COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 8, No. 3, pp. 185-200, 2008

001

WIDE-SENSE ESTIMATION ON THE SPECIAL ORTHOGONAL GROUP*

ALESSANDRO CHIUSO†, GIORGIO PICCI‡, AND STEFANO SOATTO§

Abstract. In this paper we consider a simple estimation problem on the special orthogonal group $SO(n)$ and indicate a possible way to construct approximate filters which is much in the same spirit of the "wide sense" approach to linear filtering theory. Our interest is mainly motivated by applications to computer vision.

1. Introduction. Estimation on differentiable manifolds arises in a variety of applications such as target tracking, robotics and computer vision, see e.g. [14, 19, 20, 6]. In general the estimation problem on a differentiable manifold can be seen as a nonlinear estimation problem once (local) coordinates are introduced. Most of times the "solution" involves a stochastic partial differential equation for the conditional density whose solution is extremely difficult.

Finite dimensional filters may exist in the rare cases where invariant solutions of the PDE for the conditional density exist within a finitely parametrized class of densities. In such situations the partial differential equation can be reduced to an ordinary one whose solution is normally much simpler. This relates also to the existence of finite dimensional estimation algebras associated with the filtering problem, see [4], [1], [24] and references therein.

Although many attempts have been made in the past to obtain approximate solutions to the general nonlinear filtering problem, (see e.g. [3] and the references therein) the special case of the orthogonal group offers a rich structure and at least a chance of a deeper understanding of the problem.

We shall follow a similar approach to that initiated in the references [8] and [7], which consider estimation on a sphere. We shall see that the very special structure of the orthogonal group can be used to obtain a solution which is similar to the one obtained for the sphere.

2. Langevin Densities on the Special Orthogonal Group. First, let us recall what we mean by "probability density" on a manifold. Recall that a measure on an n-dimensional manifold is an n-form (volume form), which can be expressed in

*This paper is dedicated to Roger Brockett on the occasion of his 70th birthday.

†Department of Management and Engineering, Università di Padova, Stradella S. Nicola, 3, 36100 Vicenza, Italy. E-mail: chiuso@dei.unipd.it. This work has been supported in part by the national project *New methods and algorithms for Bayesian estimation, identification, adaptive and distributed control* funded by MIUR.

‡Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy.

§Department of Computer Science, UCLA, Los Angeles - CA.

local coordinates with respect to the Lebesgue measure in $\mathbb{R}^n$. A random variable with values on a manifold induces a (finite) measure on the manifold. When some regularity conditions are satisfied (absolute continuity), this measure can be expressed as the integral of a density function with respect to the base measure on the manifold (Radon-Nikodym).

It is known that every compact Lie Group $\mathcal{G}$ admits a unique bi-invariant (i.e. left and right invariant) measure μ such that $\mu(G) = 1$ [2, p. 247]. This is called the *Haar* or *base measure* of the group G. Densities in $SO(n)$ will be expressed with respect to the Haar measure.

Our aim is to develop a sort of "wide-sense" setting for doing estimation on the orthogonal group. To this purpose we need to introduce a family of distributions on the orthogonal group which should play a similar role to the Gaussian in Euclidean spaces. This family can also be seen as an instance of the Gibbs-Boltzmann distributions encountered in statistical physics.

It is reasonable to assume that observable quantities related to the orthogonal group are actions of the (elements of) the group on $\mathbb{S}^{n-1}$, i.e. objects of the form Ru, $u \in \mathbb{S}^{n-1}$. Take u_i, $i = 1, .., n$ to be an orthonormal basis and suppose that we would like to measure how far the rotation R is from a "reference rotation" M, which maps u_i into $v_i \doteq Mu_i$. For future use define $U \doteq [u_1, .., u_n]$ and $V \doteq [v_1, .., v_n]$. It is reasonable to measure how close the unit vectors v_i and $R_i u_i$ are by the inner products $\sigma_i \doteq \langle v_i, Ru_i \rangle$. Assume now some side information is available which assigns to us the average values $\bar{\sigma}_i \doteq E\sigma_i$, $i = 1, .., n$. It is well known from the theory of exponential families [10] that the maximum entropy distribution $p(R)$ which admits first moments $E\sigma_i = \bar{\sigma}_i$, takes the Gibbs-Boltzmann form

$$p(R) = \frac{1}{Z(\underline{k})} \exp\left(\sum_{i=1}^{n} k_i \langle v_i, Ru_i \rangle \right) \tag{1}$$

where the *partition function* $Z(\underline{k})$ is a function of certain parameters $\underline{k} \doteq [k_1, .., k_n]$ which are bijectively related to the mean values $\bar{\sigma}_i$, $i = 1, .., n$.

Defining $K \doteq \mathrm{diag}\{\underline{k}\}$, the function $p(R)$ can be written in the more convenient form as

$$\begin{aligned} p(R) &= \tfrac{1}{Z(\underline{k})} \exp\{\mathrm{Tr}\left(KV^\top RU\right)\} \\ &= \tfrac{1}{Z(\underline{k})} \exp\{\mathrm{Tr}\left(UKU^\top (VU^\top)^\top R\right)\} \\ &= \tfrac{1}{Z(\underline{k})} \exp\{\mathrm{Tr}\left(\Sigma M^\top R\right)\} \end{aligned} \tag{2}$$

where $\Sigma \doteq UKU^\top$ and $M \doteq VU^\top$. Note that, when $k_1 = ... = k_n = k$, the function $\mathrm{Tr}\left(\Sigma M^\top R\right) = k\mathrm{Tr}\left(M^\top R\right)$ is, up to a multiplicative constant, the Hilbert-Schmidt inner product $\langle M\,, R \rangle_{HS}$, of matrices.

Without loss of generality we shall assume that the diagonal elements of K are ordered, $|k_1| \geq |k_2| \geq ... \geq |k_n|$; it is also possible to show that one can, without

loss of generality, take $k_1 \geq k_2 \geq ... \geq k_{n-1} \geq 0$ while k_n can either be positive or negative[1].

It is natural to ask whether considering a larger number of observables $\langle u_i, Rv_i \rangle$ may give rise to a different distribution. To this purpose consider the pairs of vectors (x_i, y_i), $i = 1, ..., m$, $m > n$, $x_i, y_i \in \mathbb{R}^n$, and the associated density of the form

$$p(R) = \frac{1}{Z(\underline{\mu})} \exp\left(\sum_{i=1}^{m} \mu_i \langle y_i, Rx_i \rangle \right), \underline{\mu} \doteq [\mu_1, ..., \mu_m]. \tag{3}$$

Consider the Singular Value Decomposition

$$[x_1, ..., x_m][\mu_1 y_1, ..., \mu_n y_m]^\top = U_{xy} K_{xy} V_{xy}^\top \tag{4}$$

where K_{xy} is the diagonal matrix of singular values. We shall from now on assume that the decomposition (4) has been adjusted so that $\det(U_{xy}) = \det(V_{xy}) = 1$. This can be achieved by changing the sign of the last column of V_{xy} and of the element in position (n, n) in the diagonal matrix K.

By an argument similar to the one used in (2), it is easy to see that the maximum entropy distribution (3) has the form

$$p(R) = \frac{1}{Z(\underline{k}_{xy})} \exp\{ \mathrm{Tr}\left(\Sigma_{xy} M_{xy}^\top R \right) \} \tag{5}$$

where $\underline{k}_{xy}$ is the vector of singular values; i.e. $K_{xy} = \mathrm{diag}(\underline{k}_{xy})$ and

$$\Sigma_{xy} = U_{xy} K_{xy} U_{xy}^\top \quad M_{xy} = V_{xy} U_{xy}^\top .$$

Remarkably (5) is of the same form as (1).

The above motivates considering the exponential family of densities on $SO(n)$

$$\mathcal{L}(M, \Sigma) = \{ \ell(R) \propto \exp\{ \mathrm{Tr}\left(\Sigma M^T R \right) \ R \in SO(n) \} \tag{6}$$

where

$$\Sigma = \Sigma^\top = UKU^\top, K = \mathrm{diag}\{k_1, .., k_n\}, k_1 \geq ... \geq k_{n-1} \geq 0, |k_n| \leq k_{n-1}, M \in SO(n)$$

which we shall call *Langevin* densities for their analogy with the Langevin densities on the unit sphere [22] [8].

EXAMPLE 2.1. Let $n = 2$ and let $M = I$; parametrize $R(\theta) \in SO(2)$ as

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

[1] After the k_i's have been ordered by absolute value, one can make k_i $(i < n)$ positive, by changing its sign and by simultaneously changing the sign of v_i. The sign of k_n however is determined by the constraint $\det(M) = \det(UV^\top) = 1$.

then the density $p(R)$ takes the form

$$p(R) \propto \exp\{(k_1 + k_2)\cos(\theta)\}$$

which is the well-known Langevin (also known as Fisher or Von Mises) distribution on $\mathbb{S}^1$ [22]. This clearly depends only on the sum $k = k_1 + k_2$, which is nonnegative and plays the role of concentration parameter.

It is fairly easy to check that the matrix parameter M is the mode of the density and the eigenvalues of the symmetric matrix $\Sigma = UKU^\top$ are a measure of concentration. The normalization factor of the density, which we denoted by $Z(\underline{k})$, depends only on the singular value matrix K as can be easily seen by making use of bi-invariance (i.e. making left and right translations by $M^\top$ and U respectively.)

The following lemma clarifies the role of the concentration parameters $\underline{k} = [k_1, ..., k_n]$.

LEMMA 2.1. *Assume the random rotation $R \in SO(n)$ is distributed according to $\mathcal{L}(M, \Sigma)$. If $k_1 \to \infty$, then with probability one,*

$$R = M \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & [\ R_{n-1}\ ] & \\ 0 & & & \end{bmatrix} U^\top \tag{7}$$

i.e. R can be parametrized by an element $R_{n-1} \in SO(n-1)$. Moreover, R_{n-1} is distributed as $\mathcal{L}(I, K_{n-1})$, $K_{n-1} = \mathrm{diag}\{k_2, ..., k_n\}$. Furthemore, if $n - 1 = 2$ one can as well take $K_2 = \mathrm{diag}\{k, k\}$, $k = \frac{k_2+k_3}{2}$. Under this circumstance, the normalization constant $Z(\underline{k})$ depends only upon the sum $k_2 + k_3$ (which is necessarily non-negative).

The lemma of course can be iterated, letting k_1,..,k_{n-2} go to infinity. It is worth observing that when $k_{n-2} \to \infty$ the density depends only upon the average $k = \frac{k_{n-1}+k_n}{2}$. Hence, if also $k_{n-1} \to \infty$ the density collapses to a delta function on $SO(n)$ giving all mass to the mode matrix M. Obviously, when all $k_i's$ go to zero the density becomes uniform.

An appealing property of the Langevin densities, which is the analog of a property enjoyed by Langevin distributions on the unit sphere is stated in the following proposition:

PROPOSITION 2.1. *Among all probability densities p on the special orthogonal group having the same mean vector*

$$\hat{R} = \int_{SO(n)} R\, p(R)\, d\mu(R) \tag{8}$$

the Langevin distribution $\mathcal{L}(M, \Sigma)$ is the one of maximal entropy. The parameters Σ and M are recovered from the singular value decomposition[2] $\hat{R} = VSU^\top$ as $M =$

[2]As in (4) the sign of $K(n, n)$ is adjusted so that $\det(U) = \det(V) = 1$.

$VU^\top$, $\Sigma = UK_SU^\top$ where K_S is a diagonal matrix whose diagonal elements are functions of S.

Proof. Assume (8) holds with $\hat{R} = VSU^\top$, $U \in SO(n)$, $V \in SO(n)$. Consider $\bar{R} \doteq V^\top RU$. Using bi-invariance it follows that $E[\bar{R}] = S$. This can be rephrased in the form

$$E[\,\langle e_i, \bar{R}e_j\rangle\,] = S(i,j) = S(i,i)\delta_{ij}$$

where e_i is the $i-th$ canonical vector in $\mathbb{R}^n$. By well-known reasoning (see e.g. [10]) it follows that the maximum entropy density has the form

$$p(\bar{R}) \propto \exp\{\sum_{i,j} k_{ij}\langle e_i, \bar{R}e_j\rangle\}\,.$$

It is now a simple check based on symmetry arguments to show that $k_{ij} = k_{ji} = \delta_{ij}k_{ii}$ must hold[3]. This implies that

$$p(\bar{R}) \propto \exp\{\mathrm{Tr}\left(K\bar{R}\right)\}.$$

Using now bi-invariance of the Haar measure the probability density function of $R = U\bar{R}V^\top$ is

$$\begin{aligned} p(R) &= p(\bar{R})_{|\bar{R}=V^\top RU} \\ &\propto \exp\{\mathrm{Tr}\left(KV^\top RU\right)\} \\ &\propto \exp\{\mathrm{Tr}\left(KU^\top UV^\top RU\right)\} \\ &\propto \exp\{\mathrm{Tr}\left(UKU^\top UV^\top R\right)\} \\ &\propto \exp\{\mathrm{Tr}\left(\Sigma M^\top R\right)\} \end{aligned}$$

with $\Sigma \doteq UKU^\top$ and $M \doteq VM^\top$. This concludes the proof. □

3. A Simple Approximation Result. Proposition 2.1 above has a strong connection with the following approximation problem:

PROBLEM 3.1. *Let $p(R)$ be a probability density on $SO(n)$ with respect to the Haar measure; find a best approximation in the class of Langevin densities with respect to the Kullback-Leibler divergence as a criterion of fit.*

The approximation would then allow to represent approximately the original density $p(R)$ by a finite number of parameters, which is the basis for wide-sense estimation. Problem 3.1 could actually be rephrased in more general terms substituting the Langevin class with a general exponential family of densities in $SO(n)$.

We will see that in principle the approximation is very simple, modulo the computation of a normalization constant.

[3]The numbers k_{ij} should be determined by the condition that $E_p\bar{R} = S$. Therefore it is sufficient to verify that when $k_{ij} = k_{ji} = \delta_{ij}k_{ii}$, the mean $E_p\bar{R}$ is a diagonal matrix.

Let $\ell(R) \in \mathcal{L}(M, \Sigma)$, we want to find

$$\hat{\ell} = \arg\min_{\ell \in \mathcal{L}} K(p, \ell) = \arg\min_{\ell \in \mathcal{L}} E_p \log \frac{p}{\ell}. \tag{9}$$

Of course, since $p(R)$ is fixed, the minimum Kullback-Leibler distance problem (9) can be equivalently be rephrased as a

$$\hat{\ell} = \arg\max_{\ell \in \mathcal{L}} E_p \log \ell \,. \tag{10}$$

This can be made explicit as follows:

$$(\hat{M}, \hat{\Sigma}) = \arg\max_{M \in SO(n), \Sigma = \Sigma^\top \in \mathbb{R}^{n \times n}} - \log Z(\underline{k}) + \mathrm{Tr}\left(\Sigma M^\top E_p R\right), \tag{11}$$

where $E_p R = \hat{R} = V_R S_R U_R^\top$, $V_R, U_R \in SO(n)$. Hence, we need only to know the "mean" to compute the best (in the Kullback-Leibler sense) approximation in the Langevin class. Since $Z(\underline{k})$ does not depend on U and M, these matrices can be determined by solving

$$(\hat{U}, \hat{M}) = \arg\max_{U, M \in SO(n)} \mathrm{Tr}\left(U K U^\top M^\top V_R S_R U_R^\top\right). \tag{12}$$

It is very simple to see that $\hat{U} = U_R$ and $\hat{M} = V_R U_R^\top$. Determining K now boils down to solving:

$$\hat{K} = \arg\max_{K = \mathrm{diag}(\underline{k})} - \log Z(\underline{k}) + \mathrm{Tr}\,(K S_R) \tag{13}$$

and the final result is given by

$$(\hat{M}, \hat{\Sigma}) = (V_R U_R^\top, U_R \hat{K} U_R^\top).$$

The diagonal elements of K could also be determined by enforcing

$$S_R = \frac{1}{Z(\underline{k})} \int_{SO(n)} R \exp\left[\mathrm{Tr}(KR)\right] d\mu(R). \tag{14}$$

Unfortunately the solution of (14) (or (13)) requires the analytical expression of the partition function $Z(\underline{k})$; to the best of our knowledge this is not available in general. Special cases in which $Z(\underline{k})$ can be computed are: (i) $n = 2$, i.e. $R \in SO(2)$ (ii) $n = 3$, i.e. $R \in SO(3)$ and $k_1 = k_2 = k_3 = k$ (which we shall, for obvious reasons, call "isotropic" case later on). The case $R \in SO(3)$ but $K \neq kI$ is more involved. A series expansion of $Z(\underline{k}) = Z(k_1, k_2, k_3)$ has been derived in this case. However, further work is needed to understand whether this series expansion can be utilized to solve (14) (or (13)). In this paper we follow a different route; we shall approximate

the map attaching the diagonal matrix S_R to $\underline{k}$ using Markov Chain Monte Carlo techniques. The details are found in Section 7.

In computer vision and robotics applications one is mainly interested in $R \in SO(3)$. Since, as mentioned above, only the case $K = kI$ can be handled explicitly, we formulate below a version of Problem 3.1 in which the class of densities is restricted to have the form $\ell \in \mathcal{L}(M, kI)$.

PROBLEM 3.2. *Let $p(R)$ be a probability density on $SO(n)$ with respect to the Haar measure; find a best approximation in the class of Langevin densities $\ell \in \mathcal{L}(M, kI)$ with respect to the Kullback-Leibler divergence as a criterion of fit.*

The solution of the problem

$$\hat{\ell} = \arg\min_{\ell \in \mathcal{L}(M,kI)} K(p, \ell) \tag{15}$$

can be rephrased as

$$(\hat{M}, \hat{k}) = \arg\max_{M \in SO(n), k \geq 0} -\log Z(k) + \mathrm{Tr}\left(kM^\top E_p R\right), \tag{16}$$

where $E_p R = \hat{R} = V_R S_R U_R^\top$, $V_R, U_R \in SO(n)$. Using the same argument as above, the solution is given by

$$\begin{aligned} \hat{M} &= V_R U_R^\top \\ \hat{k} &= \arg\max_{k \geq 0} -\log Z(k) + k\mathrm{Tr}\left(S_R\right). \end{aligned} \tag{17}$$

Note that $\hat{k}$ depends only on the trace of S_R. Hence two random variables having different mean vectors $\hat{R}_1 = S_{R_1}$ and $\hat{R}_2 = S_{R_2}$ but such that $\mathrm{Tr}(S_{R_1}) = \mathrm{Tr}(S_{R_2})$ will have the same approximation in $\mathcal{L}(M, kI)$. If R itself is distributed according to $\mathcal{L}(M, kI)$, its mean vector satisfies the equality

$$\hat{R} = V_R S_R U_R^\top = s_R V_R U_R^\top = s_R M$$

that is, $S_R = s_R I$. There is a one to one correspondence between s_R and k. We shall come back to the relation between k and s_R in the next Section.

PROPOSITION 3.1. *Given the mean value $\hat{R} = V_R S_R U_R^\top$, the best approximation $\ell(\hat{M}, \hat{k}I)$ in the isotropic Langevin class $\mathcal{L}(M, kI)$ can be obtained by setting $\hat{M} = V_R U_R^\top$, and by choosing $\hat{k}$ as the value of k such that*

$$E_{\ell(\hat{k},\hat{M})} R = \mathrm{Tr}(S_R) V_R^\top U_R^\top .$$

4. Partition Function and Mean Values. We have seen that the computation of the approximating density in the Langevin class requires the knowledge of the partition function $Z(\underline{k})$. In this section we specialize to the case $n = 3$.

In order to compute the integral

$$Z(k_1, k_2, k_3) \doteq \int_{SO(3)} \exp\{Tr(KR)\} d\mu(R) \tag{18}$$

we make use of exponential coordinates:

$$R = \exp(\theta\wedge) \quad \theta \in \mathbb{R}^3, \quad \theta\wedge \doteq \begin{bmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{bmatrix}.$$

The direction $\alpha \doteq \theta/\|\theta\| \in \mathbb{S}^2$ of the coefficient vector $\theta = [\theta_1\, \theta_2\, \theta_3]^\top$ is the axis of rotation while its length $\rho \doteq \|\theta\| \in \mathbb{R}^+$ is the rotation angle. The well-known Rodrigues formula provides a closed form expression for R as a function of θ:

$$\begin{aligned} R = \exp(\theta\wedge) &= I + \sin(\|\theta\|)\tfrac{\theta\wedge}{\|\theta\|} + (1-\cos(\|\theta\|))\tfrac{(\theta\wedge)^2}{\|\theta\|^2} \\ &= I + \sin(\rho)\alpha\wedge + (1-\cos(\rho))(\alpha\wedge)^2 \end{aligned} \tag{19}$$

if $\theta \neq 0$, and $R = I$ otherwise. The formula can be made bijective by restricting $\|\theta\| \leq \pi$. With this notation, it is easy to check that r_{ii}, $i = 1, 2, 3$, the diagonal elements of R, have the expressions

$$r_{ii} = 1 + (\alpha_i^2 - 1)(1 - \cos(\rho)) \qquad i = 1, 2, 3.$$

Let us now define $\mathcal{B}_\pi \doteq \{\theta \in \mathbb{R}^3 : \rho \leq \pi, \alpha \in \mathbb{S}^2\}$, the ball of radius π in $\mathbb{R}^3$. This set provides (up to zero Lebesgue measure sets) a covering of $SO(3)$. In order to compute the integral (18) we need to express the Haar measure in exponential coordinates; this is a standard computation which gives:

$$d\mu(R(\theta)) = \frac{\sin^2(\|\theta\|/2)}{\|\theta\|^2} d\theta_1 d\theta_2 d\theta_3.$$

Introducing polar coordinates for $\alpha = \alpha(\psi, \phi)$, the vector $\theta \in \mathcal{B}_\pi$ is parametrized by $\rho = \|\theta\|$, $\psi \in [0, \pi]$, $\phi \in [0, 2\pi]$, and the Haar measure takes the form

$$d\mu(R(\rho, \psi, \phi)) = \sin^2\left(\frac{\rho}{2}\right) \sin(\psi) d\rho d\psi d\phi\,. \tag{20}$$

This yields:

$$\begin{aligned} &Z(k_1, k_2, k_3) = \textstyle\int_{SO(3)} \exp\{\mathrm{Tr}(KR)\} d\mu(R) \\ &= \textstyle\int_0^\pi \int_0^\pi \int_0^{2\pi} \exp\{\sum_{i=1}^3 k_1 r_{ii} \sin^2\left(\frac{\rho}{2}\right) \sin(\psi) d\rho d\psi d\phi \\ &= \textstyle\int_0^\pi \int_0^\pi \int_0^{2\pi} \exp\left(\sum_{i=1}^3 k_i \alpha_i^2\right) \exp\left(N(\alpha, \underline{k}) \cos(\rho)\right) \sin^2\left(\frac{\rho}{2}\right) \sin(\psi) d\rho d\psi d\phi \end{aligned} \tag{21}$$

where $N(\alpha, \underline{k}) \doteq \sum_{i=1}^3 \left(k_i - k_i \alpha_i^2\right)$. It is now convenient to recall that $\sin^2\left(\frac{\rho}{2}\right) = \frac{1}{2}(1 - \cos(\rho))$. Using well known expressions for the modified Bessel functions of the first kind [23] the inner integral, with respect to ρ, in (21) is easily seen to be

$$\frac{\pi}{2}\frac{1}{\pi}\int_0^\pi \exp\left(N(\alpha, K)\cos(\rho)\right)(1 - \cos(\rho))\, d\rho = \frac{\pi}{2}\left(I_0(N(\alpha, \underline{k})) - I_1(N(\alpha, \underline{k}))\right) \tag{22}$$

where I_0 and I_1 are respectively the zeroth and first order modified Bessel function of the first kind.

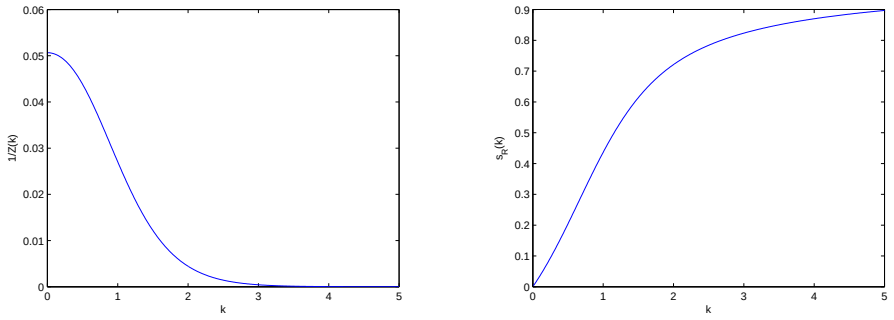


FIG. 1. *Left: normalization constant* $\frac{1}{Z(k)}$ *as a function of* k. *Right: "mean value"* $s_R(k)$ *as a function of* k.

Unfortunately, integrating (22) on the unit sphere turns out to be difficult, and we have not been able to come up with an explicit analytic expression. Note that, when $k_1 = k_2 = k_3 = k$, $N(\alpha, (k,k,k)) = 2k$ holds; in this case the expression simplifies significantly; in fact, under this assumption, the radial integral (22) reduces to

$$\frac{\pi}{2}\frac{1}{\pi}\int_0^\pi \exp\left(N(\alpha,K)cos(\rho)\right)(1-\cos(\rho))\ d\rho = \frac{\pi}{2}\left(I_0(2k)) - I_1(2k)\right)$$

which does not depend on α. Hence, using also $\exp\left(\sum_{i=1}^3 k\alpha_i^2\right) = \exp(k)$,

$$Z(k) = 2\pi^2 \exp\left(k\right)\left(I_0(2k) - I_1(2k)\right). \tag{23}$$

Hence, in the "isotropic" case $K = kI$ the Langevin distribution on $SO(3)$ has the form

$$p(R) = \frac{1}{2\pi^2 \exp(k)\left(I_0(2k) - I_1(2k)\right)} \exp\{k\mathrm{Tr}(M^\top R)\}.$$

Note that, as $k \to 0$ the Langevin density $p(R) \to \frac{1}{2\pi^2}$, the uniform density[4] on $SO(3)$. It is very simple to check that, for $M = I$,

$$\hat{R} \doteq \int_{SO(3)} R \frac{1}{2\pi^2 \exp(k)\left(I_0(2k) - I_1(2k)\right)} \exp\{k\mathrm{Tr}(R)\} d\mu(R) = S_R(k) = s_R(k) I \tag{24}$$

holds. Equation (24) provides a correspondence between k and s_R. The function $s_R(k)$ is plotted in Figure 1, right panel.

5. Stochastic Differential Equations on the Orthogonal Group. Let $\mathbf{b}(t) = [\mathbf{b}_1(t)\, \mathbf{b}_2(t)\, \ldots, \mathbf{b}_m(t)]^\top$ be standard Brownian motion in $\mathbb{R}^m$ and consider the stochastic differential equation (in the Itô sense)

$$dR(t) = f(R)dt + \sum_{i=1}^m g_i(R) d\mathbf{b}_i(t)\,. \tag{25}$$

[4]Note that the volume of $SO(3)$ with the Haar measure given in (20) is $2\pi^2$.

The conditions under which $R(t) \in SO(n),\ \forall t \geq 0$ have been derived by Brockett, [4], by imposing that

$$\begin{array}{c} d\left(R^\top(t)R(t)\right) = 0 = d\left(R(t)R^\top(t)\right) \\ R(0) \in SO(n) \end{array} \tag{26}$$

where the differential is in the Itô sense. Recalling that, by Itô rule, $d\mathbf{b}_i(t)d\mathbf{b}_j(t) = \delta_{ij}dt$, this is equivalent to requiring that

$$\begin{array}{rcl} 0 & = & f(R)R^{-1}(t) + R(t)f^\top(t) + \sum_{i=1}^m g_i(R)g_i^\top(R) \\ 0 & = & g_i(R)R^{-1}(t) + R(t)g_i^\top(R) \quad i = 1, .., m. \end{array} \tag{27}$$

Without loss of generality we can write $f(R) = A(R,t)R(t)$ and $g_i(R) = B_i(R,t)R(t)$ so that we may re-write (25) in the form

$$dR(t) = A(R,t)R(t)dt + \sum_{i=1}^m B_i(R,t)d\mathbf{b}_i(t)R(t) \tag{28}$$

where, by (27),

$$\begin{array}{rcl} 0 & = & A(R,t) + A^\top(R,t) + \sum_{i=1}^m B_i(R,t)B_i^\top(R,t) \\ 0 & = & B_i(R,t) + B_i^\top(R,t) \quad i = 1, .., m \end{array} \tag{29}$$

must hold. This shows that the B_i's and $A(R,t) + \frac{1}{2}\sum_{i=1}^m B_i(R,t)B_i^\top(R,t)$ must be skew symmetric matrices, i.e. must belong to $so(n)$, the Lie Algebra of $SO(n)$. A matrix in this Lie algebra can be expressed as a linear combination of the elements of a basis of elementary rotations $\{E_k\, k = 1, ..., N = \frac{n(n-1)}{2}\}$[5] so that every $R \in SO(n)$ takes the form $R = \exp\{\sum_k \theta_k E_k\}$. The direction $\theta/\|\theta\|$ of the coefficient vector $\theta = [\theta_1\, \theta_2\, \ldots\, \theta_N]^\top$ is the axis of rotation while its length $\|\theta\|$ is the rotation angle. We shall use the wedge (or "external") product notation:

$$\theta\wedge \doteq \sum_k \theta_k E_k\,.$$

The simplest form of the model (28) is obtained when A and B_i's are constant. Denoting $A + \frac{1}{2}\sum_{i=1}^m B_i B_i^\top$ by $\omega\wedge$ and introducing:

$$B_i = L_i\wedge = \sum_{j=1}^N l_{ij}E_j \qquad \Delta_L \doteq \sum_{i=1}^m (L_i\wedge)(L_i\wedge)^\top \tag{30}$$

[5]The elementary rotations for $SO(3)$ are

$$E_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \quad E_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

we obtain

$$A = (\omega\wedge) - \frac{1}{2}\sum_{i=1}^{m}(L_i\wedge)(L_i\wedge)^\top = (\omega\wedge) - \frac{1}{2}\Delta_L \tag{31}$$

which leads to the stochastic differential equation

$$dR(t) = \left(\omega \wedge -\frac{1}{2}\Delta_L\right) R(t)dt + \sum_{i=1}^{m} d\mathbf{b}_i(t) L_i \wedge R(t). \tag{32}$$

This can also be rewritten in the more suggestive form

$$dR(t) = d\Theta(t)R(t) - \frac{1}{2}\Delta_L R(t)dt \tag{33}$$

with $d\Theta \doteq d\theta\wedge$, the infinitesimal random rotation of instantaneous angular velocity ω modeled by

$$d\theta = \omega dt + Ld\mathbf{b}(t) \tag{34}$$

where the infinitesimal stochastic component $Ld\mathbf{b}(t) \doteq \sum_{i=1}^{m} L_i d\mathbf{b}_i(t)$, is induced by a standard Brownian motion $\mathbf{b}(t)$ in $\mathbb{R}^N$.

This is indeed the form of general infinitesimal random rotation which has been named "Skew Brownian Motion" in $SO(n)$ by McKean [13].

Note finally that $-\frac{1}{2}\Delta_L R(t)dt$ is the Ito's correction term which is needed to keep the path $t \mapsto R(t)$ in $SO(n)$.

6. Approximate Filtering for the "Skew Brownian Motion" under Linear Observations. In this section we shall assume that $R(t)$ is a rotation process induced by a skew Brownian motion, of the form (32). We shall postulate that the initial condition $R(0)$ has a Langevin distribution

$$R(0) \sim \ell(R; \Sigma_0, M_0) = \frac{1}{Z(K_0)} e^{\mathrm{Tr}(\Sigma_0 M_0^\top R)} \tag{35}$$

and that at each sample instant $t = kT$, are collected m linear noisy observations of the form

$$\mathbf{y}_j(kT) = R_{\epsilon_j(kT)} R(kT) x_j \quad , \quad x_j \in \mathbb{R}^3 \quad j = 1, \ldots, m \tag{36}$$

where the "rotational noise" terms $R_{\epsilon_j(kT)}$ are independent Langevin-distributed and hence induce a conditional density of the form

$$p(\mathbf{y}_1(kT), \mathbf{y}_2(kT), \ldots, \mathbf{y}_m(kT) \mid R(kT)) = \prod_j \frac{k_{0j}}{4\pi \sinh(k_{0j})} e^{k_{0j} x_j^\top R^\top(kT) \mathbf{y}_j(kT)} . \tag{37}$$

At time $t = 0$, the a posteriori density is given by [6]:

[6] We have dropped the time index for notational convenience

$$
\begin{aligned}
p\left(R \mid \mathbf{y}_j, j=1,\dots,m\right) &\propto p(R)\prod_{j=1}^m p(\mathbf{y}_j \mid R) \\
&\propto \exp\left(\mathrm{Tr}\left[\left(\Sigma_0 M_0^\top + \textstyle\sum_{i=1}^m k_{0j} x_j \mathbf{y}_j^\top\right) R\right]\right).
\end{aligned} \tag{38}
$$

Hence the a posteriori density is still Langevin; i.e. the Langevin on $SO(n)$ is a conjugate prior [18] to the Langevin likelihood on $\mathbb{S}^{n-1}$. Computing the Singular Value Decomposition

$$
U_p S_p V_p^\top = \Sigma_0 M_0^\top + \sum_{i=1}^m k_{0j} x_j \mathbf{y}_j^\top \tag{39}
$$

we obtain the parameters

$$
\begin{aligned}
\hat{M} &= V_p U_p^\top \\
\hat{\Sigma} &= U_p S_p U_p^\top.
\end{aligned} \tag{40}
$$

Clearly the procedure can be iterated if the "target" $R(t)$ is fixed ($\omega = 0 \quad L = 0$). This is rather trivial and we shall not further discuss this case.

Instead, we consider the more interesting case in which both ω and L are not necessarily zero. Let us first assume that, at time $t = kT$, the conditional density of $R(t)$ given the (strict) past of the measurements, $\mathcal{Y}^{k-1} \doteq \{\mathbf{y}_i(0), \dots, \mathbf{y}_i((k-1)T), i = 1, .., m\}$, is Langevin-distributed; i.e.

$$
p(R(kT) \mid \mathcal{Y}^{k-1}) = \mathcal{L}(\hat{M}(k \mid k-1), \hat{\Sigma}(k \mid k-1)).
$$

By the same argument used above, it it fairly easy to see that $p(R(kT) \mid \mathcal{Y}^k)$ is still Langevin with parameters $\hat{M}(k \mid k)$ and $\hat{\Sigma}(k \mid k)$ which can be computed by a SVD step as in (38), (39) and (40). We shall denote the measurement update step abstractly by two maps Φ, Ψ, as follows

$$
\begin{aligned}
\hat{M}(k \mid k) &= \Phi(\hat{M}(k \mid k-1), \hat{\Sigma}(k \mid k-1), \mathbf{y}_1(k(T)), \dots, \mathbf{y}_m(kT)) \\
\hat{\Sigma}(k \mid k) &= \Psi(M(k \mid k-1), \hat{\Sigma}(k \mid k-1), \mathbf{y}_1(k(T)), \dots, \mathbf{y}_m(kT)).
\end{aligned} \tag{41}
$$

The "prediction" step is implemented by two maps, which we abstractly denote as

$$
\begin{aligned}
\hat{M}(k+1|k) &= \Gamma(\hat{M}(k \mid k), \hat{\Sigma}(k \mid k)) \\
\hat{\Sigma}(k+1|k) &= \Xi(\hat{M}(k \mid k), \hat{\Sigma}(k \mid k)).
\end{aligned} \tag{42}
$$

This step is a little more involved. In fact, assuming that an estimate $\hat{R}(kT \mid kT)$ is available, the prediction $\hat{R}(kT+T \mid kT)$ is readily computed integrating the (ordinary) differential equation:

$$
d\hat{R}(t|kT) = \left(\omega \wedge -\frac{1}{2}\Delta_L\right) \hat{R}(t|kT)\, dt \tag{43}
$$

with initial condition $\hat{R}(kT \mid kT)$. This equation is obtained by simply taking the (conditional) expectation of (32) since the martingale term vanishes under expectation

and only the drift survives. Unfortunately however the Langevin class is not invariant under the evolution of (32) and at this point one needs to resort to approximations.

As stated in Problem 3.1, one can approximate the posterior density by a distribution in the Langevin class. This only requires computing the mean vector. Yet, the computation of the parameters $\hat{R}(kT \mid kT)$ from $\hat{M}(k \mid k)$ and $\hat{\Sigma}(k \mid k)$ and of $\hat{M}(k+1 \mid k)$ and $\hat{\Sigma}(k+1 \mid k)$ given $\hat{R}(kT+T \mid kT)$ is not entirely trivial. This is due to the fact, already pointed out in Section 4, that the map relating the parameters M, Σ of $p(R) = \ell(R; M, \Sigma)$ to $E_p[R]$, is available in closed form only in the isotropic case $\Sigma = kI$. In the next Section we shall discuss a computational procedure which allows to approximate these maps for general Σ.

7. Approximation of the Map $(M, \Sigma) \to E_p[R]$**.** The map attaching M, Σ to $E_p[R]$ can be approximated arbitrarily well using Monte Carlo techniques. As discussed in Proposition 2.1 it suffices to address the case $M = I$ and $\Sigma = \text{diag}\{k_1, k_2, k_3\}$; the general case follows making use of left and right translations (left and right multiplications in $SO(3)$).

As discussed in Proposition 2.1, if $p(R) \propto \exp\{Tr(\text{diag}\{k_1, k_2, k_3\}R)\} = \exp\{Tr(\text{diag}\{\underline{k}\}R)\}$, then $S = E_p[R]$ is also diagonal; namely

$$E_p[R] = S = \text{diag}\{s_1, s_2, s_3\} = \text{diag}\{\underline{s}\}$$

therefore, it is sufficient to compute the map which links $\underline{k}$ to $\underline{s}$.

This computation can be performed off-line once and for all. It needs not be repeated during the actual on-line signal processing. Therefore, the computational load and/or computing time of the procedure need not concern us.

A simple yet effective solution would be to compute $\underline{s}$ for a number of values $\underline{k}$; this can be done with arbitrary accuracy by Monte Carlo techniques. It is in fact sufficient to simulate from $p(R) \propto \exp\{Tr(\text{diag}\{\underline{k}\}R)\}$ using a Metropolis-Hasting type of algorithm (note that the normalization constant needs not be known) and then compute sample averages.

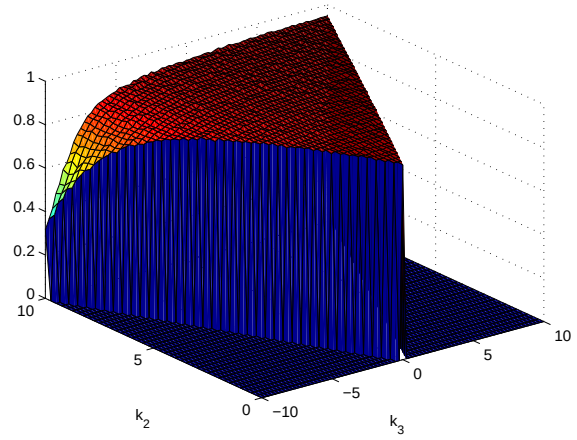
Since the function $\underline{s}(\underline{k})$ is smooth (see Figure 2), provided $\underline{s}(\underline{k})$ is known on a sufficiently fine grid $\underline{k}_i$, $i = 1, .., K$, the function $\underline{s}(\underline{k})$ can be obtained via interpolation in a variety of ways, e.g. using RKHS techniques [21]. For the sake of exposition, we have computed $\underline{s}$ on a grid of values in the set

$$\underline{k_i} = \{k_1 = 10,\ k_2 \in [0, 5],\ k_1 \in [-k_2, k_2]\}. \tag{44}$$

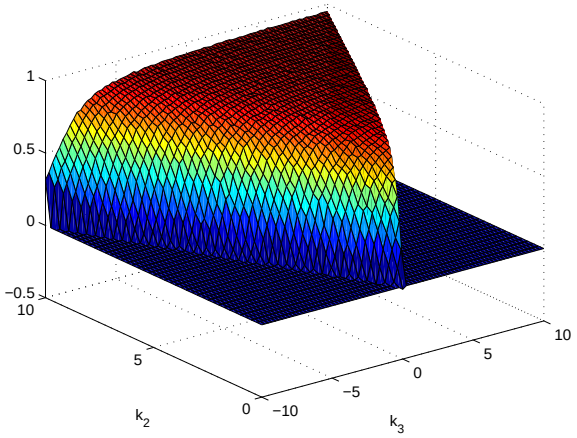
The value of k_1 has been kept fixed only for presentation of the results, which are reported in Figure 2.

The results have been obtained by running a standard Metropolis-Hasting chain for $N = 50000$ steps. Average values are then taken. For sake of comparison, we report in Figure 3 the results for the isotropic case. The estimate obtained by the

$s_1 = E_p R(1,1)$



$s_2 = E_p R(2,2)$



$s_3 = E_p R(3,3)$

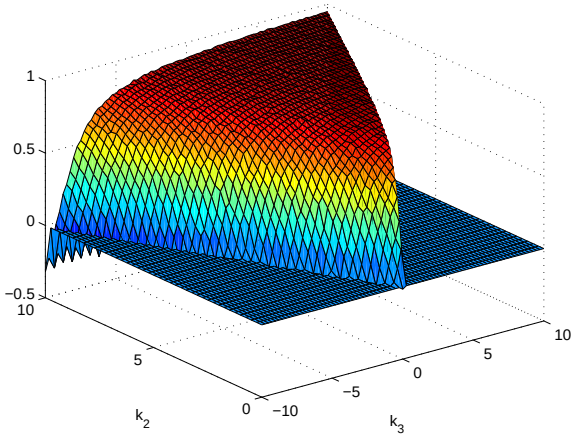


FIG. 2. *Estimates of the diagonal elements of the mean value* $\hat{R} = E_p(R)$ *for* $M = I$ *and* $\Sigma = \mathrm{diag}\{\underline{k}\}$ *obtained via Markov Chain Monte Carlo. Left to right:* $s_1 = E_p R(1,1)$, $s_2 = E_p R(2,2)$ *and* $s_3 = E_p R(3,3)$. *The value of* k_1 *has been kept fixed to* 10, *while* $k_2 \in [0, 10]$ *and* $k_3 \in [-k_2, k_2]$ *(the functions* s_1, s_2, s_3 *are conventionally set to zero for* $|k_3| > k_2$.*)*

$$s_R(k) = \mathrm{Tr}\{E_p R\}/3$$

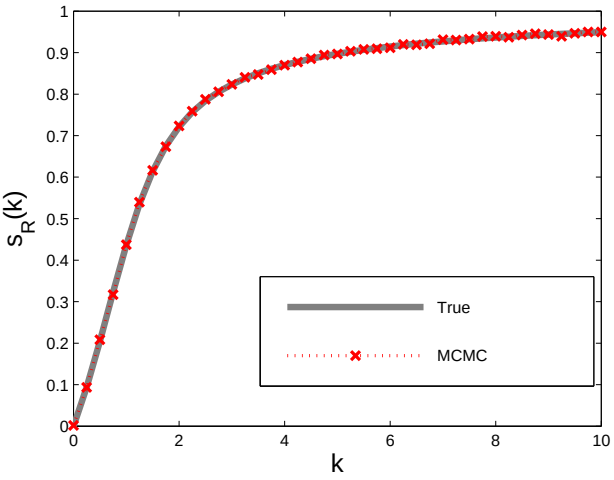


FIG. 3. *MCMC estimates vs. true values (see Section 4) of* $s_R(k) = \mathrm{Tr}\{E_p R\}/3$ *in the isotropic case* $\Sigma = kI$.

Monte Carlo method of $s(k)$ is within about 0.5% error of the theoretical value (note that the error is roughly of the order $1/\sqrt{50000}$). The results have been obtained in a few hours of computing time on a standard laptop computer; no interpolation nor smoothing have been performed; of course much more accurate results could be obtained, for instance, constraints on the monotonicity of $\underline{s}(\underline{k})$ could be enforced.

Conclusions. In this paper we have discussed a simple Bayesian estimation problem in $SO(n)$ which has applications to attitude estimation and to pose reconstruction problems in computer vision. Sequential estimation of a fixed rotation in space can be implemented by a simple closed-form recursive MAP estimator. For a general stochastic rotation process in $SO(n)$ described by a stochastic differential equation driven by "rotational Brownian motion" the problem has no exact closed form solution but we have indicated how "wide-sense" approximate filters can be constructed.

REFERENCES

[1] V. BENES, *Exact finite dimensional filters for certain diffusions with nonlinear drift*, Stochastics, 5(1981) , pp. 65-92.

[2] W. M. BOOTHBY, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, 1986.

[3] D. BRIGO, *Fitering by Projection on the Manifold of Exponential Densities*, Ph. D. thesis, Department of Mathematics, Vrije Universiteit, Amsterdam, 1996.

[4] R. W. BROCKETT, *Lie Algebras and Lie Groups in Control Theory*, in: Geometric Methods in Control, R. W. Brockett and D. Mayne eds. Reidel, Dordrecht, 1973.

[5] R. W. Brockett, *Notes on Stochastic Processes on Manifolds*, in: Control Theory in the 21st Century, C.I Byrnes, C. Martin, B. Datta eds. Birkhauser, 1997.

[6] A. Chiuso, R. W. Brockett, and S. Soatto, *Optimal Structure From Motion: Local Ambiguities and Global Estimates*, Int. Journal of Computer Vision, 39:3(2000), pp. 195-228.

[7] A. Chiuso and G. Picci, *Visual Tracking of Points as Estimation on the Unit Sphere*, in: The Confluence of Vision and Control, D. Kriegman, G. Hager and S. Morse eds. Springer-Verlag Lecture Notes in Control and Information Systems (LNCIS) n. 237, pp. 90-105, 1998.

[8] A. Chiuso and G. Picci, *A wide-sense estimation theory on the unit sphere*, in: Proceedings of the 1998 Conference on Decision and Control, Tampa, Florida, paper n. FM02-5, p. 3745-3750, 1998.

[9] R. A. Fisher, *Dispersion on a sphere*, Proc. Royal Soc. London, A 217(1953), pp. 295-305.

[10] E. T. Jaynes, *Information theory and statistical mechanics*, Physical Review, 106(1957), pp. 620-630.

[11] P. Langevin, *Magnetisme et theorie des electrons*, Ann. de Chim et de Phys., 5(1905), pp. 70-127.

[12] J. Lo and A. Willsky, *Estimation for rotational processes with one degree of freedom, parts I, II, III*, IEEE Transactions on Automatic Control, AC-20(1975), pp. 10-33.

[13] H. P. McKean, *Brownian Motion on the Three-Dimensional Rotation Group*, Mem. Coll. Sci. University of Kyoto, Series A, XXXIII:1(1960), pp. 25-38.

[14] R. M. Murray, S. S. Sastry, and Z. Li, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, 1994.

[15] B. Øksendal, *Stochastic Differential Equations*, Springer, 1990.

[16] F. Perrin, *Étude Mathématique du Mouvement Brownien de Rotation*, Ann. Ecole Normale Superieure, XLV:3(1928), pp. 1-51, 1928.

[17] G. Picci, *Dynamic Vision and Estimation on Spheres*, in: Proceedings of the 36th IEEE Conf. on Decision and Control, San Diego Ca., pp. 1140-1145, 1997.

[18] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University, 1961.

[19] S. Soatto, R. Frezza, and P. Perona, *Motion estimation via dynamic vision*. IEEE Trans. on Automatic Control, 41(1996), pp. 393-413.

[20] S. Soatto, *A Geometric Framework for Dynamic Vision*. Dr. Sc. Thesis, California Institute of Technology, 1996.

[21] G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.

[22] G. S. Watson, *Statistics on Spheres*, Wiley, N.Y., 1983.

[23] G. N. Watson, *A Treatise on the Theory of Bessel Functions, 2nd Edition*, Cambridge Mathematical Library, Cambridge University Press, 1995.

[24] W. S. Wong and S.S.T. Yau, *The Estimation Algebra of Nonlinear Filtering Sytems*, in: Mathematical Control Theory, J. Baillieul and J.C. Willems eds. pp. 33-65, Springer, N.Y., 1999.

[25] A. H. Jazwinski, *Stochastic processes and Filtering Theory*, Academic Press, New York, 1970.