Particle-based methods for parameter estimation and tracking: Numerical experiments

Johann Fichou, François Le Gland, Laurent Mevel

I R I S A

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

PUBLICATION INTERNE
N° 1604

PARTICLE–BASED METHODS FOR PARAMETER ESTIMATION AND TRACKING : NUMERICAL EXPERIMENTS

JOHANN FICHOU, FRANÇOIS LE GLAND, LAURENT MEVEL

ISSN 1166-8687

IRISA
CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

Particle–based Methods for Parameter Estimation and Tracking : Numerical Experiments

Johann Fichou* , François Le Gland* , Laurent Mevel*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SIGMA2 (σ^2)

Publication interne n1604 — Février 2004 — 59 pages

Abstract: The purpose of this work is to obtain as much intuition as possible, through numerical experiments in a simple case where exact solutions are explicitly available, about the particle approximation of finite signed measures. A prototypical example of a finite signed measure is the derivative, w.r.t. a parameter of the model, of some probability distributions related with a hidden Markov chain. This includes prior, prediction, filtering probability distributions, etc. Two points of view are considered here, to feel the quality of the approximation, at least in a qualitative manner :

(i) how accurate is the particle approximation of the finite signed measure, in view of an histogram representation of the weighted particle system ?

(ii) considering the log–likelihood function and the score function, how close is the approximate expression provided by the particle approximation to the exact expression ?

These two questions seem closely related, however the numerical experiments presented in this work show that one of the two particle approximation schemes fails to satisfy the first criteria (quality of the approximation of the finite signed measure), and that both schemes satisfy the second criteria (quality of the approximation of the statistics).

Key-words: monitoring, mechanical system, modal parameter, recursive maximum likelihood, estimation, tracking, particle filter, linear tangent particle filter.

(Résumé : tsvp)

Ce travail correspond au rapport final d'un stage effectué dans le cadre du DESS *Mathématiques Appliqués* de l'université de Rennes 1, dans la filière *Statistique*. Il a été soutenu en partie par le projet Eurêka 2419 *Flight Test Easy* (FLITE).

* IRISA / INRIA, Campus de Beaulieu, 35042 RENNES Cédex, France. — {legland,lmevel}@irisa.fr

Centre National de la Recherche Scientifique
(UPRESSA 6074) Université de Rennes 1 – Insa de Rennes

Institut National de Recherche en Informatique
et en Automatique – unité de recherche de Rennes

Méthodes particulaires
pour l'estimation et la poursuite de paramètres :
Expériences numériques

Résumé : L'objectif de ce travail est de mieux comprendre l'approximation particulaire de mesures signées finies, au travers de quelques expériences numériques menées dans un cas simple où les solutions exactes sont connues de manière explicite. Un exemple typique de mesure signée finie est la dérivée, par rapport à un paramètre du modèle, de distributions de probabilité associées à une chaîne de Markov cachée. Cela inclut la distribution a priori, le prédicteur, le filtre, etc. Deux points de vue sont considérés ici pour évaluer la qualité de l'approximation, au moins dans un sens qualitatif :

- quelle est la précision de l'approximation particulaire de la mesure signée finie, au vu d'une représentation sous forme d'histogramme du système de particules pondérées ?
- si on s'intéresse seulement à la fonction de log–vraisemblance ou à la fonction score, quel est l'écart entre l'expression fournie par l'approximation particulaire et l'expression exacte de ces quantités ?

Ces deux questions sont évidemment liées, mais les expériences numériques présentées dans ce travail montrent que l'un des deux schémas d'approximation particulaire proposés ne répond pas de manière satisfaisante au premier critère (qualité de l'approximation de la mesure signée finie), et que les deux schémas proposés donnent une bonne approximation pour le second critère (précison de l'approximation des statistiques).

Mots clés : surveillance, système mécanique, paramètre modal, maximum de vraisemblance récursif, estimation, poursuite, filtre particulaire, filtre particulaire linéaire tangent.

1 Introduction

The purpose of this work is to obtain as much intuition as possible, through numerical experiments in a simple case where exact solutions are explicitly available, about the particle approximation of finite signed measures. A prototypical example of a finite signed measure is the derivative, w.r.t. a parameter of the model, of some probability distributions related with a hidden Markov chain. This includes prior, prediction, filtering probability distributions, etc. Two points of view are considered here, to feel the quality of the approximation, at least in a qualitative manner :

(i) how accurate is the particle approximation of the finite signed measure, in view of an histogram representation of the weighted particle system ?

(ii) considering the log–likelihood function and the score function, how close is the approximate expression provided by the particle approximation to the exact expression ?

These two questions seem closely related, however the numerical experiments presented in this work show that one of the two particle approximation schemes fails to satisfy the first criteria (quality of the approximation of the finite signed measure), and that both schemes satisfy the second criteria (quality of the approximation of the statistics).

The initial motivation for this work was provided by an application to monitoring the integrity of structural and mechanical systems. Detecting and localizing damages for monitoring the integrity of structural and mechanical systems is a topic of growing interest, due to the aging of many engineering constructions and machines and to increased safety norms. Automatic global vibration–based monitoring techniques turn out to be useful alternatives to visual inspections or local non destructive (e.g. ultrasonic) evaluations performed manually.

Health monitoring techniques based on processing vibration measurements basically handle two types of characteristics: the *structural parameters* (mass, stiffness, flexibility, damping) and the *modal parameters* (modal frequencies, and associated damping values and mode–shapes), see [22, 9, 21]. A central question for monitoring is to compute *changes* in those characteristics and to assess their *significance*. For the *frequencies*, crucial issues are then: how to compute the changes, to assess that the changes are significant, to handle *correlations* among individual changes. A related issue is how to compare the changes in the frequencies obtained from experimental data with the sensitivity of modal parameters obtained from an analytical model. Furthermore, it has been widely acknowledged that changes in frequencies bear useful information for damage *detection*.

Our contribution in this work is to design a particle filtering method to track the modal parameters. Particle filtering techniques are a set of powerful and versatile simulation–based methods to perform optimal state estimation in nonlinear non–Gaussian state–space models, and we consider here an approach combining particle filtering and gradient algorithm to perform recursive maximum likelihood parameter estimation and tracking. In the next section, the modeling issues are introduced and some key parameterizations are discussed. Section 3 details the particle approximation mechanisms. Section 4 is devoted to particle approximation of finite signed measures, with numerical experiments. In Section 5, the particle implementation of the recursive maximum likelihood (RML) algorithm is described.

2 Mechanical model

2.1 Dynamical model and structural parameters

It is assumed that the behavior of the mechanical system can be described by a stationary linear dynamical system, and that, in the frequency range of interest, the input forces can be modeled as a non-stationary white noise. This results in :

$$\left\{ \begin{array}{rcl} M\ddot{\mathcal{Z}}(t) + C\dot{\mathcal{Z}}(t) + K\mathcal{Z}(t) &=& \nu(t) \\ Y(t) &=& L\mathcal{Z}(t) \end{array} \right. \tag{1}$$

where t denotes continuous time, M, C and K are the mass, damping and stiffness matrices respectively, the (high dimensional) vector $\mathcal{Z}$ collects the displacements of the degrees of freedom of the structure, the external (non measured) force ν is modeled as a non–stationary white noise with time–varying covariance matrix $Q_\nu(t)$, measurements are collected in the (often, low dimensional) vector Y, and the matrix L indicates which components of the state vector are actually measured, i.e. where the sensors are located.

The modes or eigenfrequencies, denoted generically by μ, are solutions of :

$$\det(\mu^2 \, M + \mu \, C + K) = 0 \; . \tag{2}$$

The frequency f and the damping coefficient d are recovered from a *continuous* eigenvalue μ through :

$$f = \frac{\Im(\mu)}{2\,\pi} \qquad \text{and} \qquad d = -\Re(\mu) \; . \tag{3}$$

Some comments are in order on parameterizations of interest for damage detection and localization. Since a local damage in the structure reduces the stiffness and increases the damping, many damage detection techniques have been proposed which monitor the stiffness matrix K. Monitoring its inverse K^{-1}, namely the flexibility matrix, has proven more tractable and computationally feasible [22, 13, 5]. In some cases, other structural parameterizations such as volumic mass and Young elasticity modulus may be preferable [14, 22]. Also, several methods in the literature are based on a transmissibility matrix [21, 23], which involves the processing of input–output data. However, in the case of non measured input excitation, processing output–only data is mandatory [17, 3]. On the other hand, a reduced stiffness and an increased damping result in decreased natural frequencies. Thus, monitoring the modal parameters is relevant.

2.2 State–space model and parameterization

Sampling model (1) at rate $1/\Delta$ yields the discrete time model in state space form [12, 18] :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & F \, X_k + W_k \\ Y_k & = & H \, X_k \end{array} \right. \tag{4}$$

where the state and the output are :

$$X_k = \begin{pmatrix} \mathcal{Z}(k\,\Delta) \\ \dot{\mathcal{Z}}(k\,\Delta) \end{pmatrix} \qquad \text{and} \qquad Y_k = Y(k\,\Delta) \; , \tag{5}$$

the state transition and observation matrices are :

$$F = \exp(A\,\Delta) \qquad \text{where} \qquad A = \begin{pmatrix} 0 & I \\ -M^{-1}\,K & -M^{-1}\,C \end{pmatrix}$$

and

$$H = \begin{pmatrix} L & 0 \end{pmatrix} \; . \tag{6}$$

The measurement equation in (4) with H as in (6) implicitly assumes that the available sensors measure the (relative) displacements of the degrees of freedom themselves. The nature of the sensors used only influences the observation matrix H. In (4), the *unmeasured* state noise $\{W_k \, , \, k \geq 0\}$ is assumed to be a zero–mean Gaussian white noise, with covariance matrix :

$$Q_k \stackrel{\triangle}{=} \mathbb{E}[W_k \; W_k^*] = \int_{k\,\Delta}^{(k+1)\,\Delta} \exp(A\,s) \; \widetilde{Q}(s) \; \exp(A^*\,s) \, ds$$

and

$$\widetilde{Q}(s) = \begin{pmatrix} 0 & 0 \\ 0 & M^{-1} \, Q_\nu(s) \, M^{-*} \end{pmatrix} \; .$$

The state X and the observed output Y have dimensions $2m$ and r respectively, with r (often much) smaller than $2m$ in practice.

Let λ be the modes of the state transition matrix F, namely :

$$\det(F - \lambda I) = 0 \ . \tag{7}$$

The continuous modes μ in (2) can be deduced from the discrete modes λ in (7) using :

$$\exp(\Delta\,\mu) = \lambda \ .$$

The frequency f and the damping coefficient d are recovered from a discrete eigenvalue λ through :

$$f = \frac{\arg(\lambda)}{2\,\pi\,\Delta} \qquad \text{and} \qquad d = -\frac{1}{\Delta}\,\log|\lambda| \ . \tag{8}$$

Because of the structure of the state in (5), the discrete modes λ are pairwise complex conjugate. If F and H are unknown matrices but we are able to track the coefficients of these two matrices, then it is easy to obtain the frequency and the damping coefficient using (8).

Remark 2.1. If we suppose noisy measurements, the model becomes :

$$\begin{cases} X_{k+1} &= F\,X_k + W_k \\ Y_k &= H\,X_k + V_k \end{cases} \tag{9}$$

where $\{V_k\,,\,k \geq 0\}$ is an unmeasured Gaussian white noise with zero mean. It is essential to note that, with this assumption, the measurement noise does not affect the eigenstructure of (9).

3 About particle approximation

In this section, we will first detail particle approximation of the filter and the derivative of the filter w.r.t. the parameter of interest, following [6, 16]. Then, we will see different resampling schemes and their efficiency.

3.1 Hidden Markov model

The state sequence $\{X_k\,,\,k \geq 0\}$ is a Markov chain taking values in the space $E = \mathbb{R}^d$, with transition kernel $Q(x, dx')$ (which is assumed time independent for simplicity), i.e.

$$\mathbb{P}[X_{k+1} \in dx' \mid X_k = x] = Q(x, dx') \ .$$

The kernel $Q(x, dx')$ could depend on a parameter, that should be either estimated, or monitored (i.e. changes w.r.t. a nominal value should be detected), however the dependence w.r.t. the parameter is not written explicitly, so as to avoid intricated notations. The following assumption is made

> **It is *easy* to *simulate* a r.v. X with probability distribution $Q(x, dx')$, even though the analytical expression of the kernel $Q(x, dx')$ is not known, or is so complicated that it is *pratically impossible* to *compute* such integrals as**

$$Q\,\phi(x) = \int_E Q(x, dx')\,\phi(x') \qquad \textbf{or} \qquad Q\,\mu(dx') = \int_E \mu(dx)\,Q(x, dx') \ .$$

The state sequence $\{X_k\,,\,k \geq 0\}$ is not observed, but instead an observation sequence $\{Y_k\,,\,k \geq 0\}$ is available, which has the following property : given the hidden states $\{X_k\,,\,k \geq 0\}$, the observations $\{Y_k\,,\,k \geq 0\}$ are mutually independent, and the conditionnal probability distribution of Y_k (which is assumed time independent for simplicity) depends only on the hidden state X_k at the same time instant, and by definition

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = g(x, y)\,\lambda(dy) \qquad \text{and} \qquad \Psi_k(x) = g(x, Y_k) \ .$$

Notice that when x varies, all the conditionnal probability distributions $\mathbb{P}[Y_k \in dy \mid X_k = x]$ are assumed absolutely continuous w.r.t. a nonnegative measure $\lambda(dy)$ which does not depend on x (with densities $g(x, y)$ which do depend on x).

Example This *memoryless channel* assumption is satisfied for instance in the case where the hidden state is observed in an additive white noise sequence, not necessarily Gaussian, i.e. in our model where the observation Y_k is related to the hidden state X_k by the relation

$$Y_k = H\,X_k + V_k \ ,$$

where $\{V_k\,,\,k \geq 0\}$ is a white noise sequence (i.e. a sequence of mutually independent r.v.'s) with probability distribution $q(v)\,dv$ (which is assumed time independent for simplicity), independent of $\{X_k\,,\,k \geq 0\}$. In this case

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q(y - H\,x)\,dy \qquad \text{and} \qquad \Psi_k(x) = q(Y_k - H\,x) \ .$$

3.2 Particle approximation of the filter

Given observations, the objective is to estimate the hidden states, and to this effect the probability distributions

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_0, \cdots, Y_k] \qquad \text{and} \qquad \mu_{k|k-1}(dx) = \mathbb{P}[X_k \in dx \mid Y_0, \cdots, Y_{k-1}] \ ,$$

are introduced. The evolution of the sequence $\{\mu_k\,,\,k \geq 0\}$ taking values in the space of probability distributions on E, is very easily described by the following steps

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = Q\,\mu_{k-1} \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1} \ ,$$

where

$$\mu_{k|k-1}(dx') = Q\,\mu_{k-1}(dx') = \int_E \mu_{k-1}(dx)\,Q(x,dx') \ ,$$

can happen to be *difficult* (if not just impossible) to *compute*, and where $\cdot$ denotes the projective product, i.e.

$$\mu_k(dx) = \Psi_k \cdot \mu_{k|k-1}(dx) = \frac{\Psi_k(x)\,\mu_{k|k-1}(dx)}{\langle \mu_{k|k-1}, \Psi_k \rangle} \ .$$

In view of the key assumption that it is on the other hand *easy* to *simulate* r.v.'s with probability distribution $Q(x,dx')$, the idea is to approximate the predictor $\mu_{k|k-1}$ with the empirical probability distribution associated with an N–sample, i.e.

$$\mu_{k|k-1} \approx \mu^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k|k-1}} \ .$$

This approximation is completely characterized by the set $\{\xi^i_{k|k-1}\,,\,i = 1, \cdots, N\}$ of particles, and the algorithm is completely described by the mechanism which builds $\{\xi^i_{k+1|k}\,,\,i = 1, \cdots, N\}$ from $\{\xi^i_{k|k-1}\,,\,i = 1, \cdots, N\}$. This mechanism is as follows :

(i) the correction step is applied *exactly* to $\mu^N_{k|k-1}$, which results in

$$\mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} = \sum_{i=1}^N \frac{\Psi_k(\xi^i_{k|k-1})\,\delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} = \sum_{i=1}^N \omega^i_k\,\delta_{\xi^i_{k|k-1}} \ ,$$

i.e. particles $\{\xi^i_{k|k-1}\,,\,i = 1, \cdots, N\}$ are now weighted, with weights $\{\omega^i_k\,,\,i = 1, \cdots, N\}$ which are more heavy for those particles which are more *consistent* with the current observation Y_k,

(ii) instead of trying to *compute* $Q\,\mu_k^N$, the following particle approximation

$$\mu_{k+1|k}^N = S^N(Q\,\mu_k^N) = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{k+1|k}^i} \;,$$

is used, where the N–sample $\{\xi_{k+1|k}^i\,,\; i=1,\cdots,N\}$ has precisely the probability distribution $Q\,\mu_k^N$, i.e. N independent r.v.'s are *simulated* with common probability distribution $Q\,\mu_k^N$, which can be achieved in the following manner : independently for any $i=1,\cdots,N$

$$\xi_k^i \sim \mu_k^N(dx) \quad \Longleftarrow \quad \textbf{easy, since the probability distribution } \mu_k^N \textbf{ is discrete,}$$

$$\xi_{k+1|k}^i \sim Q(\xi_k^i, dx') \Longleftarrow \textbf{easy, by assumption.}$$

3.3 Linear tangent kernel / extended kernel, etc.

If the transition kernel $Q(x,dx')$ depends on a parameter, then the filter μ_k depends also on the parameter, and one would like to compute the linear tangent filter w_k, i.e. the derivative of the filter μ_k w.r.t. the parameter. To this end, one needs first to study the linear tangent kernel $\Gamma(x,dx')$, i.e. the derivative of the transition kernel $Q(x,dx')$ w.r.t. the parameter, and the following assumption is made

Assumption AC : The following probabilistic representation holds for the linear tangent kernel $\Gamma(x,dx')$

$$\Gamma\,\phi(x) = \int_E \Gamma(x,dx')\,\phi(x') = \mathbb{E}[\phi(X_{k+1})\,\Xi_{k+1} \mid X_k = x] \;,$$

where $\{\,(X_k,\Xi_k)\,,\; k\geq 0\}$ is a Markov chain taking values in the product space $E\times F$, such that

$$\mathbb{P}[X_{k+1}\in dx', \Xi_{k+1}\in ds' \mid X_k = x, \Xi_k = s]$$

$$= \mathbb{P}[X_{k+1}\in dx', \Xi_{k+1}\in ds' \mid X_k = x] = K(x,dx',ds') \;.$$

The following assumption, which extends the similar assumption introduced in Section 3.1, is made

It is *easy* to *simulate* a r.v. (X,Ξ) with probability distribution $K(x,dx',ds')$, even though the analytical expression of the kernel $K(x,dx',ds')$ is not known, or is so complicated that it is *pratically impossible* to *compute* such integrals as

$$\Gamma\,\phi(x) = \int_{E\times F} s'\,\phi(x')\,K(x,dx',ds') \qquad \textbf{or} \qquad \Gamma\,\mu(dx') = \int_{E\times F} \mu(dx)\,s'\,K(x,dx',ds') \;.$$

Example In our model, the Markov chain $\{X_k\,,\; k\geq 0\}$ taking values in $E=\mathbb{R}^d$, is defined by

$$X_{k+1} = F\,X_k + W_k \;,$$

where only the matrix F depends on the parameter, and where $\{W_k\,,\; k\geq 0\}$ is a sequence of independent r.v.'s taking values in $\mathbb{R}^d$ with probability distribution $p(w)\,dw$ (coefficients are assumed time independent for simplicity). For any $x\in\mathbb{R}^d$, the transition kernel $Q(x,dx')$ is given by

$$Q(x,dx') = p(x'-F\,x)\,dx' \;,$$

and one can show directly that

$$\Gamma(x,dx') = -p'(x'-F\,x)\,\frac{\partial F}{\partial\theta}\,x\,dx' = \frac{-p'}{p}(x'-F\,x)\,\frac{\partial F}{\partial\theta}\,x\,Q(x,dx') \;.$$

It follows that

$$\begin{aligned}\Gamma\,\phi(x) = \int_E \Gamma(x,dx')\,\phi(x') &= \int_E \phi(x')\,\frac{-p'}{p}(x' - F\,x)\,\frac{\partial F}{\partial \theta}\,x\,Q(x,dx') \\ &= \mathbb{E}[\phi(X_{k+1})\,\frac{-p'}{p}(X_{k+1} - F\,X_k)\,\frac{\partial F}{\partial \theta}\,X_k \mid X_k = x]\ ,\end{aligned}$$

i.e. Assumption AC is satisfied, with

$$\Xi_{k+1} = \frac{-p'}{p}(W_k)\,\frac{\partial F}{\partial \theta}\,X_k\ .$$

Notice that in the above example, the r.v. Ξ_{k+1} depends only on (X_k, W_k), in which case it does not seem necessary to simulate Ξ_{k+1} in addition to W_k. This apparently very particular situation is actually very general, as the following result shows.

Lemma 3.1. *Under Assumption AC*

$$\Gamma(x,dx') = I(x,x')\,Q(x,dx')\ ,$$

with

$$I(x,x') = \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x']\ ,$$

for any $x, x' \in E$.

Proof. For any probability distribution μ on E, and any pair B, B' of Borel subsets of E

$$\mathbb{E}_\mu[\Xi_{k+1}\,\mathbf{1}_{(X_{k+1} \in B',\, X_k \in B)}] = \int_{B \times B'} \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x']\,\mu(dx)\,Q(x,dx')\ ,$$

and

$$\begin{aligned}\mathbb{E}_\mu[\Xi_{k+1}\,\mathbf{1}_{(X_{k+1} \in B')}\,\mathbf{1}_{(X_k \in B)}] &= \int_B \mathbb{E}[\Xi_{k+1}\,\mathbf{1}_{(X_{k+1} \in B')} \mid X_k = x]\,\mu(dx) \\ &= \int_B \Gamma(x,B')\,\mu(dx)\ ,\end{aligned}$$

hence taking $B = E$ yields

$$\int_E \Gamma(x,B')\,\mu(dx) = \int_E \{\int_{B'} \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x']\,Q(x,dx')\}\,\mu(dx)\ ,$$

and since the probability distribution μ is arbitrary, it holds

$$\Gamma(x,B') = \int_{B'} \mathbb{E}[\Xi_{k+1} \mid X_k = x, X_{k+1} = x']\,Q(x,dx')\ ,$$

which proves the result. □

By definition

$$\Gamma\,\phi(x) = \int_{E \times F} \phi(x')\,s'\,K(x,dx',ds') \qquad \text{and} \qquad Q\,\phi(x) = \int_{E \times F} \phi(x')\,K(x,dx',ds')\ ,$$

hence

$$\Gamma(x,dx') = \int_F s'\,K(x,dx',ds') \qquad \text{and} \qquad Q(x,dx') = \int_F K(x,dx',ds')\ .$$

On the product space $E \times E \times F$, define the projection $\pi_0 \;:(x, x', s') \longmapsto x$ on the (first) space E, the projection $\pi \;:(x, x', s') \longmapsto x'$ on the (second) space E and the projection $\pi_F \;:(x, x', s') \longmapsto s'$ on the auxiliary space F. For any probability distribution μ on the space E, the probability distribution $\mu \otimes K$ is defined on the product space $E \times E \times F$ by

$$(\mu \otimes K)(dx, dx', ds') = \mu(dx)\; K(x, dx', ds') \; .$$

It follows that

$$Q\,\mu(dx') = \int_{E \times F} \mu(dx)\, K(x, dx', ds') = (\mu \otimes K) \circ \pi^{-1}(dx') \; ,$$

and

$$\begin{aligned} \Gamma\,\mu(dx') \;&=\; \int_{E \times F} \mu(dx)\, s'\, K(x, dx', ds') \\ &=\; \int_{E \times F} \pi_F(x, x', s')\, (\mu \otimes K)(dx, dx', ds') = (\pi_F\, (\mu \otimes K)) \circ \pi^{-1}(dx') \; , \end{aligned}$$

and if the finite signed measure w is absolutely continuous w.r.t. μ, then

$$\begin{aligned} Q\, w(dx') \;&=\; \int_{E \times F} w(dx)\, K(x, dx', ds') \\ &=\; \int_{E \times F} \frac{dw}{d\mu}(x)\, \mu(dx)\, K(x, dx', ds') \\ &=\; \int_{E \times F} (\frac{dw}{d\mu} \circ \pi_0)(x, x', s')\, (\mu \otimes K)(dx, dx', ds') = ((\frac{dw}{d\mu} \circ \pi_0)\; (\mu \otimes K)) \circ \pi^{-1}(dx') \; , \end{aligned}$$

i.e.

$$Q\,\mu = (\mu \otimes K) \circ \pi^{-1} \qquad \text{and} \qquad \Gamma\,\mu = (\pi_F\, (\mu \otimes K)) \circ \pi^{-1} \; ,$$

and

$$(\, w \ll \mu \quad \Longrightarrow \quad Q\, w = ((\frac{dw}{d\mu} \circ \pi_0)\; (\mu \otimes K)) \circ \pi^{-1}\,) \; .$$

Lemma 3.2. *Under Assumption AC, $\Gamma\,\mu \ll Q\,\mu$ for any probability distribution μ on E, with Radon–Nikodym derivative (which depends on μ)*

$$\frac{d(\Gamma\,\mu)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\Xi_{n+1} \mid X_{n+1} = x'] \; .$$

Proof. For any Borel subset B' of E, it holds

$$\begin{aligned} \Gamma\,\mu(B') \;&=\; \int_E \mu(dx)\; \Gamma(x, B') \\ &=\; \int_E \mu(dx)\, \mathbb{E}[\Xi_{k+1}\, \mathbf{1}_{(X_{k+1} \in B')} \mid X_k = x] \\ &=\; \mathbb{E}_\mu[\Xi_{k+1}\, \mathbf{1}_{(X_{k+1} \in B')}] = \int_{B'} \mathbb{E}_\mu[\Xi_{k+1} \mid X_{k+1} = x']\; Q\,\mu(dx') \; , \end{aligned}$$

hence $\Gamma\,\mu$ is a signed measure, absolutely continuous w.r.t. $Q\,\mu$, with Radon–Nikodym derivative

$$\frac{d(\Gamma\,\mu)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\Xi_{k+1} \mid X_{k+1} = x']$$

□

For completeness, the following elementary property is recalled

Lemma 3.3. *If the finite signed measure w is absolutely continuous w.r.t. the probability distribution μ, then $Q\,w \ll Q\,\mu$, with Radon–Nikodym derivative*

$$\frac{d(Q\,w)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\frac{dw}{d\mu}(X_k) \mid X_{k+1} = x'] \ .$$

Proof. For any Borel subset B' of E, it holds

$$\begin{aligned} Q\,w(B') &= \int_E w(dx)\,Q(x,B') \\ &= \int_E \frac{dw}{d\mu}(x)\,\mu(dx)\,Q(x,B') \\ &= \mathbb{E}_\mu[\frac{dw}{d\mu}(X_k)\,\mathbf{1}_{(X_{k+1} \in B')}] = \int_{B'} \mathbb{E}_\mu[\frac{dw}{d\mu}(X_k) \mid X_{k+1} = x']\;Q\,\mu(dx') \ , \end{aligned}$$

hence $Q\,w$ is a signed measure, absolutely continuous w.r.t. $Q\,\mu$, with Radon–Nikodym derivative

$$\frac{d(Q\,w)}{d(Q\,\mu)}(x') = \mathbb{E}_\mu[\frac{dw}{d\mu}(X_k) \mid X_{k+1} = x']$$

□

The explicit expression of the Radon–Nikodym derivatives will not be used in the sequel : only the qualitative properties

$$\Gamma\,\mu \ll Q\,\mu \qquad \text{and} \qquad (\,w \ll \mu \quad \Longrightarrow \quad Q\,w \ll Q\,\mu\,) \ ,$$

will be used.

By definition

$$F_k(\mu)\,w = \frac{\Psi_k\,w}{\langle \mu, \Psi_k \rangle} - \frac{\langle w, \Psi_k \rangle}{\langle \mu, \Psi_k \rangle}\,\frac{\Psi_k\,\mu}{\langle \mu, \Psi_k \rangle} \ ,$$

is the derivative at point μ and in the direction w, of the mapping $\mu \longmapsto \Psi_k \cdot \mu$. The following elementary property holds

Lemma 3.4. *If the finite signed measure w is absolutely continuous w.r.t. the probability distribution μ, then $F_k(\mu)\,w \ll \Psi_k \cdot \mu$, with Radon–Nikodym derivative*

$$\frac{d(F_k(\mu)\,w)}{d(\Psi_k \cdot \mu)}(x) = \frac{dw}{d\mu}(x) - \langle \Psi_k \cdot \mu, \frac{dw}{d\mu} \rangle \ .$$

3.4 Particle approximation of some finite signed measures

With the notations of the previous section, it easily seen that the probability distribution $Q\,\mu$ and the finite signed measures $\Gamma\,\mu$ and $Q\,w$ can be put in the general form $(r\,(\mu \otimes K)) \circ \pi^{-1}$ for some appropriate choice of the weight function r, namely $r \equiv 1$, $r = \pi_F$ and $r = \dfrac{dw}{d\mu} \circ \pi_0$ respectively. The weighted particle approximation of a finite signed measure of the general form $r\,(\mu \otimes K)$ is defined by

$$r\,(\mu \otimes K) \approx r\,S^N(\mu \otimes K) = \frac{1}{N} \sum_{i=1}^N r(\xi_0^i, \xi^i, \Xi^i)\;\delta_{(\xi_0^i, \xi^i, \Xi^i)} \ ,$$

where the N–sample $\{\xi_0^i, \xi^i, \Xi^i\,,\ i = 1, \cdots, N\}$ has precisely the probability distribution $\mu \otimes K$, i.e. one *simulate* N independent r.v.'s with common probability distribution $\mu \otimes K$, which can be achieved in the following manner : independently for any $i = 1, \cdots, N$

$$\xi_0^i \sim \mu(dx) \qquad \text{and} \qquad (\xi^i, \Xi^i) \sim K(\xi_0^i, dx', ds')\ ,$$

and the corresponding particle approximation for the marginal measure $(r\,(K \otimes \mu)) \circ \pi^{-1}$ is defined by

$$(r\,(\mu \otimes K)) \circ \pi^{-1} \approx (r\,S^N(\mu \otimes K)) \circ \pi^{-1} = \frac{1}{N} \sum_{i=1}^N r(\xi_0^i, \xi^i, \Xi^i)\ \delta_{\xi^i}\ .$$

In particular for the weight functions $r \equiv 1$, $r = \pi_F$ and $r = \dfrac{dw}{d\mu} \circ \pi_0$, it holds

$$Q\,\mu = (\mu \otimes K) \circ \pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i}\ ,$$

$$\Gamma\,\mu = (\pi_F\,(\mu \otimes K)) \circ \pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \Xi^i\ \delta_{\xi^i}\ ,$$

and

$$Q\,w = ((\frac{dw}{d\mu} \circ \pi_0)\,(\mu \otimes K)) \circ \pi^{-1} \approx \frac{1}{N} \sum_{i=1}^N \frac{dw}{d\mu}(\xi_0^i)\ \delta_{\xi^i}\ ,$$

respectively. For any test function ϕ defined on E, it holds

$$\begin{aligned}
&\mathbb{E}|\,\frac{1}{N} \sum_{i=1}^N r(\xi_0^i, \xi^i, \Xi^i)\ \phi(\xi^i) - \langle (r\,(\mu \otimes K)) \circ \pi^{-1}, \phi \rangle\,| \\
&\quad \leq \frac{1}{\sqrt{N}}\ \{\int_{E \times E \times F} |\phi(x')|^2\ |r(x, x', s')|^2\ \mu(dx)\ K(x, dx', ds')\,\}^{1/2} \\
&\quad \leq \frac{1}{\sqrt{N}}\ \{\int_{E \times E \times F} |r(x, x', s')|^2\ \mu(dx)\ K(x, dx', ds')\,\}^{1/2}\ \|\phi\|\ ,
\end{aligned}$$

and in particular for the weight functions $r \equiv 1$, $r = \pi_F$ and $r = \dfrac{dw}{d\mu} \circ \pi_0$, it holds

$$\sup_{\|\phi\|=1} \mathbb{E}|\,\frac{1}{N} \sum_{i=1}^N \phi(\xi^i) - \langle Q\,\mu, \phi \rangle\,| \leq \frac{1}{\sqrt{N}}\ ,$$

$$\sup_{\|\phi\|=1} \mathbb{E}|\,\frac{1}{N} \sum_{i=1}^N \Xi^i\ \phi(\xi^i) - \langle \Gamma\,\mu, \phi \rangle\,| \leq \frac{1}{\sqrt{N}}\ \{\sup_{x \in E} \int_{E \times F} |s'|^2\ K(x, dx', ds')\,\}^{1/2}\ ,$$

and

$$\sup_{\|\phi\|=1} \mathbb{E}|\,\frac{1}{N} \sum_{i=1}^N \frac{dw}{d\mu}(\xi_0^i)\ \phi(\xi^i) - \langle Q\,w, \phi \rangle\,| \leq \frac{1}{\sqrt{N}}\ \{\int_E |\frac{dw}{d\mu}(x)|^2\,\mu(dx)\,\}^{1/2}\ ,$$

respectively.

3.5 Joint particle approximation of the filter and the linear tangent filter

Recall that the evolution of the sequence $\{\mu_k\,,\,k \geq 0\}$ taking values in the space of probability distributions on E, is described by the following two steps

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = Q\,\mu_{k-1} \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1}\ .$$

As we want to estimate H, Ψ_k also depends on θ. If w_k denotes at each time instant the linear tangent filter, i.e. the derivative of the filter μ_k w.r.t. the parameter, then the evolution of the sequence $\{w_k\,,\,k \geq 0\}$ taking values in the linear tangent space to the space of probability distributions on E, i.e. taking values in the space of finite signed measures on E with zero total mass, is described by the following two steps, which are linear tangent versions of the prediction step and correction step respectively

$$w_{k-1} \xrightarrow[\text{prediction}]{\text{linear tangent}} w_{k|k-1} = Q\,w_{k-1} + \Gamma\,\mu_{k-1}$$

$$\xrightarrow[\text{correction}]{\text{linear tangent}} w_k = F_k(\mu_{k|k-1})\,w_{k|k-1} + G_k(\mu_{k|k-1})\ .$$

where

$$G_k(\mu) = \frac{\dfrac{\partial \Psi_k}{\partial \theta}\,\mu}{\langle \mu, \Psi_k \rangle} - \frac{\langle \mu, \dfrac{\partial \Psi_k}{\partial \theta} \rangle}{\langle \mu, \Psi_k \rangle}\,\frac{\Psi_k\,\mu}{\langle \mu, \Psi_k \rangle} = [\,\frac{\partial \log \Psi_k}{\partial \theta} - \langle \Psi_k \cdot \mu, \frac{\partial \log \Psi_k}{\partial \theta} \rangle\,]\ \Psi_k \cdot \mu\ .$$

Under Assumption AC, it is easily seen by induction, and using Lemmas 3.2, 3.3 and 3.4, that at each time instant $w_{k|k-1} \ll \mu_{k|k-1}$ and $w_k \ll \mu_k$.

In view of this absolute continuity property, and of the key assumption that it is *easy* to *simulate* r.v.'s with probability distribution $K(x, dx', ds')$, the idea is to jointly approximate the predictor $\mu_{k|k-1}$ and its derivative $w_{k|k-1}$ w.r.t. the parameter with the empirical probability distribution and with a weighted empirical distribution associated with the same and unique N–sample, i.e.

$$\mu_{k|k-1} \approx \mu^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k|k-1}} \qquad \text{and} \qquad w_{k|k-1} \approx w^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \rho^i_{k|k-1}\,\delta_{\xi^i_{k|k-1}}\ .$$

With this definition $w^N_{k|k-1} \ll \mu^N_{k|k-1}$, with Radon–Nikodym derivative

$$r^N_{k|k-1}(x) = \frac{dw^N_{k|k-1}}{d\mu^N_{k|k-1}}(x) = \frac{1}{|I^N_{k|k-1}(x)|} \sum_{i \in I^N_{k|k-1}(x)} \rho^i_{k|k-1}\ ,$$

where $I^N_{k|k-1}(x) = \{i = 1, \cdots, N\ :\ \xi^i_{k|k-1} = x\}$, for any x in the support $\operatorname{supp}\,\mu^N_{k|k-1}$ of the discrete probability distribution $\mu^N_{k|k-1}$. Indeed

$$\mu^N_{k|k-1} = \frac{1}{N} \sum_{x \in \operatorname{supp}\,\mu^N_{k|k-1}} |I^N_{k|k-1}(x)|\ \delta_x\ ,$$

and

$$w^N_{k|k-1} = \frac{1}{N} \sum_{x \in \operatorname{supp}\,\mu^N_{k|k-1}} [\sum_{i \in I^N_{k|k-1}(x)} \rho^i_{k|k-1}\,]\ \delta_x\ .$$

Notice that in most cases, the particle locations $\{\xi^i_{k|k-1}\,,\,i = 1, \cdots, N\}$ turn out to be all distinct, and the much simpler relation

$$r^N_{k|k-1}(\xi^i_{k|k-1}) = \rho^i_{k|k-1}\ ,$$

holds for any $i = 1, \cdots, N$.

This approximation is completely characterized by the set $\{\xi^i_{k|k-1}, \rho^i_{k|k-1}, \, i = 1, \cdots, N\}$ of particles and weights, and the algorithm is completely described by the mechanism which builds $\{\xi^i_{k+1|k}, \rho^i_{k+1|k}, \, i = 1, \cdots, N\}$ from $\{\xi^i_{k|k-1}, \rho^i_{k|k-1}, \, i = 1, \cdots, N\}$. This mechanism is as follows :

(i) the correction step is applied *exactly* to $\mu^N_{k|k-1}$, which results in

$$\mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} = \sum_{i=1}^N \frac{\Psi_k(\xi^i_{k|k-1}) \, \delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} = \sum_{i=1}^N \omega^i_k \, \delta_{\xi^i_{k|k-1}} \; ,$$

as previously, and the linear tangent correction step is applied *exactly* to $w^N_{k|k-1}$, which results in

$$\begin{aligned} w^N_k &= F_k(\mu^N_{k|k-1}) \, w^N_{k|k-1} + G_k(\mu^N_{k|k-1}) \\ &= [\, r^N_{k|k-1} - \langle \Psi_k \cdot \mu^N_{k|k-1}, r^N_{k|k-1} \rangle \,] \; \Psi_k \cdot \mu^N_{k|k-1} + [\, \frac{\partial \log \Psi_k}{\partial \theta} - \langle \Psi_k \cdot \mu^N_{k|k-1}, \frac{\partial \log \Psi_k}{\partial \theta} \rangle \,] \; \Psi_k \cdot \mu^N_{k|k-1} \\ &= [\, r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta} - \langle \mu^N_k, r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta} \rangle \,] \; \mu^N_k \; , \end{aligned}$$

(ii) instead of trying to *compute* $Q \, \mu^N_k$, the following particle approximation

$$\mu^N_{k+1|k} = S^N(Q \, \mu^N_k) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k+1|k}} \; ,$$

is used as previously, instead of trying to *compute*

$$\begin{aligned} Q \, w^N_k &= Q((r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta}) \, \mu^N_k) - \langle \mu^N_k, r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta} \rangle \; Q \, \mu^N_k \\ &= (\, ((r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta}) \circ \pi_0) \; (\mu^N_k \otimes Q) \,) \circ \pi^{-1} - \langle \mu^N_k, r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta} \rangle \; Q \, \mu^N_k \; , \end{aligned}$$

the following weighted particle approximation

$$\begin{aligned} &(\, ((r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta}) \circ \pi_0) \; S^N(\mu^N_k \otimes Q) \,) \circ \pi^{-1} - \langle S^N(\mu^N_k), r^N_{k|k-1} + \frac{\partial \log \Psi_k}{\partial \theta} \rangle \; S^N(Q \, \mu^N_k) \\ &\quad = \frac{1}{N} \sum_{i=1}^N [\, r^N_{k|k-1}(\xi^i_k) + \frac{\partial \log \Psi_k}{\partial \theta}(\xi^i_k) - \frac{1}{N} \sum_{j=1}^N [\, r^N_{k|k-1}(\xi^j_k) + \frac{\partial \log \Psi_k}{\partial \theta}(\xi^j_k) \,] \,] \; \delta_{\xi^i_{k+1|k}} \; , \end{aligned}$$

is used, and instead of trying to *compute*

$$\Gamma \, \mu^N_k = (\pi_F \, (\mu^N_k \otimes K)) \circ \pi^{-1} \; ,$$

the following weighted particle approximation

$$(\pi_F \; S^N(\mu^N_k \otimes K)) \circ \pi^{-1} = \frac{1}{N} \sum_{i=1}^N \Xi^i_{k+1} \; \delta_{\xi^i_{k+1|k}} \; ,$$

is used, hence finally the weighted particle approximation

$$w^N_{k+1|k} = \frac{1}{N} \sum_{i=1}^N \underbrace{[\, r^N_{k|k-1}(\xi^i_k) + \frac{\partial \log \Psi_k}{\partial \theta}(\xi^i_k) - \frac{1}{N} \sum_{j=1}^N [\, r^N_{k|k-1}(\xi^j_k) + \frac{\partial \log \Psi_k}{\partial \theta}(\xi^j_k) \,] + \Xi^i_{k+1} \,]}_{\rho^i_{k+1|k}} \; \delta_{\xi^i_{k+1|k}} \; ,$$

where the N–sample $\{ \xi_k^i, \xi_{k+1|k}^i, \Xi_{k+1}^i \, , \, i = 1, \cdots, N \}$ has precisely the probability distribution $\mu_k^N \otimes K$, i.e. N independent r.v.'s are *simulated* with common probability distribution $\mu_k^N \otimes K$, which can be achieved in the following manner : independently for any $i = 1, \cdots, N$

$$\xi_k^i \sim \mu_k^N(dx) \qquad \Longleftarrow \textbf{ easy, since the probability distribution } \mu_k^N \textbf{ is discrete,}$$

$$(\xi_{k+1|k}^i, \Xi_{k+1}^i) \sim K(\xi_k^i, dx', ds') \Longleftarrow \textbf{ easy, by assumption.}$$

The proposed particle approximation of the linear tangent optimal filter is especially attractive, since it uses the same particle system already used in the approximation of the optimal filter : only one–dimensional weights are needed in addition. Other closely related particle approximation schemes will be presented in Sections 4.1, 4.2, 4.3 and 4.4 on some simple models where comparison with explicit exact expressions provided by Kalman filtering is possible.

3.6 Different redistribution schemes

In practice, after a few iterations all but one particle will have negligible weights, and a large computational effort is devoted to updating particles whose contribution is almost zero. To avoid this phenomenon of *degeneracy of particle weights*, we resort to resampling. The basic idea of resampling, and more generally of redistribution, is to eliminate particles with small weights and to replicate particles with large weights.

3.6.1 Resampling

In the selection step, we want to generate a N–sample $\{\xi_{k+1}^i \, , \, i = 1, \cdots, N\}$ with discrete probability distribution $\displaystyle\sum_{i=1}^N \omega_k^i \delta_{\xi_k^i}$. One of the most direct method is based on an inversion method and consists in generating a uniform r.v. U on $[0,1]$, and if

$$\omega_k^1 + \cdots \omega_k^j \leq U < \omega_k^1 + \cdots \omega_k^{j+1}$$

then choose ξ_k^j. The new weigths are reset to $1/N$, hence the following algorithm :

Resampling

Let $\{\xi_k^i, \omega_k^i \, , \, i = 1, \cdots, N\}$ be the particle system we want to resample from.
Construct cumulative distribution function (CDF) :

$c_1 = 0.$

For $i = 2, \cdots, N$, set $c_i = c_{i-1} + \omega_k^i$.

For $i = 1, \cdots, N$,

$U \sim U[0,1]$.

Find α_i so that $c_{\alpha_i} \leq U < c_{\alpha_i + 1}$.

The new particle system is given by :

For $i = 1, \cdots, N$, set $\xi_{k+1}^i = \xi_k^{\alpha_i}$ and $\omega_{k+1}^i = 1/N$.

Some heuristic measures have been proposed to decide when it is necessary to resample. In this work, we choose to resample at each time iteration.

3.6.2 Comb method

A more determistic way to resample is to multiply $\lfloor N\,\omega_k^i \rfloor$ times the particle ξ_k^i. Then, we use a *comb method* to select $\sum_{i=1}^{N}(N\,\omega_k^i - \lfloor N\,\omega_k^i \rfloor)$ particles as described in the following algorithm :

Comb method

Let $\{\xi_k^i, \omega_k^i\,,\, i = 1, \cdots, N\}$ be the particle system we want to redistribute from.
Let σ be a random permutation on $\{1, \cdots, N\}$.
For $i = 1, \cdots, N$, replace i with $\sigma(i)$.
Construct CDF :

Set $\alpha_1 = \omega_k^{\sigma(1)} - \dfrac{1}{N}\,\lfloor N\,\omega_k^{\sigma(1)} \rfloor$.

For $i = 2, \cdots, N$, compute $\alpha_i = \alpha_{i-1} + \omega_k^{\sigma(i)} - \dfrac{1}{N}\,\lfloor N\,\omega_k^{\sigma(i)} \rfloor$.

The new particle system is given by :

For $i = 1, \cdots, N-1$, replicate the particle $\xi_k^{\sigma(i)}$ a number of times equal to $1 + \lfloor N\,\omega_k^{\sigma(i)} \rfloor$ if $\alpha_i < i/N \leq \alpha_{i+1}$, and to $\lfloor N\,\omega_k^{\sigma(i)} \rfloor$ otherwise, and keep the particle $\xi_k^{\sigma(N)}$.

For $i = 1, \cdots, N$, set $\omega_{k+1}^i = 1/N$.

Notice that the last particle is always selected in this algorithm, and the random permutation is introduced so as to avoid this undesirable systematic effect.

3.6.3 An alternative comb method

Another deterministic algorithm has been proposed :

Alternative comb method

Let $\{\xi_k^i, \omega_k^i\ ,\ i = 1, \cdots, N\}$ be the particle system we want to redistribute from.
Let σ be a random permutation on $\{1, \cdots, N\}$.
For $i = 1, \cdots, N$, replace i with $\sigma(i)$.
Construct CDF :

Set $\alpha_1 = \omega_k^{\sigma(1)}$.

For $i = 2, \cdots, N$, compute $\alpha_i = \alpha_{i-1} + \omega_k^{\sigma(i)}$.

For $i = 1, \cdots, N$, let β_i denote the nearest integer to $N\,\alpha_i$.
Construct :

Set $\gamma_1 = \beta_1$.

For $i = 2, \cdots, N$, set $\gamma_i = \beta_i - \beta_{i-1}$.

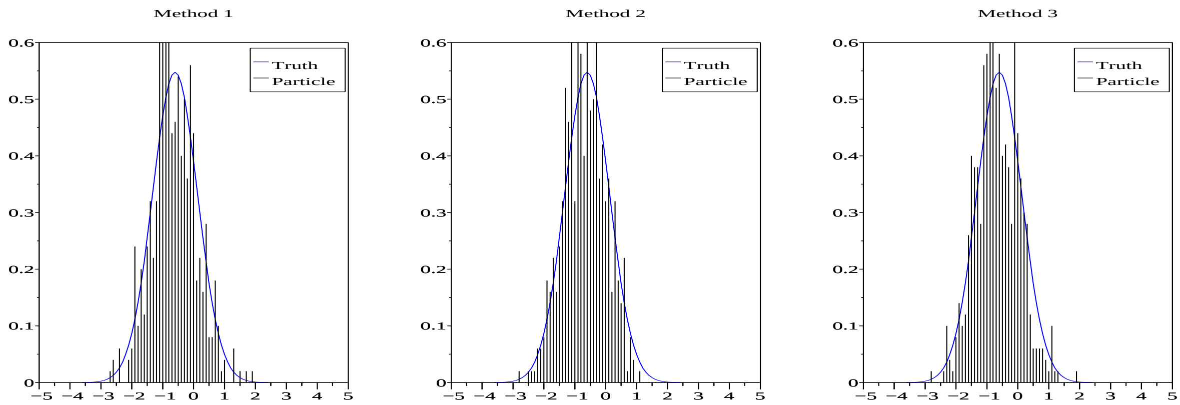
The new particle system is given by :

For $i = 1, \cdots, N$, replicate the particle $\xi_k^{\sigma(i)}$ a number of times equal to γ_i, and set $\omega_{k+1}^i = 1/N$.

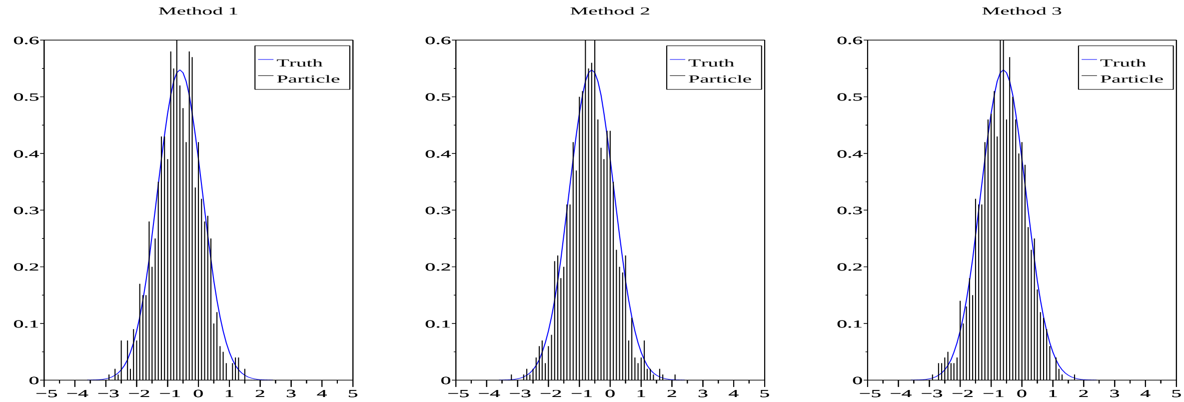
3.6.4 Conclusions

We have seen several methods to redistribute our particle system. The first one seems very natural but is slower than the two others. In practice, we will use a deterministic method since the particle approximations produced by all three methods are very similar, as the following experimental results show :

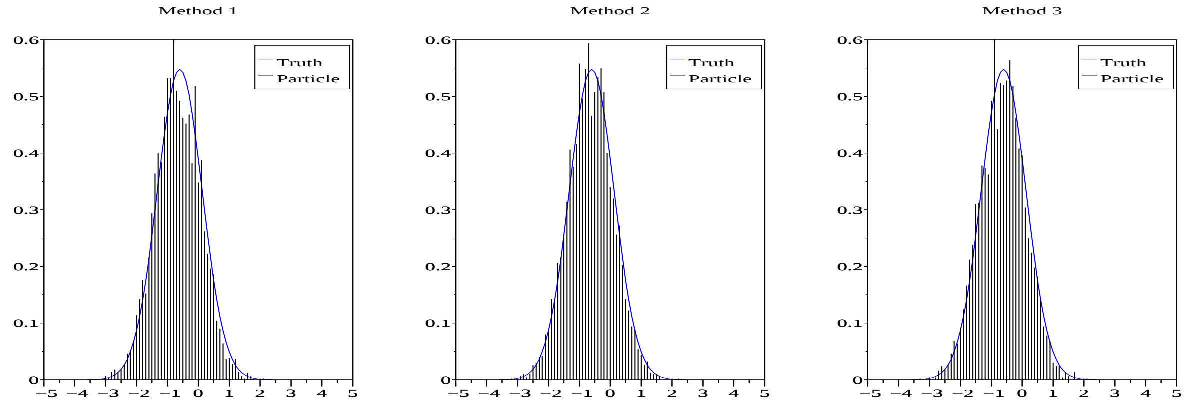
Particle approximation with 500 particles



Particle approximation with 1000 particles



Particle approximation with 5000 particles



4 Particle approximation of finite signed measures, with numerical experiments

The purpose of this section is to obtain as much intuition as possible, through numerical experiments in simple cases where exact solutions are explicitly available, about the particle approximation of finite signed measures. A prototypical example of a finite signed measure is the derivative, w.r.t. a parameter of the model, of some probability distribution related with a Markov chain. This includes prior, prediction, filtering probability distributions, and the like. Two points of view are considered here, to feel the quality of the approximation, at least in a qualitative manner :

(i) how accurate is the particle approximation of the finite signed measure, in view of an histogram representation of the weighted particle system ?

(ii) considering the log–likelihood function and the score function, how close is the approximate expression provided by the particle approximation to the exact expression ?

These two questions seem closely related, however the numerical experiments presented in this section show that one of the two particle approximation schemes fails to satisfy the first criteria (quality of the approximation of the finite signed measure), and that both schemes satisfy the second criteria (quality of the approximation of the statistics). Explanations about the graphical outputs are provided at the end of this section.

4.1 Part A : AR(1) model

The model considered in this example is the simplest scalar AR(1) model

$$X_{k+1} = a\, X_k + \sigma\, W_k \;, \qquad X_0 \sim \mathcal{N}(\bar{X}_0, \sigma_0^2) \;,$$

i.e.

$$\mu_0(dx) = \mathbb{P}[X_0 \in dx] = \underbrace{\frac{1}{\sqrt{2\,\pi}\,\sigma_0}\; \exp\{-\frac{1}{2\,\sigma_0^2}\,(x - \bar{X}_0)^2\}}_{p_0(x)}\; dx \;,$$

where $\{W_k \,,\, k \geq 0\}$ is a standard Gaussian white noise sequence, and where only a is considered as an unknown parameter, i.e. the data $\bar{X}_0$ and σ^2 are known and σ_0^2 could depend on a. It follows from the model that

$$X_{k+1} \mid X_k = x \sim \mathcal{N}(a\,x, \sigma^2) \;,$$

i.e.

$$Q(x, dx') = \mathbb{P}[X_{k+1} \in dx' \mid X_k = x] = \underbrace{\frac{1}{\sqrt{2\,\pi}\,\sigma}\; \exp\{-\frac{1}{2\,\sigma^2}\,(x' - a\,x)^2\}}_{q(x,x')}\; dx' \;.$$

The goal here is to design a particle approximation scheme for the derivative, with respect to the parameter a, of the probability distribution of the state.

4.1.1 Exact expressions

Obviously, in such a simple model, an exact expression can easily be obtained. Indeed

$$X_k \sim \mathcal{N}(\bar{X}_k, \sigma_k^2) \;,$$

i.e.

$$\mu_k(dx) = \mathbb{P}[X_k \in dx] = \underbrace{\frac{1}{\sqrt{2\,\pi}\,\sigma_k}\; \exp\{-\frac{1}{2\,\sigma_k^2}\,(x - \bar{X}_k)^2\}}_{p_k(x)}\; dx \;,$$

where

$$\bar{X}_{k+1} = a\,\bar{X}_k \qquad \text{and} \qquad \sigma_{k+1}^2 = a^2\,\sigma_k^2 + \sigma^2 \;.$$

The log–density of the r.v. X_k is

$$\log p_k(x) = \text{cste} - \frac{1}{2}\,\log \sigma_k^2 - \frac{1}{2\,\sigma_k^2}\,(x - \bar{X}_k)^2 \;,$$

hence the logarithmic derivative w.r.t. the parameter a

$$\frac{\partial \log p_k}{\partial a}(x) = -\frac{1}{2\,\sigma_k^2}\,\frac{\partial \sigma_k^2}{\partial a} + \frac{1}{2\,\sigma_k^4}\,\frac{\partial \sigma_k^2}{\partial a}\,(x - \bar{X}_k)^2 + \frac{1}{\sigma_k^2}\,\frac{\partial \bar{X}_k}{\partial a}\,(x - \bar{X}_k) \;,$$

and

$$w_k(dx) = \frac{\partial \mu_k}{\partial a}(dx) = [\,\frac{1}{2\,\sigma_k^4}\,\frac{\partial \sigma_k^2}{\partial a}\,[\,(x - \bar{X}_k)^2 - \sigma_k^2\,] + \frac{1}{\sigma_k^2}\,\frac{\partial \bar{X}_k}{\partial a}\,(x - \bar{X}_k)\,]\;\mu_k(dx) \;,$$

where

$$\frac{\partial \bar{X}_{k+1}}{\partial a} = \bar{X}_k + a\, \frac{\partial \bar{X}_k}{\partial a} \qquad \text{and} \qquad \frac{\partial \sigma_{k+1}^2}{\partial a} = 2\, a\, \sigma_k^2 + a^2\, \frac{\partial \sigma_k^2}{\partial a} \; .$$

In the special case where $\bar{X}_0 = 0$, it holds $\bar{X}_k = 0$, $\dfrac{\partial \bar{X}_k}{\partial a} = 0$, hence

$$w_k(dx) = \frac{1}{2\, \sigma_k^4}\, \frac{\partial \sigma_k^2}{\partial a}\, (x^2 - \sigma_k^2)\; \mu_k(dx) \; ,$$

and the support of w_k^+, the positive part of w_k, is the set $\{x \; : \; \dfrac{dw_k}{d\mu_k}(x) > 0\} = \{x \; : \; |x| > \sigma_k\}$. In the stationary case where $|a| < 1$ and $\bar{X}_0 = 0$, it holds

$$\sigma_k^2 = \sigma_0^2 = \frac{\sigma^2}{1 - a^2} \qquad \text{and} \qquad \frac{\partial \sigma_0^2}{\partial a} = \frac{\partial \sigma_k^2}{\partial a} = \frac{2\, a\, \sigma^2}{(1-a^2)^2} \; ,$$

hence

$$\mu_k(dx) = \mu_0(dx) \qquad \text{and} \qquad w_k(dx) = w_0(dx) = \frac{a}{\sigma^2}\; (x^2 - \sigma_0^2)\; \mu_0(dx) \; .$$

4.1.2 Preliminary computations

The transition log–density is

$$\log q(x, x') = \text{cste} - \frac{1}{2}\, \log \sigma^2 - \frac{1}{2\, \sigma^2}\, (x' - a\, x)^2 \; ,$$

hence the logarithmic derivative w.r.t. the parameter a

$$\frac{\partial \log q}{\partial a}(x, x') = \frac{x}{\sigma^2}\, (x' - a\, x) \; ,$$

and

$$\Gamma(x, dx') = \frac{\partial Q}{\partial a}(x, dx') = \frac{x}{\sigma^2}\, (x' - a\, x)\; Q(x, dx') \; .$$

Differentiating with respect to the parameter a, throughout the recursion

$$\mu_{k+1}(dx') = Q\, \mu_k(dx') = \int_{-\infty}^{\infty} \mu_k(dx)\, Q(x, dx') \; , \tag{10}$$

yields

$$w_{k+1}(dx') = \frac{\partial \mu_{k+1}}{\partial a}(dx') = \int_{-\infty}^{\infty} [\, \frac{\partial \mu_k}{\partial a}(dx)\, Q(x, dx') + \mu_k(dx)\, \frac{\partial Q}{\partial a}(x, dx')\,] \; ,$$

hence

$$w_{k+1}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, w_k(dx) + \frac{x}{\sigma^2}\, (x' - a\, x)\; \mu_k(dx)\,]\; Q(x, dx') \; . \tag{11}$$

Since $w_k \ll \mu_k$, it holds

$$w_{k+1}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, \frac{dw_k}{d\mu_k}(x) + \frac{x}{\sigma^2}\, (x' - a\, x)\,]\; \mu_k(dx)\, Q(x, dx') \; ,$$

i.e. w_{k+1} can be interpreted as the marginal of the finite signed measure $w_{k,k+1}$ defined on the product space $\mathbb{R} \times \mathbb{R}$ by

$$\begin{aligned} w_{k,k+1}(dx, dx') \; &= \; w_k(dx)\, Q(x, dx') + \mu_k(dx)\, \Gamma(x, dx') \\ &= \; [\, \frac{dw_k}{d\mu_k}(x) + \frac{x}{\sigma^2}\, (x' - a\, x)\,]\; \mu_k(dx)\, Q(x, dx') \; . \end{aligned} \tag{12}$$

4.1.3 A particle approximation scheme

Assuming that the following particle approximations

$$\mu_k \approx \mu_k^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_k^i} \qquad \text{and} \qquad w_k \approx w_k^N = \frac{1}{N} \sum_{i=1}^N \rho_k^i \; \delta_{\xi_k^i} \; ,$$

are available at time index k, and plugging these approximations into equations (10) and (11), yields

$$Q \, \mu_k^N(dx') = \frac{1}{N} \sum_{i=1}^N Q(\xi_k^i, dx') \; , \tag{13}$$

and

$$Q \, w_k^N(dx') + \Gamma \, \mu_k^N(dx') = \frac{1}{N} \sum_{i=1}^N [\, \rho_k^i + \frac{\xi_k^i}{\sigma^2} \, (x' - a \, \xi_k^i) \,] \; Q(\xi_k^i, dx') \; , \tag{14}$$

which can be interpreted as the marginals of the finite signed measures $m = (m^i \, , \, i = 1, \cdots, N)$ and $s = (s^i \, , \, i = 1, \cdots, N)$ defined on the product space $E^N = \{1, \cdots, N\} \times \mathbb{R}$ by

$$s^i(dx') = \underbrace{[\, \rho_k^i + \frac{\xi_k^i}{\sigma^2} \, (x' - a \, \xi_k^i) \,]}_{r^i(x')} \; \underbrace{\frac{1}{N} \, Q(\xi_k^i, dx')}_{m^i(dx')} \; ,$$

respectively. Introducing the particle approximation

$$m \approx m^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\tau_k^i, \, \xi_{k+1}^i)}$$

and the weighted particle approximation

$$s \approx s^N = \frac{1}{N} \sum_{i=1}^N r^{\tau_k^i}(\xi_{k+1}^i) \; \delta_{(\tau_k^i, \, \xi_{k+1}^i)} \; ,$$

where independently for any $i = 1, \cdots, N$, the pair (τ_k^i, ξ_{k+1}^i) is jointly distributed according to the probability distribution $m = (m^i \, , \, i = 1, \cdots, N)$, or using systematic sampling alternatively

$$\tau_k^i = i \qquad \text{and} \qquad \xi_{k+1}^i \sim Q(\xi_k^{\tau_k^i}, dx') \; ,$$

and marginalizing, yields

$$\mu_{k+1} \approx \mu_{k+1}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1}^i} \; ,$$

and

$$w_{k+1} \approx w_{k+1}^N = \frac{1}{N} \sum_{i=1}^N r^i(\xi_{k+1}^i) \; \delta_{\xi_{k+1}^i} = \frac{1}{N} \sum_{i=1}^N \rho_{k+1}^i \; \delta_{\xi_{k+1}^i} \; ,$$

where independently for any $i = 1, \cdots, N$

$$\xi_{k+1}^i \sim Q(\xi_k^i, dx') \; .$$

Notice that the weight

$$\rho_{k+1}^i = r^i(\xi_{k+1}^i) = \rho_k^i + \frac{\xi_k^i}{\sigma^2}\,(\xi_{k+1}^i - a\,\xi_k^i)\ ,$$

does not depend only on the position of ξ_{k+1}^i, but on the position of the pair (ξ_k^i, ξ_{k+1}^i). An approximation of the support of w_{k+1}^+, the positive part of w_{k+1}, is provided by the set $\{\xi_{k+1}^i\ ,\ i \in I_{k+1}^{N,+}\}$ of particles with positive weight, where

$$I_{k+1}^{N,+} = \{i = 1, \cdots, N\ :\ \rho_k^i + \frac{\xi_k^i}{\sigma^2}\,(\xi_{k+1}^i - a\,\xi_k^i) > 0\}\ .$$

4.1.4 An alternate particle approximation scheme

Equations (13) and (14) read

$$Q\,\mu_k^N(dx') = \frac{1}{N}\sum_{i=1}^N Q(\xi_k^i, dx') = [\,\frac{1}{N}\sum_{i=1}^N q(\xi_k^i, x')\,]\ dx'\ ,$$

and

$$\begin{aligned} Q\,w_k^N(dx') + \Gamma\,\mu_k^N(dx') &= \frac{1}{N}\sum_{i=1}^N [\,\rho_k^i + \frac{\xi_k^i}{\sigma^2}\,(x' - a\,\xi_k^i)\,]\ Q(\xi_k^i, dx') \\ &= [\,\frac{1}{N}\sum_{i=1}^N [\,\rho_k^i + \frac{\xi_k^i}{\sigma^2}\,(x' - a\,\xi_k^i)\,]\ q(\xi_k^i, x')\,]\ dx' \\ &= r_{k+1}^N(x')\ Q\,\mu_k^N(dx')\ , \end{aligned}$$

where

$$\begin{aligned} r_{k+1}^N(x') &= \frac{\displaystyle\sum_{j=1}^N [\,\rho_k^j + \frac{\xi_k^j}{\sigma^2}\,(x' - a\,\xi_k^j)\,]\ q(\xi_k^j, x')}{\displaystyle\sum_{j=1}^N q(\xi_k^j, x')} \\ &= \frac{\displaystyle\sum_{j=1}^N [\,\rho_k^j + \frac{\xi_k^j}{\sigma^2}\,(x' - a\,\xi_k^j)\,]\ \exp\{-\tfrac{1}{2}\,(x' - a\,\xi_k^j)^2\}}{\displaystyle\sum_{j=1}^N \exp\{-\tfrac{1}{2}\,(x' - a\,\xi_k^j)^2\}}\ . \end{aligned}$$

Using systematic sampling, yields the particle approximation

$$Q\,\mu_k^N \approx \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{k+1}^i}\ ,$$

where independently for any $i = 1, \cdots, N$

$$\xi_{k+1}^i \sim Q(\xi_k^i, dx')\ ,$$

and the weighted particle approximation

$$Q\,w_k^N + \Gamma\,\mu_k^N \approx w_{k+1}^N = \frac{1}{N}\sum_{i=1}^N r_{k+1}^N(\xi_{k+1}^i)\ \delta_{\xi_{k+1}^i} = \frac{1}{N}\sum_{i=1}^N \rho_{k+1}^i\ \delta_{\xi_{k+1}^i}\ ,$$

where the weight

$$\rho^i_{k+1} = r^N_{k+1}(\xi^i_{k+1}) = \frac{\displaystyle\sum_{j=1}^N [\, \rho^j_k + \frac{\xi^j_k}{\sigma^2}\, (\xi^i_{k+1} - a\, \xi^j_k)\,]\, \exp\{-\tfrac{1}{2}\, (\xi^i_{k+1} - a\, \xi^j_k)^2\}}{\displaystyle\sum_{j=1}^N \exp\{-\tfrac{1}{2}\, (\xi^i_{k+1} - a\, \xi^j_k)^2\}} \; ,$$

depends only on the position of ξ^i_{k+1}. An approximation of the support of w^+_{k+1}, the positive part of w_{k+1}, is provided by the set $\{\xi^i_{k+1} \, , \, i \in I^{N,+}_{k+1}\}$ of particles with positive weight, where

$$I^{N,+}_{k+1} = \{i = 1, \cdots , N \; : \; r^N_{k+1}(\xi^i_{k+1}) > 0\} \; .$$

4.2 Part B : HMM situation

Consider now the HMM situation, where the state, still described by the same scalar AR(1) model as above, is not observed directly and where observations are available instead, which are related to the hidden state by

$$Y_k = c\, X_k + s\, V_k \; ,$$

where $\{V_k \, , \, k \geq 0\}$ is a another standard Gaussian white noise sequence, independent of $\{W_k \, , \, k \geq 0\}$, and where only a is considered as an unknown parameter, i.e. the data $\bar{X}_0$, σ_0, σ, c and s are known. It follows from the model that

$$Y_k \mid X_k = x \sim \mathcal{N}(c\, x, s^2) \; ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = \underbrace{\frac{1}{\sqrt{2\,\pi}\, s}\; \exp\{-\frac{1}{2\,\sigma^2}\, (y - c\, x)^2\}}_{g(x,y)} \; dy \; ,$$

and let

$$\Psi_k(x) = g(x, Y_k) \; ,$$

denote the *likelihood function*. The goal here is to design a particle approximation scheme for the derivative, with respect to the parameter a, of the conditional probability distribution of the hidden state, given the observations.

4.2.1 Exact expressions (Kalman filter)

In such a simple model, an exact expression can easily be obtained, via the prediction / correction steps of the Kalman filter framework. Indeed, the prediction step reads

$$X_k \mid Y_1, \cdots , Y_{k-1} \sim \mathcal{N}(\hat{X}_{k|k-1}, \sigma^2_{k|k-1}) \; ,$$

i.e.

$$\mu_{k|k-1}(dx) = \mathbb{P}[X_k \in dx \mid Y_1, \cdots , Y_{k-1}] = \underbrace{\frac{1}{\sqrt{2\,\pi}\, \sigma_{k|k-1}}\; \exp\{-\frac{1}{2\, \sigma^2_{k|k-1}}\, (x - \hat{X}_{k|k-1})^2\}}_{p_{k|k-1}(x)} \; dx \; ,$$

where

$$\hat{X}_{k|k-1} = a\, \hat{X}_{k-1} \qquad \text{and} \qquad \sigma^2_{k|k-1} = a^2\, \sigma^2_{k-1} + \sigma^2 \; .$$

Proceeding exactly as above yields

$$w_{k|k-1}(dx) = \frac{\partial \mu_{k|k-1}}{\partial a}(dx) \;=\; [\, \frac{1}{2\,\sigma^4_{k|k-1}} \, \frac{\partial \sigma^2_{k|k-1}}{\partial a} \, [\, (x - \hat{X}_{k|k-1})^2 - \sigma^2_{k|k-1} \,]$$

$$+ \frac{1}{\sigma^2_{k|k-1}} \, \frac{\partial \hat{X}_{k|k-1}}{\partial a} \, (x - \hat{X}_{k|k-1}) \,] \; \mu_{k|k-1}(dx) \;,$$

where

$$\frac{\partial \hat{X}_{k|k-1}}{\partial a} = \hat{X}_{k-1} + a \, \frac{\partial \hat{X}_{k-1}}{\partial a} \qquad \text{and} \qquad \frac{\partial \sigma^2_{k|k-1}}{\partial a} = 2\, a\, \sigma^2_{k-1} + a^2 \, \frac{\partial \sigma^2_{k-1}}{\partial a} \;.$$

The correction step reads

$$X_k \mid Y_1, \cdots, Y_k \sim \mathcal{N}(\hat{X}_k, \sigma^2_k) \;,$$

i.e.

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_k] = \underbrace{\frac{1}{\sqrt{2\,\pi}\,\sigma_k} \; \exp\{-\frac{1}{2\,\sigma^2_k} \, (x - \hat{X}_k)^2\}}_{p_k(x)} \; dx \;,$$

where the Kalman gain is defined by

$$K_k = \frac{c\,\sigma^2_{k|k-1}}{c^2\,\sigma^2_{k|k-1} + s^2} \qquad \text{hence} \qquad 1 - K_k\, c = \frac{s^2}{c^2\,\sigma^2_{k|k-1} + s^2} \;,$$

and where

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k \, (Y_k - c\,\hat{X}_{k|k-1}) = \hat{X}_{k|k-1} + \frac{c\,\sigma^2_{k|k-1}}{c^2\,\sigma^2_{k|k-1} + s^2} \, (Y_k - c\,\hat{X}_{k|k-1}) \;,$$

and

$$\sigma^2_k = (1 - K_k\, c)\, \sigma^2_{k|k-1} = \frac{s^2\,\sigma^2_{k|k-1}}{c^2\,\sigma^2_{k|k-1} + s^2} \;.$$

Therefore

$$w_k(dx) = \frac{\partial \mu_k}{\partial a}(dx) = [\, \frac{1}{2\,\sigma^4_k} \, \frac{\partial \sigma^2_k}{\partial a} \, [\, (x - \hat{X}_k)^2 - \sigma^2_k \,] + \frac{1}{\sigma^2_k} \, \frac{\partial \hat{X}_k}{\partial a} \, (x - \hat{X}_k) \,] \; \mu_k(dx) \;,$$

where

$$\frac{\partial \hat{X}_k}{\partial a} \;=\; \frac{s^2}{c^2\,\sigma^2_{k|k-1} + s^2} \, \frac{\partial \hat{X}_{k|k-1}}{\partial a} + \frac{c\, s^2}{(c^2\,\sigma^2_{k|k-1} + s^2)^2} \, \frac{\partial \sigma^2_{k|k-1}}{\partial a} \, (Y_k - c\,\hat{X}_{k|k-1}) \;,$$

and

$$\frac{\partial \sigma^2_k}{\partial a} = \frac{s^4}{(c^2\,\sigma^2_{k|k-1} + s^2)^2} \, \frac{\partial \sigma^2_{k|k-1}}{\partial a} \;.$$

From the decomposition

$$Y_k = c\,\hat{X}_{k|k-1} + I_k \;,$$

where the innovation

$$I_k = Y_k - c\,\hat{X}_{k|k-1} = c\,(X_k - \hat{X}_{k|k-1}) + V_k \sim \mathcal{N}(0, s_k^2) \qquad \text{with} \qquad s_k^2 = c^2\,\sigma_{k|k-1}^2 + s^2 \ ,$$

is independent of the past observations $Y_1, \cdots, Y_{k-1}$, it follows that

$$Y_k \mid Y_1, \cdots, Y_{k-1} \sim \mathcal{N}(c\,\hat{X}_{k|k-1}, s_k^2) \ ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] = \underbrace{\frac{1}{\sqrt{2\,\pi}\,s_k} \ \exp\{-\frac{1}{2\,s_k^2}\,(y - c\,\hat{X}_{k|k-1})^2\}}_{g_k(y)} \ dy \ .$$

Using the straightforward identity

$$\mathbb{P}[Y_1 \in dy_1, \cdots, Y_n \in dy_n] = \prod_{k=1}^n \mathbb{P}[Y_k \in dy_k \mid Y_1 = y_1, \cdots, Y_{k-1} = y_{k-1}] = \prod_{k=1}^n g_k(y_k)\,dy_k \ ,$$

yields the following expression for the (suitably normalized) log–likelihood function

$$\ell_n = \frac{1}{n}\ \log \prod_{k=1}^n g_k(Y_k) = \frac{1}{n}\ \sum_{k=1}^n \log g_k(Y_k) = -\frac{1}{n}\ \sum_{k=1}^n [\,\tfrac{1}{2}\ \log(2\pi s_k^2) + \frac{1}{2\,s_k^2}\,(Y_k - c\,\hat{X}_{k|k-1})^2\,] \ ,$$

and for the score function

$$\frac{\partial \ell_n}{\partial a} = \frac{1}{n}\ \sum_{k=1}^n [-\frac{1}{2\,s_k^2}\,\frac{\partial s_k^2}{\partial a} + \frac{1}{2\,s_k^4}\,\frac{\partial s_k^2}{\partial a}\,(Y_k - c\,\hat{X}_{k|k-1})^2 + \frac{c}{s_k^2}\,\frac{\partial \hat{X}_{k|k-1}}{\partial a}\,(Y_k - c\,\hat{X}_{k|k-1})\,] \ ,$$

where

$$\frac{\partial s_k^2}{\partial a} = c^2\,\frac{\partial \sigma_{k|k-1}^2}{\partial a} \ .$$

Notice that in full generality

$$\begin{aligned}
\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] &= \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy, X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\
&= \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy \mid X_k = x]\ \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\
&= \underbrace{[\int_{-\infty}^{\infty} g(x,y)\ \mu_{k|k-1}(dx)\,]}_{g_k(y)}\ dy \ ,
\end{aligned}$$

hence the following equivalent expression for the log–likelihood function

$$\ell_n = \frac{1}{n}\ \sum_{k=1}^n \log g_k(Y_k) = \frac{1}{n}\ \sum_{k=1}^n \log \int_{-\infty}^{\infty} \Psi_k(x)\ \mu_{k|k-1}(dx) \ , \tag{15}$$

and for the score function

$$\frac{\partial \ell_n}{\partial a} = \frac{1}{n}\ \sum_{k=1}^n \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x)\ \frac{\partial \mu_{k|k-1}}{\partial a}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x)\ \mu_{k|k-1}(dx)} = \frac{1}{n}\ \sum_{k=1}^n \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x)\ w_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x)\ \mu_{k|k-1}(dx)} \ . \tag{16}$$

4.2.2 Preliminary computations

Differentiating with respect to the parameter a, throughout the recursions

$$\mu_k(dx) = (\Psi_k \cdot \mu_{k|k-1})(dx) = \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \;, \tag{17}$$

and

$$\mu_{k+1|k}(dx) = Q\, \mu_k(dx') = \int_{-\infty}^{\infty} \mu_k(dx)\, Q(x, dx') \;, \tag{18}$$

yields

$$w_k(dx) = \frac{\partial \mu_k}{\partial a}(dx) = \frac{\Psi_k(x)\, \dfrac{\partial \mu_{k|k-1}}{\partial a}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \frac{\partial \mu_{k|k-1}}{\partial a}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \;,$$

hence

$$\begin{aligned} w_k(dx) &= (F_k(\mu_{k|k-1}) w_{k|k-1})(dx) \\ &= \frac{\Psi_k(x)\, w_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, w_{k|k-1}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \;, \end{aligned} \tag{19}$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, w_k(dx) + \frac{x}{\sigma^2}\, (x' - a\, x)\; \mu_k(dx) \,]\; Q(x, dx') \;. \tag{20}$$

Since $w_{k|k-1} \ll \mu_{k|k-1}$, it holds $F_k(\mu_{k|k-1}) w_{k|k-1} \ll \Psi_k \cdot \mu_{k|k-1}$, with Radon–Nikodym derivative

$$\frac{dw_k}{d\mu_k}(x) = \frac{d(F_k(\mu_{k|k-1}) w_{k|k-1})}{d(\Psi_k \cdot \mu_{k|k-1})}(x) = \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x) - \int_{-\infty}^{\infty} \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x')\; (\Psi_k \cdot \mu_{k|k-1})(dx') \;,$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, \frac{dw_k}{d\mu_k}(x) + \frac{x}{\sigma^2}\, (x' - a\, x) \,]\; \mu_k(dx)\; Q(x, dx') \;.$$

4.2.3 A particle approximation scheme

Assuming that the following particle approximations

$$\mu_{k|k-1} \approx \mu_{k|k-1}^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{k|k-1}^i} \qquad \text{and} \qquad w_{k|k-1} \approx w_{k|k-1}^N = \frac{1}{N} \sum_{i=1}^{N} \rho_{k|k-1}^i\; \delta_{\xi_{k|k-1}^i} \;, \tag{21}$$

are available at time index k, and plugging these approximations into equations (17) and (19), yields

$$\mu_k^N = \Psi_k \cdot \mu_{k|k-1}^N = \frac{\displaystyle\sum_{i=1}^{N} \Psi_k(\xi_{k|k-1}^i)\; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^{N} \Psi_k(\xi_{k|k-1}^j)} = \sum_{i=1}^{N} \omega_k^i\; \delta_{\xi_{k|k-1}^i} \;,$$

and

$$w_k^N = F_k(\mu_{k|k-1}^N) w_{k|k-1}^N \; = \; \frac{\displaystyle\sum_{i=1}^N \rho_{k|k-1}^i \; \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} - \frac{\displaystyle\sum_{j=1}^N \rho_{k|k-1}^j \; \Psi_k(\xi_{k|k-1}^j)}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} \; \frac{\displaystyle\sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)}$$

$$= \sum_{i=1}^N \rho_{k|k-1}^i \; \omega_k^i \; \delta_{\xi_{k|k-1}^i} - [\sum_{j=1}^N \rho_{k|k-1}^j \; \omega_k^j \,] \; \sum_{i=i}^N \omega_k^i \; \delta_{\xi_{k|k-1}^i}$$

$$= \sum_{i=1}^N [\, \rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j \; \omega_k^j \,]\,] \; \omega_k^i \; \delta_{\xi_{k|k-1}^i} \; .$$

Plugging these expressions into equations (18) and (20), yields

$$Q \, \mu_k^N(dx') = \sum_{i=1}^N \omega_k^i \; Q(\xi_{k|k-1}^i, dx') \; , \tag{22}$$

and

$$Q \, w_k^N(dx') + \Gamma \, \mu_k^N(dx') = \sum_{i=1}^N [\, \rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j \; \omega_k^j \,] + \frac{\xi_{k|k-1}^i}{\sigma^2} \, (x' - a \, \xi_{k|k-1}^i) \,] \; \omega_k^i \; Q(\xi_{k|k-1}^i, dx') \; , \tag{23}$$

which can be interpreted as the marginals of the finite signed measures $m = (m^i \, , \, i = 1, \cdots, N)$ and $s = (s^i \, , \, i = 1, \cdots, N)$ defined on the product space $E^N = \{1, \cdots, N\} \times \mathbb{R}$ by

$$s^i(dx') = [\, \underbrace{\rho_{k|k-1}^i + \frac{\xi_{k|k-1}^i}{\sigma^2} \, (x' - a \, \xi_{k|k-1}^i)}_{r^i(x')} - [\sum_{j=1}^N \rho_{k|k-1}^j \; \omega_k^j \,]\,] \; \underbrace{\omega_k^i \; Q(\xi_{k|k-1}^i, dx')}_{m^i(dx')} \; .$$

Notice that

$$\sum_{j=1}^N \int_{-\infty}^{\infty} r^j(x') \; m^j(dx') = \sum_{j=1}^N \rho_{k|k-1}^j \; \omega_k^j \; ,$$

is just a normalizing constant. Introducing the particle approximation

$$m \approx m^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\tau_k^i, \, \xi_{k+1|k}^i)} \; ,$$

and the weighted particle approximation

$$s \approx s^N = \frac{1}{N} \sum_{i=1}^N [\, r^{\tau_k^i}(\xi_{k+1}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1}^j) \,]\,] \; \delta_{(\tau_k^i, \, \xi_{k+1|k}^i)} \; ,$$

where independently for any $i = 1, \cdots, N$, the pair $(\tau_k^i, \xi_{k+1|k}^i)$ is jointly distributed according to the probability distribution $m = (m^i \, , \, i = 1, \cdots, N)$, i.e.

$$\tau_k^i \sim (\omega_k^j \, , \, j = 1, \cdots, N) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\xi_{k|k-1}^{\tau_k^i}, dx') \; ,$$

and marginalizing, yields

$$\mu_{k+1|k} \approx \mu^N_{k+1|k} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i_{k+1|k}} \ .$$

$$w_{k+1|k} \approx w^N_{k+1|k} = \frac{1}{N} \sum_{i=1}^{N} [\, r^{\tau^i_k}(\xi^i_{k+1|k}) - [\, \frac{1}{N} \sum_{j=1}^{N} r^{\tau^j_k}(\xi^j_{k+1|k})\,]\,]\; \delta_{\xi^i_{k+1}} = \frac{1}{N} \sum_{i=1}^{N} \rho^i_{k+1|k}\; \delta_{\xi^i_{k+1|k}} \ .$$

Notice that the weight

$$\rho^i_{k+1|k} = r^{\tau^i_k}(\xi^i_{k+1|k}) - [\, \frac{1}{N} \sum_{j=1}^{N} r^{\tau^j_k}(\xi^j_{k+1|k})\,] \ ,$$

where

$$r^{\tau^i_k}(\xi^i_{k+1|k}) = \rho^{\tau^i_k}_{k|k-1} + \frac{\xi^{\tau^i_k}_{k|k-1}}{\sigma^2}\, (\xi^i_{k+1|k} - a\, \xi^{\tau^i_k}_{k|k-1}) \ ,$$

does not depend only on the position $\xi^i_{k+1|k}$, but on the position of the pair $(\xi^{\tau^i_k}_{k|k-1}, \xi^i_{k+1|k})$.

Statistics Plugging *any* joint particle approximation of the form (21) into equations (15) and (16), yields the following approximation

$$\ell_n \approx \frac{1}{n} \sum_{k=1}^{n} \log[\, \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi^i_{k|k-1})\,] \ ,$$

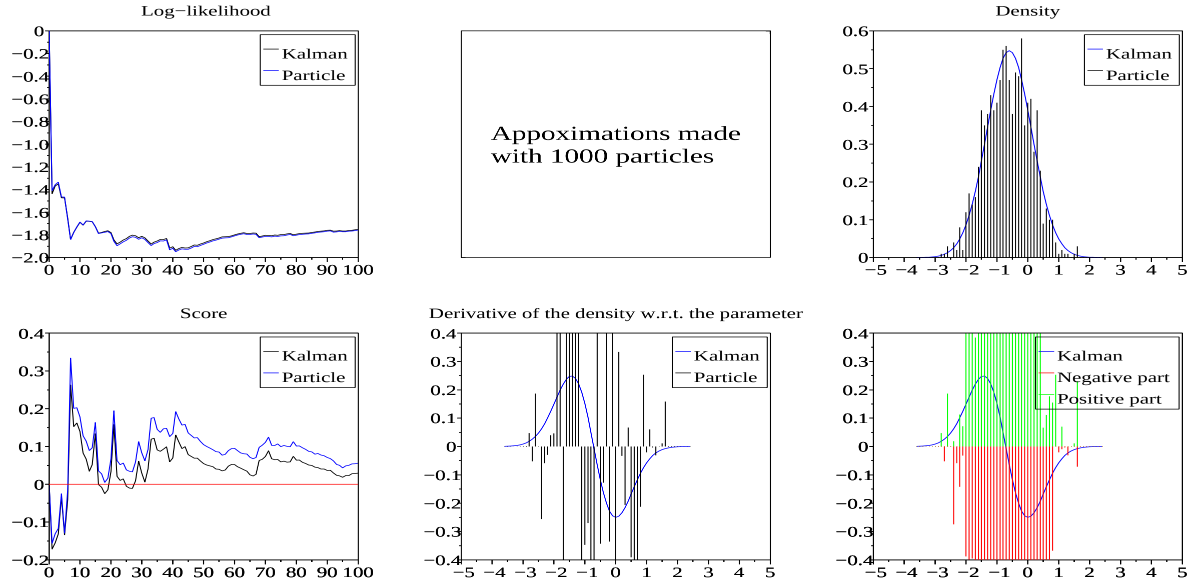
for the log–likelihood function, and the following approximation

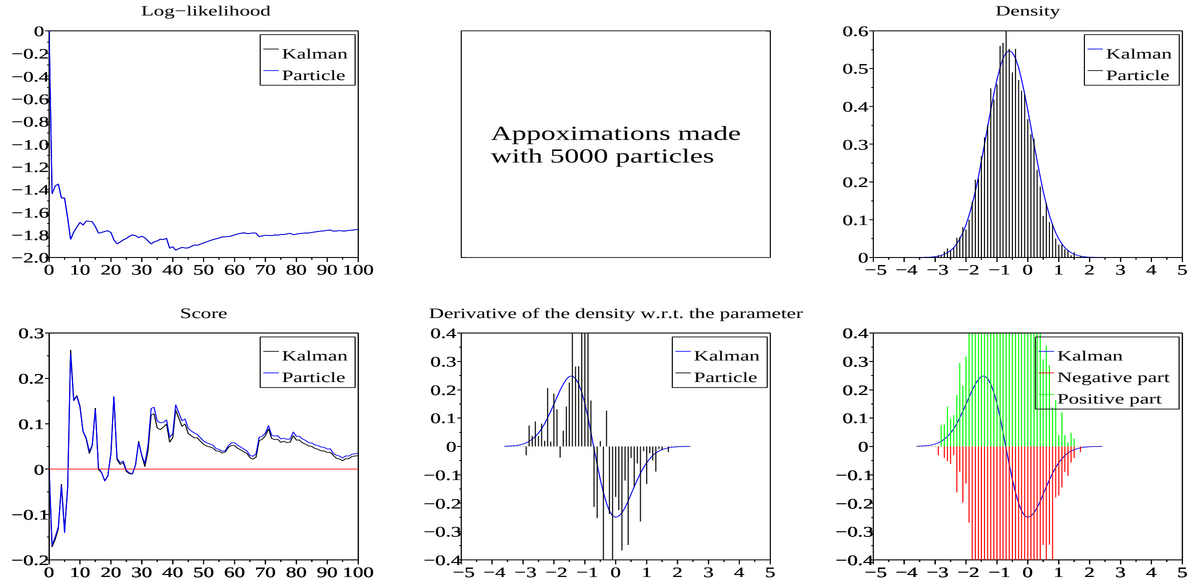
$$\frac{\partial \ell_n}{\partial \theta} \approx \frac{1}{n} \sum_{k=1}^{n} \left[\frac{\displaystyle \frac{1}{N} \sum_{i=1}^{N} \rho^i_{k|k-1}\; \Psi_k(\xi^i_{k|k-1})}{\displaystyle \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi^i_{k|k-1})} \right] = \frac{1}{n} \sum_{k=1}^{n} [\sum_{i=1}^{N} \rho^i_{k|k-1}\; \omega^i_k\,] \ ,$$

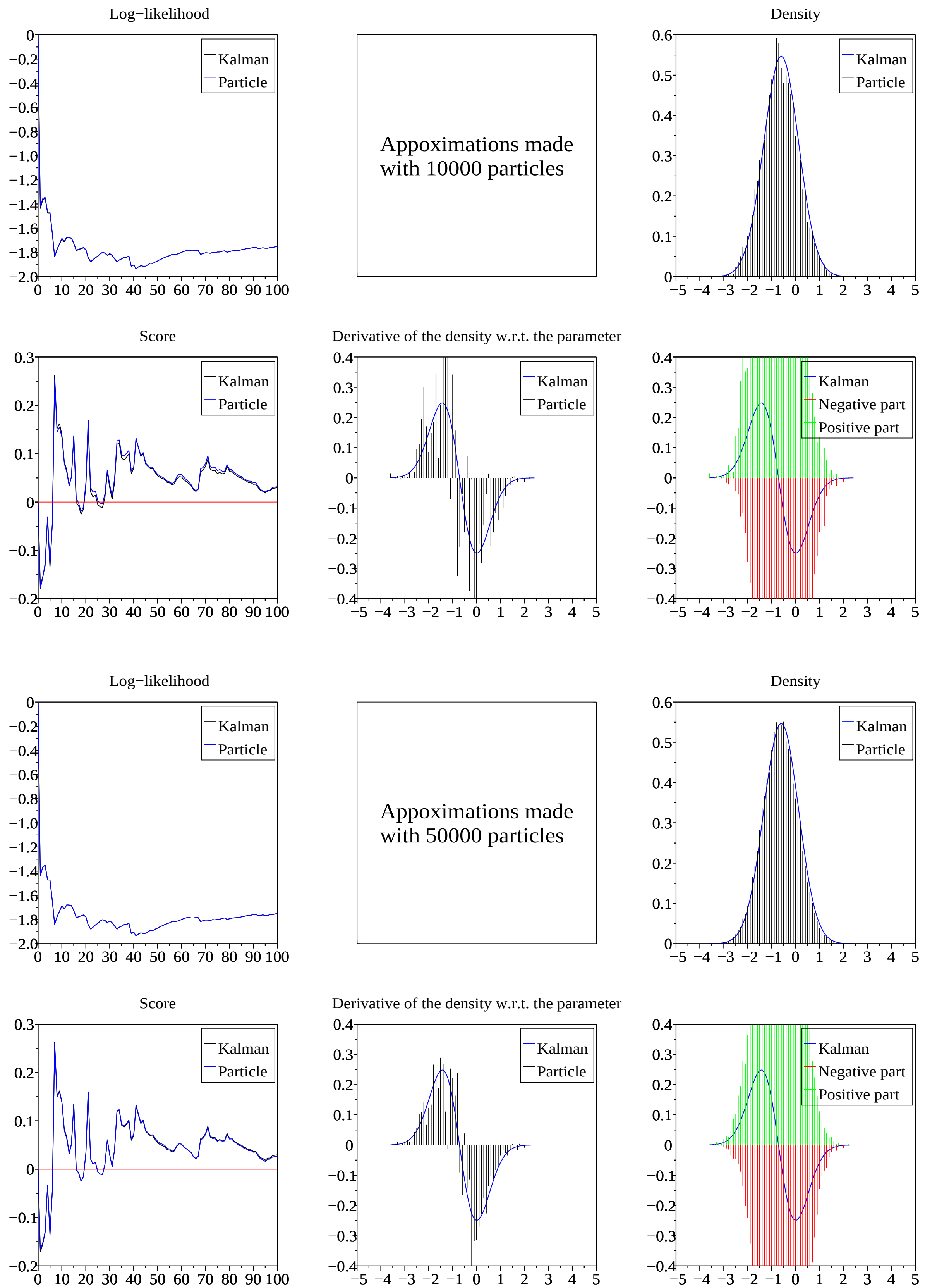
for the score function, respectively.

Numerical results

$\bar{X}_0$	σ_0^2	a	σ^2	c	s^2	N (# particles)
0.0	1.0	0.5	1.0	1.0	1.0	1000, 5000, 10000, 50000


Log−likelihood
Kalman
Particle
Appoximations made
with 1000 particles
Density
Kalman
Particle
Score
Kalman
Particle
Derivative of the density w.r.t. the parameter
Kalman
Particle
Kalman
Negative part
Positive part


Log−likelihood
Kalman
Particle
Appoximations made
with 5000 particles
Density
Kalman
Particle
Score
Kalman
Particle
Derivative of the density w.r.t. the parameter
Kalman
Particle
Kalman
Negative part
Positive part



Log−likelihood
Kalman
Particle
Appoximations made
with 10000 particles
Density
Kalman
Particle
Score
Kalman
Particle
Derivative of the density w.r.t. the parameter
Kalman
Particle
Kalman
Negative part
Positive part
Log−likelihood
Kalman
Particle
Appoximations made
with 50000 particles
Density
Kalman
Particle
Score
Kalman
Particle
Derivative of the density w.r.t. the parameter
Kalman
Particle
Kalman
Negative part
Positive part

4.2.4 An alternate particle approximation scheme

Equations (22) and (23) read

$$Q\,\mu_k^N(dx') = \sum_{i=1}^N \omega_k^i\; Q(\xi_{k|k-1}^i, dx') = [\sum_{i=1}^N \omega_k^i\; q(\xi_{k|k-1}^i, x')]\; dx' \;,$$

and

$$\begin{aligned} Q\,w_k^N(dx') + \Gamma\,\mu_k^N(dx') &= \sum_{i=1}^N [\,\rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j\; \omega_k^j\,] + \frac{\xi_{k|k-1}^i}{\sigma^2}\,(x' - a\,\xi_{k|k-1}^i)\,]\; \omega_k^i\; Q(\xi_{k|k-1}^i, dx') \\ &= [\sum_{i=1}^N [\,\rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j\; \omega_k^j\,] + \frac{\xi_{k|k-1}^i}{\sigma^2}\,(x' - a\,\xi_{k|k-1}^i)\,]\; \omega_k^i\; q(\xi_{k|k-1}^i, x')\,]\; dx' \\ &= r_{k+1}^N(x')\; Q\,\mu_k^N(dx') \;, \end{aligned}$$

where

$$\begin{aligned} r_{k+1}^N(x') &= \frac{\displaystyle\sum_{j=1}^N [\,\rho_{k|k-1}^j + \frac{\xi_{k|k-1}^j}{\sigma^2}\,(x' - a\,\xi_{k|k-1}^j)\,]\; \omega_k^j\; q(\xi_{k|k-1}^j, x')}{\displaystyle\sum_{j=1}^N \omega_k^j\; q(\xi_{k|k-1}^j, x')} - [\sum_{j=1}^N \rho_{k|k-1}^j\; \omega_k^j\,] \\ &= \frac{\displaystyle\sum_{j=1}^N [\,\rho_{k|k-1}^j + \frac{\xi_{k|k-1}^j}{\sigma^2}\,(x' - a\,\xi_{k|k-1}^j)\,]\; \omega_k^j\; \exp\{-\tfrac{1}{2}\,(x' - a\,\xi_{k|k-1}^j)^2\}}{\displaystyle\sum_{j=1}^N \omega_k^j\; \exp\{-\tfrac{1}{2}\,(x' - a\,\xi_{k|k-1}^j)^2\}} - [\sum_{j=1}^N \rho_{k|k-1}^j\; \omega_k^j\,] \;. \end{aligned}$$

Resampling yields the particle approximation

$$Q\,\mu_k^N \approx \mu_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1|k}^i} \;,$$

where independently for any $i = 1, \cdots, N$

$$\xi_{k+1|k}^i \sim Q\,\mu_k^N(dx') \;,$$

which can be achieved for instance by taking

$$\widehat{\xi}_k^i \sim \mu_k^N(dx) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx') \;,$$

or even more explicitly by taking

$$\tau_k^i \sim (\omega_k^j\,,\, j = 1, \cdots, N) \;, \qquad \widehat{\xi}_k^i = \xi_{k|k-1}^{\tau_k^i} \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx') \;,$$

and the weighted particle approximation

$$Q\,w_k^N + \Gamma\,\mu_k^N \approx w_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N r_{k+1}^N(\xi_{k+1|k}^i)\; \delta_{\xi_{k+1|k}^i} = \frac{1}{N} \sum_{i=1}^N \rho_{k+1|k}^i\; \delta_{\xi_{k+1|k}^i} \;,$$

where the weight

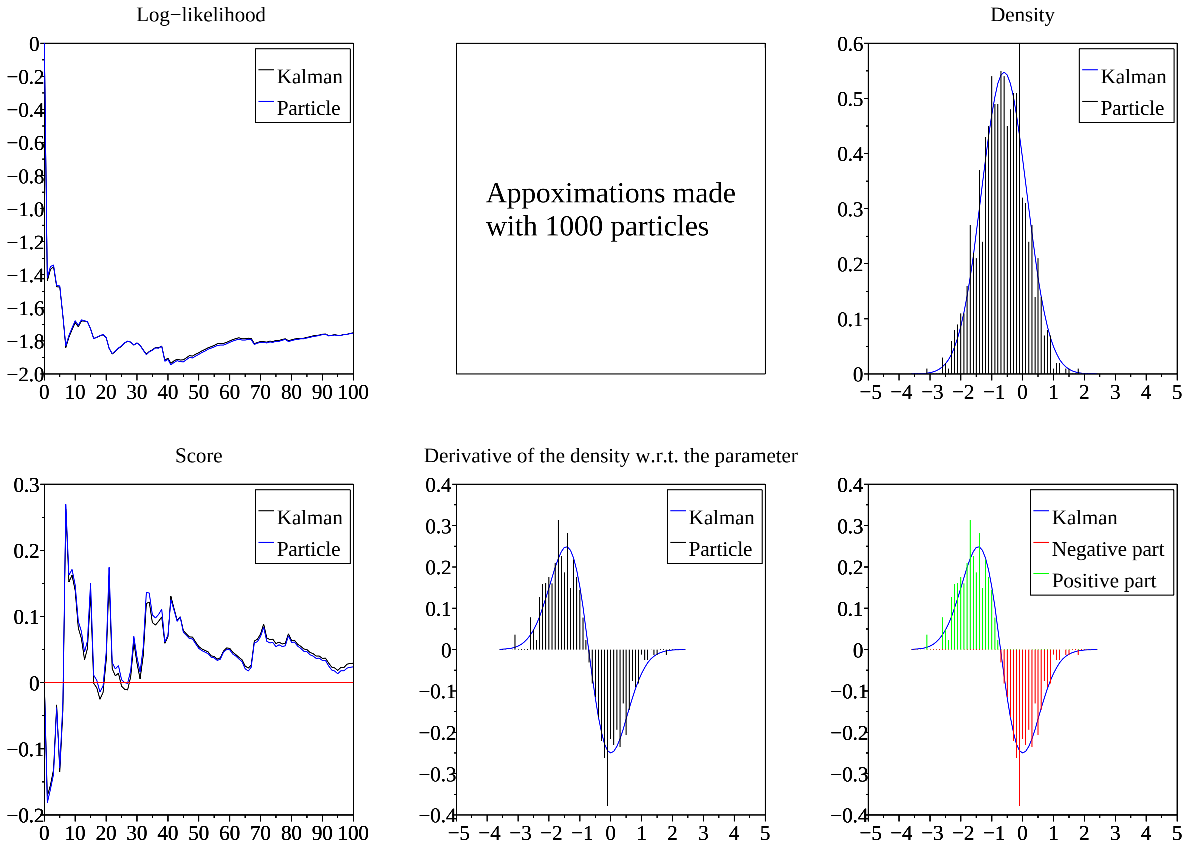
$$\rho^i_{k+1|k} = r^N_{k+1}(\xi^i_{k+1|k}) \;=\; \frac{\displaystyle\sum_{j=1}^{N} [\, \rho^j_{k|k-1} + \frac{\xi^j_{k|k-1}}{\sigma^2} \, (\xi^i_{k+1|k} - a\, \xi^j_{k|k-1}) \,] \; \omega^j_k \;\; \exp\{-\tfrac{1}{2}\, (\xi^i_{k+1|k} - a\, \xi^j_{k|k-1})^2\}}{\displaystyle\sum_{j=1}^{N} \omega^j_k \;\; \exp\{-\tfrac{1}{2}\, (\xi^i_{k+1|k} - a\, \xi^j_{k|k-1})^2\}}$$

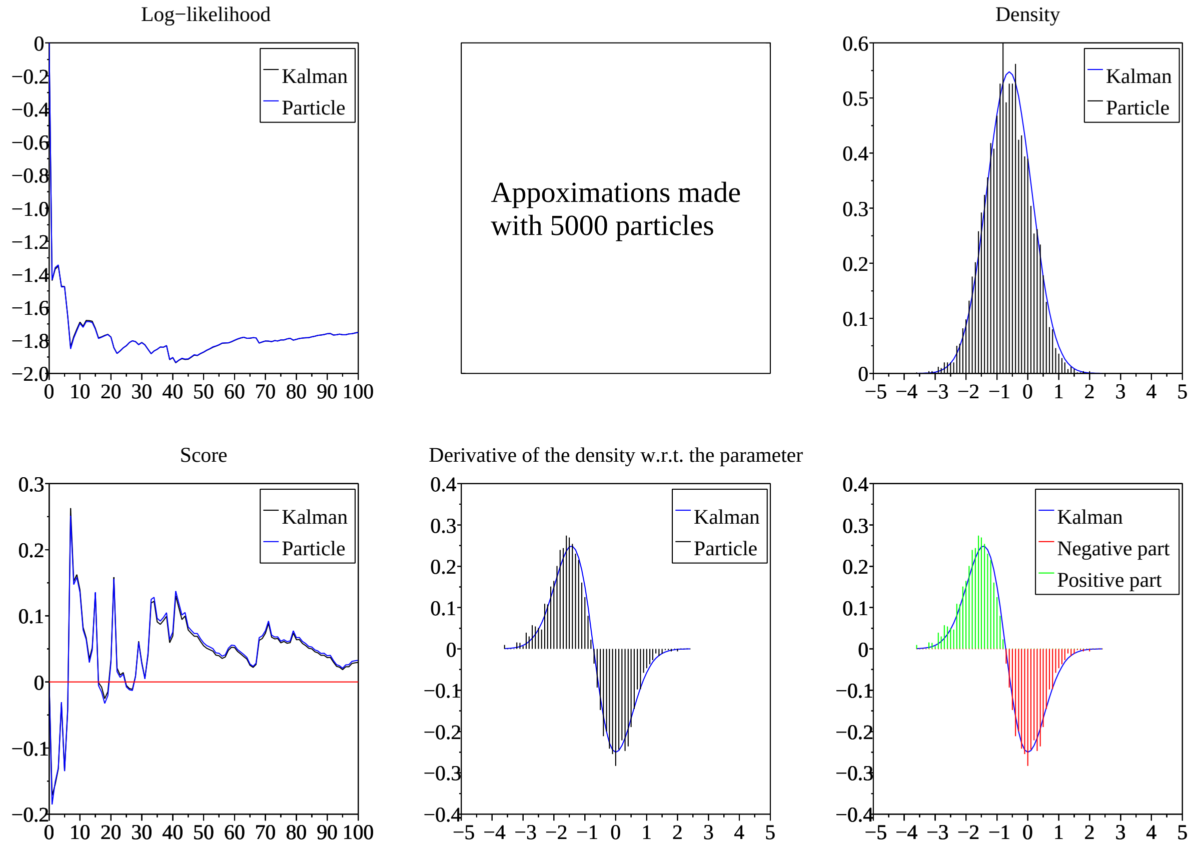
$$- \,[\, \sum_{j=1}^{N} \rho^j_{k|k-1} \; \omega^j_k \,] \;,$$

depends only on the position of $\xi^i_{k+1|k}$.

Numerical results

X_0	σ_0^2	a	σ^2	c	s^2	N (# particles)
0.0	1.0	0.5	1.0	1.0	1.0	1000, 5000





4.3 Part C : 2–dimensional case

We consider now the following model :

$$\begin{cases} X_{k+1} = F\, X_k + W_k \\ Y_k = H\, X_k + V_k \ , \end{cases} \tag{24}$$

where $\{W_k\, ,\, k \geq 0\}$ and $\{V_k\, ,\, k \geq 0\}$ are standard Gaussian independent white noise sequences, with covariance matrix Σ and S respectively. Here

$$F = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

is considered as an unknown parameter and $X_0 \sim \mathcal{N}(\bar{X}_0, \Sigma_0)$. It follows from the model that

$$X_{k+1} \mid X_k = x \sim \mathcal{N}(F\, x, \Sigma) \ ,$$

i.e.

$$Q(x, dx') = \mathbb{P}[X_{k+1} \in dx' \mid X_k = x] = \underbrace{\frac{1}{2\,\pi\,\sqrt{\det\,\Sigma}}\ \exp\{-\frac{1}{2}\,(x' - F\,x)^*\,\Sigma^{-1}\,(x' - F\,x)\}}_{q(x,x')}\ dx' \ ,$$

and

$$Y_k \mid X_k = x \sim \mathcal{N}(H\, x, S) \ ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = \underbrace{\frac{1}{2\,\pi\,\sqrt{\det\, S}}\;\exp\{-\frac{1}{2}\,(y - H\,x)^*\, S^{-1}\,(y - H\,x)\}}_{g(x,y)}\; dy \;,$$

and let

$$\Psi_k(x) = g(x, Y_k) \;,$$

denote the *likelihood function*. The goal here is to design a particle approximation scheme for the derivative, with respect to the parameters $f_{11}, f_{12}, f_{21}, f_{22}$, of the conditional probability distribution of the hidden state, given the observations.

4.3.1 Exact expressions (Kalman filter)

In such a simple model, an exact expression can easily be obtained, via the prediction / correction steps of the Kalman filter framework. Indeed, the prediction step reads

$$X_k \mid Y_1, \cdots, Y_{k-1} \sim \mathcal{N}(\hat{X}_{k|k-1}, \Sigma_{k|k-1}) \;,$$

i.e.

$$\begin{aligned} \mu_{k|k-1}(dx) &= \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\ &= \underbrace{\frac{1}{2\,\pi\,\sqrt{\det\, \Sigma_{k|k-1}}}\;\exp\{-\frac{1}{2}\,(x - \hat{X}_{k|k-1})^*\, \Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1})\}}_{p_{k|k-1}(x)}\; dx \;, \end{aligned}$$

where

$$\hat{X}_{k|k-1} = F\,\hat{X}_{k-1} \qquad \text{and} \qquad \Sigma_{k|k-1} = F\,\Sigma_{k-1}\,F^* + \Sigma \;.$$

The log–density of the r.v. $X_k \mid Y_1, \cdots, Y_{k-1}$ is

$$\log p_{k|k-1}(x) = \text{cste} - \tfrac{1}{2}\,\log\,\det\,\Sigma_{k|k-1} - \tfrac{1}{2}\,(x - \hat{X}_{k|k-1})^*\,\Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1}) \;,$$

hence the logarithmic derivative w.r.t. the parameter θ where $\theta \in \{f_{11}, f_{12}, f_{21}, f_{22}\}$

$$\begin{aligned} \frac{\partial \log p_{k|k-1}}{\partial \theta}(x) = \tfrac{1}{2} \Bigg[&- \frac{\dfrac{\partial \det\,\Sigma_{k|k-1}}{\partial \theta}}{\det\,\Sigma_{k|k-1}} + (\frac{\partial \hat{X}_{k|k-1}}{\partial \theta})^*\,\Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1}) \\ &- (x - \hat{X}_{k|k-1})^*\,\frac{\partial \Sigma_{k|k-1}^{-1}}{\partial \theta}\,(x - \hat{X}_{k|k-1}) + (x - \hat{X}_{k|k-1})^*\,\Sigma_{k|k-1}^{-1}\,\frac{\partial \hat{X}_{k|k-1}}{\partial \theta} \Bigg] \;, \end{aligned}$$

and

$$w_{k|k-1}(dx) = \frac{\partial \mu_{k|k-1}}{\partial \theta}(dx) = \frac{\partial \log p_{k|k-1}}{\partial \theta}(x)\,\mu_{k|k-1}(dx) \;,$$

where

$$\begin{aligned} \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} &= \frac{\partial F}{\partial \theta}\,\hat{X}_{k-1} + F\,\frac{\partial \hat{X}_{k-1}}{\partial \theta} \;, \\ \frac{\partial \Sigma_{k|k-1}}{\partial \theta} &= \frac{\partial F}{\partial \theta}\,\Sigma_{k-1} F^* + F\,\frac{\partial \Sigma_{k-1}}{\partial \theta}\,F^* + F\,\Sigma_{k-1}\,(\frac{\partial F}{\partial \theta})^* \;, \\ \frac{\dfrac{\partial \det\,\Sigma_{k|k-1}}{\partial \theta}}{\det\,\Sigma_{k|k-1}} &= \text{trace}[\,\frac{\partial \Sigma_{k|k-1}}{\partial \theta}\,\Sigma_{k|k-1}^{-1}\,] \;, \end{aligned}$$

and

$$\frac{\partial \Sigma_{k|k-1}^{-1}}{\partial \theta} = -\Sigma_{k|k-1}^{-1} \, \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \, \Sigma_{k|k-1}^{-1} \ .$$

The last equality comes from the differentiation of $\Sigma_{k|k-1} \, \Sigma_{k|k-1}^{-1} = I$. We have used the following result :

Proposition. *If U is an invertible $d \times d$ matrix, then*

$$\frac{\dfrac{\partial \det U}{\partial \theta}}{\det U} = \mathrm{trace}[\, \frac{\partial U}{\partial \theta} \, U^{-1} \,] \ .$$

Proof. Writing $U = (\vec{u}_1, \cdots, \vec{u}_d)$, we have

$$\vec{u}_1 \wedge \cdots \wedge \vec{u}_d = \det U \; (\vec{e}_1 \wedge \cdots \wedge \vec{e}_d) \ ,$$

where $(\vec{e}_1, \cdots, \vec{e}_d)$ is the canonical basis. Differentiating w.r.t. the parameter yields

$$\begin{aligned} \frac{\partial}{\partial \theta}(\vec{u}_1 \wedge \cdots \wedge \vec{u}_d) &= \frac{\partial \det U}{\partial \theta} \; (\vec{e}_1 \wedge \cdots \wedge \vec{e}_d) \\ &= \sum_{i=1}^{d} \vec{u}_1 \wedge \cdots \wedge \vec{u}_{i-1} \wedge \frac{\partial \vec{u}_i}{\partial \theta} \wedge \vec{u}_{i+1} \wedge \cdots \wedge \vec{u}_d \ . \end{aligned}$$

Let $\displaystyle\frac{\partial \vec{u}_i}{\partial \theta} = \sum_{j=1}^{n} v_{ij} \, \vec{u}_j$, hence

$$\begin{aligned} \frac{\partial}{\partial \theta}(\vec{u}_1 \wedge \cdots \wedge \vec{u}_d) &= \sum_{i,j=1}^{d} v_{ij} \, (\vec{u}_1 \wedge \cdots \wedge \vec{u}_{i-1} \wedge \vec{u}_i \wedge \vec{u}_j \wedge \vec{u}_{i+1} \wedge \cdots \wedge \vec{u}_d \\ &= [\sum_{i=1}^{n} v_{ii} \,] \; \det U \; (\vec{e}_1 \wedge \cdots \wedge \vec{e}_d) \ , \end{aligned}$$

and it follows that

$$\frac{\partial \det U}{\partial \theta} = [\sum_{i=1}^{n} v_{ii} \,] \; \det U \ .$$

Writing $\displaystyle\vec{u}_i = \sum_{k=1}^{d} u_{ij} \, \vec{e}_k$, and $\displaystyle\vec{e}_k = \sum_{j=1}^{d} u^{kj} \, \vec{u}_j$, with $U^{-1} = (u^{kj})_{1 \leq j,k \leq d}$. we have

$$\frac{\partial \vec{u}_i}{\partial \theta} = \sum_{k=1}^{d} \frac{\partial u_{ik}}{\partial \theta} \, \vec{e}_k = \sum_{j=1}^{d} [\sum_{k=1}^{d} \frac{\partial u_{ik}}{\partial \theta} \, u^{kj} \,] \, \vec{u}_j \ ,$$

hence

$$v_{ij} = \sum_{k=1}^{d} \frac{\partial u_{ij}}{\partial \theta} \, u^{kj} \qquad \text{and} \qquad \sum_{i=1}^{n} v_{ii} = \sum_{i,k=1}^{d} \frac{\partial u_{ik}}{\partial \theta} \, u^{ki} = \mathrm{trace}[\, \frac{\partial U}{\partial \theta} \, U^{-1} \,] \ .$$

□

The correction step reads

$$X_k \mid Y_1, \cdots, Y_k \sim \mathcal{N}(\hat{X}_k, \Sigma_k) \ ,$$

i.e.

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_k] = \underbrace{\frac{1}{2\,\pi\,\sqrt{\det\,\Sigma_k}}\ \exp\{-\frac{1}{2}\,(x-\hat{X}_k)^*\,\Sigma_k^{-1}\,(x-\hat{X}_k)\}}_{p_k(x)}\ dx\ ,$$

where the Kalman gain is defined by

$$K_k = \Sigma_{k|k-1}\,H^*\left[H\,\Sigma_{k|k-1}\,H^* + \Sigma\right]^{-1}$$

and where

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k\,(Y_k - H\,\hat{X}_{k|k-1})$$

and

$$\Sigma_k = (I - K_k\,H)\,\Sigma_{k|k-1}\ .$$

Therefore

$$\begin{aligned}\frac{\partial \log p_k}{\partial\theta}(x) \;=\; \tfrac{1}{2}\Bigg[& -\frac{\dfrac{\partial \det\,\Sigma_k}{\partial\theta}}{\det\,\Sigma_k} + (\frac{\partial\hat{X}_k}{\partial\theta})^*\,\Sigma_k^{-1}\,(x-\hat{X}_k) \\ & - (x-\hat{X}_k)^*\,\frac{\partial\Sigma_k^{-1}}{\partial\theta}\,(x-\hat{X}_k) + (x-\hat{X}_k)^*\,\Sigma_k^{-1}\,\frac{\partial\hat{X}_k}{\partial\theta}\Bigg]\ ,\end{aligned}$$

and

$$w_k(dx) = \frac{\partial\mu_k}{\partial\theta}(dx) = \frac{\partial\log p_k}{\partial\theta}(x)\ \mu_k(dx)\ ,$$

where

$$\begin{aligned}\frac{\partial\hat{X}_k}{\partial\theta} &= (I-K_k\,H)\,\frac{\partial\hat{X}_{k|k-1}}{\partial\theta} + \frac{\partial K_k}{\partial\theta}\,(Y_k - H\,\hat{X}_{k|k-1})\ ,\\ \frac{\partial\Sigma_k^{-1}}{\partial\theta} &= -\Sigma_k^{-1}\,\frac{\partial\Sigma_k}{\partial\theta}\,\Sigma_k^{-1}\ ,\\ \frac{\dfrac{\partial\det\,\Sigma_k}{\partial\theta}}{\det\,\Sigma_k} &= \mathrm{trace}[\,\frac{\partial\Sigma_k}{\partial\theta}\,\Sigma_k^{-1}\,]\ ,\\ \frac{\partial\Sigma_k}{\partial\theta} &= (I-K_k\,H)\,\frac{\partial\Sigma_{k|k-1}}{\partial\theta}\ ,\end{aligned}$$

and

$$\frac{\partial K_k}{\partial\theta} = -[\,\frac{\partial\Sigma_k}{\partial\theta}\,\Sigma_{k|k-1}^{-1} + \Sigma_k\,\frac{\partial\Sigma_{k|k-1}}{\partial\theta}\,]\,H^{-1}\ .$$

The last equality is obvious since $\Sigma_k\,\Sigma_{k|k-1}^{-1} = I - K_k\,H$, and we have used the following result :

Proposition. *If Q and R are two symmetric, square and positive definite matrices, then*

$$(H^*\,R^{-1}\,H + Q^{-1})^{-1} = Q - Q\,H^*\,(H^*\,Q\,H + R)^{-1}\,H\,Q\ .$$

Proof. We have

$$H^* Q H + R \geq R \qquad \text{and} \qquad H^* R^{-1} H + Q^{-1} \geq Q^{-1}$$

hence $H^* Q H + R$ and $H^* R^{-1} H + Q^{-1}$ are invertible. It is now easy to check that

$$\begin{aligned} & [Q - Q H^* (H^* Q H + R)^{-1} H Q] \, [H^* R^{-1} H + Q^{-1}] \\ = & \; Q H^* R^{-1} H + I - Q H^* (H^* Q H + R)^{-1} (H^* Q H + R - R) R^{-1} H - Q H^* (H^* Q H + R)^{-1} H \\ = & \; I \; . \end{aligned}$$

□

We have

$$\Sigma_k = \Sigma_{k|k-1} - \Sigma_{k|k-1} H^* (H \Sigma_{k|k-1} H^* + S)^{-1} H \Sigma_{k|k-1} \; ,$$

hence

$$\Sigma_k^{-1} = H^* S^{-1} H + \Sigma_{k|k-1}^{-1} \; ,$$

and

$$\frac{\partial \Sigma_k^{-1}}{\partial \theta} = \frac{\partial \Sigma_{k|k-1}^{-1}}{\partial \theta} = -\Sigma_{k|k-1}^{-1} \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \Sigma_{k|k-1}^{-1} \; .$$

It follows from

$$\Sigma_k \Sigma_{k|k-1}^{-1} = I - K_k H \; ,$$

that

$$\frac{\partial \Sigma_k}{\partial \theta} = (I - K_k H) \frac{\partial \Sigma_{k|k-1}}{\partial \theta} (I - K_k H)^* \; .$$

From the decomposition

$$Y_k = H \hat{X}_{k|k-1} + I_k \; ,$$

where the innovation

$$I_k = Y_k - H \hat{X}_{k|k-1} = H (X_k - \hat{X}_{k|k-1}) + V_k \sim \mathcal{N}(0, S_k) \qquad \text{with} \qquad S_k = H \Sigma_{k|k-1} H^* + S \; ,$$

is independent of the past observations $Y_1, \cdots, Y_{k-1}$, it follows that

$$Y_k \mid Y_1, \cdots, Y_{k-1} \sim \mathcal{N}(H \hat{X}_{k|k-1}, S_k) \; ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] = \underbrace{\frac{1}{2 \pi \sqrt{\det S_k}} \; \exp\{-\tfrac{1}{2} (y - H \hat{X}_{k|k-1})^* S_k^{-1} (y - H \hat{X}_{k|k-1})\}}_{g_k(y)} \; dy \; .$$

Using the straightforward identity

$$\mathbb{P}[Y_1 \in dy_1, \cdots, Y_n \in dy_n] = \prod_{k=1}^{n} \mathbb{P}[Y_k \in dy_k \mid Y_1 = y_1, \cdots, Y_{k-1} = y_{k-1}] = \prod_{k=1}^{n} g_k(y_k) \, dy_k \; ,$$

yields the following expression for the (suitably normalized) log–likelihood function

$$\ell_n = \frac{1}{n} \log \prod_{k=1}^{n} g_k(Y_k) = \frac{1}{n} \sum_{k=1}^{n} \log g_k(Y_k)$$
$$= \text{cste} - \frac{1}{n} \sum_{k=1}^{n} [\, \tfrac{1}{2} \log \det S_k + \tfrac{1}{2} (y - H \hat{X}_{k|k-1})^* S_k^{-1} (y - H \hat{X}_{k|k-1}) \,] \; ,$$

and for the score function

$$\frac{\partial \ell_n}{\partial \theta} = \frac{1}{2n} \sum_{k=1}^{n} \Bigg[-\frac{\dfrac{\partial \det S_k}{\partial \theta}}{\det S_k} + (H \frac{\partial \hat{X}_{k|k-1}}{\partial \theta})^* S_k^{-1} (y - H \hat{X}_{k|k-1})$$
$$- (y - H \hat{X}_{k|k-1})^* \frac{\partial S_k^{-1}}{\partial \theta} (y - H \hat{X}_{k|k-1}) + (y - H \hat{X}_{k|k-1})^* S_k^{-1} H \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} \Bigg] \; ,$$

where

$$\frac{\partial S_k}{\partial \theta} = H \frac{\partial \Sigma_{k|k-1}}{\partial \theta} H^* \qquad \text{and} \qquad \frac{\partial S_k^{-1}}{\partial \theta} = -S_k^{-1} \frac{\partial S_k}{\partial \theta} S_k^{-1} \; .$$

Notice that in full generality

$$\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] = \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy, X_k \in dx \mid Y_1, \cdots, Y_{k-1}]$$
$$= \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy \mid X_k = x] \, \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_{k-1}]$$
$$= \underbrace{[\int_{-\infty}^{\infty} g(x,y) \, \mu_{k|k-1}(dx) \,]}_{g_k(y)} \, dy \; ,$$

hence the following equivalent expression for the log–likelihood function

$$\ell_n = \frac{1}{n} \sum_{k=1}^{n} \log g_k(Y_k) = \frac{1}{n} \sum_{k=1}^{n} \log \int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx) \; , \tag{25}$$

and for the score function

$$\frac{\partial \ell_n}{\partial \theta} = \frac{1}{n} \sum_{k=1}^{n} \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \frac{\partial \mu_{k|k-1}}{\partial \theta}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx)} = \frac{1}{n} \sum_{k=1}^{n} \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, w_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx)} \; . \tag{26}$$

4.3.2 Preliminary computations

We have to design a particle approximation scheme for the derivative w.r.t. $f_{11}, f_{12}, f_{21}, f_{22}$. The transition log–density is

$$\log q(x, x') = \text{cste} - \tfrac{1}{2} \log \det \Sigma - \tfrac{1}{2} (x' - F\, x)^* \Sigma^{-1} (x' - F\, x) \; ,$$

with $\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $x' = \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$, hence the logarithmic derivative w.r.t. the parameter θ is

$$I(x, x') = \frac{\partial \log q}{\partial \theta}(x, x') = \tfrac{1}{2} [\, (\frac{\partial F}{\partial \theta} x)^* \Sigma^{-1} (x' - F\, x) + (x' - F\, x)^* \Sigma^{-1} \frac{\partial F}{\partial \theta} x \,] \; ,$$

and w.r.t. the parameters $f_{11}, f_{12}, f_{21}, f_{22}$

$$\begin{aligned}
\frac{\partial \log q}{\partial f_{11}}(x,x') &= x_1\, \Sigma_{11}\, (x'_1 - f_{11}\, x_1 - f_{12}\, x_2) + x_1\, \Sigma_{12}\, (x'_2 - f_{21}\, x_1 - f_{22}\, x_2)\ , \\
\frac{\partial \log q}{\partial f_{12}}(x,x') &= x_2\, \Sigma_{11}\, (x'_1 - f_{11}\, x_1 - f_{12}\, x_2) + x_2\, \Sigma_{12}\, (x'_2 - f_{21}\, x_1 - f_{22}\, x_2)\ , \\
\frac{\partial \log q}{\partial f_{21}}(x,x') &= x_1\, \Sigma_{12}\, (x'_1 - f_{11}\, x_1 - f_{12}\, x_2) + x_1\, \Sigma_{22}\, (x'_2 - f_{21}\, x_1 - f_{22}\, x_2)\ , \\
\frac{\partial \log q}{\partial f_{22}}(x,x') &= x_2\, \Sigma_{12}\, (x'_1 - f_{11}\, x_1 - f_{12}\, x_2) + x_2\, \Sigma_{22}\, (x'_2 - f_{21}\, x_1 - f_{22}\, x_2)\ ,
\end{aligned}$$

and

$$\Gamma(x,dx') = \frac{\partial Q}{\partial \theta}(x,dx') = I(x,x')\; Q(x,dx')$$

where $\theta \in \{f_{11}, f_{12}, f_{21}, f_{22}\}$. Differentiating with respect to the parameter θ, throughout the recursions

$$\mu_k(dx) = (\Psi_k \cdot \mu_{k|k-1})(dx) = \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}\ , \tag{27}$$

and

$$\mu_{k+1|k}(dx) = Q\, \mu_k(dx') = \int_{-\infty}^{\infty} \mu_k(dx)\, Q(x,dx')\ , \tag{28}$$

yields

$$w_k(dx) = \frac{\partial \mu_k}{\partial \theta}(dx) = \frac{\Psi_k(x)\, \dfrac{\partial \mu_{k|k-1}}{\partial \theta}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \dfrac{\partial \mu_{k|k-1}}{\partial \theta}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}\; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}\ ,$$

hence

$$\begin{aligned}
w_k(dx) &= (F_k(\mu_{k|k-1}) w_{k|k-1})(dx) \\
&= \frac{\Psi_k(x)\, w_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, w_{k|k-1}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}\; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}\ ,
\end{aligned} \tag{29}$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, w_k(dx) + I(x,x')\; \mu_k(dx)\,]\; Q(x,dx')\ . \tag{30}$$

Since $w_{k|k-1} \ll \mu_{k|k-1}$, it holds $F_k(\mu_{k|k-1}) w_{k|k-1} \ll \Psi_k \cdot \mu_{k|k-1}$, with Radon–Nikodym derivative

$$\frac{dw_k}{d\mu_k}(x) = \frac{d(F_k(\mu_{k|k-1}) w_{k|k-1})}{d(\Psi_k \cdot \mu_{k|k-1})}(x) = \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x) - \int_{-\infty}^{\infty} \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x')\; (\Psi_k \cdot \mu_{k|k-1})(dx')\ ,$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, \frac{dw_k}{d\mu_k}(x) + I(x,x')\,]\; \mu_k(dx)\; Q(x,dx')\ .$$

4.3.3 A particle approximation scheme

Assuming that the following particle approximations

$$\mu_{k|k-1} \approx \mu^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i_{k|k-1}} \qquad \text{and} \qquad w_{k|k-1} \approx w^N_{k|k-1} = \frac{1}{N} \sum_{i=1}^N \rho^i_{k|k-1} \, \delta_{\xi^i_{k|k-1}} \; , \tag{31}$$

are available at time index k, and plugging these approximations into equations (27) and (29), yields

$$\mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} = \frac{\displaystyle\sum_{i=1}^N \Psi_k(\xi^i_{k|k-1}) \, \delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} = \sum_{i=1}^N \omega^i_k \, \delta_{\xi^i_{k|k-1}} \; ,$$

and

$$\begin{aligned} w^N_k = F_k(\mu^N_{k|k-1}) w^N_{k|k-1} &= \frac{\displaystyle\sum_{i=1}^N \rho^i_{k|k-1} \, \Psi_k(\xi^i_{k|k-1}) \, \delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} - \frac{\displaystyle\sum_{j=1}^N \rho^j_{k|k-1} \, \Psi_k(\xi^j_{k|k-1})}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} \, \frac{\displaystyle\sum_{i=1}^N \Psi_k(\xi^i_{k|k-1}) \, \delta_{\xi^i_{k|k-1}}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi^j_{k|k-1})} \\ &= \sum_{i=1}^N \rho^i_{k|k-1} \, \omega^i_k \, \delta_{\xi^i_{k|k-1}} - [\sum_{j=1}^N \rho^j_{k|k-1} \, \omega^j_k] \, \sum_{i=i}^N \omega^i_k \, \delta_{\xi^i_{k|k-1}} \\ &= \sum_{i=1}^N [\, \rho^i_{k|k-1} - [\sum_{j=1}^N \rho^j_{k|k-1} \, \omega^j_k]\,] \, \omega^i_k \, \delta_{\xi^i_{k|k-1}} \; . \end{aligned}$$

Plugging these expressions into equations (28) and (30), yields

$$Q \, \mu^N_k(dx') = \sum_{i=1}^N \omega^i_k \, Q(\xi^i_{k|k-1}, dx') \; , \tag{32}$$

and

$$Q \, w^N_k(dx') + \Gamma \, \mu^N_k(dx') = \sum_{i=1}^N [\, \rho^i_{k|k-1} - [\sum_{j=1}^N \rho^j_{k|k-1} \, \omega^j_k] + I(\xi^i_{k|k-1}, x')\,] \, \omega^i_k \, Q(\xi^i_{k|k-1}, dx') \; , \tag{33}$$

which can be interpreted as the marginals of the finite signed measures $m = (m^i \, , \, i = 1, \cdots, N)$ and $s = (s^i \, , \, i = 1, \cdots, N)$ defined on the product space $E^N = \{1, \cdots, N\} \times \mathbb{R}$ by

$$s^i(dx') = [\underbrace{\rho^i_{k|k-1} + I(\xi^i_{k|k-1}, x')}_{r^i(x')} - [\sum_{j=1}^N \rho^j_{k|k-1} \, \omega^j_k]\,] \, \underbrace{\omega^i_k \, Q(\xi^i_{k|k-1}, dx')}_{m^i(dx')} \; .$$

Notice that

$$\sum_{j=1}^N \int_{-\infty}^{\infty} r^j(x') \, m^j(dx') = \sum_{j=1}^N \rho^j_{k|k-1} \, \omega^j_k \; ,$$

is just a normalizing constant. Introducing the particle approximation

$$m \approx m^N = \frac{1}{N} \sum_{i=1}^N \delta_{(\tau^i_k, \, \xi^i_{k+1|k})} \; ,$$

and the weighted particle approximation

$$s \approx s^N = \frac{1}{N} \sum_{i=1}^N [\, r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,]\,]\; \delta_{(\tau_k^i,\, \xi_{k+1|k}^i)} \ ,$$

where independently for any $i = 1, \cdots, N$, the pair $(\tau_k^i, \xi_{k+1|k}^i)$ is jointly distributed according to the probability distribution $m = (m^i \, , \, i = 1, \cdots, N)$, i.e.

$$\tau_k^i \sim (\omega_k^j \, , \, j = 1, \cdots, N) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\xi_{k|k-1}^{\tau_k^i}, dx') \ ,$$

and marginalizing, yields

$$\mu_{k+1|k} \approx \mu_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1|k}^i} \ .$$

$$w_{k+1|k} \approx w_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N [\, r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,]\,]\; \delta_{\xi_{k+1|k}^i} = \frac{1}{N} \sum_{i=1}^N \rho_{k+1|k}^i \; \delta_{\xi_{k+1|k}^i} \ .$$

Notice that the weight

$$\rho_{k+1|k}^i = r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,] \ ,$$

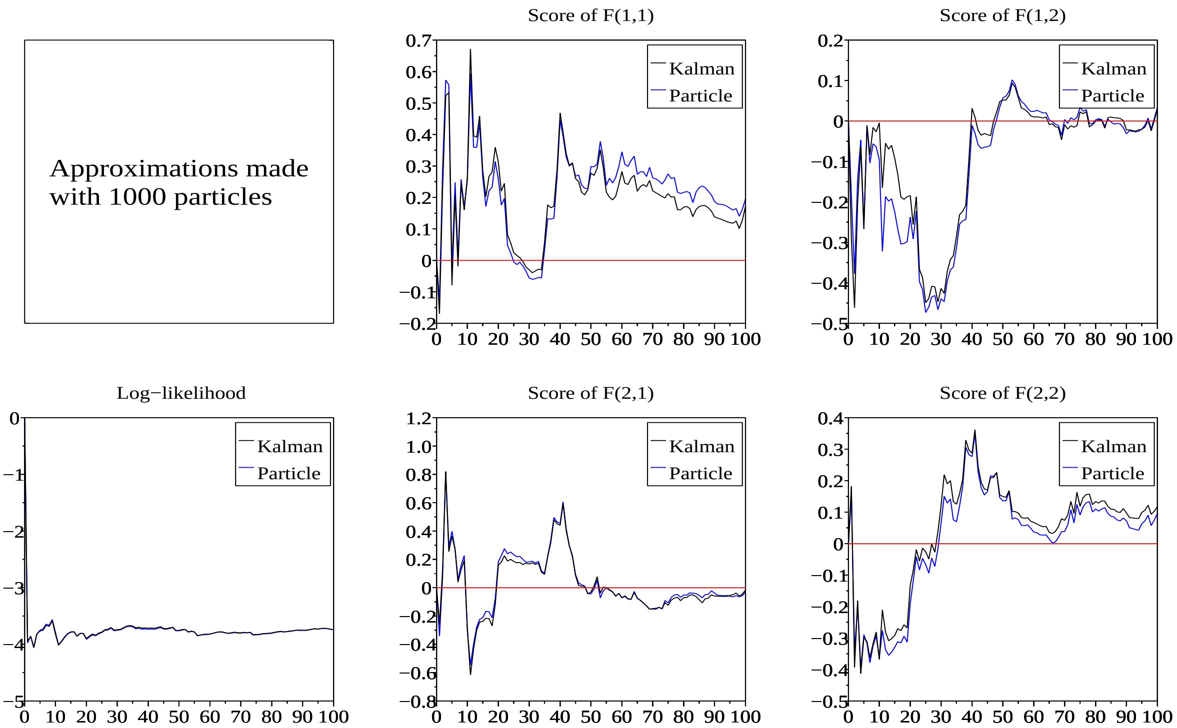
where

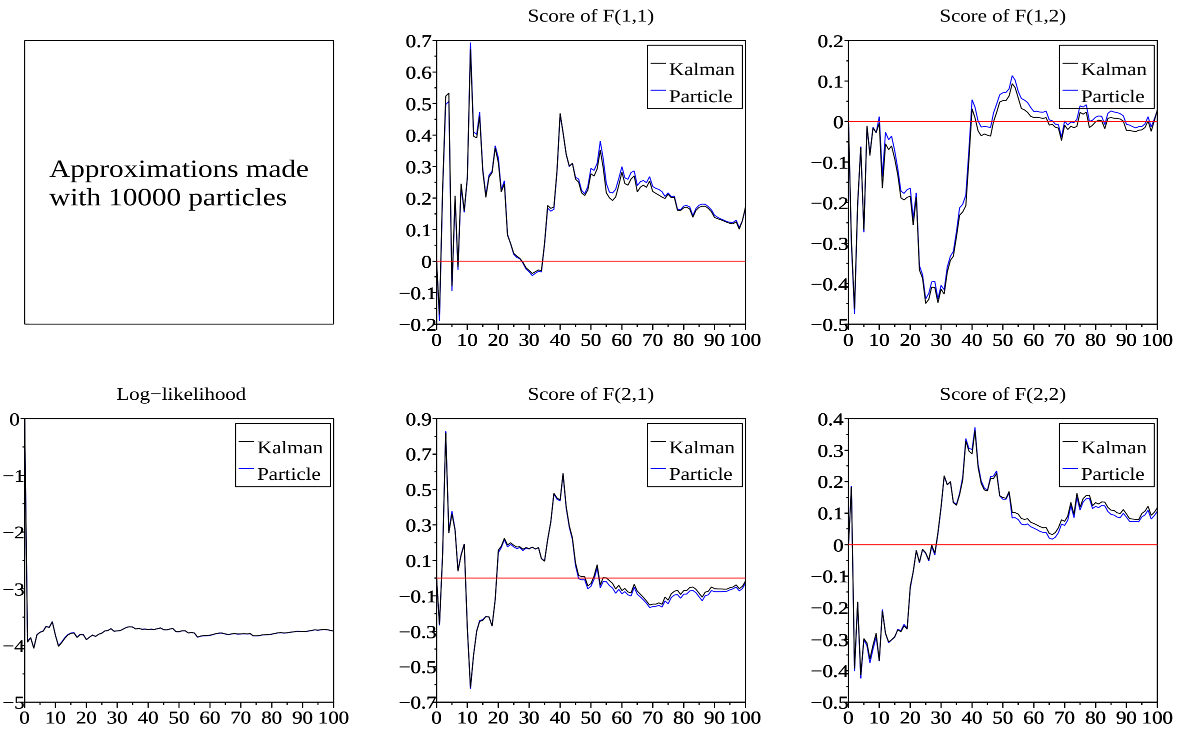
$$r^{\tau_k^i}(\xi_{k+1|k}^i) = \rho_{k|k-1}^{\tau_k^i} + I(\xi_{k|k-1}^{\tau_k^i}, \xi_{k+1|k}^i) \ ,$$

does not depend only on the position $\xi_{k+1|k}^i$, but on the position of the pair $(\xi_{k|k-1}^{\tau_k^i}, \xi_{k+1|k}^i)$.

Numerical results

$\bar{X}_0$	Σ_0	F	Σ	H	S	N (# particles)
$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.7 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1000, 5000, 10000, 50000


Approximations made with 1000 particles
Score of F(1,1)
Score of F(1,2)
Log−likelihood
Score of F(2,1)
Score of F(2,2)
Kalman
Particle


Approximations made with 10000 particles
Score of F(1,1)
Score of F(1,2)
Log−likelihood
Score of F(2,1)
Score of F(2,2)
Kalman
Particle

4.3.4 An alternate particle approximation scheme

Equations (32) and (33) read

$$Q\,\mu_k^N(dx') = \sum_{i=1}^N \omega_k^i\; Q(\xi_{k|k-1}^i, dx') = [\sum_{i=1}^N \omega_k^i\; q(\xi_{k|k-1}^i, x')]\; dx'\ ,$$

and

$$\begin{aligned} Q\,w_k^N(dx') + \Gamma\,\mu_k^N(dx') \;&=\; \sum_{i=1}^N [\,\rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j\;\omega_k^j\,] + I(\xi_{k|k-1}^i, x')\,]\;\omega_k^i\; Q(\xi_{k|k-1}^i, dx') \\ &=\; [\sum_{i=1}^N [\,\rho_{k|k-1}^i - [\sum_{j=1}^N \rho_{k|k-1}^j\;\omega_k^j\,] + I(\xi_{k|k-1}^i, x')\,]\;\omega_k^i\; q(\xi_{k|k-1}^i, x')\,]\; dx' \\ &=\; r_{k+1}^N(x')\; Q\,\mu_k^N(dx')\ , \end{aligned}$$

where

$$\begin{aligned} r_{k+1}^N(x') \;&=\; \frac{\displaystyle\sum_{j=1}^N [\,\rho_{k|k-1}^j + I(\xi_{k|k-1}^j, x')\,]\;\omega_k^j\; q(\xi_{k|k-1}^j, x')}{\displaystyle\sum_{j=1}^N \omega_k^j\; q(\xi_{k|k-1}^j, x')} - [\sum_{j=1}^N \rho_{k|k-1}^j\;\omega_k^j\,] \\ &=\; \frac{\displaystyle\sum_{j=1}^N [\,\rho_{k|k-1}^j + I(\xi_{k|k-1}^j, x')\,]\;\omega_k^j\;\exp\{-\tfrac{1}{2}\,(x' - F\,\xi_{k|k-1}^j)^*\,\Sigma^{-1}\,(x' - F\,\xi_{k|k-1}^j)\}}{\displaystyle\sum_{j=1}^N \omega_k^j\;\exp\{-\tfrac{1}{2}\,(x' - F\,\xi_{k|k-1}^j)^*\,\Sigma^{-1}\,(x' - F\,\xi_{k|k-1}^j)\}} - [\sum_{j=1}^N \rho_{k|k-1}^j\;\omega_k^j\,]\ . \end{aligned}$$

Resampling yields the particle approximation

$$Q\,\mu_k^N \approx \mu_{k+1|k}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{k+1|k}^i}\ ,$$

where independently for any $i = 1, \cdots, N$

$$\xi_{k+1|k}^i \sim Q\,\mu_k^N(dx')\ ,$$

which can be achieved for instance by taking

$$\widehat{\xi}_k^i \sim \mu_k^N(dx) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx')\ ,$$

or even more explicitly by taking

$$\tau_k^i \sim (\omega_k^j\,,\, j = 1, \cdots, N)\ , \qquad \widehat{\xi}_k^i = \xi_{k|k-1}^{\tau_k^i} \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx')\ ,$$

and the weighted particle approximation

$$Q\,w_k^N + \Gamma\,\mu_k^N \approx w_{k+1|k}^N = \frac{1}{N}\sum_{i=1}^N r_{k+1}^N(\xi_{k+1|k}^i)\;\delta_{\xi_{k+1|k}^i} = \frac{1}{N}\sum_{i=1}^N \rho_{k+1|k}^i\;\delta_{\xi_{k+1|k}^i}\ ,$$

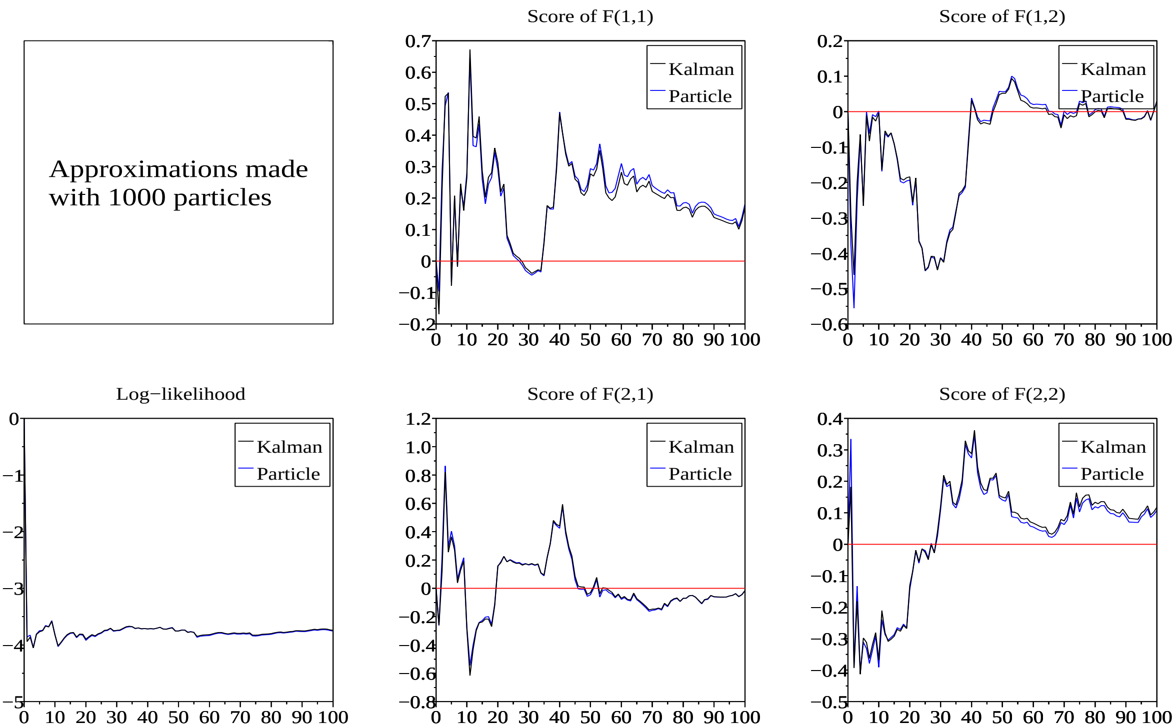
where the weight

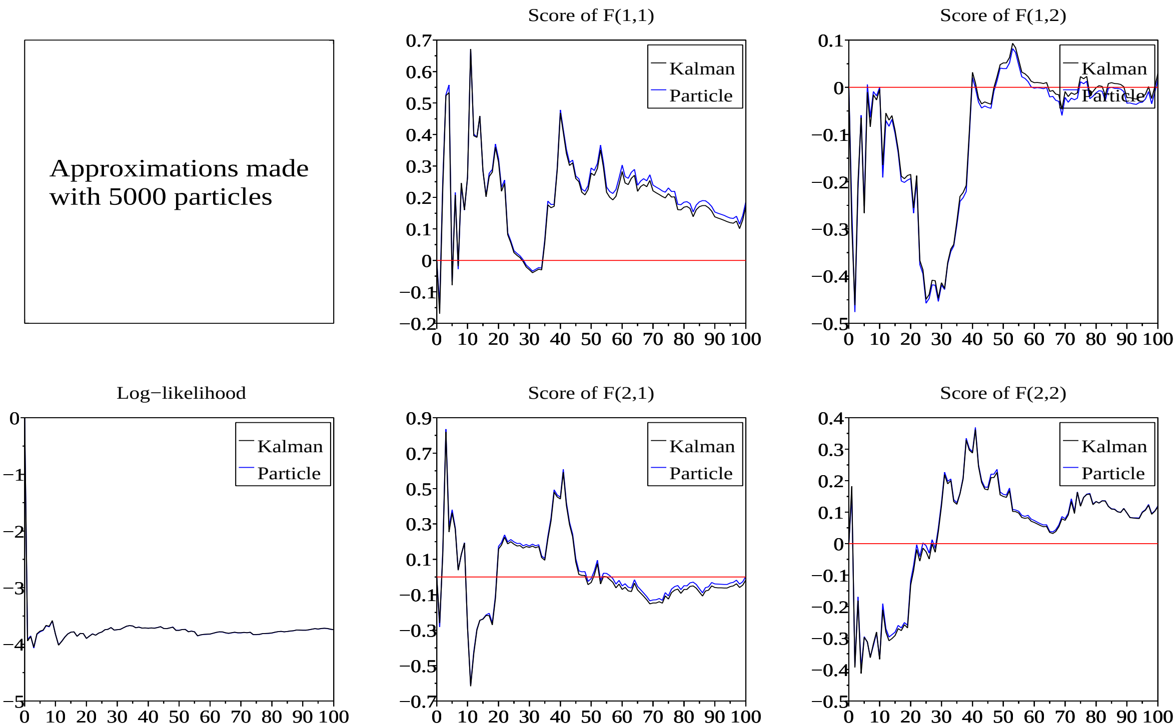
$$
\begin{aligned}
\rho^i_{k+1|k} &= r^N_{k+1}(\xi^i_{k+1|k}) \\
&= \frac{\displaystyle\sum_{j=1}^{N} [\,\rho^j_{k|k-1} + I(\xi^j_{k|k-1}, \xi^i_{k+1|k})\,]\; \omega^j_k \;\exp\{-\tfrac{1}{2}\,(\xi^i_{k+1|k} - F\,\xi^j_{k|k-1})^*\,\Sigma^{-1}\,(\xi^i_{k+1|k} - F\,\xi^j_{k|k-1})\}}{\displaystyle\sum_{j=1}^{N} \omega^j_k \;\exp\{-\tfrac{1}{2}\,(\xi^i_{k+1|k} - F\,\xi^j_{k|k-1})^*\,\Sigma^{-1}\,(\xi^i_{k+1|k} - F\,\xi^j_{k|k-1})\}} \\
&\qquad - [\sum_{j=1}^{N} \rho^j_{k|k-1}\;\omega^j_k\,]\;,
\end{aligned}
$$

depends only on the position of $\xi^i_{k+1|k}$.

Numerical results

X_0	Σ_0	F	Σ	H	S	N (# particles)
$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.7 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1000, 5000





4.4 Part D : General case

In this last section, we consider a more general model :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & F\,X_k + W_k \\ Y_k & = & H\,X_k + V_k\ , \end{array} \right. \tag{34}$$

where $\{W_k\,,\,k \geq 0\}$ and $\{V_k\,,\,k \geq 0\}$ are standard Gaussian independent white noise sequences, with covariance matrix Σ and S respectively. Here

$$F = \begin{pmatrix} f_{11} & \cdots & f_{1m'} \\ \vdots & \ddots & \vdots \\ f_{m'1} & \cdots & f_{m'm'} \end{pmatrix} \qquad \text{and} \qquad H = \begin{pmatrix} h_{11} & \cdots & h_{1m} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h_{r1} & \cdots & h_{rm} & 0 & \cdots & 0 \end{pmatrix}$$

are considered as unknown parameters and $X_0 \sim \mathcal{N}(\bar{X}_0, \Sigma_0)$. As X is a vector with $2m$ rows, we set $m' = 2m$. It follows from the model that

$$X_{k+1} \mid X_k = x \sim \mathcal{N}(F\,x, \Sigma)\ ,$$

i.e.

$$Q(x, dx') = \mathbb{P}[X_{k+1} \in dx' \mid X_k = x] = \underbrace{\frac{1}{(2\,\pi)^{m'/2}\,\sqrt{\det\,\Sigma}}\ \exp\{-\frac{1}{2}\,(x' - F\,x)^*\,\Sigma^{-1}\,(x' - F\,x)\}}_{q(x,x')}\ dx'\ ,$$

and

$$Y_k \mid X_k = x \sim \mathcal{N}(H\,x, S)\ ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = \underbrace{\frac{1}{(2\,\pi)^{r/2}\,\sqrt{\det\,S}}\ \exp\{-\frac{1}{2}\,(y - H\,x)^*\,S^{-1}\,(y - H\,x)\}}_{g(x,y)}\ dy\ ,$$

and let

$$\Psi_k(x) = g(x, Y_k) \ ,$$

denote the *likelihood function.* The goal here is to design a particle approximation scheme for the derivative, with respect to the parameters $f_{11}, \cdots, f_{m'm'}, h_{11}, \cdots, h_{rm}$, of the conditional probability distribution of the hidden state, given the observations.

4.4.1 Exact expressions (Kalman filter)

In such a simple model, an exact expression can easily be obtained, via the prediction / correction steps of the Kalman filter framework. Indeed, the prediction step reads

$$X_k \mid Y_1, \cdots, Y_{k-1} \sim \mathcal{N}(\hat{X}_{k|k-1}, \Sigma_{k|k-1}) \ ,$$

i.e.

$$\begin{aligned} \mu_{k|k-1}(dx) &= \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\ &= \underbrace{\frac{1}{(2\,\pi)^{m'/2}\,\sqrt{\det\,\Sigma_{k|k-1}}}\;\exp\{-\frac{1}{2}\,(x - \hat{X}_{k|k-1})^*\,\Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1})\}}_{p_{k|k-1}(x)}\;dx \ , \end{aligned}$$

where

$$\hat{X}_{k|k-1} = F\,\hat{X}_{k-1} \qquad \text{and} \qquad \Sigma_{k|k-1} = F\,\Sigma_{k-1}\,F^* + \Sigma \ .$$

The log–density of the r.v. $X_k \mid Y_1, \cdots, Y_{k-1}$ is

$$\log p_{k|k-1}(x) = \text{cste} - \tfrac{1}{2}\,\log\,\det\,\Sigma_{k|k-1} - \tfrac{1}{2}\,(x - \hat{X}_{k|k-1})^*\,\Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1}) \ ,$$

hence the logarithmic derivative w.r.t. the parameter θ where $\theta \in \{f_{11}, \cdots, f_{rm'}, h_{11}, \cdots, h_{rm}\}$

$$\begin{aligned} \frac{\partial \log p_{k|k-1}}{\partial\theta}(x) &= \tfrac{1}{2}\,\Bigg[-\frac{\dfrac{\partial \det\,\Sigma_{k|k-1}}{\partial\theta}}{\det\,\Sigma_{k|k-1}} + (\frac{\partial \hat{X}_{k|k-1}}{\partial\theta})^*\,\Sigma_{k|k-1}^{-1}\,(x - \hat{X}_{k|k-1}) \\ &\quad - (x - \hat{X}_{k|k-1})^*\,\frac{\partial \Sigma_{k|k-1}^{-1}}{\partial\theta}\,(x - \hat{X}_{k|k-1}) + (x - \hat{X}_{k|k-1})^*\,\Sigma_{k|k-1}^{-1}\,\frac{\partial \hat{X}_{k|k-1}}{\partial\theta} \Bigg] \ , \end{aligned}$$

and

$$w_{k|k-1}(dx) = \frac{\partial \mu_{k|k-1}}{\partial\theta}(dx) = \frac{\partial \log p_{k|k-1}}{\partial\theta}(x)\,\mu_{k|k-1}(dx) \ ,$$

where

$$\begin{aligned} \frac{\partial \hat{X}_{k|k-1}}{\partial\theta} &= \frac{\partial F}{\partial\theta}\,\hat{X}_{k-1} + F\,\frac{\partial \hat{X}_{k-1}}{\partial\theta} \ , \\ \frac{\partial \Sigma_{k|k-1}}{\partial\theta} &= \frac{\partial F}{\partial\theta}\,\Sigma_{k-1}\,F^* + F\,\frac{\partial \Sigma_{k-1}}{\partial\theta}\,F^* + F\,\Sigma_{k-1}\,(\frac{\partial F}{\partial\theta})^* \ , \\ \frac{\dfrac{\partial \det\,\Sigma_{k|k-1}}{\partial\theta}}{\det\,\Sigma_{k|k-1}} &= \text{trace}\,[\,\frac{\partial \Sigma_{k|k-1}}{\partial\theta}\,\Sigma_{k|k-1}^{-1}\,] \ , \end{aligned}$$

and

$$\frac{\partial \Sigma_{k|k-1}^{-1}}{\partial \theta} = -\Sigma_{k|k-1}^{-1} \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \Sigma_{k|k-1}^{-1} .$$

The last equality comes from the differentiation of $\Sigma_{k|k-1} \Sigma_{k|k-1}^{-1} = I$.

The correction step reads

$$X_k \mid Y_1, \cdots, Y_k \sim \mathcal{N}(\hat{X}_k, \Sigma_k) ,$$

i.e.

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_k] = \underbrace{\frac{1}{(2\pi)^{m'/2}\sqrt{\det \Sigma_k}} \exp\{-\frac{1}{2} (x - \hat{X}_k)^* \Sigma_k^{-1} (x - \hat{X}_k)\}}_{p_k(x)} dx ,$$

where the Kalman gain is defined by

$$K_k = \Sigma_{k|k-1} H^* \left[H \Sigma_{k|k-1} H^* + \Sigma\right]^{-1}$$

and where

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k (Y_k - H \hat{X}_{k|k-1})$$

and

$$\Sigma_k = (I - K_k H) \Sigma_{k|k-1} .$$

Therefore

$$\begin{aligned} \frac{\partial \log p_k}{\partial \theta}(x) \;=\; \tfrac{1}{2} \Bigg[& -\frac{\dfrac{\partial \det \Sigma_k}{\partial \theta}}{\det \Sigma_k} + (\frac{\partial \hat{X}_k}{\partial \theta})^* \Sigma_k^{-1} (x - \hat{X}_k) \\ & - (x - \hat{X}_k)^* \frac{\partial \Sigma_k^{-1}}{\partial \theta} (x - \hat{X}_k) + (x - \hat{X}_k)^* \Sigma_k^{-1} \frac{\partial \hat{X}_k}{\partial \theta} \Bigg] , \end{aligned}$$

and

$$w_k(dx) = \frac{\partial \mu_k}{\partial \theta}(dx) = \frac{\partial \log p_k}{\partial \theta}(x) \mu_k(dx) ,$$

where

$$\begin{aligned} \frac{\partial \hat{X}_k}{\partial \theta} &= (I - K_k H) \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} + \frac{\partial K_k}{\partial \theta} (Y_k - H \hat{X}_{k|k-1}) - K_k \frac{\partial H}{\partial \theta} \hat{X}_{k|k-1} , \\ \frac{\partial \Sigma_k^{-1}}{\partial \theta} &= -\Sigma_k^{-1} \frac{\partial \Sigma_k}{\partial \theta} \Sigma_k^{-1} , \\ \frac{\dfrac{\partial \det \Sigma_k}{\partial \theta}}{\det \Sigma_k} &= \mathrm{trace}[\frac{\partial \Sigma_k}{\partial \theta} \Sigma_k^{-1}] , \\ \frac{\partial \Sigma_k}{\partial \theta} &= (I - K_k H) \frac{\partial \Sigma_{k|k-1}}{\partial \theta} - [\frac{\partial K_k}{\partial \theta} H + K_k \frac{\partial H}{\partial \theta}] \Sigma_{k|k-1} , \end{aligned}$$

and

$$\frac{\partial K_k}{\partial \theta} = \Sigma_{k|k-1} (\frac{\partial H}{\partial \theta})^* \, [H \, \Sigma_{k|k-1} \, H^* + \Sigma]^{-1} + \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \, H^* \, [H \, \Sigma_{k|k-1} \, H^* + \Sigma]^{-1}$$
$$- \Sigma_{k|k-1} \, H^* \, [H \, \Sigma_{k|k-1} \, H^* + \Sigma]^{-1}] \, \Big[\frac{\partial H}{\partial \theta} \, \Sigma_{k|k-1} \, H^* + H \, \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \, H^* + H \, \Sigma_{k|k-1} \, (\frac{\partial H}{\partial \theta})^* \Big]$$
$$[H \, \Sigma_{k|k-1} \, H^* + \Sigma]^{-1} \; .$$

From the decomposition

$$Y_k = H \, \hat{X}_{k|k-1} + I_k \; ,$$

where the innovation

$$I_k = Y_k - H \, \hat{X}_{k|k-1} = H \, (X_k - \hat{X}_{k|k-1}) + V_k \sim \mathcal{N}(0, S_k) \qquad \text{with} \qquad S_k = H \, \Sigma_{k|k-1} \, H^* + S \; ,$$

is independent of the past observations $Y_1, \cdots, Y_{k-1}$, it follows that

$$Y_k \mid Y_1, \cdots, Y_{k-1} \sim \mathcal{N}(H \, \hat{X}_{k|k-1}, S_k) \; ,$$

i.e.

$$\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] = \underbrace{\frac{1}{(2\,\pi)^{r/2} \, \sqrt{\det \, S_k}} \; \exp\{-\tfrac{1}{2} \, (y - H \, \hat{X}_{k|k-1})^* \, S_k^{-1} \, (y - H \, \hat{X}_{k|k-1})\}}_{g_k(y)} \; dy \; .$$

Using the straightforward identity

$$\mathbb{P}[Y_1 \in dy_1, \cdots, Y_k \in dy_k] = \prod_{k=1}^{n} \mathbb{P}[Y_k \in dy_k \mid Y_1 = y_1, \cdots, Y_{k-1} = y_{k-1}] = \prod_{k=1}^{n} g_k(y_k) \, dy_k \; ,$$

yields the following expression for the (suitably normalized) log–likelihood function

$$\ell_n = \frac{1}{n} \; \log \prod_{k=1}^{n} g_k(Y_k) \; = \; \frac{1}{n} \; \sum_{k=1}^{n} \log g_k(Y_k)$$
$$= \; \text{cste} - \frac{1}{2n} \; \sum_{k=1}^{n} [\, \log(\det \, S_k) + (y - H \, \hat{X}_{k|k-1})^* \, S_k^{-1} \, (y - H \, \hat{X}_{k|k-1}) \,] \; ,$$

and for the score function

$$\frac{\partial \ell_n}{\partial \theta} = \frac{1}{2n} \sum_{k=1}^{n} \Bigg[- \frac{\dfrac{\partial \det \, S_k}{\partial \theta}}{\det \, S_k} + (H \, \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} + \frac{\partial H}{\partial \theta} \hat{X}_{k|k-1})^* \, S_k^{-1} \, (y - H \, \hat{X}_{k|k-1})$$
$$- (y - H \, \hat{X}_{k|k-1})^* \, \frac{\partial S_k^{-1}}{\partial \theta} \, (y - H \, \hat{X}_{k|k-1})$$
$$+ (y - H \, \hat{X}_{k|k-1})^* \, S_k^{-1} \, (H \, \frac{\partial \hat{X}_{k|k-1}}{\partial \theta} + \frac{\partial H}{\partial \theta} \, \hat{X}_{k|k-1}) \Bigg] \; ,$$

where

$$\frac{\partial S_k}{\partial \theta} = \frac{\partial H}{\partial \theta} \, \Sigma_{k|k-1} \, H^* + H \, \frac{\partial \Sigma_{k|k-1}}{\partial \theta} \, H^* + H \, \Sigma_{k|k-1} \, (\frac{\partial H}{\partial \theta})^* \qquad \text{and} \qquad \frac{\partial S_k^{-1}}{\partial \theta} = -S_k^{-1} \, \frac{\partial S_k}{\partial \theta} \, S_k^{-1} \; .$$

Notice that in full generality

$$\begin{aligned}
\mathbb{P}[Y_k \in dy \mid Y_1, \cdots, Y_{k-1}] &= \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy, X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\
&= \int_{-\infty}^{\infty} \mathbb{P}[Y_k \in dy \mid X_k = x] \, \mathbb{P}[X_k \in dx \mid Y_1, \cdots, Y_{k-1}] \\
&= \underbrace{[\int_{-\infty}^{\infty} g(x,y) \, \mu_{k|k-1}(dx)]}_{g_k(y)} \, dy \; ,
\end{aligned}$$

hence the following equivalent expression for the log–likelihood function

$$\ell_n = \frac{1}{n} \sum_{k=1}^{n} \log g_k(Y_k) = \frac{1}{n} \sum_{k=1}^{n} \log \int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx) \; , \tag{35}$$

and for the score function

$$\begin{aligned}
\frac{\partial \ell_n}{\partial \theta} &= \frac{1}{n} \sum_{k=1}^{n} \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \frac{\partial \mu_{k|k-1}}{\partial \theta}(dx) + \frac{\partial \Psi_k(x)}{\partial \theta} \, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx)} \\
&= \frac{1}{n} \sum_{k=1}^{n} \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, w_{k|k-1}(dx) + \frac{\partial \Psi_k(x)}{\partial \theta} \, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x) \, \mu_{k|k-1}(dx)} \; .
\end{aligned}$$

4.4.2 Preliminary computations

We have to design a particle approximation scheme for the derivative w.r.t. $f_{11}, \cdots, f_{m'm'}, h_{11}, \cdots, h_{rm}$. The transition log–density is

$$\log q(x,x') = \text{cste} - \tfrac{1}{2} \log(\det \Sigma) - \tfrac{1}{2} (x' - F\,x)^* \Sigma^{-1} (x' - F\,x) \; ,$$

hence the logarithmic derivative w.r.t. the parameter θ is

$$I(x,x') = \frac{\partial \log q}{\partial \theta}(x,x') = \tfrac{1}{2} \, [\, (\frac{\partial F}{\partial \theta} x)^* \, \Sigma^{-1} \, (x' - F\,x) + (x' - F\,x)^* \, \Sigma^{-1} \, \frac{\partial F}{\partial \theta} x \,] \; ,$$

and

$$\Gamma(x, dx') = \frac{\partial Q}{\partial \theta}(x, dx') = I(x,x') \; Q(x, dx')$$

where $\theta \in \{f_{11}, \cdots, f_{m'm'}, h_{11}, \cdots, h_{rm}\}$. Differentiating with respect to the parameter θ, throughout the recursions

$$\mu_k(dx) = (\Psi_k \cdot \mu_{k|k-1})(dx) = \frac{\Psi_k(x) \, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x') \, \mu_{k|k-1}(dx')} \; , \tag{36}$$

and

$$\mu_{k+1|k}(dx) = (Q\,\mu_k)(dx') = \int_{-\infty}^{\infty} \mu_k(dx) \, Q(x, dx') \; , \tag{37}$$

yields

$$w_k(dx) = \frac{\partial \mu_k}{\partial \theta}(dx) = \frac{\Psi_k(x)\, \dfrac{\partial \mu_{k|k-1}}{\partial \theta}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \dfrac{\partial \mu_{k|k-1}}{\partial \theta}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')}$$
$$+ \frac{\dfrac{\partial \log \Psi_k}{\partial \theta}(x)\, \Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \dfrac{\partial \log \Psi_k}{\partial \theta}(x')\, \Psi_k(x')\, \mu_{k|k-1}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \,,$$

hence, with $S_k(x) = \dfrac{\partial \log \Psi_k}{\partial \theta}(x)$

$$\begin{aligned} w_k(dx) &= (F_k(\mu_{k|k-1}) w_{k|k-1})(dx) + (G_k(\mu_{k|k-1}))(dx) \\ &= \frac{\Psi_k(x)\, w_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, w_{k|k-1}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \\ &\quad + \frac{S_k(x)\, \Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} - \frac{\displaystyle\int_{-\infty}^{\infty} S_k(x')\, \Psi_k(x')\, \mu_{k|k-1}(dx')}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \; \frac{\Psi_k(x)\, \mu_{k|k-1}(dx)}{\displaystyle\int_{-\infty}^{\infty} \Psi_k(x')\, \mu_{k|k-1}(dx')} \,. \end{aligned}$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, w_k(dx) + I(x,x')\, \mu_k(dx) \,]\; Q(x,dx') \,. \tag{38}$$

Since $w_{k|k-1} \ll \mu_{k|k-1}$, it holds $F_k(\mu_{k|k-1}) w_{k|k-1} \ll \Psi_k \cdot \mu_{k|k-1}$, and $G_k(\mu_{k|k-1}) \ll \Psi_k \cdot \mu_{k|k-1}$ with Radon–Nikodym derivative

$$\frac{dw_k}{d\mu_k}(x) = \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x) + S_k(x) - \int_{-\infty}^{\infty} [\, \frac{dw_{k|k-1}}{d\mu_{k|k-1}}(x') + S_k(x') \,]\; (\Psi_k \cdot \mu_{k|k-1})(dx') \,,$$

and

$$w_{k+1|k}(dx') = Q\, w_k(dx') + \Gamma\, \mu_k(dx') = \int_{-\infty}^{\infty} [\, \frac{dw_k}{d\mu_k}(x) + I(x,x') \,]\; \mu_k(dx)\; Q(x,dx') \,.$$

4.4.3 A particle approximation scheme

Assuming that the following particle approximations

$$\mu_{k|k-1} \approx \mu_{k|k-1}^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{k|k-1}^i} \qquad \text{and} \qquad w_{k|k-1} \approx w_{k|k-1}^N = \frac{1}{N} \sum_{i=1}^{N} \rho_{k|k-1}^i\; \delta_{\xi_{k|k-1}^i} \,, \tag{39}$$

are available at time index k, and plugging these approximations into equations (36) and (38), yields

$$\mu_k^N = \Psi_k \cdot \mu_{k|k-1}^N = \frac{\displaystyle\sum_{i=1}^{N} \Psi_k(\xi_{k|k-1}^i)\; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^{N} \Psi_k(\xi_{k|k-1}^j)} = \sum_{i=1}^{N} \omega_k^i\; \delta_{\xi_{k|k-1}^i} \,,$$

and

$$
\begin{aligned}
w_k^N \;=\; & \frac{\displaystyle\sum_{i=1}^N \rho_{k|k-1}^i \; \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} - \frac{\displaystyle\sum_{j=1}^N \rho_{k|k-1}^j \; \Psi_k(\xi_{k|k-1}^j)}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} \; \frac{\displaystyle\sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} \\
+ & \frac{\displaystyle\sum_{i=1}^N S_k(\xi_{k|k-1}^i) \; \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} - \frac{\displaystyle\sum_{j=1}^N S_k(\xi_{k|k-1}^j) \; \Psi_k(\xi_{k|k-1}^j)}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} \; \frac{\displaystyle\sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \; \delta_{\xi_{k|k-1}^i}}{\displaystyle\sum_{j=1}^N \Psi_k(\xi_{k|k-1}^j)} \\
= & \sum_{i=1}^N [\, \rho_{k|k-1}^i + S_k(\xi_{k|k-1}^i) - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \,] \; \omega_k^i \; \delta_{\xi_{k|k-1}^i} \; .
\end{aligned}
$$

Plugging these expressions into equations (37) and (38), yields

$$
Q\, \mu_k^N(dx') = \sum_{i=1}^N \omega_k^i \; Q(\xi_{k|k-1}^i, dx') \; , \tag{40}
$$

and

$$
\begin{aligned}
Q\, w_k^N(dx') + \Gamma\, \mu_k^N(dx') \;=\; & \sum_{i=1}^N [\, \rho_{k|k-1}^i + S_k(\xi_{k|k-1}^i) - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \\
& + I(\xi_{k|k-1}^i, x') \,] \; \omega_k^i \; Q(\xi_{k|k-1}^i, dx') \; ,
\end{aligned} \tag{41}
$$

which can be interpreted as the marginals of the finite signed measures $m = (m^i \, , \, i = 1, \cdots, N)$ and $s = (s^i \, , \, i = 1, \cdots, N)$ defined on the product space $E^N = \{1, \cdots, N\} \times \mathbb{R}$ by

$$
s^i(dx') = [\, \underbrace{\rho_{k|k-1}^i + S_k(\xi_{k|k-1}^i) + I(\xi_{k|k-1}^i, x')}_{r^i(x')} - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \,] \; \underbrace{\omega_k^i \; Q(\xi_{k|k-1}^i, dx')}_{m^i(dx')} \; .
$$

Notice that

$$
\sum_{j=1}^N \int_{-\infty}^{\infty} r^j(x') \; m^j(dx') = \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \; ,
$$

is just a normalizing constant. Introducing the particle approximation

$$
m \approx m^N = \frac{1}{N} \; \sum_{i=1}^N \delta_{(\tau_k^i, \, \xi_{k+1|k}^i)} \; ,
$$

and the weighted particle approximation

$$
s \approx s^N = \frac{1}{N} \; \sum_{i=1}^N [\, r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \; \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,] \,] \; \delta_{(\tau_k^i, \, \xi_{k+1|k}^i)} \; ,
$$

where independently for any $i = 1, \cdots, N$, the pair $(\tau_k^i, \xi_{k+1|k}^i)$ is jointly distributed according to the probability distribution $m = (m^i \,,\, i = 1, \cdots, N)$, i.e.

$$\tau_k^i \sim (\omega_k^j \,,\, j = 1, \cdots, N) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\xi_{k|k-1}^{\tau_k^i}, dx') \;,$$

and marginalizing, yields

$$\mu_{k+1|k} \approx \mu_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1|k}^i} \;.$$

$$w_{k+1|k} \approx w_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N [\, r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,] \,] \; \delta_{\xi_{k+1|k}^i} = \frac{1}{N} \sum_{i=1}^N \rho_{k+1|k}^i \; \delta_{\xi_{k+1|k}^i} \;.$$

Notice that the weight

$$\rho_{k+1|k}^i = r^{\tau_k^i}(\xi_{k+1|k}^i) - [\, \frac{1}{N} \sum_{j=1}^N r^{\tau_k^j}(\xi_{k+1|k}^j) \,] \;,$$

where

$$r^{\tau_k^i}(\xi_{k+1|k}^i) = \rho_{k|k-1}^{\tau_k^i} + S_k(\xi_{k|k-1}^{\tau_k^i}) + I(\xi_{k|k-1}^{\tau_k^i}, \xi_{k+1|k}^i) \;,$$

does not depend only on the position $\xi_{k+1|k}^i$.

4.4.4 An alternate particle approximation scheme

Equations (40) and (41) read

$$Q\,\mu_k^N(dx') = \sum_{i=1}^N \omega_k^i \; Q(\xi_{k|k-1}^i, dx') = [\, \sum_{i=1}^N \omega_k^i \; q(\xi_{k|k-1}^i, x') \,] \; dx' \;,$$

and

$$\begin{aligned} Q\, w_k^N(dx') + \Gamma\, \mu_k^N(dx') \;&=\; [\, \sum_{i=1}^N [\, \rho_{k|k-1}^i + S_k(\xi_{k|k-1}^i) - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \\ &\qquad + I(\xi_{k|k-1}^i, x') \,] \; \omega_k^i \; q(\xi_{k|k-1}^i, x') \,] \; dx' = r_{k+1}^N(x') \; Q\,\mu_k^N(dx') \;, \end{aligned}$$

where

$$\begin{aligned} r_{k+1}^N(x') \;&=\; \frac{\displaystyle\sum_{j=1}^N [\, \rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j) + I(\xi_{k|k-1}^j, x') \,] \; \omega_k^j \; q(\xi_{k|k-1}^j, x')}{\displaystyle\sum_{j=1}^N \omega_k^j \; q(\xi_{k|k-1}^j, x')} - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \\ &=\; \frac{\displaystyle\sum_{j=1}^N [\, \rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j) + I(\xi_{k|k-1}^j, x') \,] \; \omega_k^j \; \exp\{-\tfrac{1}{2}\,(x' - F\,\xi_{k|k-1}^j)^* \, \Sigma^{-1} \, (x' - F\,\xi_{k|k-1}^j)\}}{\displaystyle\sum_{j=1}^N \omega_k^j \; \exp\{-\tfrac{1}{2}\,(x' - F\,\xi_{k|k-1}^j)^* \, \Sigma^{-1} \, (x' - F\,\xi_{k|k-1}^j)\}} \\ &\qquad - [\, \sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j)) \; \omega_k^j \,] \;. \end{aligned}$$

Resampling yields the particle approximation

$$Q\,\mu_k^N \approx \mu_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{k+1|k}^i} \; ,$$

where independently for any $i = 1, \cdots, N$

$$\xi_{k+1|k}^i \sim Q\,\mu_k^N(dx') \; ,$$

which can be achieved for instance by taking

$$\widehat{\xi}_k^i \sim \mu_k^N(dx) \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx') \; ,$$

or even more explicitly by taking

$$\tau_k^i \sim (\omega_k^j \, , \, j = 1, \cdots, N) \; , \qquad \widehat{\xi}_k^i = \xi_{k|k-1}^{\tau_k^i} \qquad \text{and} \qquad \xi_{k+1|k}^i \sim Q(\widehat{\xi}_k^i, dx') \; ,$$

and the weighted particle approximation

$$Q\,w_k^N + \Gamma\,\mu_k^N \approx w_{k+1|k}^N = \frac{1}{N} \sum_{i=1}^N r_{k+1}^N(\xi_{k+1|k}^i)\,\delta_{\xi_{k+1|k}^i} = \frac{1}{N} \sum_{i=1}^N \rho_{k+1|k}^i\,\delta_{\xi_{k+1|k}^i} \; ,$$

where the weight

$$\begin{aligned} \rho_{k+1|k}^i &= r_{k+1}^N(\xi_{k+1|k}^i) \\ &= \frac{\displaystyle\sum_{j=1}^N [\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j) + I(\xi_{k|k-1}^j, \xi_{k+1|k}^i)]\,\omega_k^j\,\exp\{-\tfrac{1}{2}(\xi_{k+1|k}^i - F\,\xi_{k|k-1}^j)^*\,\Sigma^{-1}\,(\xi_{k+1|k}^i - F\,\xi_{k|k-1}^j)\}}{\displaystyle\sum_{j=1}^N \omega_k^j\,\exp\{-\tfrac{1}{2}(\xi_{k+1|k}^i - F\,\xi_{k|k-1}^j)^*\,\Sigma^{-1}\,(\xi_{k+1|k}^i - F\,\xi_{k|k-1}^j)\}} \\ &\quad - [\sum_{j=1}^N (\rho_{k|k-1}^j + S_k(\xi_{k|k-1}^j))\,\omega_k^j\,] \end{aligned}$$

depends only on the position of ξ_{k+1}^i.

4.5 About the graphical outputs

The figures presented in Part B, where the HMM situation is considered, correspond to the particle approximation of the prediction density and its derivative w.r.t. the parameter a, and to the re–sampled particle approximation of the correction density and its derivative w.r.t. the parameter a.

(i) the first graphic shows the exact trajectory of the log–likelihood function, displayed in black solid line, as provided by the Kalman filter, and the approximate trajectory, displayed in blue solid line, as provided by the particle filter approximation,

(ii) similarly, the second graphic shows the exact trajectory of the score function, displayed in black solid line, as provided by the Kalman filter, and the approximate trajectory, displayed in blue solid line, as provided by the particle filter approximation,

(iii) the third graphic shows the exact density, displayed in blue solid line, and an histogram of the particle approximation associated with the particle system $\{\xi_{k|k-1}^i \, , \, i = 1, \cdots, N\}$, i.e. for each subinterval A of length $|A|$, a black solid bar of height

$$\frac{1}{|A|}\,[\,\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{(\xi_{k|k-1}^i \in A)}\,] \; ,$$

is drawn upwards at the center of the interval A,

(iv) the fourth graphic shows the exact derivative of the density w.r.t. the parameter a, displayed in blue solid line, and an histogram of the weighted particle approximation associated with the weighted particle system $\{(\xi^i_{k|k-1}, \rho^i_{k|k-1})\ ,\ i = 1, \cdots, N\}$, i.e. for each subinterval A of length $|A|$, a black solid bar of (either positive or negative) height

$$\frac{1}{|A|}\, [\, \frac{1}{N} \sum_{i=1}^{N} \rho^i_{k|k-1}\ \mathbf{1}(\xi^i_{k|k-1} \in A)\,]\ ,$$

is drawn (either upwards or downwards) at the center of the interval A.

Even if the second graphic shows a good match between the blue solid line and the black histogram bars, which implies in particular a good match between the region where the green solid line lies above the zero axis and the region where the red solid bars are drawn upwards, it could nevertheless happen that some particles with a negative weight fall within this region, at the expense of having their contribution compensated by the contribution of particles with a positive weight, in such a way that the balance over each subinterval of this region has the right sign and magnitude : to detect this situation, which denotes an unefficient allocation of signed weights to particles, the third graphic shows the exact derivative of the density w.r.t. the parameter a, displayed in blue solid line, and separately an histogram of the positive and negative parts of the weighted particle approximation associated with the weighted particle system $\{(\xi^i_{k|k-1}, \rho^i_{k|k-1})\ ,\ i = 1, \cdots, N\}$, i.e. for each subinterval A of length $|A|$, a red solid bar of positive height

$$\frac{1}{|A|}\, [\, \frac{1}{N} \sum_{i=1}^{N} \rho^i_{k|k-1}\ \mathbf{1}(\rho^i_{k|k-1} \geq 0)\ \mathbf{1}(\xi^i_{k|k-1} \in A)\,]\ ,$$

is drawn upwards at the center of the interval A, and a green solid bar of negative height

$$\frac{1}{|A|}\, [\, \frac{1}{N} \sum_{i=1}^{N} \rho^i_{k|k-1}\ \mathbf{1}(\rho^i_{k|k-1} \leq 0)\ \mathbf{1}(\xi^i_{k|k-1} \in A)\,]\ ,$$

is drawn downwards at the center of the interval A,

(v) the fifth graphic shows the weighted particle system $\{(\xi^i_{k|k-1}, \rho^i_{k|k-1})\ ,\ i = 1, \cdots, N\}$, i.e. an (either green or red) solid bar of (either positive or negative) height $\rho^i_{k|k-1}$, is drawn (either upwards or downwards) at each particle location $\xi^i_{k|k-1}$.

In Part C, we have graphics of the score function for each parameter and the log–likelihood function with the same conventions as above.

4.6 Conclusions

In this section, we have seen two different methods to approximate the derivative of the filter w.r.t. the parameter. The second particle approximation scheme provides better results with less particles but the first scheme is faster and very easy to compute. As our goal is to design a particle approximation scheme to study real data with a high sampling rate, we will use the first scheme.

Moreover, the simulations show that the deterministic resampling step described before leads to very good approximations of the optimal log–likelihood and score functions.

5 Particle implementation of the RML algorithm

The algorithm presented in this section is a recursive version of the gradient algorithm usually referred to as recursive maximum likelihood (RML). It requires essentially the ability to compute the optimal filter and the derivative of this filter with respect to the parameter of interest. We apply our method to approximate the derivative of the filter to perform RML.

5.1 Particle filter and parameter estimation

In this section, the parameter is denoted by θ and dependence w.r.t. the parameter appears explicitly in the notation for the transition kernel $Q^\theta(x, dx')$, and for the linear tangent kernel $K^\theta(x, dx', ds')$. It is well–known that in such a parametric model, the log–likelihood function for the estimation of the parameter θ can be written as

$$\ell_n^\theta = \frac{1}{n} \sum_{k=0}^{n} \log \langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle \ ,$$

and the score function, i.e. the derivative of the log–likelihood function w.r.t. the parameter, can be written as

$$\frac{\partial \ell_n^\theta}{\partial \theta} = \frac{1}{n} \sum_{k=0}^{n} \frac{\langle w_{k|k-1}^\theta, \Psi_k^\theta \rangle + \langle \mu_{k|k-1}^\theta, \Psi_k^\theta \, S_k^\theta \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k^\theta \rangle} \ ,$$

where the filter $\{\mu_k^\theta \, , \, k \geq 0\}$ and the linear tangent filter $\{w_k^\theta \; k \geq 0\}$ satisfy

$$\mu_{k-1}^\theta \xrightarrow{\text{prediction}} \mu_{k|k-1}^\theta = Q^\theta \, \mu_{k-1}^\theta \xrightarrow{\text{correction}} \mu_k^\theta = \Psi_k^\theta \cdot \mu_{k|k-1}^\theta \ ,$$

and

$$w_{k-1}^\theta \xrightarrow{\text{adaptive linear tangent prediction}} w_{k|k-1}^\theta = Q^\theta \, w_{k-1}^\theta + \Gamma^\theta \, \mu_{k-1}^\theta$$

$$\xrightarrow{\text{linear tangent correction}} w_k^\theta = F_k^\theta(\mu_{k|k-1}^\theta) \, w_{k|k-1}^\theta + G_k^\theta(\mu_{k|k-1}^\theta) \ ,$$

respectively. It is natural to consider the RML algorithm to identify the parametric model, which is defined by the following relation

$$\widehat{\theta}_k = \widehat{\theta}_{k-1} + \gamma_k \, \frac{\langle \widehat{w}_{k|k-1}, \Psi_k^{\widehat{\theta}_{k-1}} \rangle + \langle \widehat{\mu}_{k|k-1}, \Psi_k^{\widehat{\theta}_{k-1}} \, S_k^{\widehat{\theta}_{k-1}} \rangle}{\langle \widehat{\mu}_{k|k-1}, \Psi_k^{\widehat{\theta}_{k-1}} \rangle} \ , \tag{42}$$

where typically $\gamma_k \simeq k^{-2/3}$, and the averaged estimator (which achieves the minimum variance of the estimation error) is obtained by post–processing

$$\overline{\theta}_k = \overline{\theta}_{k-1} + \frac{1}{k} \, (\widehat{\theta}_k - \overline{\theta}_{k-1}) \qquad \text{i.e.} \qquad \overline{\theta}_n = \frac{1}{n} \sum_{k=1}^{n} \widehat{\theta}_k \ ,$$

where the adaptive filter $\{\widehat{\mu}_k \, , \, k \geq 0\}$ and the adaptive linear tangent filter $\{\widehat{w}_k, \, , \, k \geq 0\}$ satisfy the same equations as the filter and the linear tangent filter respectively, in which the value of the parameter is adapted at each time instant according to equation (42), i.e.

$$\widehat{\mu}_{k-1} \xrightarrow{\text{adaptive prediction}} \widehat{\mu}_{k|k-1} = Q^{\widehat{\theta}_{k-1}} \, \widehat{\mu}_{k-1} \xrightarrow{\text{correction}} \widehat{\mu}_k = \Psi_k^{\widehat{\theta}_{k-1}} \cdot \widehat{\mu}_{k|k-1} \ ,$$

and

$$\widehat{w}_{k-1} \xrightarrow{\text{adaptive linear tangent prediction}} \widehat{w}_{k|k-1} = Q^{\widehat{\theta}_{k-1}} \, \widehat{w}_{k-1} + \Gamma^{\widehat{\theta}_{k-1}} \, \widehat{\mu}_{k-1}$$

$$\xrightarrow{\text{linear tangent correction}} \widehat{w}_k = F_k^{\widehat{\theta}_{k-1}}(\widehat{\mu}_{k|k-1}) \, \widehat{w}_{k|k-1} + G_k^{\widehat{\theta}_{k-1}}(\widehat{\mu}_{k|k-1}) \ ,$$

respectively. The particle implementation of the RML equation (42) is

$$\widehat{\theta}^N_k = \widehat{\theta}^N_{k-1} + \gamma_k \sum_{i=1}^N \omega^i_k \, [\, \rho^i_{k|k-1} + S_k^{\widehat{\theta}^N_{k-1}}(\xi^i_{k|k-1})\,] \qquad \text{and} \qquad \overline{\theta}^N_k = \overline{\theta}^N_{k-1} + \frac{1}{k}\,(\widehat{\theta}^N_k - \overline{\theta}^N_{k-1}) \; ,$$

and the corresponding algorithm is described below.

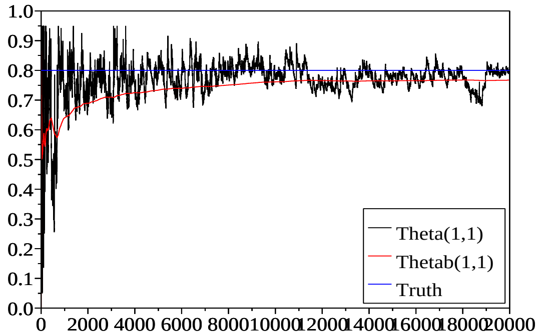
Particle implementation of the RML algorithm

$k = 0$ [initialisation]
pick $\widehat{\theta}^N_{-1}$
[simulation]
for $i = 1, \cdots, N$ independently **do**
  $\xi^i_{0|-1} \sim \mu_{0|-1}(dx)$ and $\rho^i_{0|-1} = 0$
end for
loop
  [weighting]
  for $i = 1, \cdots, N$ **do**
    $\omega^i_k = c^N_k \; \Psi_k^{\widehat{\theta}^N_{k-1}}(\xi^i_{k|k-1})$ [c^N_k such that $\sum_{i=1}^N \omega^i_k = 1$] and $\rho^i_k = \rho^i_{k|k-1} + S_k^{\widehat{\theta}^N_{k-1}}(\xi^i_{k|k-1})$
  end for
  [parameter update]
  $\widehat{\theta}^N_k = \widehat{\theta}^N_{k-1} + \gamma_k \sum_{i=1}^N \rho^i_k \, \omega^i_k$ [with $\gamma_k \simeq (k+1)^{-2/3}$] and $\overline{\theta}^N_k = \overline{\theta}^N_{k-1} + \frac{1}{k+1}\,(\widehat{\theta}^N_k - \overline{\theta}^N_{k-1})$
  [selection]
  for $i = 1, \cdots, N$ independently **do**
    $\widehat{\xi}^i_k = \xi^{\tau^i_k}_{k|k-1}$ and $\widehat{\rho}^i_k = \rho^{\tau^i_k}_k$ [with $\tau^i_k \sim (\omega^1_k, \cdots, \omega^N_k)$]
  end for
  [mutation]
  for $i = 1, \cdots, N$ independently **do**
    $\xi^i_{k+1|k} \sim Q^{\widehat{\theta}^N_k}(\widehat{\xi}^i_k, dx')$ and $\rho^i_{k+1|k} = \widehat{\rho}^i_k + I^{\widehat{\theta}^N_k}(\widehat{\xi}^i_k, \xi^i_{k+1|k}) - a^N_{k+1}$
    [a^N_{k+1} such that $\sum_{i=1}^N \rho^i_{k+1|k} = 0$]
  end for
  $k \leftarrow k+1$ [time iteration]
end loop

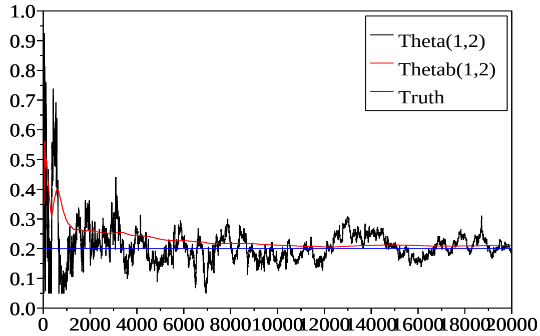
The mathematical analysis of the asymptotic properties of the estimator $\widehat{\theta}^N_k$ as $k \to \infty$ and $N \to \infty$ is far beyond the scope of this work, and would rely on joint stability properties of the filter and the linear tangent filter, which is a very difficult question. Even the asymptotic properties of the estimator $\widehat{\theta}_k$ as $k \to \infty$ are difficult to prove, unless some mixing assumption holds for the transition kernels $Q^\theta(x, dx')$ and the linear tangent kernels $\Gamma^\theta(x, dx')$, which practically implies that the state–space E should be compact, see e.g. [10] where only the non–recursive ML is studied.

Numerical results

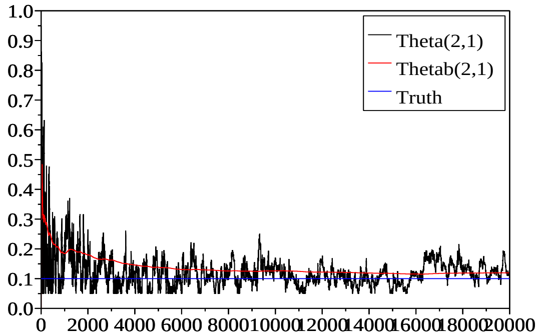
Estimation of F(1,1) with 1000 particles



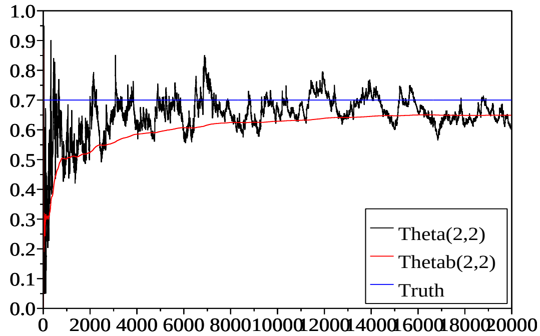
Estimation of F(1,2) with 1000 particles



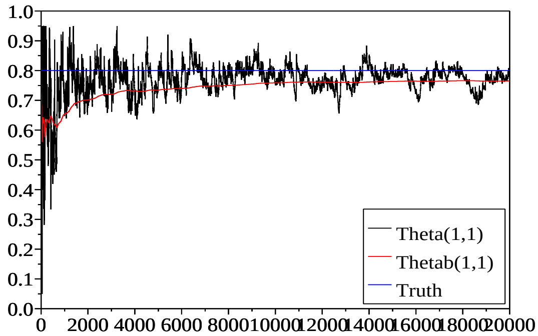
Estimation of F(2,1) with 1000 particles



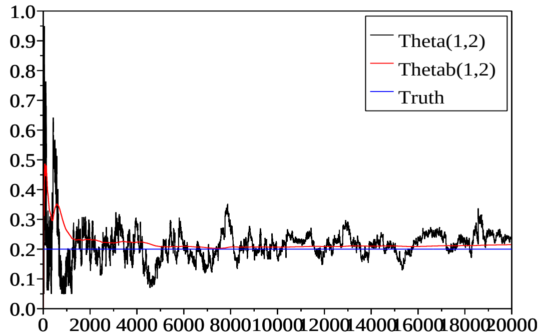
Estimation of F(2,2) with 1000 particles



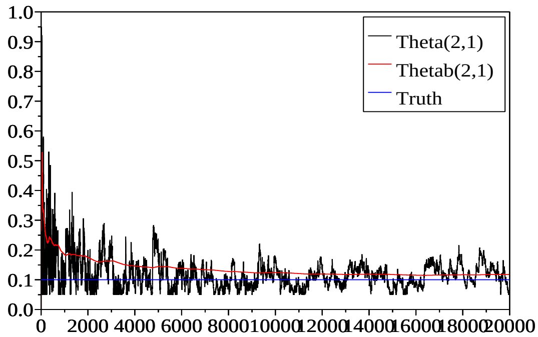
Estimation of F(1,1) with 5000 particles



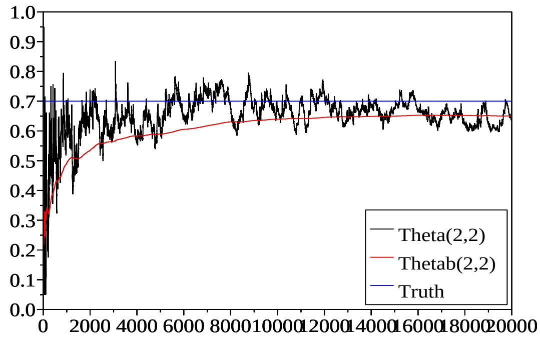
Estimation of F(1,2) with 5000 particles

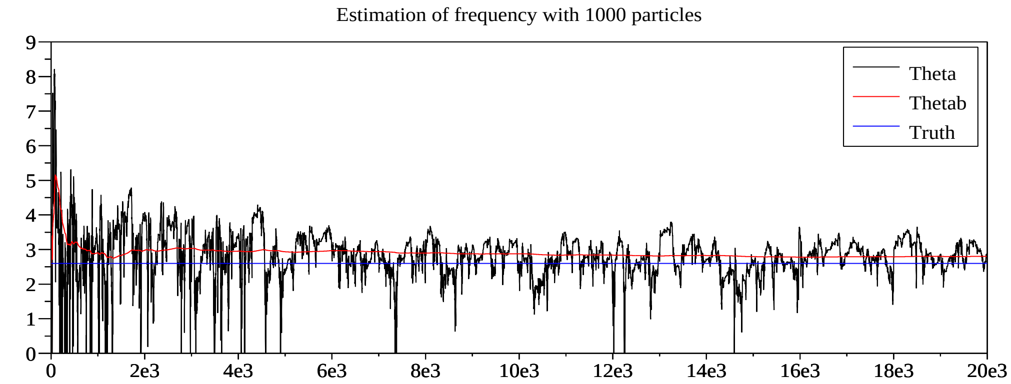


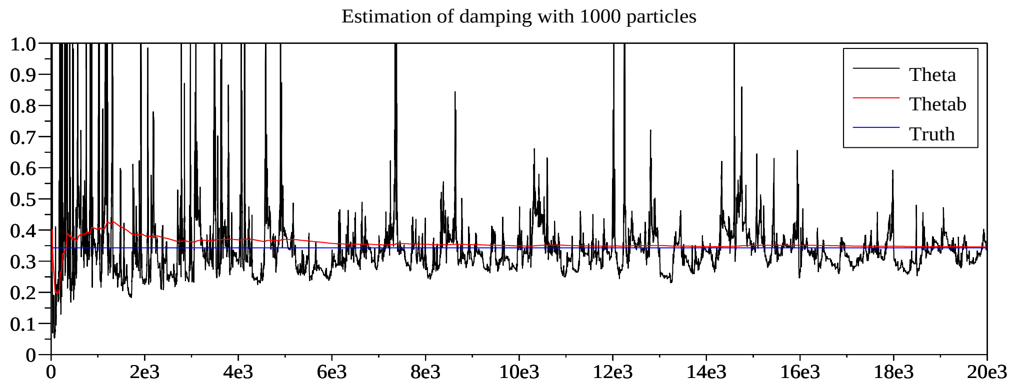
Estimation of F(2,1) with 5000 particles

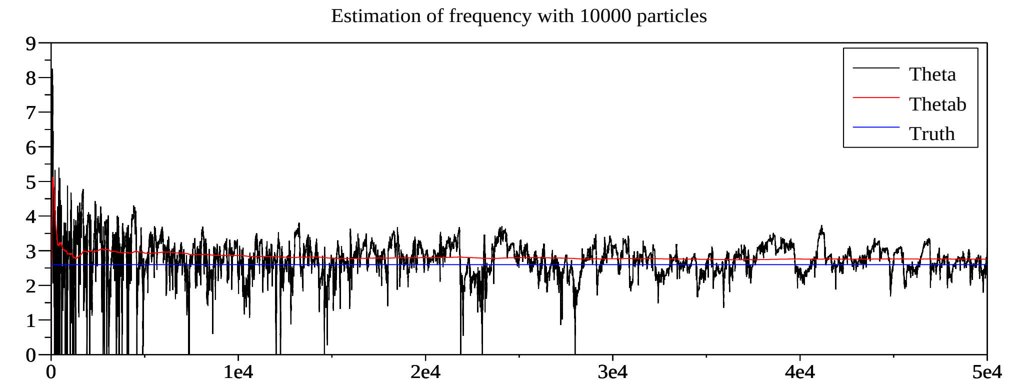


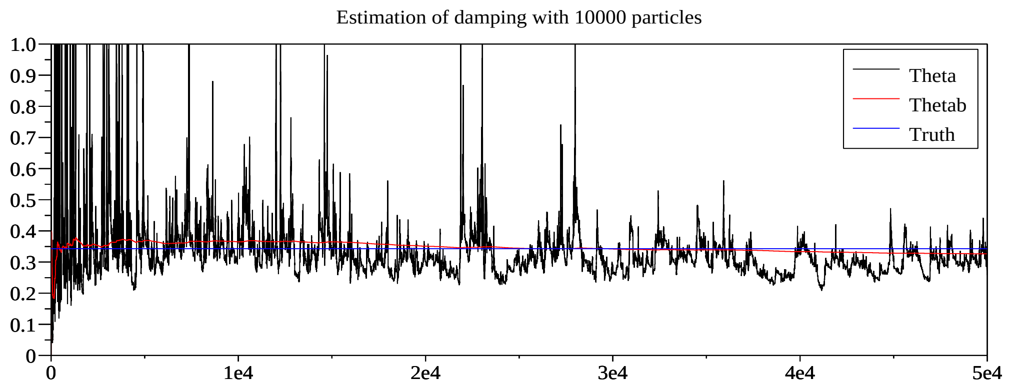
Estimation of F(2,2) with 5000 particles




Estimation of frequency with 1000 particles
Theta
Thetab
Truth
0
1
2
3
4
5
6
7
8
9
0
2e3
4e3
6e3
8e3
10e3
12e3
14e3
16e3
18e3
20e3


Estimation of damping with 1000 particles
Theta
Thetab
Truth
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
2e3
4e3
6e3
8e3
10e3
12e3
14e3
16e3
18e3
20e3

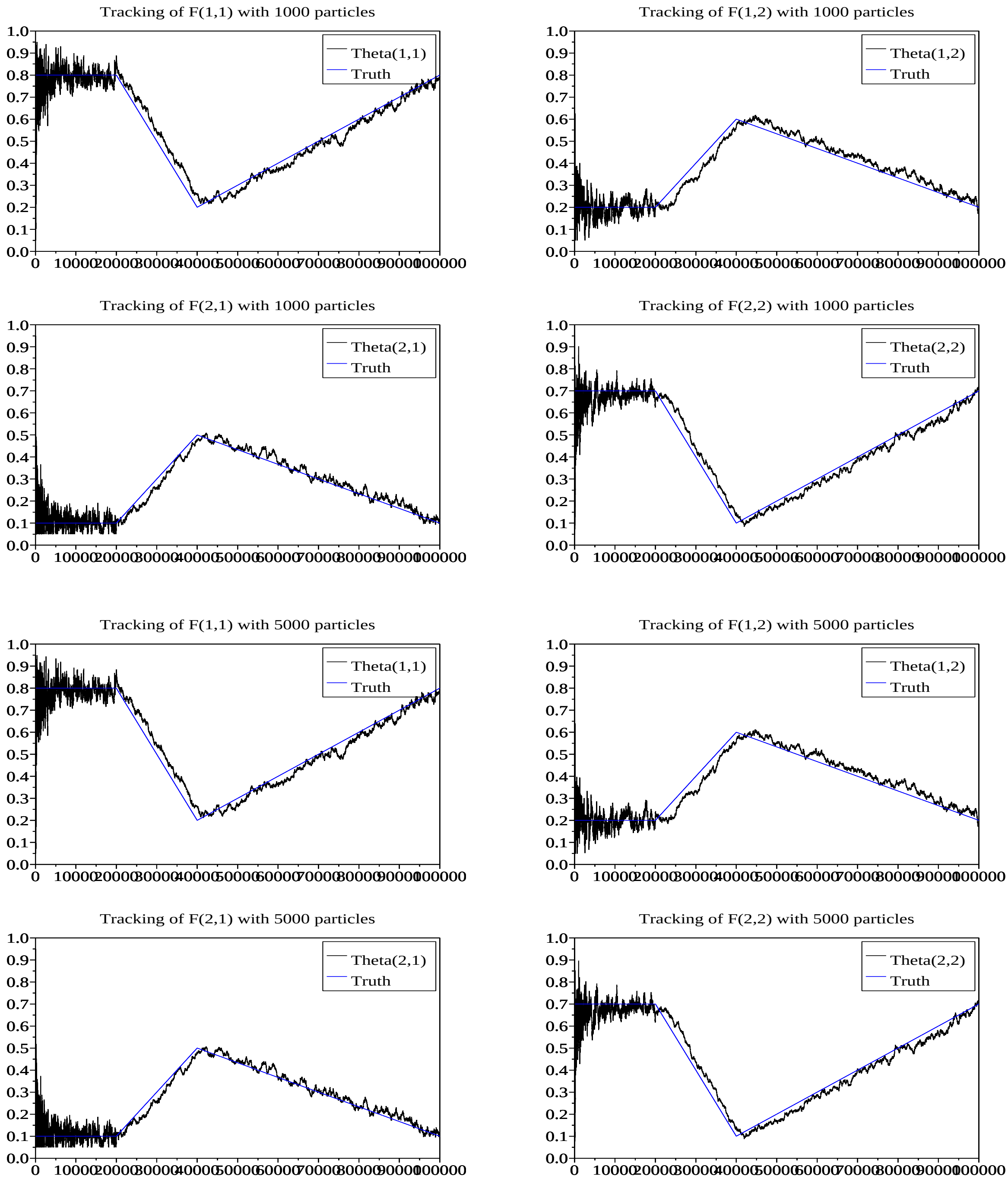

Estimation of frequency with 10000 particles
Theta
Thetab
Truth
0
1
2
3
4
5
6
7
8
9
0
1e4
2e4
3e4
4e4
5e4


Estimation of damping with 10000 particles
Theta
Thetab
Truth
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
1e4
2e4
3e4
4e4
5e4

5.2 Particle filter and parameter tracking

If the parameter θ is actually time–varying but one does not have a dynamic model for its evolution, a standard approach to *track* this parameter consists of using the recursive algorithm presented in Section 5.1 using a fixed–step size γ instead of a decreasing sequence γ_k. Selecting the step size is a difficult problem. If γ is too large, the statistical fluctuations around the parameter are too large. If γ is too small, the algorithm loses its tracking ability.

Numerical results



In this section, the true parameter symbolized by a solid blue line is approximated by the estimator θ_k provided by the RML in solid black line and the averaged estimator $\overline{\theta}_k$ in red blue line.

We have used a reprojection method in order to ensure the convergence of our algorithm. In fact, the parameter $\theta = (\theta_1, \cdots, \theta_k) \in \Theta$, where Θ is a compact subset of $\mathbb{R}^k$. The updating step does not ensure that $\theta_{k+1} \in \Theta$ even if $\theta_k \in \Theta$. A standard approach in stochastic approximation to prevent divergence consists of reprojecting θ_{k+1} inside Θ whenever the value obtained, say $\widetilde{\theta}_{k+1}$, does not belong to Θ. One has $\Theta = \prod_{i=1}^{k} [\theta_{i,\min}, \theta_{i,\max}]$ and the reprojection procedure simply consists of setting $\theta_{i,k+1} = \theta_{i,\min}$ if $\theta_{i,k+1} < \theta_{i,\min}$ and $\theta_{i,k+1} = \theta_{i,\max}$ if $\theta_{i,k+1} > \theta_{i,\max}$. For further results and asymptotic theory, see [8].

5.3 Conclusions

We have seen the efficiency of particle approximation through several examples. Indeed, we are able to estimate and track a multidimensional parameter using good approximations of optimal filters and its derivatives. In the estimation problem, the true parameter is well approximated by the estimator using particle methods even in the two–dimensional case where the four coefficients of a square matrix are estimated at the same time. The tracking problem has also been studied, with good results.

References

[1] M. Sanjeev Arulampalam, Simon Maksell, Neil J. Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear / non–Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, SP–50(2 (Special issue on Monte Carlo Methods for Statistical Signal Processing)):174–188, February 2002.

[2] Michèle Basseville. On–board component fault detection and isolation using the statistical local approach. *Automatica*, 34(11):1391–1416, November 1998.

[3] Michèle Basseville, Albert Benveniste, Maurice Goursat, Luc Hermans, Laurent Mevel, and Herman van der Auweraer. Output–only subspace–based structural identification : from theory to industrial testing practice. *Journal of Dynamic Systems, Measurement, and Control*, 123(4 (Special issue on Identification of Mechanical Systems)):668–676, December 2001.

[4] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes : Theory and Application.* Information and System Sciences Series. Prentice–Hall, Englewood Cliffs, 1993.

[5] D. Bernal. Load vectors for damage localization. *ASME Journal of Mechanical Engineering*, 128(1):7–14, 2002.

[6] Frédéric Cérou, François Le Gland, and Nigel J. Newton. Stochastic particle methods for linear tangent filtering equations. In José-Luis Menaldi, Edmundo Rofman, and Agnès Sulem, editors, *Optimal Control and Partial Differential Equations. In honour of professor Alain Bensoussan's 60th birthday*, pages 231–240. IOS Press, Amsterdam, 2001.

[7] Dan Crişan, Jessica G. Gaines, and Terry J. Lyons. Convergence of a branching particle method to the solution of the Zakai equation. *SIAM Journal on Applied Mathematics*, 58(5):1568–1590, October 1998.

[8] Bernard Delyon. Stochastic approximation with decreasing gain : Convergence and asymptotic theory. Unpublished lecture notes, université de Rennnes 1, 2001.

[9] Scott W. Doebling, Charles R. Farrar, Michael B. Prime, and Daniel W. Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics : A literature review. Technical Report LA–13070–MS, Los Alamos National Laboratory, May 1996.

[10] Randal Douc and Catherine Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7(3):381–420, June 2001.

[11] Arnaud Doucet. On sequential simulation–based methods for Bayesian filtering. Technical Report CUED/F-INFENG/TR310, Department of Engineering, University of Cambridge, February 1998. `http://www-sigproc.eng.cam.ac.uk/~ad2/tr310.ps`.

[12] D. J. Ewins. *Modal testing : Theory, Practice and Application*, volume 12. Research Studies Press, Ltd., Baldock, 2000. 2nd edition.

[13] Charles R. Farrar, Scott W. Doebling, and David A. Nix. Vibration–based structural damage identification. *Philosophical Transactions of the Royal Society of London, Series A*, 359(1778):131–149, January 2001.

[14] Béatrice Gach-Devauchelle. *Diagnostic Mécanique des Fatigues sur les Structures Soumises à des Vibrations en Ambiance de Travail*. Thèse de Doctorat, Université de Paris–Dauphine, Paris, December 1991.

[15] Neil J. Gordon, David J. Salmond, and Adrian F. M. Smith. Novel approach to nonlinear / non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, April 1993.

[16] Arnaud Guyader, François Le Gland, and Nadia Oudjane. A particle implementation of the recursive MLE for partially observed diffusions. In *Proceedings of the 13th Symposium on System Identification (SYSID), Rotterdam*, pages 1305–1310. IFAC / IFORS, August 2003.

[17] Luc Hermans and Herman van der Auweraer. Modal testing and analysis of structures under operational conditions : industrial applications. *Mechanical Systems and Signal Processing*, 13(2):193–216, March 1999.

[18] Jer-Nan Juang. *Applied System Identification*. Prentice–Hall, Englewood Cliffs, NJ, 1994.

[19] Jun S. Liu and Rong Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, September 1998.

[20] Laurent Mevel, Michèle Basseville, Albert Benveniste, and Maurice Goursat. Merging sensor data from multiple measurement setups for nonstationary subspace–based modal analysis. *Journal of Sound and Vibration*, 249(4):719–741, January 2002.

[21] Hans Günther Natke and Czesław Cempel. *Model–Aided Diagnosis of Mechanical Systems : Fundamentals, Detection, Localization, and Assessment*. Springer Verlag, Berlin, 1997.

[22] A. K. Pandey and M. Biswas. Damage detection in structures using changes in flexibility. *Journal of Sound and Vibration*, 169(1):3–17, January 1994.

[23] R. P. C. Sampaio, N. M. M. Maia, J. M. M. Silva, and A. M. R. Ribeiro. On the use of transmissibility for damage detection and locatio. In J. A. Güemes, editor, *European COST F3 Conference on System Identification and Structural Health Monitoring, Madrid*, volume 1, pages 363–376, June 2000.

Contents