Particle and cell approximations for nonlinear filtering

Fabien Campillo, Frédéric Cérou, François Le Gland, Rivo Rakotozafy

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Particle and cell approximations for nonlinear filtering

Fabien Campillo, Frédéric Cérou, François Le Gland, Rivo Rakotozafy

N° 2567

Juin 1995

PROGRAMME 5

apport de recherche

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Particle and cell approximations for nonlinear filtering

Fabien Campillo, Frédéric Cérou, François Le Gland, Rivo Rakotozafy

Programme 5 — Traitement du signal, automatique et productique
Projet Mefisto

Rapport de recherche n° 2567 — Juin 1995 — 58 pages

Abstract: We consider the nonlinear filtering problem for systems with noise–free state equation. First, we study a particle approximation of the a posteriori probability distribution, and we give an estimate of the approximation error. Then we show, and we illustrate with numerical examples, that this approximation can produce a non consistent estimation of the state of the system when the measurement noise tends to zero. Hence, we propose a histogram–like modification of the particle approximation, which is always consistent. Finally, we present an application to target motion analysis.

Key-words: Nonlinear filtering, particle approximation, cell approximation, target motion analysis.

(Résumé : tsvp)

Approximations particulaire et cellulaire pour le filtrage non linéaire

Résumé : Nous considérons le problème de filtrage non–linéaire pour les systèmes sans bruit de dynamique. Nous étudions d'abord une approximation particulaire de la loi a posteriori, et nous donnons une estimation de l'erreur d'approximation. Nous mettons ensuite en évidence, et nous illustrons à l'aide d'exemples numériques, le fait que cette approximation peut donner un estimateur non–consistant de l'état du système, quand le bruit d'observation tend vers zéro. Nous proposons alors une modification de l'approximation particulaire, de type histogramme, qui est toujours consistante. Nous présentons enfin une application à la trajectographie passive.

Mots-clé : Filtrage non-linéaire, approximation particulaire, approximation cellulaire, trajectographie passive.

Contents

1 Introduction

Consider the following nonlinear filtering problem

$$\begin{aligned} \dot{X}_t &= b(X_t) \;, \qquad\qquad X_0 \text{ unknown,} \\ z_k &= h(X_{t_k}) + v_k \;, \end{aligned} \tag{1}$$

where $t_0 < t_1 < \cdots < t_k < \cdots$ is a strictly increasing sequence of observation times, and $\{v_k \;, \; k \geq 0\}$ is a Gaussian white noise with non singular covariance matrix R.

In this model, only the initial condition X_0 is unknown. The problem is to estimate X_0, at any instant t_k, given the past measurements $z_1, \cdots, z_k$.

We can expect that, for this particular model, the nonlinear filtering problem will reduce to a parameter estimation problem of the unknown parameter X_0. In this case, $x_0 \in \mathbb{R}^m$ will denote the true value of the parameter.

Whether we take a Bayesian approach or not, the goal is to compute :

Bayesian : the conditional probability distribution

$$\mu_0^k(dx) = P(X_0 \in dx | \mathcal{Z}_k) \;, \tag{2}$$

where

$$\mathcal{Z}_k \triangleq \sigma(z_1, \cdots, z_k) \;, \tag{3}$$

Non Bayesian : the likelihood function $\Xi_k(x)$ corresponding to the estimation of the unknown parameter $X_0 \in \mathbb{R}^m$.

In this work, we focus on problems where the a priori available information on the initial condition X_0 is quite poor, for example $X_0 \in K$, where K is a compact subset of $\mathbb{R}^m$. In this case, the two points of view — Bayesian and non Bayesian — are rather close.

We can study the asymptotic behavior (consistency, convergence rate, etc.) when

(i) the number k of observations tends to ∞ (long time asymptotics),

(ii) or when the noise covariance matrix R tend to 0 (small noise asymptotics).

We also want to study the case where the system (1) is not identifiable. In this case the conditional probability distribution $\mu_0^k(dx)$ does not concentrate around the true value $x_0 \in \mathbb{R}^m$, neither in the long time asymptotics, nor in the small noise asymptotics, but it concentrates around the subset $M(x_0) \subset \mathbb{R}^m$ of those points which are indistinguishable from the true value parameter. In this context, the relevant statistics that we should compute from the conditional probability distribution $\mu_0^k(dx)$ is not the usual pointwise Bayesian estimator of the *conditional mean* type :

$$\widehat{X}_0^k \triangleq \int x \; \mu_0^k(dx) \;.$$

We rather use the concept of *confidence region* :

Definition 1.1 (Confidence region) *For any given level* $0 < \alpha \leq 1$, *a* confidence region of level α *is defined by :*

$$\widehat{D}_k^\alpha \in \operatorname{Arg} \min_{D \in \mathbf{D}_k^\alpha} \lambda(D) \ , \quad \textit{with} \quad \mathbf{D}_k^\alpha \stackrel{\triangle}{=} \{D \subset \mathbb{R}^m \ : \ \mu_0^k(D) \geq \alpha\} \ , \tag{4}$$

where λ *denote the Lebesgue measure on* $\mathbb{R}^m$.

This is an extension of the maximum a posteriori (MAP) estimator which can be seen as the limit, when the level $\alpha \downarrow 0$, of the sequence of decreasing confidence regions. Our goal is to numerically compute the conditional probability distribution $\mu_0^k(dx)$.

In Section 2, we recall some properties of this nonlinear filtering problem, especially the crucial importance of the flow of diffeomorphisms $\{\xi_{s,t}(\cdot) \, , \, 0 \leq s \leq t\}$ associated with the differential equation $\dot{X}_t = b(X_t)$, i.e.

$$X_t = \xi_{s,t}(X_s) \ , \quad 0 \leq s \leq t \ ,$$

which allows an explicit formulation of the conditional probability distributions $\mu_0^k(dx) = P(X_0 \in dx \mid \mathcal{Z}_k)$ and $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$. As an example, we first consider the case where the initial probability distribution $\mu_0(dx) = P(X_0 \in dx)$ is discrete, and the case where it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$.

Then we develop two numerical methods for the approximation of the conditional probability distribution $\mu_k(dx)$.

Section 3 is devoted to a particle–like approximation algorithm

$$\mu_k(dx) \simeq \mu_k^H(dx) \ , \quad \text{with} \quad \mu_k^H = \sum_{i \in I} a_k^i \ \delta_{x_k^i} \ ,$$

as a convex linear combination of Dirac measures, called *particles*. This kind of approximation has been introduced and studied by Raviart [13] for the first order deterministic PDE's.

In the context of nonlinear filtering, this approximation technique has already been applied to various real case studies, see Campillo–Le Gland [2], Le Gland–Pardoux [10] and Cérou–Rakotozafy [5].

The problem is to define, at each time t_k, the positions $\{x_k^i \, , \, i \in I\}$ and the weights $\{a_k^i \, , \, i \in I\}$ of the particles. The more natural choice is to define the approximation $\mu_k^H(dx)$ as the conditional probability distribution of X_{t_k} given $\mathcal{Z}_k$, with $\mu_0^H(dx)$ instead of $\mu_0(dx)$ as an initial probability distribution. With this choice, the algorithm is the following

$$x_k^i = \xi_{t_{k-1},t_k}(x_{k-1}^i) \quad \text{and} \quad a_k^i = c_k \ \Psi_k(x_k^i) \ a_{k-1}^i \ , \tag{5}$$

for all $i \in I$, where, by definition, $\Psi_k(x)$ is the likelihood function for the estimation, given the measurement z_k, of the parameter $X_{t_k} \in \mathbb{R}^m$, that is

$$\Psi_k(x) \stackrel{\triangle}{=} \exp\left\{-\tfrac{1}{2}\|z_k - h(x)\|_{R^{-1}}^2\right\} \quad \text{with} \quad \|x\|_{R^{-1}}^2 \stackrel{\triangle}{=} x^* \, R^{-1} \, x \tag{6}$$

and c_k is a normalizing factor. The only error is in the approximation $\mu_0(dx) \simeq \mu_0^H(dx)$, and it is then important to choose appropriately the initial points $\{x_0^i \, , \, i \in I\}$ and weights $\{a_0^i \, , \, i \in I\}$. For this purpose, we use the approximation proposed in Florchinger–Le Gland [6].

The main points of this section are the following :

(i) From the numerical point of view, we obtain an estimate of the error

$$\mu_k(dx) - \mu_k^H(dx) \ ,$$

in terms of the discretization parameter H, see Theorem 3.1 below.

(ii) From the point–estimation point of view, the particle algorithm consists in restricting the parameter set to a set $G_H = \{x_0^i \, , \, i \in I\} \subset \mathbb{R}^m$ of possible initial conditions : this is a misspecified estimation problem (i.e. the true value x_0 is not necessarily in G_H). In case where the model (1) is not identifiable the setwise Bayesian estimator (4) based on the particle approximation could be non consistent in both the long time asymptotics and the small noise asymptotics (for a fixed discretization parameter H). This phenomenon will be illustrated by simulation results.

Because of this possible non–consistency, we will study in Section 4 another approximation technique, where at time t_k, instead of evaluating as in (5) the value of the likelihood function at the point x_k^i, we evaluate a generalized likelihood function on a neighborhood B_k^i of the particle position x_k^i, see definition (8) below. In this way, we introduce a cell–like approximation algorithm

$$\mu_k(dx) \simeq \bar{p}_k(x)\, dx \ , \quad \text{with} \quad \bar{p}_k(x) = \sum_{i \in I_k} \frac{\bar{\mu}_k^i}{\lambda_k^i} \, \mathbf{1}_{B_k^i}(x) \ ,$$

where, for all $i \in I_k$, λ_k^i is the Lebesgue measure of the *cell* B_k^i, and $\bar{\mu}_k^i$ is an approximation of the conditional probability $\mu_k^i = P(X_{t_k} \in B_k^i \mid \mathcal{Z}_k)$. This kind of approximation was proposed by James–Le Gland [7], for the approximation of nonlinear filters and observers. Here, the problem is to define, at each time t_k, the cells $\{B_k^i \, , \, i \in I_k\}$ and the approximate conditional probabilities $\{\bar{\mu}_k^i \, , \, i \in I_k\}$. Among many possible choices, we will focus on the following :

$$B_k^i = \xi_{t_{k-1},t_k}(B_{k-1}^i) \quad \text{and} \quad \bar{\mu}_k^i = c_k \, R_k^i \, \bar{\mu}_{k-1}^i \ , \tag{7}$$

for all $i \in I_k$, where, by definition, $I_k = I$ does not depend on k, R_k^i is the generalized likelihood function for the estimation, given the observation z_k, of the parameter $i \in I$ such that $\{X_{t_k} \in B_k^i\}$, that is :

$$R_k^i = \max_{x \in B_k^i} \Psi_k(x) \ , \tag{8}$$

and c_k is a normalization constant.

The main points of this section are the following :

(i) From the numerical point of view, we obtain an estimate of the error

$$\mu_k(dx) - \bar{p}_k(x)\,(dx) \ ,$$

in terms of the discretization parameter H, see Theorem 4.1 below.

(*ii*) From the point of view of estimating the cell containing the initial condition $X_0 \in \mathbb{R}^m$, we show the consistency of the Bayesian parameter in the small noise asymptotics (for a fixed discretization parameter H). This example is also illustrated by the problem already studied at Section 3 for the particle approximation.

The *long time asymptotics* is more difficult to handle. Actually, in the purely theoretical setup (without approximation) we have recently obtained some results concerning the convergence of the filter to the true value, when the identifiability hypothesis is fulfilled, see Cérou [3, 4].

2 Problem setting

We consider a particular nonlinear filtering problem where there is no noise input in the state equation. As a result, the only unknown quantity is the initial state of the system.

We can tackle this problem in two ways. Because the only unknown parameter is the initial condition X_0, we can expect that, in this case, the nonlinear filtering problem reduces to the problem of estimating the parameter X_0. We can consider the maximum likelihood estimator or the Bayesian estimator, and study the consistency properties, the rate of convergence, etc.

On the other hand, we can study the consequences of this particular setting on the nonlinear filtering equations and their numerical solution.

In the sequel, we study the following model

$$\begin{aligned} \dot{X}_t &= b(X_t) \;, \qquad\qquad X_0 \text{ unknown}, \\ z_k &= h(X_{t_k}) + v_k \;, \end{aligned} \tag{9}$$

where $t_0 < t_1 < \cdots < t_k < \cdots$ is a strictly increasing sequence of observation times, and $\{v_k \, , \, k \geq 0\}$ is a i.i.d. sequence of centered Gaussian random variables with covariance matrix R. Throughout this paper, we assume for simplicity that the sequence of observation times is uniform, i.e. $t_{k+1} - t_k = \Delta$ for all $k = 0, 1, \cdots$, and we make the following

Hypothesis 2.1 *The covariance matrix R is non singular. In the case where X_0 is a random variable, we suppose that it is independent of $\{v_k \, , \, k \geq 0\}$.*

2.1 Nonlinear filtering approach

Let $\mathcal{Z}_k$ denotes the σ–algebra generated by the observations up to time t_k

$$\mathcal{Z}_k \stackrel{\triangle}{=} \sigma(z_1, \cdots, z_k) \;,$$

and suppose that X_0 is a r.v. with probability distribution :

$$\mu_0(dx) \stackrel{\triangle}{=} P(X_0 \in dx) \;.$$

Our goal is to compute the conditional probability distribution $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$ of X_{t_k} given $\mathcal{Z}_k$.

Definition 2.2 (Conditional probability distributions) *We introduce the following notation :*

(i) $\mu_k(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_k)$, *is the conditional probability distribution of* X_{t_k} *given* $\mathcal{Z}_k$.

(ii) $\mu_k^-(dx) = P(X_{t_k} \in dx \mid \mathcal{Z}_{k-1})$, *is the conditional probability distribution of* X_{t_k} *given* $\mathcal{Z}_{k-1}$.

(iii) For $t_k \leq t \leq t_{k+1}$, $\mu_t^k(dx) = P(X_t \in dx \mid \mathcal{Z}_k)$ *is the conditional probability distribution of* X_t *given* $\mathcal{Z}_k$. *We have* $\mu_{t_k}^k = \mu_k$ *and* $\mu_{t_{k+1}}^k = \mu_{k+1}^-$.

Proposition 2.3 (Optimal nonlinear filter) *The sequence $\{\mu_k\,,\,k\geq 0\}$ satisfies a recurrence equation, and the iteration $\mu_k \to \mu_{k+1}$ splits in two steps : prediction, and correction.*

Prediction step : *From t_k to t_{k+1}, $\mu_t^k(dx)$ satisfies, in a weak sense, the Fokker–Planck equation*

$$\frac{\partial \mu_t^k}{\partial t} = L^* \mu_t^k \ , \tag{10}$$

where

$$L \triangleq \sum_{i=1}^m b^i \frac{\partial}{\partial x_i}$$

is the partial differential operator associated with the state equation in model (9).

Correction step : *At time t_{k+1}, the a priori information $\mu_{k+1}^-(dx)$, is combined with the new observation z_{k+1}, according to the Bayes formula*

$$\mu_{k+1}(dx) = c_{k+1}\ \Psi_{k+1}(x)\ \mu_{k+1}^-(dx)\ , \tag{11}$$

where by definition $\Psi_{k+1}(x)$ is the likelihood function for the estimation of then parameter $X_{t_{k+1}} \in \mathbb{R}^m$ given the observation z_{k+1}

$$\Psi_{k+1}(x) = \exp\left\{-\tfrac{1}{2}\|z_{k+1} - h(x)\|^2_{R^{-1}}\right\} \tag{12}$$

and c_{k+1} is a normalization constant.

Proof For $t \geq t_k$, and for any test function φ defined on $\mathbb{R}^m$, we have

$$\varphi(X_t) = \varphi(X_{t_k}) + \int_{t_k}^t L\varphi(X_s)\,ds\ ,$$

and so

$$E[\varphi(X_t)|\mathcal{Z}_k] = E[\varphi(X_{t_k})|\mathcal{Z}_k] + \int_{t_k}^t E[L\varphi(X_s)|\mathcal{Z}_k]\,ds\ ,$$

or

$$\langle \mu_t^k, \varphi\rangle = \langle \mu_k, \varphi\rangle + \int_{t_k}^t \langle \mu_s^k, L\varphi\rangle\,ds\ .$$

which proves that $\{\mu_t^k\,,\,t_k \leq t \leq t_{k+1}\}$ satisfies the Fokker–Planck equation (10) in a weak sense.

The correction step is obvious. □

2.2 Parametric estimation approach

We can reformulate the state parameter estimation for the partially observed system (9).

Using the flow of diffeomorphisms $\xi_{s,t}(\cdot)$, we get $X_{t_k} = \xi_{0,t_k}(X_0)$, for all $k \geq 1$, and the observation z_k reads

$$z_k = h(\xi_{0,t_k}(X_0)) + v_k\ . \tag{13}$$

This is a standard statistical model for the estimation of the unknown parameter X_0. Actually, we have to choose among trajectories

$$\{\xi_{0,t}(x) \, , \, t \geq 0\}$$

for different initial conditions $x \in \mathbb{R}^m$ at time 0.

Maximum likelihood estimate

The initial condition X_0 (or the state X_t at a given time $t \geq 0$) is considered as a parameter of $\mathbb{R}^m$, without a priori information. The likelihood function for the estimation of the unknown parameter X_0 in the statistical model defined in (13) above, given the observations $\{z_1, \cdots, z_k\}$ is

$$\Xi_k(x) = \exp\left\{ -\frac{1}{2} \sum_{l=1}^{k} \|z_l - h(\xi_{0,t_l}(x))\|^2_{R^{-1}} \right\} = \prod_{l=1}^{k} \Psi_l(\xi_{0,t_l}(x)) \; , \tag{14}$$

where, for all $l = 1, \cdots, k$, the function $\Psi_l(\cdot)$ is defined by (12).

The maximum likelihood estimator $\widehat{X}_0$ is given by :

$$\widehat{X}_0 \in \text{Arg} \max_{x \in \mathbb{R}^m} \Xi_k(x) \; .$$

Actually, the estimator is not only $\widehat{X}_0$, but also the trajectory which satisfies the state equation and starts from $\widehat{X}_0$ at time 0.

Bayesian estimator

In this section, we have an a priori information on the initial state X_0, represented as a probability distribution $\mu_0(dx)$ on $\mathbb{R}^m$. This a priori information can be translated, through the state equation and the associated flow of diffeomorphisms, into an a priori information on the state X_{t_k} at time t_k.

We can get an explicit expression for the conditional probability distribution $\mu_k(dx)$, using the flow of diffeomorphisms $\xi_{s,t}(\cdot)$ associated with the state equation :

Proposition 2.4 *For any Borel set $A \subset \mathbb{R}^m$, we have*

$$\mu_k(A) = c_k \int_{\xi^{-1}_{0,t_k}(A)} \Xi_k(x) \, \mu_0(dx) \; ,$$

where the normalization constant c_k is given by

$$c_k = \int_{\mathbb{R}^m} \Xi_k(x) \, \mu_0(dx) \; ,$$

and $\Xi_k(x)$ is defined by (14).

Proof First, we translate — via $\xi_{0,t_k}(\cdot)$ — the a priori information on the initial condition X_0 into an a priori information on the state X_{t_k} at time t_k.

Indeed, $X_{t_k} = \xi_{0,t_k}(X_0)$ and for any test function φ defined on $\mathbb{R}^m$

$$E[\varphi(X_{t_k})] = E[\varphi(\xi_{0,t_k}(X_0))] = \int \varphi(\xi_{0,t_k}(x))\, \mu_0(dx) \ .$$

This relation defines the probability distribution $\mu_0^k(dx) = P(X_{t_k} \in dx)$ of the state X_{t_k}, in the following way : for any test function φ defined on $\mathbb{R}^m$

$$\langle \mu_0^k, \varphi \rangle = \int \varphi(x)\, \mu_0^k(dx) = \int \varphi(\xi_{0,t_k}(x))\, \mu_0(dx) \ .$$

Actually, $\mu_0^k(dx)$ is the image of the probability distribution $\mu_0(dx)$ under the diffeomorphism $\xi_{0,t_k}(\cdot)$. Equivalently, for any Borel set $A \subset \mathbb{R}^m$, we have

$$\mu_0^k(A) = \mu_0(\xi_{0,t_k}^{-1}(A)) \ .$$

From the Bayes rule, the conditional probability distribution $\mu_k(dx)$ of the state X_{t_k}, given observations $\mathcal{Z}_k$, is given — up to a normalization factor — as the product of the a priori probability distribution $\mu_0^k(dx)$ and the likelihood function $\Xi_k(\xi_{0,t_k}^{-1}(x))$ for the estimation of the parameter X_{t_k} (or the corresponding initial state $\xi_{0,t_k}^{-1}(X_{t_k})$), that is

$$\mu_k(dx) = c_k\ \Xi_k(\xi_{0,t_k}^{-1}(x))\ \mu_0^k(dx) \ .$$

Hence, for any test function φ defined on $\mathbb{R}^m$

$$\begin{aligned} \langle \mu_k, \varphi \rangle &= \int \varphi(x)\, \mu_k(dx) = c_k \int \varphi(x)\, \Xi_k(\xi_{0,t_k}^{-1}(x))\, \mu_0^k(dx) \\ &= c_k \int \varphi(\xi_{0,t_k}(x))\, \Xi_k(x)\, \mu_0(dx) \ , \end{aligned}$$

and for any Borel set $A \subset \mathbb{R}^m$, we have the desired formula. □

2.3 A posteriori probability distribution

The computation of the a posteriori (i.e. given the observations) conditional probability distribution $\mu_k(dx)$ shows off two interesting particular cases, depending on the form of the probability distribution $\mu_0(dx)$ of the random variable X_0 :

- (i) If $\mu_0(dx)$ has a density with respect to the Lebesgue measure on $\mathbb{R}^m$, then the conditional probability distribution $\mu_k(dx)$ has also a density.
- (ii) On the other hand, if $\mu_0(dx)$ is a discrete probability distribution (i.e. a linear and convex combination of Dirac measures), then the conditional probability distribution $\mu_k(dx)$ is also discrete.

Definition 2.5 *We define the operator Q_k which relates the conditional probability distribution $\mu_k(dx)$ with the probability distribution $\mu_0(dx)$ of the random variable X_0 :*

$$Q_k\, \mu_0 \stackrel{\triangle}{=} \mu_k \ ,$$

that is

$$\langle Q_k \mu_0 \, , \, \varphi \rangle = c_k \int \varphi(\xi_{0,t_k}(x)) \, \Xi_k(x) \, \mu_0(dx) \ , \tag{15}$$

for any test function φ defined on $\mathbb{R}^m$, or equivalently

$$Q_k \mu_0(A) = c_k \int_{\xi_{0,t_k}^{-1}(A)} \Xi_k(x) \, \mu_0(dx) \ ,$$

for any Borel set $A \subset \mathbb{R}^m$.

Proposition 2.6 (μ_0 absolutely continuous) *Suppose that the initial probability distribution $\mu_0(dx)$ has a density $p_0(x)$ with respect to the Lebesgue measure, $\mu_0(dx) = p_0(x)\, dx$. Then $\mu_k(dx)$ has also a density $p_k(x)$: $\mu_k(dx) = p_k(x)\, dx$, defined by*

$$p_k(x) = c_k \, [J_k(\xi_{0,t_k}^{-1}(x))]^{-1} \, \Xi_k(\xi_{0,t_k}^{-1}(x)) \, p_0(\xi_{0,t_k}^{-1}(x)) \ ,$$

where J_k denotes the Jacobian determinant associated with the diffeomorphism $\xi_{0,t_k}(\cdot)$.

Proof By definition of the operator Q_k, we have :

$$\langle \mu_k \, , \, \varphi \rangle = \langle Q_k \mu_0 \, , \, \varphi \rangle = c_k \int \varphi(\xi_{0,t_k}(x)) \, \Xi_k(x) \, p_0(x) \, dx \ .$$

Taking $\varphi = f \circ \xi_{0,t_k}^{-1}$, where f is any test function defined on $\mathbb{R}^m$, we get :

$$\langle \mu_k \, , \, \varphi \rangle = c_k \int f(x) \, \Xi_k(x) \, p_0(x) \, dx \ . \tag{16}$$

On the other hand

$$\langle \mu_k \, , \, \varphi \rangle = \int \varphi(x) \, p_k(x) \, dx = \int f(\xi_{0,t_k}^{-1}(x)) \, p_k(x) \, dx \ .$$

The change of variable : $x' = \xi_{0,t_k}^{-1}(x)$, gives :

$$\langle \mu_k \, , \, \varphi \rangle = \int f(x') \, p_k(\xi_{0,t_k}(x')) \, J_k(x') \, dx' \ . \tag{17}$$

From (16) and (17) we deduce

$$p_k(\xi_{0,t_k}(x)) \ J_k(x) = c_k \ \Xi_k(x) \ p_0(x) \ ,$$

which gives the probability density function $p_k(x)$. □

Proposition 2.7 (μ_0 discrete) *Suppose that the initial probability distribution $\mu_0(dx)$ is a linear combination of Dirac measures :*

$$\mu_0 = \sum_{i \in I} a_0^i \; \delta_{x_0^i} \; ,$$

where $\{x_0^i \, , \, i \in I\}$ and $\{a_0^i \, , \, i \in I\}$ are respectively the positions and the weights of the particles. Then the conditional probability distribution $\mu_k(dx)$ is also a linear combination of Dirac measures :

$$\mu_k = \sum_{i \in I} a_k^i \; \delta_{x_k^i} \; ,$$

with

$$x_k^i = \xi_{0,t_k}(x_0^i) \quad \text{and} \quad a_k^i = c_k \; a_0^i \; \Xi_k(x_0^i) \; .$$

Proof By definition of the operator Q_k we have :

$$\langle \mu_k \, , \, \varphi \rangle = \langle Q_k \mu_0 \, , \, \varphi \rangle = c_k \; \sum_{i \in I} a_0^i \; \varphi(\xi_{0,t_k}(x_0^i)) \; \Xi_k(x_0^i) \; ,$$

for any test function φ defined on $\mathbb{R}^m$, i.e.

$$\langle \mu_k \, , \, \varphi \rangle = \sum_{i \in I} a_k^i \; \varphi(x_k^i) \; .$$

□

3 Particle approximation

In this section, we consider the case where the probability distribution $\mu_0(dx)$ has a density $p_0(x)$ with respect to the Lebesgue measure on $\mathbb{R}^m$.

We introduce $\mu_0^H(dx)$, an approximation of $\mu_0(dx)$, as a linear and convex combination of Dirac measures, called *particles*. The probability distribution $\mu_k^H(dx) = Q_k\ \mu_0^H(dx)$ is an approximation of the probability distribution $\mu_k(dx) = Q_k\ \mu_0(dx)$, and we study the associated approximation error. This kind of approximation was proposed by Raviart [13] for deterministic, first order PDE's.

Then, we will see that a coarse approximation, with respect to the covariance matrix R of the observation noise, can produce a non–consistent approximated Bayesian estimator.

3.1 Choice of the approximation

Let $\mu_0^H(dx)$ be the approximation of the initial probability distribution $p_0(x)\,dx$. We have :

$$p_0(x)\,dx = \mu_0(dx) \sim \mu_0^H(dx)\ , \quad \text{with} \quad \mu_0^H = \sum_{i\in I} a_0^i\ \delta_{x_0^i}\ ,$$

where $\{a_0^i\ , \ i \in I\}$ are the weights of the particles, and $\{x_0^i\ , \ i \in I\}$ are the positions of the particles.

First, we fix $\varepsilon > 0$. Then there exist a compact set $K' \subset \mathbb{R}^m$ such that

$$\mu_0(K') \geq 1 - \varepsilon\ .$$

We introduce a covering of K' consisting of bounded and convex Borel sets $\{B^i\ , \ i \in I\}$ with mutually disjoint interiors, e.g. cubes. We set $K = \bigcup_{i\in I} B^i \supset K'$. A fortiori

$$\mu_0(K) \geq \mu_0(K') \geq 1 - \varepsilon\ . \tag{18}$$

We define the weights in the following way :

$$a_0^i \stackrel{\triangle}{=} \mu_0(B^i) = \int_{B^i} p_0(x)\,dx, \quad a_0^i > 0\ . \tag{19}$$

We can always suppose that $a_0^i > 0$, otherwise we replace K by $K \setminus B^i$.

Then, we define the position the following way :

$$x_0^i \stackrel{\triangle}{=} \frac{1}{a_0^i} \int_{B^i} x\, p_0(x)\,dx\ . \tag{20}$$

Because B^i is convex, we have $x_0^i \in B^i$. For all $i \in I$, let δ_i be the diameter of the bounded subset B^i, and let H be the maximum of the diameters $\{\delta_i\ , \ i \in I\}$. In particular, for all $i \in I$ we get

$$\sup_{x\in B^i} |x - x_0^i| \leq H\ . \tag{21}$$

This approximation of $\mu_0(dx)$ has already been considered in Florchinger–Le Gland [6].

If we use $\mu_0^H(dx)$ as an approximation of the initial probability distribution $p_0(x)\,dx$, then the conditional probability distribution $p_k(x)\,dx$ is approximated by $\mu_k^H(dx) = Q_k\;\mu_0^H(dx)$, which is a linear combination of Dirac measures, and we get :

$$p_k(x)\,dx = \mu_k(dx) \sim \mu_k^H(dx)\ , \quad \text{with} \quad \mu_k^H = \sum_{i\in I} a_0^i\;\Xi_k(x_0^i)\;\delta_{x_k^i}\ ,$$

where $x_k^i = \xi_{0,t_k}(x_0^i)$.

Algorithm

We can decompose the particle approximation in two steps

Prediction step : for all $i \in I$

$$x_{k+1}^i = \Phi_\Delta(x_k^i)\ . \tag{22}$$

i.e. x_{k+1}^i is the image of x_k^i by the diffeomorphism $\Phi_\Delta(\cdot) = \xi_{t_k,t_{k+1}}(\cdot)$.

Correction step : for all $i \in I$

$$a_{k+1}^i = c_{k+1}\;\Psi_{k+1}(x_{k+1}^i)\,a_k^i\ , \tag{23}$$

where c_{k+1} is a normalization constant and

$$\Psi_{k+1}(x) \stackrel{\triangle}{=} \exp\left\{-\tfrac{1}{2}\,\|z_{k+1} - h(x)\|_{R^{-1}}^2\right\}\ , \quad x \in \mathbb{R}^d\ .$$

3.2 Error estimate

Theorem 3.1 *Suppose that $b(\cdot)$ and $h(\cdot)$ are bounded, together with their derivatives up to order 2. Then :*

$$E\|\mu_k - \mu_k^H\|_{-2,1} = E\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_k\ ,\ f\rangle - \langle \mu_k^H\ ,\ f\rangle|}{\|f\|_{2,\infty}}\right] \leq 2\,\varepsilon + H^2(C + \frac{C'}{r})\ .$$

where H is the largest of the diameters $\{\delta_i\ ,\ i\in I\}$, and r is the smallest eigenvalue of the observation noise covariance matrix R.

First we introduce some notations. We consider the following factorization :

$$\Xi_k(x) = \exp\left\{-\tfrac{1}{2}\sum_{l=1}^k \|z_l\|_{R^{-1}}^2\right\}\;\Lambda_k(x)\ .$$

for all $x \in \mathbb{R}^m$, which defines $\Lambda_k(x)$.

Let $\mathbf{P}^\dagger$ be the probability measure under which $\{z_k\ ,\ k \geq 1\}$ is an i.i.d. sequence of centered Gaussian random variables with covariance matrix R, independent of X_0.

For all $x \in \mathbb{R}^m$, we define the probability measure $\mathbf{P}_x$ equivalent to $\mathbf{P}^\dagger$, with Radon–Nikodym derivative :

$$\left.\frac{d\mathbf{P}_x}{d\mathbf{P}^\dagger}\right|_{\mathcal{Z}_k} = \Lambda_k(x)\ . \tag{24}$$

Under the probability measure $\mathbf{P}_x$:

$$z_k = h(\xi_{0,t_k}(x)) + v_k^x \ ,$$

for all $k \geq 1$, where $\{v_k^x \ , \ k \geq 1\}$ is an i.i.d. sequence of centered Gaussian random variables, with covariance matrix R, independent of X_0.

Finally, the probability measure $\mathbf{P}$ satisfies

$$\left.\frac{d\mathbf{P}}{d\mathbf{P}^\dagger}\right|_{\mathcal{Z}_k} = \int \Lambda_k(x)\mu_0(dx) = \langle \mu_0 \ , \ \Lambda_k \rangle \ . \tag{25}$$

Then, for any test function f defined on $\mathbb{R}^m$

$$\begin{aligned} \langle \mu_k \ , \ f \rangle &= \langle Q_k \mu_0 \ , \ f \rangle = \frac{\displaystyle\int f(\xi_{0,t_k}(x)) \ \Xi_k(x) \ p_0(x) \, dx}{\displaystyle\int \Xi_k(x) \ p_0(x) \, dx} \\ &= \frac{\displaystyle\int f(\xi_{0,t_k}(x)) \ \Lambda_k(x) \ p_0(x) \, dx}{\displaystyle\int \Lambda_k(x) \ p_0(x) \, dx} = \frac{\langle \mu_0 \ , \ g_k \rangle}{\langle \mu_0 \ , \ \Lambda_k \rangle} \ , \end{aligned}$$

with $g_k(x) = f(\xi_{0,t_k}(x)) \ \Lambda_k(x)$ for all $x \in \mathbb{R}^m$. Equivalently :

$$\langle \mu_k^H \ , \ f \rangle = \langle Q_k \mu_0^H \ , \ f \rangle = \frac{\langle \mu_0^H \ , \ g_k \rangle}{\langle \mu_0^H \ , \ \Lambda_k \rangle} \ ,$$

We finally notice that :

$$\begin{aligned} \langle \mu_k, f \rangle - \langle \mu_k^H, f \rangle &= \frac{\langle \mu_0, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} - \frac{\langle \mu_0^H, g_k \rangle}{\langle \mu_0^H, \Lambda_k \rangle} \\ &= \frac{\langle \mu_0, g_k \rangle - \langle \mu_0^H, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} - \frac{\langle \mu_0^H, g_k \rangle}{\langle \mu_0^H, \Lambda_k \rangle} \ \frac{\langle \mu_0, \Lambda_k \rangle - \langle \mu_0^H, \Lambda_k \rangle}{\langle \mu_0, \Lambda_k \rangle} \\ &= \mathcal{E}_k(f) - \langle \mu_k^H, f \rangle \ \mathcal{E}_k(1) \ , \end{aligned} \tag{26}$$

where, for any test function f defined on $\mathbb{R}^m$

$$\mathcal{E}_k(f) \triangleq \frac{\langle \mu_0, g_k \rangle - \langle \mu_0^H, g_k \rangle}{\langle \mu_0, \Lambda_k \rangle} \ ,$$

with $g_k(x) = f(\xi_{0,t_k}(x)) \, \Lambda_k(x)$ for all $x \in \mathbb{R}^m$. Hence, it is sufficient to estimate $\mathcal{E}_k(f)$, which is the purpose of the following Lemmas 3.3 and 3.4. But first, we make an error on the initial probability distribution, which we have to evaluate :

Lemma 3.2 *The following estimate holds :*

$$\| \mu_0 - \mu_0^H \|_{-2,1} = \sup_{f \in W^{2,\infty}} \frac{|\langle \mu_0 \ , \ f \rangle - \langle \mu_0^H \ , \ f \rangle|}{\| f \|_{2,\infty}} \leq \varepsilon + \tfrac{1}{2} H^2 \ .$$

Proof Let f be a test function defined on $\mathbb{R}^m$. Taylor expansion of f at point x_0^i reads :

$$\begin{aligned} f(x) &= f(x_0^i) + (x - x_0^i)^* f'(x_0^i) \\ &\quad + (x - x_0^i)^* \int_0^1 (1-u)\; f''[ux + (1-u)x_0^i]\, du\; (x - x_0^i) \; . \end{aligned}$$

Moreover :

$$\langle \mu_0 \, , \, f \rangle = \int_{K^c} f(x)\; p_0(x)\, dx + \sum_{i \in I} \int_{B^i} f(x)\; p_0(x)\, dx \; ,$$

and

$$\langle \mu_0^H \, , \, f \rangle = \sum_{i \in I} a_0^i \; f(x_0^i) \; .$$

The difference $\langle \mu_0 \, , \, f \rangle - \langle \mu_0^H \, , \, f \rangle$ satisfies :

$$\begin{aligned} \langle \mu_0 \, , \, f \rangle - \langle \mu_0^H \, , \, f \rangle &= \int_{K^c} f(x)\; p_0(x)\, dx + \sum_{i \in I} \int_{B^i} [f(x) - f(x_0^i)]\; p_0(x)\, dx \\ &= \int_{K^c} f(x)\; p_0(x)\, dx \\ &\quad + \sum_{i \in I} \int_{B^i} (x - x_0^i)^* \int_0^1 (1-u)\; f''[ux + (1-u)x_0^i]\, du\; (x - x_0^i)\; p_0(x)\, dx \; , \end{aligned} \tag{27}$$

since by definition (20)

$$\int_{B^i} (x - x_0^i)^* f'(x_0^i)\; p_0(x)\, dx = 0 \; , \quad i \in I \; .$$

Hence we get :

$$\begin{aligned} |\langle \mu_0 \, , \, f \rangle - \langle \mu_0^H \, , \, f \rangle| &\leq \|f\|_{\infty, K^c} \int_{K^c} p_0(x)\, dx \\ &\quad + \tfrac{1}{2} \sum_{i \in I} \|f''\|_{\infty, B^i} \int_{B^i} |x - x_0^i|^2\; p_0(x)\, dx \; . \end{aligned}$$

Moreover, we have

$$\int_{K^c} p_0(x)\, dx = \mu_0(K^c) \leq \varepsilon \; , \tag{28}$$

and

$$\int_{B^i} |x - x_0^i|^2\; p_0(x)\, dx \leq H^2\; a_0^i \; ,$$

according to (18), (19) and (21). Hence :

$$|\langle \mu_0 \, , \, f \rangle - \langle \mu_0^H \, , \, f \rangle| \leq \varepsilon\; \|f\|_{\infty, K^c} + \tfrac{1}{2} H^2\; \|f''\|_{\infty, K} \leq (\varepsilon + \tfrac{1}{2} H^2)\; \|f\|_{2,\infty} \; .$$

□

Lemma 3.3 *The following inequality holds :*

$$E \left[\sup_{f \in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} \right] \leq \varepsilon + \tfrac{1}{2} \sum_{i \in I} \sup_{x \in B^i} E_x \left[\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \right] \; \delta_i^2 \; a_0^i \; ,$$

with $g_k(x) = f(\xi_{0,t_k}(x))\; \Lambda_k(x)$ *and* $g_k''(x) = d_k(x)\; \Lambda_k(x)$ *for all* $x \in \mathbb{R}^m$.

Proof First, according to (25)

$$E\left[\sup_{f\in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}}\right] = E\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}\,\langle \mu_0\,,\,\Lambda_k\rangle}\right] = E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] .$$

Moreover, from (27) we get

$$\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle = \int_{K^c} g_k(x)\; p_0(x)\, dx + \sum_{i\in I} \int_{B^i} (x - x_0^i)^* \int_0^1 (1-u)\; g_k''[ux + (1-u)x_0^i]\, du\; (x - x_0^i)\; p_0(x)\, dx \; .$$

The following inequality comes up :

$$\begin{aligned} |\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle| &\leq \int_{K^c} |g_k(x)|\; p_0(x)\, dx \\ &\quad + \sum_{i\in I} \int_{B^i} |(x - x_0^i) \int_0^1 (1-u)\; g_k''[ux + (1-u)x_0^i]\, du\; (x - x_0^i)|\; p_0(x)\, dx \\ &\leq \|f\|_\infty \int_{K^c} \Lambda_k(x)\; p_0(x)\, dx \\ &\quad + \sum_{i\in I} \delta_i^2 \int_{B^i} \int_0^1 (1-u)\; |d_k[ux + (1-u)x_0^i]|\; \Lambda_k[ux + (1-u)x_0^i]\, du\, p_0(x)\, dx \; , \end{aligned}$$

where δ_i is the diameter of the subset B^i. From this estimate, we deduce

$$\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}} \leq \int_{K^c} \Lambda_k(x)\; p_0(x)\, dx + \sum_{i\in I} \delta_i^2 \int_{B^i} \int_0^1 (1-u) \sup_{f\in W^{2,\infty}} \frac{|d_k[ux + (1-u)x_0^i]|}{\|f\|_{2,\infty}}\; \Lambda_k[ux + (1-u)x_0^i]\, du\, p_0(x)\, dx \; ,$$

hence

$$E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] \leq \sup_{x\in K^c} E^\dagger[\Lambda_k(x)] \int_{K^c} p_0(x)\, dx + \tfrac{1}{2} \sum_{i\in I} \delta_i^2 \sup_{x\in B^i} E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \Lambda_k(x)\right] \int_{B^i} p_0(x)\, dx \; .$$

Notice that, for all $x \in \mathbb{R}^m$, $E^\dagger[\Lambda_k(x)] = 1$ and

$$E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \Lambda_k(x)\right] = E_x\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}}\right] \; ,$$

according to (24). So, from (19) and (28), we get

$$E^\dagger\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_0\,,\, g_k\rangle - \langle \mu_0^H\,,\, g_k\rangle|}{\|f\|_{2,\infty}}\right] \leq \varepsilon + \tfrac{1}{2} \sum_{i\in I} \sup_{x\in B^i} E_x\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}}\right] \delta_i^2\, a_0^i \; ,$$

which completes the proof of the lemma, according to (29). □

Lemma 3.4 *If $f(\cdot)$, $b(\cdot)$ and $h(\cdot)$ are bounded, together with their derivatives up to order 2, then there exist $C > 0$ and $C' > 0$ such that :*

$$E_x \left[\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \right] \leq C + \frac{C'}{r}$$

where $g_k(x) = f(\xi_{0,t_k}(x)) \Lambda_k(x)$, and $g''_k(x) = d_k(x) \Lambda_k(x)$ for all $x \in \mathbb{R}^m$, and r is the smallest eigenvalue of the observation noise covariance matrix R.

Proof An explicit computation of the second derivative of $g_k(x)$, gives the following expression for the function $d_k(x)$:

$$\begin{aligned} d_k(x) \;&=\; f''(\xi_{0,t_k}(x))\,(\xi'_{0,t_k}(x))^2 + f'(\xi_{0,t_k}(x))\,\xi''_{0,t_k}(x) \\ &\quad + 2 f'(\xi_{0,t_k}(x))\,\xi'_{0,t_k}(x)\,(\log \Lambda_k)'(x) + f(\xi_{0,t_k}(x))\,(\log \Lambda_k)''(x) \\ &\quad + f(\xi_{0,t_k}(x))\,[(\log \Lambda_k)'(x)]^2 \;, \end{aligned}$$

where we suppose for simplicity that $m = 1$.

Since the functions $f(\cdot)$ and $b(\cdot)$ are bounded together with their derivatives up to order 2, there exists $C > 0$ such that :

$$\sup_{f \in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}} \leq C [1 + |(\log \Lambda_k)'(x)|^2 + |(\log \Lambda_k)''(x)|] \;.$$

Moreover

$$\begin{aligned} \log \Lambda_k(x) \;&=\; \sum_{l=1}^k z_l^*\, R^{-1}\, h(\xi_{0,t_l}(x)) - \tfrac{1}{2} \sum_{l=1}^k \|h(\xi_{0,t_l}(x))\|^2_{R^{-1}} \;, \\ (\log \Lambda_k)'(x) \;&=\; \sum_{l=1}^k [z_l - h(\xi_{0,t_l}(x))]^*\, R^{-1}\, h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x) \\ &=\; \sum_{l=1}^k [R^{-1/2} v_l^x]^*\, R^{-1/2}\, h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x) \;, \end{aligned}$$

and

$$\begin{aligned} (\log \Lambda_k)''(x) &= \sum_{l=1}^k [z_l - h(\xi_{0,t_l}(x))]^*\, R^{-1}\, [h''(\xi_{0,t_l}(x))\, (\xi'_{0,t_l}(x))^2 \\ &\quad + h'(\xi_{0,t_l}(x))\, \xi''_{0,t_l}(x)] - \sum_{l=1}^k [h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x)]^*\, R^{-1}\, h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x) \\ &= \sum_{l=1}^k [R^{-1/2} v_l^x]^*\, R^{-1/2}\, [h''(\xi_{0,t_l}(x))\, (\xi'_{0,t_l}(x))^2 + h'(\xi_{0,t_l}(x))\, \xi''_{0,t_l}(x)] \\ &\quad - \sum_{l=1}^k [h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x)]^*\, R^{-1}\, h'(\xi_{0,t_l}(x))\, \xi'_{0,t_l}(x) \end{aligned}$$

So we get :

$$E_x|(\log \Lambda_k)'(x)|^2 \le \frac{C'}{r}$$

and

$$E_x|(\log \Lambda_k)''(x)| \le C + \frac{C'}{r}$$

which leads to

$$E_x\left[\sup_{f\in W^{2,\infty}} \frac{|d_k(x)|}{\|f\|_{2,\infty}}\right] \le C + \frac{C'}{r} \; .$$

□

Proof of Theorem 3.1 According to (26), we have

$$\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_k \, , \, f\rangle - \langle \mu_k^H \, , \, f\rangle|}{\|f\|_{2,\infty}} \le \sup_{f\in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} + |\mathcal{E}_k(1)| \le 2 \sup_{f\in W^{2,\infty}} \frac{|\mathcal{E}_k(f)|}{\|f\|_{2,\infty}} \; .$$

From estimates proved in Lemmas 3.3 and 3.4, we get

$$E\left[\sup_{f\in W^{2,\infty}} \frac{|\langle \mu_k \, , \, f\rangle - \langle \mu_k^H \, , \, f\rangle|}{\|f\|_{2,\infty}}\right] \le 2\varepsilon + H^2(C + \frac{C'}{r}) \; ,$$

which proves the theorem. □

3.3 Consistency

The above error estimate shows that it is not sufficient for the discretization step H to be small. It is also necessary for H^2 to be small compared with the smallest eigenvalue r of the observation noise covariance matrix R. The aim of this section is to prove that, if this is not the case, the particle approximation can produce a coarse, but also non–consistent, estimator. To get a better view of this situation, we consider the case where $R = rI$ and r tends to 0. When $r > 0$, the likelihood function $\Xi_k(x)$ for the estimation of the initial value X_0 satisfies

$$-r \log \Xi_k(x) = \tfrac{1}{2} \sum_{l=1}^{k} \|z_l - h(\xi_{0,t_l}(x))\|^2 \; .$$

The maximum likelihood estimator $\widehat{X}_0$ is given by

$$\widehat{X}_0 \in \operatorname{Arg} \max_{x\in \mathbb{R}^m} \Xi_k(x) \; .$$

When $r \downarrow 0$, we get the following limiting expression (Kullback–Leibler information) :

$$-r \log \Xi_k(x) \longrightarrow K(x, x_0) = \tfrac{1}{2} \sum_{l=1}^{k} \|h(\xi_{0,t_l}(x_0)) - h(\xi_{0,t_l}(x))\|^2 \; ,$$

where $x_0 \in \mathbb{R}^m$ denotes the true value of the initial condition.

We introduce the set

$$\begin{aligned} M(x_0) &= \operatorname{Arg} \min_{x \in \mathbb{R}^m} K(x, x_0) \\ &= \{x \in \mathbb{R}^m : h(\xi_{0,t_l}(x)) = h(\xi_{0,t_l}(x_0)) \ , \quad \text{for all } l = 1, \cdots, k\} \end{aligned}$$

$M(x_0)$ is the set of initial values which, in the limiting deterministic system, cannot be distinguished from the true value x_0. Obviously, $x_0 \in M(x_0)$, but the system may be not identifiable, so that $M(x_0) \neq \{x_0\}$. An example of such a system is presented in Lévine–Marino [11], where a target with constant speed is tracked with angle measurements only. In this example, the set $M(x_0)$ is a one dimensional submanifold.

We have the following consistency result for the maximum likelihood estimator :

$$d(\widehat{X}_0 \, , \, M(x_0)) \to 0 \ , \quad \text{with probability one, as } r \downarrow 0 \ .$$

The particle approximation described above consists in restricting the parameter set to a finite set $G_H = \{x_0^i \, , \, i \in I\} \subset \mathbb{R}^m$ of possible initial values, and nothing can insure that the true value x_0 belongs to G_H : this is a mis–specified statistical model, see for example McKeague [12]. The maximum likelihood estimator $\widehat{X}_0^H$ is given by

$$\widehat{X}_0^H \in \operatorname{Arg} \max_{x \in G_H} \Xi_k(x) \ .$$

We define :

$$M_H(x_0) \stackrel{\triangle}{=} \operatorname{Arg} \min_{x \in G_H} K(x, x_0) \ .$$

and we get the following consistency result for the maximum likelihood estimator :

$$d(\widehat{X}_0^H \, , \, M_H(x_0)) \to 0 \ , \quad \text{with probability one, as } r \downarrow 0 \ .$$

Usually $x_0 \notin M_H(x_0)$ except if $x_0 \in G_H$. It can happen that $d(x_0, M_H(x_0))$ is large, in particular when the system is not identifiable. This phenomenon is illustrated in Figure 1, where the set $M_H(x_0)$ reduce to the single point $\widehat{X}_0^H \neq x_0$, while the set $M(x_0)$ corresponds to the continuous curve crossing the true value x_0.