), pp. 35–45, 1960.
- [2] L. A. McGee and S. F. Schmidt, "Discovery of the Kalman filter as a practical tool for aerospace and industry," tech. rep., NASA, 1985. NASA-TM-86847.
- [3] P. S. Maybeck, *Stochastic models, estimation and control*, vol. 1. Academic Press, 1979.
- [4] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 1 ed., 2000.
- [5] A. Mourikis, *Characterization and Optimization of the Accuracy of Mobile Robot Localization*. PhD thesis, University of Minnesota, 2008.
- [6] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1453–1464, Sept. 2004.
- [7] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, 2007.
- [8] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," in *IEEE Conference on Decision and Control (CDC)*, pp. 2692–2697, Dec. 2008.
- [9] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 54, pp. 386–390, Feb. 2009.
- [10] S. Kar, B. Sinopoli, and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence to a stationary distribution," 2009. arXiv:abs/0903.2890.
- [11] A. Censi, "On the performance of Kalman filtering with intermittent observations: a geometric approach with fractals," in *Proceedings of the American Control Conference (ACC)*, 2009. to appear, available at <http://purl.org/censi/2008/fractals>.
- [12] G. A. Edgar, *Integral, probability, and fractal measures*. Springer, 1998.
- [13] P. Bougerol, "Kalman filtering with random coefficients and contractions," *SIAM J. Control and Optimization*, vol. 31, no. 4, 1993.
- [14] M. Epstein, L. Shi, A. Tiwari, and R. M. Murray, "Probabilistic performance of state estimation across a lossy network," *Automatica*, vol. 44, no. 12, pp. 3046 – 3053, 2008.
- [15] G. Casella and R. L. Berger, *Statistical Inference*. Duxbury Press, 1990.
- [16] Y. Wang and G. Chirikjian, "Second-order theory of error propagation on motion groups," *Int. J. of Robotics Research*, 2008.
- [17] H. Le and D. G. Kendall, "The Riemannian structure of Euclidean shape spaces: A novel environment for statistics," *Annals of Statistics*, vol. 21, no. 3, pp. 1225–1271, 1993.
- [18] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. on Signal Processing*, vol. 53, no. 5, 2005.
- [19] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, S. Lehericy, and K. Ugurbil, "A Riemannian approach to diffusion tensor images segmentation," in *Proc. Information Processing in Medical Imagin*, pp. 591–602, 2005.
- [20] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras, "Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing," *J. Math. Imaging Vis.*, vol. 25, no. 3, pp. 423–444, 2006.

- [21] U. Grenander, *Probabilities on Algebraic Structures*. Dover Publications, 1965.
- [22] M. do Carmo, *Riemannian Geometry*. Birkhauser, 1994.
- [23] H. Karcher, “Riemannian center of mass and mollifier smoothing,” *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977.
- [24] J. M. Corcuera and W. S. Kendall, “Riemannian barycentres and geodesic convexity,” *Math. Proc. Cambridge Philos. Soc.*, vol. 127, 1999.
- [25] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM Review*, vol. 38, no. 1, 1996.
- [26] S. Amari, *Differential-Geometrical Methods in Statistics*. Springer Verlag, 1985.
- [27] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*. Chapman & Hall/CRC, 1993.
- [28] M. Barnsley and J. H. Elton, “A new class of markov processes for image encoding,,” *Adv. Appl. Prob.*, vol. 20, pp. 14–32, 1988.
- [29] M. F. Barnsley, S. G. Demko, J. H. Elton, and J. S. Geronimo, “Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities,” *Annales de l’Institut Henry Poincaré – Probabilité et Statistiques*, vol. 24, no. 3, 1988.
- [30] M. Atiyah, *Mathematical Evolutions*, ch. Mathematics in the 20th century. Mathematical Association of America, 2002.