Kalman filtering with intermittent observations: convergence for semi-Markov chains and an intrinsic performance measure

Andrea Censi, *student member*

***Abstract*—This technical note shows that the stationary distribution for the covariance of Kalman filtering with intermittent observations exists under mild conditions for a very general class of packet dropping models (semi-Markov chain). These results are proved using the geometric properties of Riccati recursions with respect to a particular Riemannian distance. Moreover, the Riemannian mean induced by that distance is always bounded, therefore it can be used for characterizing the performance of the system for regimes where the moments of the covariance do not exist. Other interesting properties of that mean include the symmetry between covariance and information matrices (averaging covariances or their inverse gives the same result), and its interpretation in information geometry as the "natural" mean for the manifold of Gaussian distributions.**

***Index Terms*—Kalman filtering, packet drops, intrinsic estimation, information geometry.**

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]) and its extensions to nonlinear problems (e.g., Extended Kalman Filter, Unscented Kalman Filter).

In recent years, since the paper by Sinopoli *et al.* [5], there has been intense interest in studying the problem of Kalman filtering with intermittent observations. Sinopoli *et al.* consider the case where observations are available intermittently with independent probability, and they show the non-obvious result 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. Mo and Sinopoli [6] and Plarre and Bullo [7] better characterize this critical probability, for which an explicit expression is still lacking. Huang and Dey [8] consider the case where the availability of observations is regulated by a Markov chain**,** which is more realistic than iid packet dropping. Xu and Hespanha [9] discuss similar results in the continuous-time case.

The way the result of Sinopoli *et al.* was often interpreted was that the system has a qualitatively different behavior above and below the critical probability. More recently, it has been pointed out by Kar *et al.* [10] that convergence of the expected covariance is not a necessary condition for convergence in distribution of the Markov process; that is, there is a lower arrival probability such that the process converges in distribution, but the expected value of the covariance is unbounded. For independent packet drops, the stationary distribution exists for all non zero arrival probabilities, as shown by Censi [11] and Kar *et al.* [12].

A different line of research is concerned with the overall description of the stationary distribution. Epstein *et al.* [13] provide bounds for the stationary cumulative distribution function (cdf) of the covariance. Censi [14] describes the fractal nature of the stationary probability distribution and provides exact expressions for the cdf in some special cases. Vakili and Hassibi [15] analyze the stationary distribution of the eigenvalues for random Lyapunov and Riccati iterations in the case of a stable system.

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

This paper addresses two open questions in the literature. The first question is whether it is possible to prove convergence in distribution for a wider class of models other than independent packet drops. This would make the system modelling more realistic; in fact, network errors are highly correlated in time. Moreover, proving that the stationary distribution exists is a necessary step before studying the system performance. This paper shows (Proposition 6) that the system converges to a unique stationary distribution when the arrival of observations is driven by a semi-Markov chain, a generalization of Markov chains in which the system persists in each state for a random period of time. This model is much more general than those considered in the literature (Huang and Dey [8] consider a Markov chain, but do not prove convergence). The techniques used are rather different than the literature: Kar *et al.* [12] use essentially *algebraic* arguments, modelling the system as an order-preserving random dynamical systems that has a strong sublinearity property. This paper, instead, models the system as an iterated function system, and derives its results from some *geometric* properties of Riccati iterations with respect to a certain distance on the space of positive definite matrices. As a result, one can obtain stronger results with less effort, by relying on established properties of iterated function systems.

The second contribution concerns defining alternative performance measures when the expected value of the covariance does not exist. Three desirable properties are discussed: existence, invariance, and covariance/information "symmetry". It is shown that the Riemannian mean, induced by the distance used in the first part, satisfies these properties (Proposition 9). Moreover, according to the *information geometry* interpretation, this is the "natural" mean to use when "averaging" Gaussian distributions.

This paper is organized as follows. Section II provides the setup. Section III proves the existence of the stationary distribution. Section IV discusses the generalization of means to Riemannian manifolds and the properties of a particular Riemannian mean.

Notation: All matrices are assumed to be real. Let $\mathbf{A}^*$ be the transpose of the matrix $\mathbf{A}$, and let $\{\lambda_i(\mathbf{A})\}$ denote the eigenvalues. Let $\mathcal{S}(n)$ be the set of symmetric $n \times n$ matrices and let $\mathcal{P}(n)$ be the set of positive definite matrices. 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. For a Lipschitz map f, define its Lipschitz constant as $\text{Lip}(f) \triangleq \sup_{x \neq y} d(f(x), f(y))/d(x, y)$. Let $\mathcal{G}(n)$ be the manifold of Gaussian distributions on $\mathbb{R}^n$ [16], 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 $\mathcal{G}(\boldsymbol{\mu}, \mathbf{P})$.

II. Problem setup

Consider the discrete-time linear dynamical system

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

with $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\omega} \in \mathbb{R}^p$, $\boldsymbol{y} \in \mathbb{R}^q$, 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 $\boldsymbol{x}_0$ is Gaussian with covariance $\mathbf{P}_0$. Moreover, assume that the observations are available intermittently, i.e., one has available the observations $\boldsymbol{y}'_k = \gamma_k \boldsymbol{y}_k$, where $\gamma_k \in \{0, 1\}$ is a random sequence.

Channel statistics: This paper consider three models for γ_k, abbreviated as LGB, LGM, and LGSM, respectively Linear-Gaussian-*Bernoulli*, *Markov*, and *Semi-Markov*. The LGSM model subsumes the other two, but it is still worth considering the simpler models separately because for them the statements of some results are more elegant and compact. In the *LGB model*, γ_k is a sequence of independent Bernoulli random variables, with $\mathbb{P}(\{\gamma_k = 1\}) = \overline{\gamma}$, where $\overline{\gamma}$ is the arrival rate. This is the setup considered by Sinopoli *et al*. [5] and the majority of the literature. In the *LGM model*, γ_k is a Markov chain. Let the transition probabilities be $\mathbb{P}(\{\gamma_{k+1} = j | \gamma_k = i\}) = p_{ij}$. There are two degrees of freedom: the failure (p_{10}) and recovery (p_{01}) rates. This is called the Gilbert-Elliott model and is the setup considered by Huang and Dey [8].

The *LGSM model* is a further generalization, where the channel state is described by a semi-Markov chain: a Markov chain that persists in a certain state for a random period of time, which constitutes an additional degree of freedom. In general, one defines a semi-Markov chain $\{\mathrm{I}_k\}_{k=0}^{\infty}$ as follows. Let $\{(s_n, \Delta_n)\}_{n=0}^{\infty}$ be a Markov renewal process, where s_n is the state and $\Delta_n \in \mathbb{Z}_+$ is the number of steps that the system persists in state s_n. We assume that s_n and Δ_n are conditionally independent given s_n:

$$\mathbb{P}(\{s_{n+1} = j, \Delta_{n+1} = k | s_n = i, \Delta_n = l\}) = p_{ij}\delta^{(i)}(k).$$

Define the time until the n-th renewal as $\tau_n = \sum_{m=1}^{n} \Delta_m$. Let $n(k)$ be the number of renewals until time k, such that $\tau_{n(k)} \leq k < \tau_{n(k)+1}$. Then we can define I_k to be a semi-Markov chain by setting $\mathrm{I}_k = s_{n(k)}$. Define t_k to be the time since last renewal: $t_k = k - \tau_{n(k)}$. One can show that $\{(\mathrm{I}_k, t_k)\}$ is a Markov chain. One recovers the LGB model by setting $\delta^{(i)}(k) = 0$ for $k > 1$.

Evolution of $\mathbf{P}_k$ as an iterated function system: Even if some observations are missing, the conditional estimate of $\boldsymbol{x}_k$, given the available observations until time k is still Gaussian [5]; let $\mathbf{P}_k$ be the corresponding covariance matrix. If the observations are always available, the evolution of $\mathbf{P}_k$ is deterministic; if $(\mathbf{A}, \mathbf{B})$ is stabilizable and $(\mathbf{A}, \mathbf{C})$ is detectable, $\mathbf{P}_k$ tends to the fixed point $\mathbf{P}_\infty$ from any $\mathbf{P}_0$ [3]. If there are missing measurements, $\mathbf{P}_k$ is a random walk in $\mathcal{P}(n)$ that is described by the recursion

$$\mathbf{P}_{k+1} = \begin{cases} h(\mathbf{P}_k), & \text{if } \gamma_k = 0, \\ g(\mathbf{P}_k), & \text{if } \gamma_k = 1. \end{cases}$$

Let $\mathbf{Q} \triangleq \mathbf{B}\mathbf{B}^*$ and $\boldsymbol{\mathcal{I}} \triangleq \mathbf{C}^*\mathbf{C}$. The two maps $g, h : \mathcal{P}(n) \to \mathcal{P}(n)$ can be written compactly as

$$h: \quad \mathbf{P} \mapsto \quad \mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q} \tag{2}$$

$$g: \quad \mathbf{P} \mapsto \quad \left((\mathbf{A}\mathbf{P}\mathbf{A}^* + \mathbf{Q})^{-1} + \boldsymbol{\mathcal{I}}\right)^{-1} \tag{3}$$

The Kalman filter and analogous variants implement the recursion with different representations for $\mathbf{P}$, and faster and more numerically stable algorithms than (3), which is used in the present analysis for convenience and compactness.

This setup is well described by the formalism of iterated function systems [17]. Formally, one assumes to have a metric space $(\mathcal{X}, d)$, and a finite number of maps $\{f_i\}$ from $\mathcal{X}$ to itself. Given a random sequence I_k, one defines a random walk as follows: $x_{k-1} = f_{\mathrm{I}_k}(x_k)$. In our case, $\mathcal{X} = \mathcal{P}$, the sequence is $\mathrm{I}_k = \gamma_k$ and the maps are $f_0 = h$, $f_1 = g$. In the following we keep using the generic names $(\mathcal{X}, d)$, $\{f_i\}$, I_k for generic statements about iterated function systems.

III. Existence of the stationary distribution

This section proves the existence of the stationary distribution of $\mathbf{P}$, under very mild conditions. The proof is simple and builds on two lines of results: some general results concerning the theory of iterated function systems [18], and some particular results about the geometric properties of Riccati iterations for a special choice of distance [19].

A. Properties of recurrent iterated function systems

Theorem 1. *(Barnsley) Let $x_{k-1} = f_{\mathrm{I}_k}(x_k)$ be an iterated function system on a complete metric space. If $\{\mathrm{I}_k\}$ is an irreducible Markov chain and there exists an integer $K \geq 1$ such that*

$$\mathbb{E}_{(\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_K)}\{\log \mathit{Lip}(f_{\mathrm{I}_1} \circ f_{\mathrm{I}_2} \circ \cdots \circ f_{\mathrm{I}_K})\} < 0, \tag{4}$$

then the Markov process $\{(\mathrm{I}_k, x_k)\}$ has a unique stationary distribution, and from any initial state, the empirical distribution tends to the stationary distribution with probability 1.

Remark 2. Let us decipher the Theorem and the condition (4). We are looking at the behavior of the process over K consecutive time steps. The expectation is with respect to all possible sequences of length K. The number $\log \mathrm{Lip}(f_{\mathrm{I}_1} \circ f_{\mathrm{I}_2} \circ \cdots \circ f_{\mathrm{I}_K})$ is the logarithm of the Lipschitz constant of the composition of the maps corresponding to the discrete state sequence; it is negative if the composition is contracting the state space. Therefore, condition (4) is a constraint on the average separation of the phase space under the uncertain dynamics. In the degenerate case in which there is only one function ($n = 1$), the condition (4) is simply $\mathrm{Lip}(f) < 1$, thus Theorem 1 can be seen as a stochastic generalization of the contraction mapping theorem. Such generalization is quite powerful: in fact, it is not assumed that any of the maps f_i be contractions by themselves. The generalization to the semi-Markov case is due to Stenflo [20].

Theorem 3. *(Stenflo) Assume that a IFS is driven by a semi-Markov chain $\{\mathrm{I}_k\}$, whose construction was presented in Section II. Assume that: the chain $\{s_n\}$ is irreducible; the chain $\{(\mathrm{I}_k, t_k)\}$ is aperiodic; there is finite expected switching time: $\mathbb{E}\{\Delta_n\} < \infty$. Moreover, assume that there exists an integer $K \geq 1$ such that* (4) *holds. Then the random walk $\{(\mathrm{I}_k, x_k)\}$ has a unique stationary distribution, and from any initial state, the empirical distribution tends to the stationary distribution with probability 1.*

B. Geometric properties of Riccati recursions

We recall some of the results developed by Bougerol [19] regarding the geometric properties of the maps h, g. These properties hold when the set of positive definite matrices $\mathcal{P}(n)$ is equipped with the following unusual distance; later, in Section IV-B, we will comment more on its significance and interpretation.

Lemma 4. *$\mathcal{P}(n)$ is a complete metric space when equipped with the distance*

$$d_r{}^2(\mathbf{P}_1, \mathbf{P}_2) = \sum_{i=1}^{n} \log^2 \left(\lambda_i(\mathbf{P}_1\mathbf{P}_2^{-1})\right). \tag{5}$$

Theorem 5. *(Bougerol) In the metric d defined by* (5),

1) *If $\mathbf{A}$ is non-singular, the maps h and g are nonexpansive:*

$$\mathit{Lip}(h) \leq 1, \qquad \mathit{Lip}(g) \leq 1. \tag{6}$$

2) *If $\mathbf{A}$ is non-singular, $(\mathbf{A}, \mathbf{B})$ is controllable, $(\mathbf{A}, \mathbf{C})$ observable, there exists an integer K such that the composition $g^K = g \circ \cdots \circ g$ of K copies of g is a strict contraction:*

$$\mathit{Lip}(g^K) < 1. \tag{7}$$

C. Convergence results

Proposition 6. *Assume $\mathbf{A}$ non-singular, $(\mathbf{A}, \mathbf{B})$ controllable, $(\mathbf{A}, \mathbf{C})$ observable, and:*

- for the LGB model*: The arrival probability is positive: $\overline{\gamma} > 0$.*
- for the LGM model*: The Markov chain $\{\gamma_k\}$ is irreducible.*

- for the LGSM model*: a) The Markov chain $\{s_n\}$ is irreducible; b) The Markov chain $\{(\gamma_k, t_k)\}$ is aperiodic; c) There is finite expected switching time: $\mathbb{E}\{\Delta_n\} < \infty$.*

Then the Markov process has a unique stationary distribution, and from any initial state, the empirical distribution tends to the stationary distribution with probability 1.

Proof: Consider the hypotheses of Theorem 1, with K as in (7). Equation (4) is a positive combination of terms of the kind $\log \mathrm{Lip}(h \circ h \circ \cdots \circ h)$, $\log \mathrm{Lip}(g \circ h \circ \cdots \circ h)$, and so on, for all possible permutations of g and h of length K. By Theorem 5, and in particular by equations (6)-(7), all those terms are nonpositive. Moreover, the term $\log \mathrm{Lip}(g^K)$ is strictly negative, due to equation (7). Therefore, the right-hand side of (4) is strictly negative, and the theorem holds. ■

Remark 7. Note that the proof would not be valid with the usual distance based on the Frobenius norm ($d_F(\mathbf{P}_1, \mathbf{P}_2) \triangleq \|\mathbf{P}_1 - \mathbf{P}_2\|_F$), because the map h would not be nonexpansive if $\mathbf{A}$ is unstable. The contraction properties (6) suggest that d_r is a convenient choice for studying random Riccati recursions.

Remark 8. In contrast to Kar *et al.* [12], who provide a proof in the LGB case, we have the additional assumption that $\mathbf{A}$ is non-singular. This, however, is a very mild assumption, and it is always true if the discrete dynamical system (1) arises from the discretization of a continuous-time system.

IV. Performance measures and existence of the Riemannian mean

The expected value of the covariance is perhaps the easiest choice for assessing the performance of the system; however, this expectation is unbounded when operating below Sinopoli's critical probability, while the process has a stationary distribution in essentially all non degenerate cases (Proposition 9). The stationary distribution, being a fractal density [14], cannot have a compact description. Therefore, it is of interest to investigate alternative performance measures, in the spirit of a "mean", that can synthesize the performance of the system.

Admittedly, there is no intrinsic "correct" choice of performance measure. However, we can find some desirable properties for a hypothetical performance measure $\mathbb{M}\{\mathbf{P}\}$:

1) *Existence*: The performance measure should exist if the stationary distribution exists.
2) *Invariance to change of coordinates*: The measure should be invariant to a linear change of coordinate ($\boldsymbol{x} \mapsto \mathbf{T}\boldsymbol{x}$) of the original system:
$$\mathbb{M}\{\mathbf{T}\mathbf{P}\mathbf{T}^*\} = \mathbf{T}\mathbb{M}\{\mathbf{P}\}\mathbf{T}^*.$$
3) *Covariance-information symmetry:* The Information Filter is the dual of the Kalman Filter; the information matrix and covariance matrix are two different parameterizations that describe the same object (the dispersion of the error distribution). If one considers this point of view, because of the one-to-one correspondence between covariance and information matrices, and the fact that they live in the same space (positive definite matrices), it is reasonable to require that averaging covariances corresponds to averaging information matrices, and then inverting:
$$\mathbb{M}\{\mathbf{P}\} = \mathbb{M}\{\mathbf{P}^{-1}\}^{-1}.$$

Moments of covariance matrices: One alternative is considering different moments of the covariance; that is, considering $\mathbb{E}\{\mathbf{P}^r\}$ for some $r > 0$. Particularly interesting is $\mathbb{E}\{\sqrt{\mathbf{P}}\}$, because it corresponds the expected error norm, instead of the expected *squared* error norm. It is implied by some results of Huang and Dey [8] that, in the LGB scalar case, each moment $\mathbb{E}\{\mathbf{P}^r\}$ ceases to exist at certain critical probabilities. Therefore, this choice does not satisfy the first existence requirement.

Averaging information matrices: Another alternative is taking the expectation of the information matrix $\mathbf{P}^{-1}$. Because the stationary distribution of $\mathbf{P}$ has support in the set $\mathbf{P} \geq \mathbf{P}_\infty$, the information matrix $\mathbf{P}^{-1}$ is bounded above by $\mathbf{P}_\infty^{-1}$. Thus $\mathbb{E}\{\mathbf{P}^{-1}\}$ is always well defined. This expectation is also invariant to a change of coordinates. However, this solution is somewhat unsatisfying, because it breaks the symmetry between covariance matrices and information matrices: $\mathbb{E}\{\mathbf{P}\} \neq \mathbb{E}\{\mathbf{P}^{-1}\}^{-1}$.

The alternative proposed in this section is based on the definition of a Riemannian mean. Section IV-A defines the notion of Riemannian mean. Section IV-B describes the properties of the Riemannian mean according to the distance (5) and proves its existence. Section IV-C discusses the interpretation of the mean in information geometry. Section IV-E shows some numerical simulations.

A. Means on Riemannian manifolds

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 $\mathbb{E}\{X\} \triangleq \int_{\mathbb{R}^n} \boldsymbol{x}\, d\mu(\boldsymbol{x})$. 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:

$$\mathbb{E}\{X\} = \arg\min_y \mathbb{E}\left\{\|X - y\|_2^2\right\}. \tag{8}$$

Equivalently, the mean is the point y where the "errors" $\mathbb{E}\{y - X\}$ balance to zero:

$$\mathbb{E}\left\{\mathbb{E}\{X\} - X\right\} = 0. \tag{9}$$

The latter is more general, because it does not require the dispersion in (8) to be bounded. Equations (8)-(9) can be used to extend the idea of mean to Riemannian manifolds.

A Riemannian manifold $(\mathcal{M}, m)$ is a differentiable manifold $\mathcal{M}$ equipped with a smooth metric m on the tangent space [21]. The "length" $\ell(c)$ of a curve $c : [0,1] \to \mathcal{M}$ is defined as $\ell(c) = \int_0^1 \sqrt{m(\dot{c}(t), \dot{c}(t))}\, dt$. Given the notion of length, the distance between two points is

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

Consider a Riemannian manifold $(\mathcal{M}, m)$ 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:

$$\mathbb{M}_d\{X\} \triangleq \arg\inf_{y \in \mathcal{M}} \mathbb{E}\left\{d^2(X, y)\right\}. \tag{10}$$

Moreover, the *dispersion* $\mathbb{E}\left\{d^2(\mathbb{M}_d\{X\}, X)\right\}$ can be taken as a generalization of the (trace of) the covariance matrix. This definition assumes that the expected squared distance is bounded; a weaker generalization can be derived from (9) by replacing the subtraction by the inverse of the exponential map, as explained in Corcuera and Kendall [22], but that is not needed for the purpose of this paper. 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).

B. A Riemannian mean for the manifold of Gaussian distributions

Firstly, we recall some facts about the distance d_r defined by (5):

- It is a Riemannian distance [24] arising from using the metric

$$m(\mathbf{X},\mathbf{Y}) = \tfrac{1}{2}\text{Trace}\left\{\mathbf{P}^{-1}\mathbf{X}\mathbf{P}^{-1}\mathbf{Y}\right\}. \tag{11}$$

- It is invariant to a linear change of coordinates of the original system [25]:

$$d_r(\mathbf{P}_1,\mathbf{P}_2) = d_r(\mathbf{T}\mathbf{P}_1\mathbf{T}^*,\mathbf{T}\mathbf{P}_2\mathbf{T}^*). \tag{12}$$

- It is the only Riemannian distance preserved under inversion [24]:

$$d_r(\mathbf{P}_1,\mathbf{P}_2) = d_r(\mathbf{P}_1^{-1},\mathbf{P}_2^{-1}). \tag{13}$$

Now consider the corresponding Riemannian mean $\mathbb{M}_{d_r}\{\mathbf{P}\}$ defined by (10). We can prove that it satisfies all the desirable properties we discussed at the beginning of the section.

Proposition 9. *Properties of the Riemannian mean* $\mathbb{M}_{d_r}\{\mathbf{P}\}$*:*

1) *(Existence)* It exists and it is unique under the hypotheses of Proposition 6; for the LGSM model, add the constraint that $\delta^{(o)}(k)$ is eventually bounded by a geometric series.
2) *(Invariance)* It is invariant to a change of coordinates of the original system:

$$\mathbb{M}_{d_r}\{\mathbf{T}\mathbf{P}\mathbf{T}^*\} = \mathbf{T}\mathbb{M}_{d_r}\{\mathbf{P}\}\mathbf{T}^*.$$

3) (*Symmetry*) It is the only Riemannian mean invariant to inversion:

$$\mathbb{M}_{d_r}\{\mathbf{P}\} = \mathbb{M}_{d_r}\{\mathbf{P}^{-1}\}^{-1}.$$

Proof: The last two properties follow directly from properties (12)-(13). As for the first property, if we prove existence, uniqueness follows from the fact that the manifold has nonpositive curvature [23]. To prove that the Riemannian mean exists, it is sufficient to show that the minimization problem (8) is feasible; i.e., there exists a $\mathbf{X}$ such that $\mathbb{E}\{{d_r}^2(\mathbf{X},\mathbf{P})\}$ is bounded; in our case it is convenient to choose $\mathbf{X} = \mathbf{P}_\infty$.

For computing the necessary expectation, we consider the covariance at a generic time instant, and we write it as a function of the past arrival events. Let $\boldsymbol{\gamma} \in \{0,1\}^{\mathbb{N}}$ be the infinite sequence of past arrivals, and let $\mathbf{P}(\boldsymbol{\gamma})$ be the covariance as a function of $\boldsymbol{\gamma}$. We will show that $\mathbb{E}_{\boldsymbol{\gamma}}\{{d_r}^2(\mathbf{P}_\infty,\mathbf{P}(\boldsymbol{\gamma}))\}$ is bounded. Let $\{0,1\}_K^{\mathbb{N}}$ be the subset of $\{0,1\}^{\mathbb{N}}$ composed by arrival sequences that contain, at some point, K consecutive "1"s, with K as in (7). Given the hypotheses, for the three models considered, the set $\{0,1\}_K^{\mathbb{N}}$ has measure 1, because, under the hypotheses, the probability that an infinite sequence does not contain K consecutive "1"s is 0. Thus, when taking the expectation, we can assume $\boldsymbol{\gamma} \in \{0,1\}_K^{\mathbb{N}}$. Define the quantity $\boldsymbol{\tau}(\boldsymbol{\gamma})$ as the number of step since the last reception of K measurements. Thus each sequence is composed by $\boldsymbol{\tau}(\boldsymbol{\gamma})$ elements, followed by K "1"s, followed by an infinite tail:

$$\begin{aligned}&\qquad\quad\downarrow\text{last received}\\ \boldsymbol{\gamma} \quad &= \underbrace{10010010\cdots001}_{\boldsymbol{\tau}(\boldsymbol{\gamma})\,\text{steps}}\underbrace{001111\cdots1111}_{\text{last } K\text{-event}}\underbrace{101010001010\cdots}_{\text{infinite tail}}.\end{aligned}$$

For any sequence $\boldsymbol{\gamma}$, let us define $\tilde{\boldsymbol{\gamma}}$ as $\boldsymbol{\gamma}$ with the first $\boldsymbol{\tau}(\boldsymbol{\gamma})$ elements set to 0:

$$\tilde{\boldsymbol{\gamma}} \quad = \underbrace{00000000\cdots000}_{\boldsymbol{\tau}(\boldsymbol{\gamma})\,\text{steps}}\underbrace{001111\cdots1111}_{\text{last } K\text{-event}}\underbrace{101010001010\cdots}_{\text{infinite tail}}.$$

From the order-preserving properties of g, h, it follows that $\mathbf{P}(\boldsymbol{\gamma}) \leq \mathbf{P}(\tilde{\boldsymbol{\gamma}})$. Moreover, we can bound $\mathbf{P}(\tilde{\boldsymbol{\gamma}})$ as follows: after K consecutive measurements, the covariance is bounded by some matrix $\overline{\mathbf{P}}$: $g^K(\cdot) \leq \overline{\mathbf{P}} < \infty$. Thus $\mathbf{P}(\tilde{\boldsymbol{\gamma}}) \leq h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right)$, and we find the bound $\mathbf{P}_\infty \leq \mathbf{P}(\boldsymbol{\gamma}) \leq h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right)$. By Lemma 10 below, we obtain that $d_r(\mathbf{P}_\infty,\mathbf{P}(\boldsymbol{\gamma})) \leq d_r(\mathbf{P}_\infty, h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right))$. By squaring and taking expectations,

$$\mathbb{E}_{\boldsymbol{\gamma}}\{{d_r}^2(\mathbf{P}_\infty,\mathbf{P}(\boldsymbol{\gamma}))\} \leq \mathbb{E}_{\boldsymbol{\tau}(\boldsymbol{\gamma})}\{{d_r}^2(\mathbf{P}_\infty, h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right))\}. \tag{14}$$

To summarize, so far we can conclude that the Riemannian mean exists if the expression $\mathbb{E}_{\boldsymbol{\tau}(\boldsymbol{\gamma})}\{{d_r}^2(\mathbf{P}_\infty, h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right))\}$ is bounded. This expression only depends on the statistics of $\boldsymbol{\tau}(\boldsymbol{\gamma})$. The critical observation is that, in all three models, the probability that $\boldsymbol{\tau}(\boldsymbol{\gamma}) = k$ decreases geometrically with k; that is, for big enough k, we have that

$$\mathbb{P}(\{\boldsymbol{\tau}(\boldsymbol{\gamma}) = k\}) \leq ar^k, \qquad 0 < r < 1. \tag{15}$$

This can be seen directly for the LGB/LGM models; for the LGSM model, it follows from the fact that $\{s_n\}$ is assumed irreducible, and that $\delta^{(0)}(k)$ is assumed to eventually decrease geometrically. At this point, the proof consists in using (15) to show that (14) is bounded. The distance d_r is invariant to change of coordinates in the original system, so we can choose a frame of reference such that $\mathbf{P}_\infty = \mathbf{I}$. It is easy to bound the distance from $\mathbf{I}$ of a matrix $\mathbf{X} \geq \mathbf{I}$:

$${d_r}^2(\mathbf{I},\mathbf{X}) = \sum_{i=1}^{n} \log^2 \lambda_i(\mathbf{X}) \leq n\,\log^2\left(\|\mathbf{X}\|\right). \tag{16}$$

Note that the second step is valid only for $\mathbf{X} \geq 1$ ($\lambda_i(\mathbf{X}) \geq 1$). In our case, $\mathbf{X} = \mathbf{T}h^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\left(\overline{\mathbf{P}}\right)\mathbf{T}^*$, for some $\mathbf{T}$ such that $\mathbf{T}\mathbf{P}_\infty\mathbf{T}^* = \mathbf{I}$; we can write it explicitly as

$$\mathbf{X} = \mathbf{T}\big(\mathbf{A}^{\boldsymbol{\tau}(\boldsymbol{\gamma})}\overline{\mathbf{P}}\left(\mathbf{A}^*\right)^{\boldsymbol{\tau}(\boldsymbol{\gamma})} + \sum_{i=0}^{\boldsymbol{\tau}(\boldsymbol{\gamma})-1}\mathbf{A}^i\mathbf{Q}\left(\mathbf{A}^*\right)^i\big)\mathbf{T}^*.$$

The norm of the matrix $\mathbf{X}$ can be bounded by $c_1{c_2}^{2\boldsymbol{\tau}(\boldsymbol{\gamma})}$ for some $c_1, c_2 > 0$ not depending on $\boldsymbol{\gamma}$. Thus $n\log^2\left(\|\mathbf{X}\|\right) \leq \boldsymbol{\tau}(\boldsymbol{\gamma})^2c_3 + \boldsymbol{\tau}(\boldsymbol{\gamma})c_4 + c_5$ for some $c_3, c_4, c_5 > 0$, and from (16),

$${d_r}^2(\mathbf{P}_\infty,\mathbf{P}(\boldsymbol{\gamma})) \leq \boldsymbol{\tau}(\boldsymbol{\gamma})^2c_3 + \boldsymbol{\tau}(\boldsymbol{\gamma})c_4 + c_5.$$

Take the expectation directly with respect to $\boldsymbol{\tau}$:

$$\begin{aligned}\mathbb{E}\{{d_r}^2(\mathbf{P}_\infty,\mathbf{P})\} \quad &\leq \quad \sum_{k=0}^{\infty}\mathbb{P}(\{\boldsymbol{\tau} = k\})(c_3k^2 + c_4k + c_5)\\ &\leq \quad c_6 + \sum_{k=0}^{\infty}ar^k(c_3k^2 + c_4k + c_5).\end{aligned}$$

Series of the kind $\sum_{k=0}^{\infty}k^\alpha x^k$ with $\alpha \geq 0$ are convergent if $|x| < 1$, hence the dispersion $\mathbb{E}\{{d_r}^2(\mathbf{P}_\infty,\mathbf{P})\}$ is always bounded and the Riemannian mean always exists. ∎

Lemma 10. *For the distance defined in* (5)*,* $\mathbf{P}_1 \leq \mathbf{P}_2 \leq \mathbf{P}_3 \;\Rightarrow\; d_r(\mathbf{P}_1,\mathbf{P}_2) \leq d_r(\mathbf{P}_1,\mathbf{P}_3)$.

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 $d_r(\mathbf{I},\mathbf{P}_2') \leq d_r(\mathbf{I},\mathbf{P}_3')$ by direct computation using (5). ∎

C. Interpretation of d_r in information geometry

Information geometry [16], [26] is a relatively new branch of statistics that studies the properties of families of probability distributions considered as Riemannian manifolds. With it, one can rigorously define the Riemannian distance between two probability distributions, and, based on the construction in Section IV-A, a notion of Riemannian mean. It is interesting to look at this theory, because, while it is debatable how one can average "generic" positive definite matrices, information geometry gives a unique answer when considering that those matrices represent Gaussian probability distributions. In particular, the distance (5) is the *natural* distance between two

Gaussian distributions with the same mean; we briefly discuss this fact.

Information geometry gives families of probability distribution a Riemannian structure by using a generalization of the Fisher Information Matrix (FIM) as a Riemannian metric. In the Gaussian case, one defines the FIM as follows [27]: if the available observations $\boldsymbol{z} \in \mathbb{R}^q$ have a Gaussian distribution whose mean and covariance are parametrized by an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^n$: $\boldsymbol{z} \sim \mathcal{G}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, then the FIM for $\boldsymbol{\theta}$ is the semidefinite positive matrix $\boldsymbol{\mathcal{I}}[\boldsymbol{\theta}]$ defined as

$$\boldsymbol{\mathcal{I}}[\boldsymbol{\theta}]_{a,b} = \frac{\partial \boldsymbol{\mu}}{\partial \theta_a}^* \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_b} + \quad (17)$$

$$\tfrac{1}{2}\mathrm{Trace}\left\{ \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_a} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_b} \right\}. \quad (18)$$

This FIM can be interpreted as 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 coincides with (11) (compare it with the second term in (17)), and therefore the natural distance between two zero-mean Gaussian distributions is exactly $d_r(\mathbf{P}_1, \mathbf{P}_2)$.

It is worth noting that $d_r(\mathbf{P}_1, \mathbf{P}_2)$ has also a *practical* significance in estimation, as 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 by Amari [16]. By contrast, the Euclidean distance $\|\mathbf{P}_1 - \mathbf{P}_2\|_F$ (implicitly used when considering $\mathbb{E}\{\mathbf{P}\}$) does not have such a property.

D. Computational issues

There are some open issues regarding the computation of the Riemannian mean of the distribution of interest. Just like a closed-form expression for the Euclidean mean $\mathbb{E}\{\mathbf{P}\}$ as a function of the system parameters and packet dropping statistics is not known, even in the LGB case, we do not have a closed-form expression for the Riemannian mean $\mathbb{M}_{d_r}\{\mathbf{P}\}$ either. In either case, part of the difficulty in obtaining a closed-form expression lies in the fact that the LGB distribution has fractal support [14].

For the Riemannian mean, there is an additional computational problem. Given a set of samples, it is immediate to compute the Euclidean mean $\mathbb{E}\{\mathbf{P}\}$. However, a closed-form expression for the Riemannian mean of a set of samples is not available in general. It is an open question whether this computational difficulty is relevant to the problem of characterizing the Riemannian mean of the LGB distribution.

For the current research efforts regarding the problem of computating the Riemannian mean of a set of samples, see the work by Fiori and Tanaka [28] for the general case, and the work of Fortunati *et al.* [29] for the particular case of positive definite matrices, and its application to the problem of target detection.

E. Numerical example

It is challenging to illustrate the behavior of the system numerically, because some phenomena of interest, such as the convergence of $\mathbb{E}\{\mathbf{P}\}$ and its moments, cannot be simulated on a computer, because, over finite trajectories, $\mathbf{P}_k$ is always bounded. Moreover, the stationary probability distribution for $\mathbf{P}$ is a fractal [14], hard to visualize. Also, one should note that there might be inaccuracies due to the floating-point representation, because for computing these statistics one has to work simultaneously with very large and very small numbers. With these caveats in mind, it is still interesting to look at the performance measures that we discussed, interpreted as empirical means over finite trajectories. Two systems are simulated: one stable, and one unstable (see the caption of Fig. 1 for simulation details). The figures show the values of $\mathbb{E}\{\mathbf{P}\}$, $\mathbb{E}\{\sqrt{\mathbf{P}}\}^2$, $\mathbb{E}\{\mathbf{P}^{-1}\}^{-1}$, and $\mathbb{M}_{d_r}\{\mathbf{P}\}$ for the empirical distributions as a function of the arrival probability, assuming the LGB model. The results are plotted twice: once as as covariances (variances), and as information matrices (scalars). Note that, due to Jensen's inequality, there is a precise order: $\mathbb{E}\{\mathbf{P}\} \geq \mathbb{E}\{\sqrt{\mathbf{P}}\}^2 \geq \mathbb{M}_{d_r}\{\mathbf{P}\} \geq \mathbb{E}\{\mathbf{P}^{-1}\}^{-1}$. In the stable case (Fig. 1a-1b), all the statistics have the same qualitative behavior, while in the unstable case $\mathbb{E}\{\mathbf{P}\}$ and $\mathbb{E}\{\sqrt{\mathbf{P}}\}^2$ are expected to diverge at their critical probabilities (in the simulations, we obtain finite but very large averages).

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–)

Riccati recursions are nonexpansive in a particular Riemannian distance for the manifold of positive definite matrices. This geometric property allows to easily prove the existence of the stationary distribution, by applying well established properties of iterated function systems, for a wide class of arrival models (Proposition 6). Given such particular choice of distance, one can define the Riemannian mean for the stationary distribution. As a performance measure, it has interesting properties (Proposition 9): it exists under very mild conditions, it is invariant to a change of coordinates in the original system, and it respects the symmetry of covariance and information matrices: the mean of the information matrices is the inverse of the mean of the covariance matrices. For some readers, it might be interesting that, according to the information geometry interpretation, this is the "natural" distance to use for averaging Gaussian distributions.

These results show that the *geometric* properties of Riccati recursions complement the well-studied *algebraic* properties (e.g., order preservation) and offer the researcher interesting tools and interpretations for studying the problem.

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.

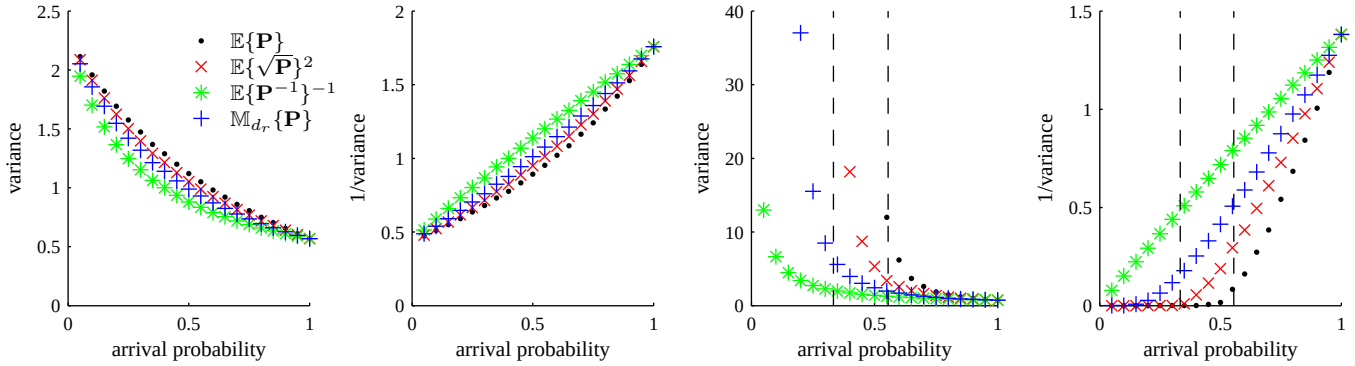
[5] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1453–1464, Sept. 2004.

[6] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," in *Proceedings of the IEEE Conference on Decision and Control (CDC)*, pp. 2692–2697, Dec. 2008.

[7] 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.

[8] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, 2007.

[9] Y. Xu and J. Hespanha, "Estimation under uncontrolled and controlled communications in networked control systems," in *Proceedings of the IEEE Conference on Decision and Control (CDC)*, pp. 842–847, Dec. 2005.



(a) $\mathbf{A} = 0.75$ (covariance) (b) $\mathbf{A} = 0.75$ (information) (c) $\mathbf{A} = 1.5$ (covariance) (d) $\mathbf{A} = 1.5$ (information)

Figure 1: The simulation show the performance for Kalman filtering with intermittent observations, for a stable ($\mathbf{A} = 0.75$) and an unstable ($\mathbf{A} = 1.5$) system with $\mathbf{Q} = \boldsymbol{\mathcal{I}} = 1$. The observations follow the LGB model. The pictures show the empirical averages of the four measures computed over 10^6 trajectories of 200 steps each as a function of the arrival probability $\overline{\gamma}$. The averages are shown as (co)variances in (a)-(c), and as the inverse of variances (information) in (b)-(d). The vertical bars show the critical probabilities for $\mathbb{E}\{\mathbf{P}\}$ and $\mathbb{E}\{\sqrt{\mathbf{P}}\}^2$.

[10] S. Kar, B. Sinopoli, and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence to a stationary distribution." arXiv:abs/0903.2890, March 2009.

[11] A. Censi, "Geometric remarks on Kalman filtering with intermittent observations." arXiv.org:0906.1637, June 2009.

[12] S. Kar and J. Moura, "Kalman filtering with intermittent observations: Weak convergence and moderate deviations." arXiv:0910.4686v1, October 2009.

[13] 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.

[14] 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.

[15] A. Vakili and B. Hassibi, "On the steady-state performance of Kalman filtering with intermittent observations for stable systems," in *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 2009.

[16] S. Amari, *Differential-Geometrical Methods in Statistics*. Springer Verlag, 1985.

[17] G. A. Edgar, *Integral, probability, and fractal measures*. Springer, 1998.

[18] M. Barnsley and J. H. Elton, "A new class of Markov processes for image encoding," *Adv. Appl. Prob*, vol. 20, 1988.

[19] P. Bougerol, "Kalman filtering with random coefficients and contractions," *SIAM J. Control and Optimization*, vol. 31, no. 4, 1993.

[20] O. Stenflo, "Iterated function systems controlled by a semi-Markov chains," *Theory of Stochastic Processes*, vol. 2, no. 18, pp. 305–313, 1996.

[21] M. do Carmo, *Riemannian Geometry*. Birkhauser, 1994.

[22] J. M. Corcuera and W. S. Kendall, "Riemannian barycentres and geodesic convexity," *Math. Proc. Cambridge Philos. Soc*, vol. 127, 1999.

[23] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. on Pure and Applied Math.*, vol. 30-5, 1977.

[24] S. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, 2005.

[25] 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.

[26] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics*. Chapman & Hall/CRC, 1993.

[27] G. Casella and R. L. Berger, *Statistical Inference*. Duxbury Press, 1990.

[28] S. Fiori and T. Tanaka, "An algorithm to compute averages on matrix lie groups," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4734–4743, 2009.

[29] S. Fortunati, F. Gini, M. S. Greco, A. Farina, A. Graziano, and S. Giompapa, "A gradient descent algorithm for evaluating the Karcher-Fréchet mean in the manifold of the symmetric positive definite matrices and its application in target detection," tech. rep., Universitá degli Studi di Pisa, SESM-FNM — SELEX Sistemi Integrati S.p.a., July 2010.

[30] M. Atiyah, *Mathematical Evolutions*, ch. Mathematics in the 20th century. Mathematical Association of America, 2002.