Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

A Wide-Sense Estimation Theory on the Unit Sphere [1]

Alessandro Chiuso and Giorgio Picci
Dipartimento di Elettronica e Informatica
Università di Padova, Italy
chiuso,picci@dei.unipd.it

Abstract

On-line estimation of the direction of feature points moving in space from noisy projections on a plane is a classical problem occuring in computer vision which has traditionally been treated by ad hoc statistical methods in the literature. In a previous paper [11] we have formulated it as a Bayesian estimation problem on the unit sphere. A natural probabilistic structure which makes this estimation problem tractable has been introduced. Within this structure, exact recursive solutions can be given for sequential observations of a fixed target point, while for a moving object in general one has to resort to approximations. In this paper an approximate ("wide-sense") solution is proposed which leads to very simple recursions similar to the Kalman Filter. In certain situations this solution may provide a substantial improvement over the traditional EKF. As an example, we discuss estimation of the direction of points whose motion is described by a simple dynamic model of the random walk type. This model is of interest in pratical situations when dealing with slowly time-varing observed feature points.

1 Introduction

The operation of perspective projection onto the image plane of an ideal (pinhole) camera can be described geometrically as the intersection of the rays (straight homogeneous lines) emanating from the optical center of the camera, connecting to the observed object in $\mathbb{R}^3$, with the image plane. In practice the projections are noisy and the detected feature points on the image plane do not correspond exactly to straight projections of the real feature points. This occurs because of distortion of the optical systems and noise of various kinds entering the signal detection and the processing of the electronic image formed on the CCD array. For these reasons, the task of reconstructing the location in 3-D of an observed object from its noisy projections on the image plane, is a non trivial problem which should be treated by appropriate statistical methods. So far only *ad hoc* estimation methods have been used (most of the times variations of the Extended Kalman filter) with generally poor performance and possible divergence problems. A sound statistical analysis of the problem has been lacking.

In the simplest case the observed feature objects are points moving unconstrained in $\mathbb{R}^3$. Since we cannot measure distances along the projecting rays and we may at most recover the feature points modulo distance from the optical center, we may, without loss of information, normalize the vectors joining the optical center to the measured projections on the image plane to unit length. In fact, both the target point and its observed projections on the image plane, may be described as *directions* and represented, say, by the coordinates of the corresponding unit vectors $\mathbf{x}$ and $\mathbf{y}_k$ lying on a unit sphere centered at the optical center of the camera. In this formulation, $\mathbf{x}$ is the true unknown direction pointing at the observed point in $\mathbb{R}^3$ and $\mathbf{y}_k,\ k = 1, \dots, m$ (vectors on the unit sphere) are noisy measurements of the "true" direction $\mathbf{x}$. Hence the problem is formulated as *estimation on the unit sphere.*

The precise nature of the observation noise will be discussed later, however it should be clear that the way the noise affects the ideal perspective projection $\mathbf{x}$ cannot be *additive* and a realistic formulation of the problem must depart sharply from the standard linear-Gaussian setup. In the next section we shall discuss a natural family of probability distributions on the unit sphere.

2 The Langevin Distribution

A family of probability distributions on the sphere which has many desirable properties is defined by the *Langevin density*

$$p(x) = \frac{\kappa}{4\pi \sinh \kappa} \exp \kappa\mu' x, \qquad \|x\| = 1 \qquad (1)$$

with respect to the spherical surface measure $d\sigma = \sin\theta d\theta d\phi$. The vector parameter $\mu \in \mathbb{S}^2$ (μ is conventionally normalized to unit length) is the mode of the distribution, while the positive number $\kappa > 0$ is called the *concentration* of the distribution. For $\kappa \to 0$ the density becomes the uniform distribution while for $\kappa \to \infty$, p tends to a Dirac distribution concentrated

[1] Research supported by grants ASI RS-103 and RS-106 from the Italian Space Agency.

at $x = \mu$. The density function (1), denoted $L(\mu, \kappa)$, was introduced by Langevin (1905) in his statistical-mechanical model of of magnetism [6]. Since then it has been rediscovered and used in statistics by a number of people, see [12]. Observe that the Langevin distributions form a one-parameter exponential family and are *invariant with respect to rotations.* The functional form is preserved under multiplication allowing a straightforward application of Bayes rule. Introducing a suitable spherical coordinate system, (1) can be rewritten in the form

$$p(\theta, \phi) = \frac{\kappa}{4\pi \sinh \kappa} \exp \kappa \cos \theta \quad 0 \le \theta \le \pi.$$

which shows that $L(\mu, \kappa)$ is rotationally symmetric around its mode μ.

The expression (1) is for a Langevin distribution on the unit sphere in $\mathbb{R}^3$. For higher dimensions, the normalization constant has a slightly more complicate expression. The distribution on $\mathbb{S}^{n-1}$, $n \ge 3$, has density

$$p(x) = \frac{\kappa^{(n/2-1)}}{(2\pi)^{n/2} I_{n/2-1}(\kappa)} \exp \kappa \mu' x, \qquad \|x\| = 1 \quad (2)$$

with respect to the spherical surface measure, where $I_{n/2-1}(x)$ is a modified Bessel function of the first kind. More generally, an arbitrary probability density functions on $\mathbb{S}^{n-1}$ can be expressed as the exponential of a finite expansion in spherical harmonics. These are discussed, for example, in [12, p. 80-88]. In this sense the Langevin density is a sort of "first order" approximation as only the first spherical harmonic, $\cos\theta$, is retained in the expansion and the others are assumed to be negligible.

Rotation-invariant distributions like the Langevin distribution are natural for describing *random rotations.* Let $\mathbf{x}$ be a fixed direction, represented as a point in $\mathbb{S}^2$, which is observed by a camera. The observation mechanism perturbs $\mathbf{x}$ in a random way (say because of lens distortion, pixel granularity etc). Since the output of the sensor, $\mathbf{y}$, is also a direction represented by a vector of unit length, the perturbation may always be seen as a random rotation described by a random rotation matrix [1] $R = R(\mathbf{p}) \in SO(3)$, where $\mathbf{p}$ is the polar vector of the rotation, i.e. $R(\mathbf{p}) := \exp\{\mathbf{p}\wedge\}$ so that

$$\mathbf{y} := R(\mathbf{p})\,\mathbf{x} \qquad (3)$$

In other words we can always model the noise affecting $\mathbf{x}$ as multiplication by a rotation matrix. The action of the "rotational observation noise" on directions $\mathbf{x} \in \mathbb{S}^2$ can in turn be described probabilistically by the *conditional density* function $p(y \,|\, \mathbf{x} = x)$ of finding the observation directed about a point y on the sphere, given that the "true" observed direction was $\mathbf{x} = x$. A very reasonable unimodal conditional distribution, rotationally symmetric around the starting direction $\mathbf{x}$ (no angular bias introduced by the observing device) is the Langevin-type density,

$$p(y \,|\, \mathbf{x}) = \frac{\kappa}{4\pi \sinh \kappa} \exp \kappa \mathbf{x}' y \qquad (4)$$

In this framework we may think of the ordinary distribution $L(\mu, \kappa)$ as a conditional density evaluated at a known point $\mathbf{x} = \mu$.

[1] The wedge $\wedge$ denotes cross product.

The Angular Gaussian Distribution

Some of the properties of the Langevin distribution are the natural analog of the properties of Gaussian distributions on an Euclidean space. There are various attempts in the literature to derive the Langevin distribution as the distribution function of some natural transformation of a Gaussian vector. Perhaps the easiest result in this direction is the observation, first made by Fisher [5], that the distribution of a normal random vector $\mathbf{x}$ with isotropic distribution $\mathcal{N}(\mu, \sigma^2 I)$, *conditional on the event* $\{\|\mathbf{x}\| = 1\}$ is Langevin with mode $\mu/\|\mu\|$ and concentration parameter $\|\mu\|/\sigma^2$.

A more useful result, discussed in [12, Appendix C] is the remarkable similarity of the so-called *Angular Gaussian* distribution to the Langevin. The angular Gaussian is the probability density of the direction vector $\mathbf{x} := \boldsymbol{\xi}/\|\boldsymbol{\xi}\|$ when $\boldsymbol{\xi}$ has an isotropic Gaussian distribution, i.e. $\boldsymbol{\xi} \sim \mathcal{N}(\mu, \sigma^2 I)$. The distribution is obtained by computing the marginal of $\mathcal{N}(\mu, \sigma^2 I)$ on the unit sphere $\|x\| = 1$. It is shown in [12, Appendix C] that the angular Gaussian is a convex combination of Langevin densities with varying concentration parameter s,

$$Ag(x) = N \int_0^{+\infty} s^{n-1} e^{-\frac{1}{2}\frac{s^2}{\alpha^2}} e^{s\lambda' x} ds,$$

where $\lambda = \frac{\mu}{\|\mu\|}$ $\alpha = \frac{\|\mu\|}{\sigma}$ and it is seen from this formula that Ag depends on μ, σ^2 only through the two parameters λ and α. We shall denote it by $Ag(\lambda, \alpha^2)$. The notation is convenient, since for either moderate or large values[2] of α, $Ag(\lambda, \alpha^2)$ is, to all practical purposes, *the same thing as* $L(\lambda, \kappa)$, where

$$\lambda = \frac{\mu}{\|\mu\|} \qquad \kappa := \alpha^2 = \frac{\|\mu\|^2}{\sigma^2}. \qquad (5)$$

Note that all distributions $\mathcal{N}(\rho\mu, \rho^2\sigma^2 I)$, $\rho > 0$, give origin to the same angular Gaussian as $\mathcal{N}(\mu, \sigma^2 I)$. (This precisely is the family of isotropic Gaussians generating the same angular distribution.)

[2] "Moderate or large" here means that $\kappa := \alpha^2$ should be greater than, say, 100 in order to have a fit within a few precent of the values of the two functions. In fact the angular Gaussian approximates a Langevin distribution also for α small, when both of them are close to uniform, but the relation between α and κ is different.

The role of the angular Gaussian in modeling directional observations can be illustrated by the following example. Let $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$ be Gaussian isotropic random vectors with $\xi \sim \mathcal{N}(\mu, \sigma^2 I)$, $\boldsymbol{\zeta} \sim \mathcal{N}(0, \sigma_z^2 I)$ and assume we observe the direction of the vector

$$\boldsymbol{\eta} = C\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{6}$$

If $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$ are independent and C is an orthogonal matrix ($CC' = I$) the distribution of $\boldsymbol{\eta}$ is isotropic Gaussian and the direction $\mathbf{y} := \boldsymbol{\eta}/\|\boldsymbol{\eta}\|$ then has an angular Gaussian (i.e. Langevin) distribution $\mathbf{y} \approx L(\mu/\|\mu\|, \frac{\|\mu\|^2}{\sigma^2+\sigma_z^2})$.

No matter how $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$ are correlated, the conditional density $p(y \mid \boldsymbol{\xi})$ is also angular Gaussian. In fact this follows since the conditional distribution of $\boldsymbol{\eta}$ given $\boldsymbol{\xi} = \xi$ is Gaussian with mean $C\xi$ and variance σ_z^2. Hence

$$p(y \mid \boldsymbol{\xi} = \xi) = Ag(C\xi/\|\xi\|, \|C\xi\|^2/\sigma_z^2) = Ag(Cx, \|\xi\|^2/\sigma_z^2)$$

where $\mathbf{x}$ is the direction vector of $\boldsymbol{\xi}$. We are interested in the conditional density $p(y \mid \mathbf{x})$. We shall state the result in a formal way as follows.

Proposition 2.1 *If the conditional variance σ_z^2, of $\boldsymbol{\eta}$ given $\boldsymbol{\xi} = \xi$ is proportional to $\|\xi\|^2$, i.e. $\sigma_z^2 = \sigma_0^2\|\xi\|^2$, then the conditional density $p(y \mid \mathbf{x})$ for the model (6) is angular Gaussian.*

Proof: Denote $r := \|\xi\|$. Then the claim follows from

$$p(y \mid x) = \int_0^\infty r^2 p(y \mid x, r) p(r \mid x) dr$$

since $p(y, r \mid x) = p(y \mid x, r)p(r \mid x) = p(y \mid \xi)p(r \mid x)$ and in the stated assumption, $p(y \mid x, r)$ does not depend on r. ■

The ξ-dependence of the conditional variance (i.e. of the power of the additive noise in the Euclidean model (6)) is a condition of "angular noise" for the directions. Note that the condition precludes independence of $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$. This agrees with the intuition of (infinitesimal) angular noise which, for each direction $\mathbf{x}$, should be represented locally as an additive vector on the tangent plane of the unit sphere at the particular point $\mathbf{x}$.

For the Langevin density, the parameters (μ, κ) can be expressed as a function of the mean vector m of the distribution. In $\mathbb{S}^2$ one has for example

$$\begin{aligned} \mu &= \frac{m}{\|m\|} \\ \frac{\cosh\kappa}{\sinh\kappa} - \frac{1}{\kappa} &= \|m\| \end{aligned} \tag{7}$$

and it can be checked generally that the formula provides a one-one correspondence between m and (μ, κ). In other words, m is a vector parameter which determines $L(\mu, \kappa)$ completely. The following is the counterpart on $\mathbb{S}^{n-1}$ of a well-known characterization of the Gaussian distribution on the Euclidean space $\mathbb{R}^n$.

Proposition 2.2 *Among all probability densities on the unit sphere having the same mean vector m, the Langevin distribution is the one of maximal entropy.*

Proof: Denote by

$$H_f := -\int_{\mathbb{S}^{n-1}} \log f(x)\, f(x)\, d\sigma_x$$

the entropy of the density f (or of the distribution $dF(x) := f(x)\, d\sigma_x$). Let $l(x)$ be the Langevin density with mean m; from the expression of l it is evident that

$$\begin{aligned} H_l &= -\textstyle\int_{\mathbb{S}^{n-1}} \log l(x)\, l(x)\, d\sigma_x \\ &= -\log \frac{\kappa^{(n/2-1)}}{(2\pi)^{n/2} I_{n/2-1}(\kappa)} - \kappa\mu' m \\ &= -\textstyle\int_{\mathbb{S}^{n-1}} \log l(x)\, f(x)\, d\sigma_x \end{aligned}$$

for an arbitrary f of mean m. It follows that the difference

$$\begin{aligned} &H_l - H_f = \\ &= -\textstyle\int_{\mathbb{S}^{n-1}} \log l(x)\, f(x)\, d\sigma_x + \int_{\mathbb{S}^{n-1}} \log f(x)\, f(x)\, d\sigma_x \\ &= \textstyle\int_{\mathbb{S}^{n-1}} \log \frac{f(x)}{l(x)}\, f(x)\, d\sigma_x \end{aligned}$$

is the Kullback-Leibler distance of f from l, which is known to be nonnegative unless $f = l$. Therefore $H_l \geq H_f$ for all f of mean m. ■

Best Approximation by a Langevin distribution
Let P be an arbitrary probability measure on the unit sphere, absolutely continuous with respect to the surface measure $d\sigma = \sin\theta\, d\theta\, d\varphi$; we want to approximate the density $f(x) = dP/d\sigma$ by means of a density of the Langevin type, i.e. by a density in the class

$$\mathcal{L} = \{\ell(x) = \frac{\kappa}{4\pi \sinh(\kappa)} \exp\{\kappa\mu^T x\} \quad , \quad \kappa \geq 0, \|\mu\| = 1\}; \tag{8}$$

using as a criterion of fit the Kullback-Leibler pseudo-distance,

$$K(f, \ell_{(\mu,\kappa)}) = E_f \log \frac{f(x)}{\ell_{(\mu,\kappa)}(x)} = \int_{\mathbb{S}^2} \log \frac{f(x)}{\ell_{(\mu,\kappa)}(x)}\, f(x)\, d\sigma_x \tag{9}$$

The problem of finding the minimum:

$$\min_{\{(\mu,\kappa)\ :\ \kappa \geq 0, \|\mu\|=1\}} K(f, \ell_{(\mu,\kappa)}) \tag{10}$$

can be solved by introducing Lagrange multipliers and taking derivatives with respect to μ and κ. It can be shown [4] that the minimum is attained for:

$$\begin{aligned} &\frac{\cosh\kappa}{\sinh\kappa} - \frac{1}{\kappa} - \mu' m = 0 \\ &\kappa m - \lambda\mu = 0 \end{aligned} \tag{11}$$

where m is the mean vector of P

$$m = \int_{\mathbb{S}^2} x\, f(x)\, d\sigma_x. \qquad (12)$$

Hence

Proposition 2.3 *The best Langevin approximant of a probabilty distribution P on the sphere in the sense of minimal Kullback-Leibler pseudo distance, is the one having the same mean vector m of P.*

and our approximation problem is solved simply by equating the mean vectors of the two distributions. In other words, to find the best Langevin approximant of P, the only thing we need to know is its mean vector. This result leads to a kind of wide-sense estimation theory on spheres with the mean parameter playing the same role of the second order statistics in the Gaussian case.

3 MAP Estimation

Assuming that the *a priori* model for $\mathbf{x}$ is of the Langevin type say,

$$\mathbf{x} =\sim L(x_0, \kappa_0)$$

and assuming independence of $\mathbf{x}$ and $\mathbf{p}$, we can form the a posteriori distribution $p(x \,|\, y)$ by Bayes rule. The joint density is

$$p(x\,,\,y) = p(y \,|\, x)p(x) = A(\kappa, \kappa_0) \exp \hat{\kappa}\, \hat{\mu}' x$$

where

$$\begin{aligned} A(\kappa, \kappa_0) &= \frac{\kappa}{4\pi \sinh \kappa} \frac{\kappa_0}{4\pi \sinh \kappa_0} \\ \hat{\kappa}\, \hat{\mu}' x &:= \kappa y' x + \kappa_0 x_0' x. \end{aligned}$$

Here $\hat{\kappa} = \hat{\kappa}(y, x_0) > 0$ and $\hat{\mu} = \hat{\mu}(y, x_0)$ are functions of y and of the a priori mode x_0 defined by

$$\hat{\mu} := \frac{\kappa y + \kappa_0 x_0}{\hat{\kappa}} \qquad \hat{\kappa} := \|\kappa y + \kappa_0 x_0\|. \qquad (13)$$

Note that $\|\hat{\mu}\| = 1$. Dividing by the marginal one obtains the *a posteriori* density

$$p(x \,|\, \mathbf{y}) = \frac{\hat{\kappa}}{4\pi \sinh \hat{\kappa}} \exp \hat{\kappa}(\mathbf{y})\, \hat{\mu}'(\mathbf{y}) x$$

which is still Langevin. The conditional mode vector $\hat{\mu}(\mathbf{y})$ (the *Bayesian Maximum a Posteriori estimate* of $\mathbf{x}$, given the observation $\mathbf{y}$) and the conditional concentration $\hat{\kappa}(\mathbf{y})$ are trivial to compute in this case and in fact still given by formula (13). These formulas can be generalized to the case of sequential observations of a fixed target point [11].

Assume we have a sequence of observations

$$\mathbf{y}(t) := R(\mathbf{p}(t))\, \mathbf{x} = \exp\{\mathbf{p}(t) \wedge\}\, \mathbf{x} \quad t = 1, 2, \ldots \quad (14)$$

where the $\mathbf{p}$'s are identically distributed independent random rotations, also independent of the random vector $\mathbf{x}$. The $\mathbf{y}(t)$'s are conditionally independent given $\mathbf{x}$, and $p(y(t) \,|\, \mathbf{x}) = L(\mathbf{x}, \kappa)$, where κ is the concentration parameter of the angular noise. Hence, denoting $\mathbf{y}^t := [\mathbf{y}(1), \ldots, \mathbf{y}(t)]'$, we may write

$$p(y^t \,|\, \mathbf{x}) = \frac{\kappa^t}{(4\pi \sinh \kappa)^t} \exp \kappa \langle \mathbf{x}, \sum_{s=1}^{t} y(s) \rangle \qquad (15)$$

where $\langle .\,, .\rangle$ denotes inner product i n $\mathbb{R}^3$. Assuming an a priori density $\mathbf{x} \sim L(x_0, \kappa_0)$, one readily obtains the a posteriori measure

$$p(x \,|\, \mathbf{y}^t) = \frac{\hat{\kappa}(t)}{(4\pi \sinh \hat{\kappa}(t))} \exp \hat{\kappa}(t) \langle \hat{\mu}(t)\,,\, x \rangle \qquad (16)$$

which is still of the Langevin class with parameters

$$\begin{aligned} \hat{\mu}(t) &= \frac{1}{\hat{\kappa}(t)} (\kappa \sum_{s=1}^{t} y(s) + \kappa_0 x_0) \qquad (17) \\ \hat{\kappa}(t) &= \|\kappa \sum_{s=1}^{t} y(s) + \kappa_0 x_0\| \qquad (18) \end{aligned}$$

These formulas can be easily updated for adjunction of the $t{+}1$-st measurement, obtaining formulas which look like a nonlinear version of the usual "Kalman-Filter" updates for the sample mean which one would obtain in the Gaussian case.

Proposition 3.1 *The MAP estimate (conditional mode) $\hat{\mu}(t)$, of the fixed random direction $\mathbf{x}$ observed corrupted by independent angular noise $\{\mathbf{p}(t)\}$ of concentration κ, propagates in time according to the recursions*

$$\begin{aligned} \hat{\mu}(t+1) &= \frac{1}{\hat{\kappa}(t+1)} (\hat{\kappa}(t)\hat{\mu}(t) + \kappa y(t+1)) \quad (19) \\ \hat{\kappa}(t+1) &= \|\hat{\kappa}(t)\hat{\mu}(t) + \kappa y(t+1)\| \qquad (20) \end{aligned}$$

with initial conditions $\hat{\mu}(0) = x_0$ and $\hat{\kappa}(0) = \kappa_0$. Moreover,as $E\hat{\mu}(t) = \mu$ and $\hat{\mu}(t) \to \mu \quad w.p.1\ ast \to \infty$.

4 Dynamic estimation

Generalizing the recursive MAP estimation to the case of a moving target point is a nontrivial problem.

Dynamic Bayes formulas

Assume the random motion of the target point $\mathbf{x}(t)$, forms a stationary Markov process on the sphere and denote by $p(x_t|\mathbf{y}^t)$ the a posteriori density given the observations $\mathbf{y}^t$. A standard application of Bayes rule, see e.g. [13, p.174], provides the formulas

$$p(x_{t+1}|\mathbf{y}^{t+1}) = \frac{1}{N} p(\mathbf{y}_{t+1}|x_{t+1}) \, p(x_{t+1}|\mathbf{y}^t) \quad (21)$$

$$p(x_{t+1}|\mathbf{y}^t) = \int_{\mathbb{S}^2} p(x_{t+1}|x_t) \, p(x_t|\mathbf{y}^t) \, d\sigma_{x_t} \quad (22)$$

where N is a normalization constant. Note that if both the observation noise model and the a priori conditional density $p(x_{t+1}|\mathbf{y}^t)$ are Langevin-like, so is the a posteriori density $p(x_{t+1}|\mathbf{y}^{t+1})$. In this ideal situation the evolution of the conditional mode $\hat{\mu}(t|t)$ of $p(x_t|\mathbf{y}^t)$ for adjunction of the $(t+1)$-st measurement is described by recursive formulas analogous to (19), (20), Unfortunately, for this to be true, the Chapman-Kolmogorov transition operator in (22) should map Langevin distributions into Langevin distributions, apparently a rather unlikely situation, if we exclude very trivial examples. Anyway, in this fortunate situation (assuming a Langevin initial distribution for $\mathbf{x}(0)$), the Bayes iteration wuold preserves the Langevin structure and an exact finite-dimensional filter would result, described completely in terms of conditional mode and conditional concentration, (22) providing an updating relation for the a priori mode of the form

$$\hat{\mu}(t+1|t) = F(\hat{\kappa}(t|t) \, \hat{\mu}(t|t)) \quad (23)$$

$$\hat{\kappa}(t+1|t) = g(\hat{\kappa}(t|t)\hat{\mu}(t|t)) \quad (24)$$

where F and g are in principle computable from the Markovian model. In reality, nontrivial examples where the Langevin distribution is preserved exactly in the prediction step, are hard to find. There are however extensive comparison studies, reported e.g. in [12] showing that some classical models (say Brownian motion on the sphere) tend in certain cases to preserve the Langevin structure, at least approximately. These examples are well-suited for the wide-sense approach based on best Langevin approximation, which we have mentioned in the previous section.

The wide-sense filter

Assume we are given a Markov model describing the motion of the target point on the Sphere. The steps of the estimation algorithm are the following

1. Let $p(y_t|\mathbf{x}(t) = x_t) \sim L(x_t, \kappa_o)$ (Langevin-distributed angular observation noise and assume $p(x_t|y^{t-1}) \sim L(\hat{\mu}_{t|t-1}, \hat{\kappa}_{t|t-1})$ is available.
2. (Measurement update) when the measurement y_t becomes available one has $p(x_t|y^t) \sim L(\hat{\mu}_{t|t}, \hat{\kappa}_{t|t})$ where

$$\hat{\mu}_{t|t} = \frac{1}{\hat{\kappa}_{t|t}}(\hat{\kappa}_{t|t-1}\hat{\mu}_{t|t-1} + \kappa_o y_t)$$

$$\hat{\kappa}_{t|t} = \|\hat{\kappa}_{t|t-1}\hat{\mu}_{t|t-1} + \kappa_o y_t\|.$$

3. (Lifting) compute the conditional mean

$$\hat{m}_x(t+1 \mid t) = E\left[\mathbf{x}(t+1) \mid y^t\right]$$

using $p(x_t|y^t)$ above and the given Markov model.

4. the best Langevin approximation of the conditional distribution of $\mathbf{x}(t+1)$ given y^t, is computed by solving

$$\hat{\mu}_{t+1|t} = \frac{\hat{m}_x(t+1 \mid t)}{\|\hat{m}_x(t+1 \mid t)\|}$$

$$\frac{\cosh \hat{\kappa}_{t+1|t}}{\sinh \hat{\kappa}_{t+1|t}} - \frac{1}{\hat{\kappa}_{t+1|t}} = \frac{1}{\|\hat{m}_x(t+1 \mid t)\|}$$

5. pretend $p(x_{t+1}|y^t)$ is Langevin. Repeat the first step when y_{t+1} is available to compute $p(x_{t+1}|y^{t+1})$, etc.

Estimation of a Brownian motion evolving on a sphere

Brownian motion on spheres can be defined axiomatically as the natural analog of the process in $\mathbb{R}^n$ and is discussed by several authors. The classical references are Perrin [10], McKean [8] and Brockett [2].

The stochastic differential equation

$$d\mathbf{x}(t) = A\mathbf{x}(t)\,dt + \sum_{i=1}^{p} B_i \mathbf{x}\,d\mathbf{b}(t) \quad \|\mathbf{x}(0)\| = 1, \quad (25)$$

where $\mathbf{b}(t)$ is p-dimensional standard Brownian motion (in $\mathbb{R}^p$) and A is the sum of a skew symmetric matrix plus a Itô "correction term", i.e.

$$A = \Omega - \frac{1}{2}\sum_{i=1}^{p} B_i' B_i \quad \Omega = -\Omega'$$

defines a homogeneous Markov process with values in $\mathbb{S}^{n-1}$ which, in fact, represents a rotational Brownian motion on the sphere. This can be seen by rewriting (25) a little more explicitely as:

$$d\mathbf{x}(t) = [\bar{\omega} dt + L d\mathbf{b}(t)] \wedge \mathbf{x}(t) - \frac{1}{2}\sum_{i=1}^{p} B_i' B_i \mathbf{x}(t)\,dt$$

where $\bar{\omega}\wedge := \Omega$ and L is defined in an obvious way. The term between square brackets is an infinitesimal random angular velocity vector $d\boldsymbol{\omega}(t)$, so that,

$$d\mathbf{x}(t) = d\boldsymbol{\omega}(t) \wedge \mathbf{x}(t) + (\text{Itô correction}) \,. \quad (26)$$

Now, assume that the observation process is governed by a conditional law of the Langevin type $p(y_t \mid x_t) \sim L(x_t, \kappa_0)$ and that after the last available measurement, $y(t_0)$, the *a posteriori* conditional distribution, $p(x_{t_0}|y^{t_0}) \sim L(\hat{\mu}(t_0|t_0), \hat{\kappa}(t_0|t_0))$ is available at time t_0. We shall compute the best Langevin approximant of the *a priori* conditional density before the next measurement, say $p(x_t|y^{t_0})$, $t > t_0$.

To this end we don't need to solve the Fokker-Planck equation to obtain $p(x_t \mid y^{t_0})$ and then approximate it via minimization of the Kullback distance; we just need to compute the conditional mean $m_x(t \mid t_0) = E(\mathbf{x}(t) \mid y^{t_0})$. The conditional mean is immediately computed from the equation (25), making use of the zero-mean property of the Itô's integral, as

$$m_x(t \mid t_0) = \exp\{(\Omega - \frac{1}{2}\sum_{i=1}^{p} B_i'B_i)(t-t_0)\}m_x(t_0 \mid t_0). \tag{27}$$

If

$$\frac{1}{2}\sum_{i=1}^{p} B_i'B_i = \sigma^2 I$$

with σ^2 a positive scalar variance paramenter,(isotropic diffusion on the sphere) we obtain

$$m_x(t \mid t_0) = e^{-\sigma^2(t-t_0)}\exp\{\Omega(t-t_0)\}m_x(t_0 \mid t_0). \tag{28}$$

which incidentally shows that the conditional mean tends to zero as $t \to \infty$, a natural phenomenon for diffusion processes. The parameters $\hat{\mu}(t \mid t_0)$ and $\hat{\kappa}(t \mid t_0)$ of the conditional Langevin distribution $L(\hat{\mu}(t \mid t_0), \hat{\kappa}(t \mid t_0))$ approximating $p(x_t \mid y^{t_0})$ are obtained from step 4 of the "wide-sense" algorithm above. Note that in order to get $\hat{\kappa}(t \mid t_0)$ we need to solve a trascendental equation. One may take advantage of the fact that for moderately large $\hat{\kappa}(t \mid t_0)$ the second equation is well approximated by :

$$1 - \frac{1}{\hat{\kappa}(t \mid t_0)} = \|m_x(t \mid t_0)\|$$

to write an approximate explicit formula for $\hat{\kappa}(t|to)$ so that

$$\hat{\kappa}(t|to) = \frac{1}{1 - \|m_x(t \mid t_0)\|} \tag{29}$$

$$\hat{\mu}(t|to) = \frac{m_x(t \mid t_0)}{\|m_x(t \mid t_0)\|} \tag{30}$$

Using this approximation and substituting (27) in the expressions above we obtain

$$\hat{\kappa}(t \mid to) = \frac{\hat{\kappa}(t_0 \mid t_0)}{\hat{\kappa}(t_0 \mid t_0)(1 - e^{-\sigma^2(t-t_0)}) + e^{-\sigma^2(t-t_0)}} \tag{31}$$

$$\hat{\mu}(t \mid to) = \exp\{\Omega(t-t_0)\}\hat{\mu}(t_0 \mid t_0) \tag{32}$$

In this way we obtain an approximate version of the conditional density $p(x_t|y^{t_0}) \simeq L(\hat{\mu}(t \mid to), \hat{\kappa}(t \mid to))$ (valid for an isotropic diffusion on the sphere and conditional concentration parameter larger than a few units). In general the Langevin approximation is fairly good for $\hat{\kappa}$ greater than 2 or 3, see [12]. Simulations (not reported here for reasons of space) show a very satisfactory perfomance of the filter for a wide range of parameter values.

Conclusions

In this paper we have discussed a simple Bayesian estimation problem on spheres related to a prototype directional reconstruction problem in computer vision. For a fixed direction in space, a simple closed-form recursive MAP estimator is derived. For a general Markovian target an approximate "wide-sense" filtering algorithm is presented which only requires the a priori updating of the conditional mean. An example of wide-sense filter tracking a Brownian trajectory on the unit sphere, has been discussed. Much work remains to be done.

References

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

[2] 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.

[3] 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.

[4] A. Chiuso and G. Picci, Tracking of point features in computer vision as estimation on the unit sphere, in *The Confluence of Vision and Control*, D. Kriegman, W. Haeger and S. Morse eds, Springer LNCIS series, to appear 1998.

[5] R. A. Fisher, Dispersion on a sphere, *Proc. Royal Soc. London*, **A 217**, p. 295-305, 1953.

[6] P. Langevin, Magnetisme et theorie des electrons, *Ann. de Chim et de Phys.*, **5**, p. 70-127, 1905.

[7] 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**, pp. 10-33, 1975.

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

[9] Oksendal, *Stochastic Differential Equations*, Springer, 1990.

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

[11] G. Picci, Dynamic Vision and Estimation on Spheres, in *Proceedings of the 36th Conf. on Decision and Control*, p. 1140-1145, IEEE Press, 1997.

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

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