Particle Filtering on Riemannian Manifolds

Hichem Snoussi[*] and Ali Mohammad-Djafari[†]

[*]*Charles Delaunay Institute,*
FRE CNRS 2848, University of Technology of Troyes
12, rue Marie Curie, 10010, Troyes France
[†]*Laboratoire des Signaux et Systèmes,*
Unité mixte de recherche 8506 (CNRS-Supélec-UPS)
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette, France

Abstract. Particle filtering is an approximate Monte Carlo method implementing the Bayesian Sequential Estimation. It consists in online estimation of the a posteriori distribution of the system state given a flow of observed data. The popularity of the particle filter method stems from its simplicity and flexibility to deal with non linear/non Gaussian dynamical models. However, this method suffers from the curse of dimensionality. In general, the system state lies in a constrained subspace which dimension is much lower than the whole space dimension. In this contribution, we propose an implementation of the particle filter with the constraint that the system state lies in a low dimensional Riemannian manifold. The sequential Bayesian updating consists in drawing state samples while moving on the manifold geodesics. An Affine Generalized Hyperbolic regression process is proposed to model the transition dynamics on the manifold. It is a parametric family able to cover a wide range of tail behaviors of real signal distributions.

Key Words: Sequential Monte Carlo, Riemannian Manifolds, Differential Geometry, Lie Groups.

INTRODUCTION

In this section, we recall some definitions from differentiable geometry related to the concept of Riemannian manifolds. For further details, please refer to [1]. First, we need to define a topological manifold as follows:

Definition 1 *A manifold $\mathcal{M}$ of dimension n, or n-manifold, is a topological space with the following properties:*
   (i) $\mathcal{M}$ is Haussdorff,
   (ii) $\mathcal{M}$ is locally Euclidean of dimension n, and
   (iii) $\mathcal{M}$ has a countable basis of open sets.

Intuitively, a topological manifold is a set of points which can be considered locally as a flat Euclidean space. In other words, each point $p \in \mathcal{M}$ has a neighborhood U homeomorphic to an n-ball in $\mathbb{R}^n$. Let ϕ be such an homeomorphism. The pair (U, ϕ) is called a coordinate neighborhood: to $p \in U$ we assign the n coordinates $\xi^1(p), \xi^2(p), ..., \xi^n(p)$ of its image $\phi(p)$ in $\mathbb{R}^n$. If p lies also in a second neighborhood V, let $\psi(p) = [\rho^1(p), \rho^2(p), ..., \rho^n(p)]$ be its correspondent coordinate system. The transfor-

mation $\psi \circ \phi^{-1}$ on $\mathbb{R}^n$ given by:

$$\psi \circ \phi^{-1} : [\xi^1, ..., \xi^n] \Longleftrightarrow [\rho^1, ..., \rho^n],$$

defines a local coordinate transformation on $\mathbb{R}^n$ from $\phi = [\xi^i]$ to $\psi = [\rho^i]$. In differential geometry, one is interested in intrinsic geometric properties which are invariant with respect to the choice of the coordinate system. This can be achieved by imposing smooth transformations between local coordinate systems (see figure 1). The following definition of a differentiable manifold formalizes this concept in a global setting:

Definition 2 *A differentiable (or smooth) manifold $\mathcal{M}$ is a topological manifold with a family $\mathcal{U} = \{U_\alpha, \phi_\alpha\}$ of coordinate neighborhoods such that:*
(1)   the U_α cover $\mathcal{M}$,
(2)   for any α, β, if the neighborhoods intersection $U_\alpha \cap U_\beta$ is non empty, then $\phi_\alpha \circ \phi_\beta^{-1}$ and $\phi_\beta \circ \phi_\alpha^{-1}$ are diffeomorphisms of the open sets $\phi_\beta(U_\alpha \cap U_\beta)$ and $\phi_\alpha(U_\alpha \cap U_\beta)$ of $\mathbb{R}^n$,
(3)   any coordinate neighborhood (V, ψ) meeting the property (2) with every $U_\alpha, \phi_\alpha \in \mathcal{U}$ is itself in $\mathcal{U}$
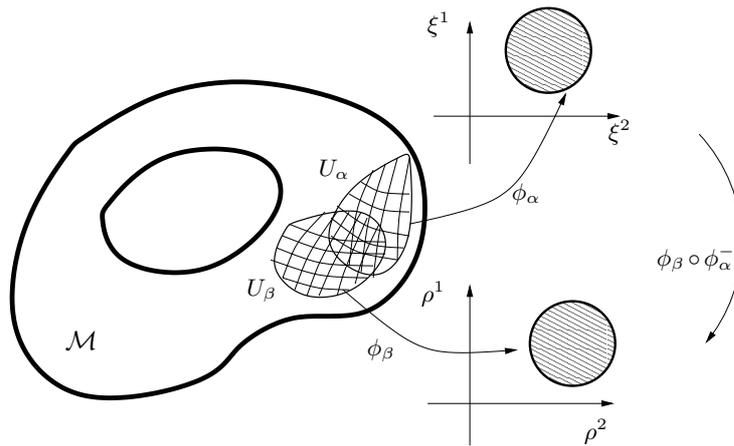

FIGURE 1.   Differentiable manifold

On a differentiable manifold, an important notion (for the remainder of this paper) is the tangent space. The tangent space $T_p(\mathcal{M})$ at a point p of the manifold $\mathcal{M}$ is the vector space of the tangent vectors to the curves passing by the point p. It is intuitively the vector space obtained by a local linearization around the point p. More formally, it is the vector space spanned by the differential operators $(\frac{\partial}{\partial \xi^i})_p$:

$$T_p(\mathcal{M}) = \left\{ c^i (\frac{\partial}{\partial \xi^i})_p \mid [c^1, ..., c^n] \in \mathbb{R}^n \right\},$$

where the differential operator $(\frac{\partial}{\partial \xi^i})_p$ can be seen geometrically as the tangent vector to the ith coordinate curve (fixing all ξ^j coordinates $j \neq i$ and varying only the value of ξ^i), see figure 2.

For each point p in $\mathcal{M}$, assume that an inner product $<,>_p$ is defined on the tangent space $T_p(\mathcal{M})$. Thus, a mapping from the points of the differentiable manifold to their
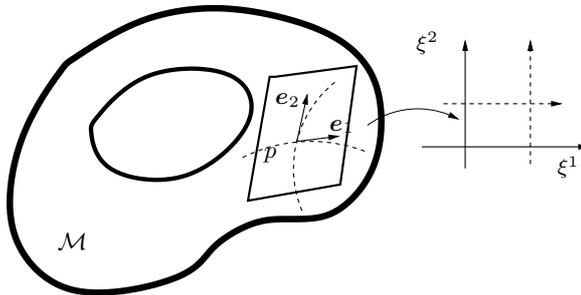
FIGURE 2. Tangent space on the manifold

inner product (bilinear form) is defined. If this mapping is smooth, then the pair $(\mathcal{M}, <, >_p)$ is called **Riemannian manifold**.

 Geodesics. A geodesic between two endpoints $\gamma(a)$ and $\gamma(b)$ on a Riemannian manifold $\mathcal{M}$ is a curve $\gamma : [a, b] \longrightarrow \mathcal{M}$ which is locally defined as the shortest curve on the manifold connecting these endpoints. More formally, the definition of a geodesic is given by:

Definition 3 *The parametrized curve $\gamma(t)$ is said to be a geodesic if its velocity (tangent vector) $d\gamma/dt$ is constant (parallel), that is if it satisfies the condition $(d/dt)(d\gamma/dt) = 0$, for $a < t < b$.*

PARTICLE FILTERING ON RIEMANNIAN MANIFOLDS

Model and objectives

 The main contribution of this paper is to propose a particle filter method for filtering in nonlinear dynamical systems, the system state being constrained to lie on a Riemannian manifold $\mathcal{M}$. Only few recent papers have tried to use the geometry of the manifold to design learning algorithms [2, 3]. Our work can be considered as an extension of [2] to deal with general stochastic models suitable to the particle filter method.

 Particle filtering is an approximate Monte Carlo method estimating, recursively in time, the marginal posterior distribution of the continuous hidden state of the system, given the observations. The particle filter provides a point mass approximation of these distributions. For more details and a comprehensive review of the particle filter on Euclidean spaces see [4].

 The observed system evolves in time according to the following nonlinear dynamics:

$$\begin{cases} \boldsymbol{x}_t & \sim \quad p_x(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, u_t) \\ \\ \boldsymbol{y}_t & \sim \quad p_y(\boldsymbol{y}_t \mid \boldsymbol{x}_t, u_t), \end{cases} \tag{1}$$

where $\boldsymbol{y}_t \in \mathbb{R}^{n_y}$ denotes the data measured at time t, $\boldsymbol{x}_t \in \mathcal{M}$ denotes the unknown continuous state, and $u_t \in \mathcal{U}$ denotes a known control signal. The probability distribution $p_x(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, u_t)$ models the stochastic transition dynamics of the hidden state. Given the

continuous state, the observations $\boldsymbol{y}_t$ follow a stochastic model $p_y(\boldsymbol{y}_t \mid \boldsymbol{x}_t, u_t)$, where the stochastic aspect reflects the observation noise.

The Bayesian filtering is based on the estimation of the posterior marginal probability $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t})$. The nonlinear and the non Gaussian aspect of the transition distributions leads to intractable integrals when evaluating the marginals. Therefore, one has to resort to Monte Carlo approximation where the joint posterior distribution $p(\boldsymbol{x}_{0:t} \mid \boldsymbol{y}_{1:t})$ is approximated by the point-mass distribution of a set of weighted samples (called particles) $\{\boldsymbol{x}_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$:

$$\hat{p}_N(\boldsymbol{x}_{0:t} \mid \boldsymbol{y}_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\boldsymbol{x}_{0:t}^{(i)}}(d\boldsymbol{x}_{0:t}),$$

where $\delta_{\boldsymbol{x}_{0:t}^{(i)}}(d\boldsymbol{x}_{0:t})$ denotes the Dirac function.

Based on the same set of particles, the marginal posterior probability (of interest) $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t})$ can also be approximated as follows:

$$\hat{p}_N(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\boldsymbol{x}_t^{(i)}}(d\boldsymbol{x}_t),$$

In the Bayesian importance sampling (IS) method, the particles $\{\boldsymbol{x}_{0:t}^{(i)}\}_{i=1}^N$ are sampled according to a proposal distribution $\pi(\boldsymbol{x}_{0:t} \mid \boldsymbol{y}_{1:t})$ and $\{w_t^{(i)}\}$ are the corresponding normalized importance weights:

$$w_t^{(i)} \propto \frac{p(\boldsymbol{y}_{1:t} \mid \boldsymbol{x}_{0:t}^{(i)})p(\boldsymbol{x}_{0:t}^{(i)})}{\pi(\boldsymbol{x}_{0:t}^{(i)} \mid \boldsymbol{y}_{1:t})}.$$

Sequential Monte Carlo (SMC) consists of propagating the trajectories $\{\boldsymbol{x}_{0:t}^{(i)}\}_{i=1}^N$ in time without modifying the past simulated particles. This is possible for the class of proposal distributions having the following form:

$$\pi(\boldsymbol{x}_{0:t} \mid \boldsymbol{y}_{1:t}) = \pi(\boldsymbol{x}_{0:t-1} \mid \boldsymbol{y}_{1:t-1})\pi(\boldsymbol{x}_t \mid \boldsymbol{x}_{0:t-1}, \boldsymbol{y}_{1:t}).$$

The normalized importance weights are then recursively computed in time as:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p_y(\boldsymbol{y}_t \mid \boldsymbol{x}_t^{(i)})p(\boldsymbol{x}_t^{(i)} \mid \boldsymbol{x}_{0:t-1}^{(i)})}{\pi(\boldsymbol{x}_t^{(i)} \mid \boldsymbol{x}_{0:t-1}^{(i)}, \boldsymbol{y}_{1:t})}. \tag{2}$$

The particle filter algorithm consists of 2 steps: the sequential importance sampling step and the selection step. The selection (resampling) step replaces the weighted particles by unweighted particles in order to avoid the collapse of the Monte Carlo approximation caused by the variance increase of the weights. It consists of selecting the trajectories $\{\boldsymbol{x}_{0:t}^{(i)}\}$ with probabilities $w_t^{(i)}$. The trajectories with weak weights are eliminated and the trajectories with strong weights are multiplied. After the selection step, all the weights are equal to $1/N$.

One of the simplest choices for the proposal distribution $\pi(\boldsymbol{x}_t \mid \boldsymbol{x}_{0:t-1}, \boldsymbol{y}_{1:t})$ is the transition prior $p_x(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, u_t)$. The weights $w_t^{(i)}$ are then proportional to the data likelihood:

$$w_t^{(i)} \propto w_{t-1}^{(i)} p_y(\boldsymbol{y}_t \mid \boldsymbol{x}_t^{(i)}). \tag{3}$$

The density distribution p_y is fixed by the observation model. Concerning the transition prior $p_x(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, u_t)$, we propose in this paper a Multivariate Affine Generalized Hyperbolic Regression model. Multivariate Hyperbolic processes have attractive analytical and statistical properties. In particular, this parametric family is able to describe the fat tails and the skewness of the regression model $p_x(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, u_t)$. In the following, after describing this statistical model, we propose an efficient algorithm to draw samples in the particle filter context.

Multivariate Affine Generalized Hyperbolic Regression (MAGH-R)

Before introducing the Multivariate Affine Generalized Hyperbolic Regression model, we briefly describe the Generalized Hyperbolic distributions and their main properties (for more details refer to Barndorff-Nielsen's original work [5] or Bibby and Sorensen [6]). Generalized hyperbolic distributions form a five parameter family $\mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ introduced by Barndorff-Nielsen (1977). If the random variable X follows the distribution $\mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ then its pdf reads,

$$\frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi} K_\lambda(\delta\gamma)} \cdot \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{(\sqrt{\delta^2 + (x-\mu)^2}/\alpha)^{\frac{1}{2}-\lambda}} \cdot e^{\beta(x-\mu)}, \; x \in \mathbb{R} \tag{4}$$

where $\gamma^2 = \alpha^2 - \beta^2$ and $K_\lambda(.)$ is the modified Bessel function of third kind:

$$K_\lambda(y) = \frac{1}{2} \int_0^\infty u^{\lambda-1} e^{-\frac{1}{2}y(u+u^{-1})} du.$$

GH distributions enjoy the property of being invariant under affine transformations. If $X \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$, then the random variable $aX + b$ follows the distribution $\mathrm{GH}(\lambda, \alpha/a, \beta/a, a\delta, a\mu + b)$. Many known subclasses can be obtained, either by fixing some parameters or by considering limiting cases: $\lambda = 1$ and $\lambda = -1/2$ respectively yield the hyperbolic and the NIG distributions (the latter being closed under convolution) ; $\lambda = 1$ with $\delta \to 0$ provides the asymmetric Laplace distribution, $\lambda = -1/2$ with $\alpha \to 0$ corresponds to the Cauchy distribution, the asymmetric scaled t-distribution is obtained for $\alpha = |\beta|$, etc. Thus, varying the parameters of the GH distributions yields a wide range of tail behaviors, from Gaussian tails to the heavy tails of the Student t-distributions. Figure 3 depicts examples of GH distributions. One can note that a wide range of tail behaviors is covered and the possibility of modeling the distribution asymmetry (via the parameter β).

Stochastic processes are usually defined on Euclidean spaces. In order to define a stochastic process on a Riemannian manifold $\mathcal{M}$, the notion of **exponential mapping** represents an interesting tool to build a bridge between an euclidean space and the
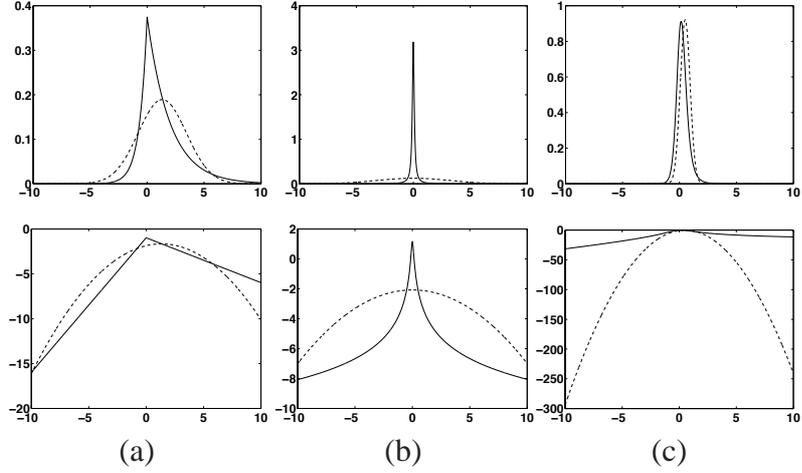
FIGURE 3. Examples of the GH distributions:
(a) hyperbolic case: $\lambda = 1, \alpha = 1, \beta = .5, \delta = .001, \mu = 0$;
(b) Cauchy case: $\lambda = -.5, \alpha = .01, \beta = .001, \delta = .01, \mu = 0$;
(c) Student case: $\lambda = 3, \alpha = 1, \beta = 1, \delta = 1, \mu = 0$.
Pdfs appear on top row, log densities on bottom row. The dashed line corresponds to the Gaussian distribution with same mean and variance.

Riemannian manifold. For a point p and a tangent vector $X \in T_p(\mathcal{M})$, let $\gamma : t \Longrightarrow \gamma(t)$ be the geodesic such that $\gamma(0) = p$ and $\frac{d\gamma}{dt}(0) = X$. The exponential mapping of X is defined as $\mathcal{E}_p(X) = \gamma(1)$. In other words, the exponential mapping assigns to the tangent vector X the endpoint of the geodesic whose velocity at time $t = 0$ is the vector X (see figure 4). It can be shown that there exists an neighborhood U of 0 in $T_p(\mathcal{M})$ and a neighborhood V of p in $\mathcal{M}$ such that $\mathcal{E}_p \mid_U$ is a diffeomorphism from U to V. Also, note that since the velocity $d\gamma/dt$ is constant along the geodesic $\gamma(t)$, its length L from p to $\mathcal{E}_p(X)$ is:

$$L = \int_0^1 \|\frac{d\gamma}{dt}\| dt = \int_0^1 \|X\| dt = \|X\|.$$

Thus, the exponential mapping $\mathcal{E}_p(X)$ corresponds to the unique point on the geodesic whose distance from p is the length of the vector X.

The Multivariate Affine Generalized Hyperbolic Regression process is the extension of the model proposed in [7] to Riemannian manifolds. It is defined as follows:

Definition 4 *An n-dimensional random process $(X_t)_{t \in \mathbb{Z}} \in \mathcal{M}$ is said to be multivariate regressive affine generalized hyperbolic process distributed with scaling matrix $\Sigma \in \mathbb{R}^{n \times n}$ and parameters $\boldsymbol{\omega} := (\omega_1, ..., \omega_n)$, $\omega_i = (\lambda_i, \alpha_i, \beta_i)$ and denoted by MAGH-R$_{\mathcal{M}}(\Sigma, \boldsymbol{\omega})$ if it has the following recursive stochastic representation:*

$$\boldsymbol{V}_t = \boldsymbol{A}^T \boldsymbol{Z}$$

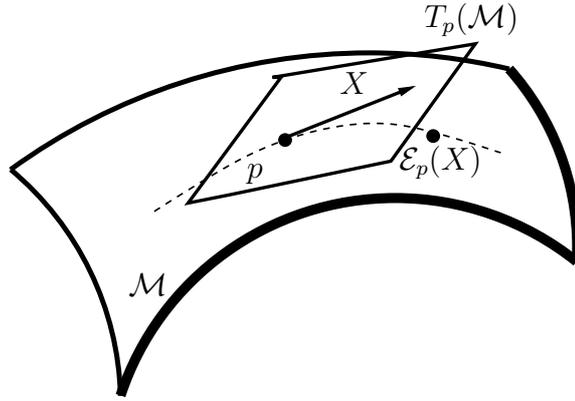$$X_t = \mathcal{E}_{X_{t-1}}\left(\sum V_t^i (\frac{\partial}{\partial \xi^i})_{X_{t-1}}\right)$$

FIGURE 4. Exponential mapping on the manifold

for some lower triangular matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{\Sigma}$ is positive definite and the random vector $\boldsymbol{Z} = (Z^1, ..., Z^n)^T$ consists of mutually independent random variables $Z^i \sim \mathrm{GH}(\lambda, \alpha, \beta, 1, 0)$.

In words, the sample X_t is obtained by first drawing a multivariate generalized hyperbolic tangent vector in the tangent space $T_{X_{t-1}}(\mathcal{M})$, then applying the exponential mapping on it.

Sampling the MAGH-R process on Riemannian Manifolds

An important feature of the GH distribution is its expression as a continuous normal mean-variance mixture:

$$\mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \int_0^\infty \mathcal{N}(x; \mu + \beta w, w) \, GIG(w; \lambda, \gamma, \delta) \, dw \qquad (5)$$

where the variance W of each Gaussian component follows a Generalized Inverse Gaussian (GIG) distribution:

$$\begin{cases} GIG(w; \lambda, \gamma, \delta) = \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} . w^{\lambda-1} . \exp\left[-\frac{1}{2}(\delta^2 w^{-1} + \gamma^2 w)\right], \\ w > 0 \end{cases}$$

In other words, the Generalized Hyperbolic process can be seen as a double stochastic process:

1. First generate $W \sim GIG(\lambda, \gamma, \delta)$.
2. Then generate $X \sim \mathcal{N}(\mu + \beta W, W)$

We turn now to the generation of the MAGH-R$_\mathcal{M}(\boldsymbol{\Sigma}, \boldsymbol{\omega})$ process on the manifold $\mathcal{M}$:
Pseudo algorithm for generating MAGH-R$_\mathcal{M}$ process:

(I)   Initialize $X_0 \sim P_0(.)$

(II)   for $t = 1, 2, ...$

        (1)   Set $\boldsymbol{\Sigma} = \boldsymbol{A}^T \boldsymbol{A}$ via Cholesky decomposition.

        (2)   Generate a random vector $\boldsymbol{Z}$ with independent $Z^i \sim \mathrm{GH}(\lambda, \alpha, \beta, 1, 0)$ (see above).

        (3)   Set the velocity $\boldsymbol{V} = \boldsymbol{A}^T \boldsymbol{Z}$ in the tangent space $T_{X_{t-1}}(\mathcal{M})$

        (4)   Return $X_t = \mathcal{E}_{X_{t-1}}(\sum V_t^i (\frac{\partial}{\partial \xi^i})_{X_{t-1}})$

Below is the pseudo algorithm of the particle filter on Riemannian manifolds:

Step 0: <u>Initialization</u>

- $X_0 \sim P_0(.)$

Step 1: For $t = 1$ to T,

a- <u>Sequential importance sampling:</u>

- For $i = 1, ..., N$, sample from the transition priors on the manifold $\mathcal{M}$:

$$\hat{X}_t^{(i)} \sim \text{MAGH-R}_{\mathcal{M}}(\boldsymbol{\Sigma}, \boldsymbol{\omega}) \text{ around } \hat{X}_{t-1}^{(i)}$$

and set

$$(\hat{X}_{0:t}^{(i)}) = (\hat{X}_t^{(i)}, X_{0:t-1}^{(i)})$$

b- <u>Update the importance weights:</u>

- For $i = 1, ..., N$, evaluate and normalize the weights:

$$w_t^{(i)} \propto p(\boldsymbol{y}_t \mid \hat{X}_t^{(i)})$$

c- <u>Resampling:</u>

- Select with replacement from $\{\hat{X}_{0:t}^{(i)}\}_{i=1}^N$ with probability $\{w_t^{(i)}\}$ to obtain N particles $\{X_{0:t}^{(i)}\}_{i=1}^N$

REFERENCES

1. W. Boothby, *An Introduction to Differential Manifolds and Riemannian Geometry*, Academic Press, Inv., 1986.

2. A. Srivastava and E. Klassen, "Bayesian and geometric subspace tracking", *Advances in Applied Probability*, vol. 36, no. 1, pp. 43–56, March 2004.

3. S. Fiori, "Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial", *Journal of Machine Learning Research*, vol. 6, pp. 743–781, 2005.

4. A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

5. O. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size", in *Proc. Roy. Soc.*, London, 1977, vol. 353, pp. 401–419.

6. B. Bibby and M. Sorensen, "Hyperbolic processes in finance", in *in Handbook of Heavy Tailed Distributions in Finance*, pp. 211–248. Rachev, S. Elsevier Science, 2003.

7. R. Schmidt, T. Hrycey, and E. Stutzle, "Multidimensional data modelling with generalized hyperbolic distributions", *Journal of Computational Statistics and Data Analysis*, vol. 50, pp. 2065–2096, 2006.