TMR Summer School
on Advanced Computational Methods
for Statistical Inference

CIRM, Luminy — September 17–21, 2001

Particle Methods for Statistical Inference of Nonlinear Non–Gaussian Systems

Frédéric Cérou, François LeGland, Bo Wang
IRISA / INRIA Rennes
e–mail : {fcerou,legland,bwang}@irisa.fr
URL : http://www.irisa.fr/sigma2/

Statistical Model

$$X_{k+1} \;=\; f_k^\theta(X_k) + g_k(X_k)\, W_k \; , \qquad X_0 = x$$

$$Y_k \;=\; h_k(X_k) + V_k \; .$$

Independent (non–Gaussian) noise sequences $\{W_k\}$ and $\{V_k\}$

$$W_k \sim p_k(w)\, dw \; , \qquad V_k \sim q_k(v)\, dv \; .$$

Two parameters : $\theta \in \Theta \subset \mathbb{R}^p$ and $x \in \mathbb{R}^m$.

Initial condition x, unknown nuisance parameter : true value $\bar{x}$.

Bayesian approach : prior probability distribution $\pi(dx)$.

Identification : estimate parameter θ, true value α.

Monitoring : known nominal value α, decide whether actually $\theta = \alpha$ or $\theta \neq \alpha$.

Either some forgetting property should hold,
or unknown initial condition should be estimated.

Assumption : $g_k(x)$ full rank,

and $\partial f_k^\theta(x)$ in the range of $g_k(x)$.

Example Higher–order model

$$X_{k+1} = a_k^\theta(X_k, \cdots, X_{k-p+1}) + b_k(X_k)\, W_k \; ,$$

with invertible $b_k(x)$. Extended state

$$X_k^{\text{ext}} = \begin{pmatrix} X_k \\ \vdots \\ X_{k-p+1} \end{pmatrix} \; ,$$

satisfies

$$\underbrace{\begin{pmatrix} X_{k+1} \\ X_k \\ \vdots \\ X_{k-p+2} \end{pmatrix}}_{X_{k+1}^{\text{ext}}} = \underbrace{\begin{pmatrix} a_k^\theta(X_k, \cdots, X_{k-p+1}) \\ X_k \\ \vdots \\ X_{k-p+2} \end{pmatrix}}_{f_k^\theta(X_k^{\text{ext}})} + \underbrace{\begin{pmatrix} b_k(X_k) \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{g_k(X_k^{\text{ext}})} W_k \; .$$

Associated hidden Markov model

Transition probability kernel

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

Conditional probability distribution of a single observation

$$\mathbb{P}^\theta[Y_k \in dy \mid X_k = x] = q_k(y - h_k(x))\, dy \; ,$$

hence likelihood function

$$\Psi_k(x) = q_k(Y_k - h_k(x)) \; .$$

Likelihood function

Direct approach

$$L_n^\theta \;=\; \int \Psi_0(x_0)\, \Psi_1(x_1) \cdots \Psi_n(x_n) \; \pi(dx_0)\, Q_1^\theta(x_0, dx_1) \cdots Q_n^\theta(x_{n-1}, dx_n) \; .$$

Alternative recursive expression based on nonlinear filtering : introducing conditional probability distributions

$$\mu_k^\theta(dx) = \mathbb{P}^\theta[X_k \in dx \mid Y_0, Y_1, \cdots, Y_k] \;,$$

$$\mu_{k|k-1}^\theta(dx) = \mathbb{P}^\theta[X_k \in dx \mid Y_0, Y_1, \cdots, Y_{k-1}] \;,$$

yields recursive expression for log–likelihood function

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

Asymptotic behaviour, as $n \to \infty$: under "suitable" conditions

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

with Kullback–Leibler information K_α^θ.

Score function

Using above representation

$$\zeta_n^\theta = \partial \log L_n^\theta = \sum_{k=0}^n \frac{\langle w_{k|k-1}^\theta, \Psi_k \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} \;,$$

where formally $w_{k|k-1}^\theta = \partial \mu_{k|k-1}^\theta$.

Asymptotic behaviour, as $n \to \infty$: under "suitable" conditions

$$\frac{1}{n}\,\zeta_n^\theta = \frac{1}{n} \sum_{k=0}^n \frac{\langle w_{k|k-1}^\theta, \Psi_k \rangle}{\langle \mu_{k|k-1}^\theta, \Psi_k \rangle} \longrightarrow -\partial K_\alpha^\theta \;,$$

and

$$\frac{1}{\sqrt{n}}\,\zeta_n^\alpha \Longrightarrow \begin{cases} \mathcal{N}(0, I_\alpha) \;, & \text{under } \mathbb{P}_n^\alpha, \\ \mathcal{N}(I_\alpha\,\Delta, I_\alpha) \;, & \text{under } \mathbb{P}_n^{\alpha+\Delta/\sqrt{n}}, \end{cases}$$

with Fisher information matrix $I_\alpha = -\partial^2 K_\alpha^\alpha$.

Recursive identification algorithm

$$\hat\theta_{n+1} = \hat\theta_n + \gamma_n \, \frac{\langle w_{k|k-1}^{\hat\theta_n}, \Psi_k \rangle}{\langle \mu_{k|k-1}^{\hat\theta_n}, \Psi_k \rangle} \;.$$

Asymptotic behaviour, as $n \to \infty$: under "suitable" conditions

$$\hat\theta_n \longrightarrow \{\theta \,:\, \partial K_\alpha^\theta = 0\} \supset \operatorname{argmin} K_\alpha^\theta \supset \{\alpha\} \;.$$

From now on, dependency w.r.t. parameter θ is dropped.

Optimal filter

Transition from μ_{k-1} to μ_k described by

$$\mu_{k-1} \xrightarrow{\text{prediction}} \mu_{k|k-1} = Q_k\, \mu_{k-1}$$

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

Bayes rule normalization : projective product

$$\Psi_k \cdot \mu_{k|k-1} = \frac{\Psi_k\, \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} \;.$$

Linear tangent optimal filter

Transition from w_{k-1} to w_k described by

$$w_{k-1} = \partial \mu_{k-1} \xrightarrow[\text{}]{\text{linear tangent prediction}} w_{k|k-1} = \partial[Q_k\, \mu_{k-1}]$$

$$\xrightarrow{\text{linear tangent correction}} w_k = \partial[\Psi_k \cdot \mu_{k|k-1}] \;.$$

Precisely

$$w_{k|k-1} = \partial[Q_k\, \mu_{k-1}] = Q_k\, w_{k-1} + [\partial\, Q_k]\, \mu_{k-1} \;,$$

and

$$\begin{aligned} w_k &= \partial\, [\Psi_k \cdot \mu_{k|k-1}] \\ &= \frac{\Psi_k\, w_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} - \frac{\Psi_k\, \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle}\, \frac{\langle w_{k|k-1}, \Psi_k \rangle}{\langle \mu_{k|k-1}, \Psi_k \rangle} \\ &= F_k(\mu_{k|k-1})\, w_{k|k-1} \;, \end{aligned}$$

where $w \longmapsto F_k(\mu)\, w$ is linear tangent mapping at point μ, associated with mapping $\mu \longmapsto \Psi_k \cdot \mu$.

Idea : a stochastic representation for derivative ∂Q_{k+1} of Q_{k+1} w.r.t. parameter θ, suitable for simulation, can be obtained by looking at a "more general" object : directional derivative of mapping $f \longmapsto Q^f_{k+1}$ at $f = f_k$, in direction γ, i.e.

$$\frac{d}{d\varepsilon}[\, Q^{f_k+\varepsilon\,\gamma}_{k+1}\,]^*\,\phi(x)\Big|_{\varepsilon=0} \; .$$

Stochastic representation result

Assumption A : For any $x \in \mathbb{R}^m$, the $m \times r$ matrix $g_k(x)$ has full–rank r, and the $r \times r$ symmetric matrix $g^*_k(x)\, g_k(x)$ is uniformly definite positive, i.e.

$$g^*_k(x)\, g_k(x) \geq \alpha_k\, I \; .$$

Assumption B : For any $x \in \mathbb{R}^m$, the m–dimensional vector $\gamma(x)$ belongs to the range of the $m \times r$ matrix $g_k(x)$, i.e. $\gamma(x) = g_k(x)\, c(x)$ for some r–dimensional vector $c(x)$.

Remark Under Assumption A, there is a unique r–dimensional vector $c(x)$ such that $\gamma(x) = g_k(x)\, c(x)$, and

$$c(x) = [g^*_k(x)\, g_k(x)]^{-1}\, g^*_k(x)\, \gamma(x) = g^*_k(x)\, a^{\ominus}_k(x)\, \gamma(x) \; ,$$

where

$$a^{\ominus}_k(x) = g_k(x)\, [g^*_k(x)\, g_k(x)]^{-2}\, g^*_k(x)$$

is pseudoinverse of $a_k(x) = g_k(x)\, g^*_k(x)$. In addition

$$|c(x)|^2 = \gamma^*(x)\, a^{\ominus}_k(x)\, \gamma(x) \; ,$$

and

$$|\gamma(x)|^2 \geq \alpha_k\, |c(x)|^2 \; .$$

Theorem Under Assumptions A and B

$$\frac{d}{d\varepsilon}[\, Q^{f_k+\varepsilon\,\gamma}_{k+1}\,]^*\,\phi(x)\Big|_{\varepsilon=0} = \mathbb{E}[\phi(X_{k+1})\; S_{k+1|k} \mid X_k = x] \; ,$$

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

$$S_{k+1|k} = \frac{-p'_k}{p_k}(W_k)\, c(X_k) = \frac{-p'_k}{p_k}(W_k)\; g^*_k(X_k)\, a^{\ominus}_k(X_k)\, \gamma(X_k) \; .$$

Éric Fournié et al. '99
Frédéric Cérou and FG '00
Frédéric Cérou, FG and Nigel Newton '00

PROOF

$$[\, Q^{f_k+\varepsilon\,\gamma}_{k+1}\,]^*\,\phi(x)$$

$$= \mathbb{E}^{f_k+\varepsilon\,\gamma}[\phi(X_{k+1}) \mid X_k = x]$$

$$= \mathbb{E}^{f_k+\varepsilon\,\gamma}[\phi(f_k(X_k) + \varepsilon\,\gamma(X_k) + g_k(X_k)\, W_k) \mid X_k = x]$$

$$= \mathbb{E}^{f_k+\varepsilon\,\gamma}[\phi(f_k(X_k) + g_k(X_k)\,(W_k + \varepsilon\, c(X_k))) \mid X_k = x]$$

$$= \int \phi(f_k(x) + g_k(x)\,(w + \varepsilon\, c(x)))\; p_k(w)\; dw$$

$$= \int \phi(f_k(x) + g_k(x)\, v)\; p_k(v - \varepsilon\, c(x))\; dv \; ,$$

hence, taking derivative w.r.t. ε yields

$$\frac{d}{d\varepsilon}[\, Q^{f_k+\varepsilon\,\gamma}_{k+1}\,]^*\,\phi(x)$$

$$= \int \phi(f_k(x) + g_k(x)\, v)\; p'_k(v - \varepsilon\, c(x))\; (-c(x))\; dv$$

$$= \int \phi(f_k(x) + g_k(x)\, v)\; \frac{-p'_k}{p_k}(v - \varepsilon\, c(x))\; c(x)\, p_k(v - \varepsilon\, c(x))\; dv \; ,$$

and letting $\varepsilon = 0$ yields directional derivative w.r.t. drift coefficient, in direction γ

$$\int \phi(f_k(x) + g_k(x)\, v)\, \frac{-p_k'}{p_k}(v)\, c(x)\, p_k(v)\, dv$$

$$= \mathbb{E}^{f_k}[\phi(X_{k+1})\, \frac{-p_k'}{p_k}(W_k)\, c(X_k) \mid X_k = x]\ . \qquad \square$$

Corollary If ∂f_k belongs to the range of g_k, then

$$[\partial\, Q_{k+1}]^*\, \phi(x) = \mathbb{E}[\phi(X_{k+1})\, \Xi_{k+1|k} \mid X_k = x]\ ,$$

for any test function ϕ defined in $\mathbb{R}^m$, where

$$\Xi_{k+1|k} = \frac{-p_k'}{p_k}(W_k)\, c(X_k) = \frac{-p_k'}{p_k}(W_k)\; g_k^*(X_k)\, a_k^{\ominus}(X_k)\, \partial f_k(X_k)\ .$$

It follows that for any probability distribution μ on $\mathbb{R}^m$, the signed measure $[\partial\, Q_{k+1}]\, \mu$ is defined by

$$\langle\, [\partial\, Q_{k+1}]\, \mu, \phi\, \rangle = \int \mathbb{E}[\phi(X_{k+1})\, \Xi_{k+1|k} \mid X_k = x]\, \mu(dx)\ ,$$

for any test function ϕ defined in $\mathbb{R}^m$.

Proposition If in addition

- ∂f_k is square–integrable w.r.t. μ,
- p_k has finite Fisher information matrix $F_k = (F_k^{ij})$, where

  $$F_k^{ij} = \int \frac{\partial \log p_k}{\partial w_i}(w)\, \frac{\partial \log p_k}{\partial w_j}(w)\, p_k(w)\, dw\ ,$$

  for any $i, j = 1, \cdots, r$,

then $[\partial\, Q_{k+1}]\, \mu \ll Q_{k+1}\, \mu$.

PROOF

Cauchy–Schwartz inequality yields

$$|\, \langle\, [\partial\, Q_{k+1}]\, \mu, \phi\, \rangle\, | \;=\; |\, \int \mathbb{E}[\phi(X_{k+1})\, \Xi_{k+1|k} \mid X_k = x]\, \mu(dx)\, |$$

$$\leq \{\, \int \mathbb{E}[\phi^2(X_{k+1}) \mid X_k = x]\, \mu(dx)\, \}^{1/2}$$

$$\{\, \int \mathbb{E}[\Xi_{k+1|k}^2 \mid X_k = x]\, \mu(dx)\, \}^{1/2}\ ,$$

for any test function ϕ defined in $\mathbb{R}^m$. Recall that

$$\Xi_{k+1|k} = \frac{-p_k'}{p_k}(W_k)\, c(X_k)\ ,$$

hence

$$\Xi_{k+1|k}^2 = c^*(X_k)\, [\frac{p_k'}{p_k}(W_k)]^*\, \frac{p_k'}{p_k}(W_k)\, c(X_k)\ ,$$

and

$$\mathbb{E}[\Xi_{k+1|k}^2 \mid X_k = x] = c^*(x)\, \underbrace{\mathbb{E}[\, [\frac{p_k'}{p_k}(W_k)]^*\, \frac{p_k'}{p_k}(W_k)]}_{F_k}\, c(x)\ ,$$

with symmetric nonnegative $r \times r$ matrix $F_k = (F_k^{ij})$ defined by

$$F_k^{ij} = \mathbb{E}[\frac{\partial \log p_k}{\partial w_i}(W_k)\, \frac{\partial \log p_k}{\partial w_j}(W_k)]$$

$$= \int \frac{\partial \log p_k}{\partial w_i}(w)\, \frac{\partial \log p_k}{\partial w_j}(w)\, p_k(w)\, dw\ .$$

Therefore

$$\mathbb{E}[\Xi_{k+1|k}^2 \mid X_k = x] \;=\; c^*(x)\, F_k\, c(x)$$

$$\leq\ [\sup_{u \neq 0} \frac{u^*\, F_k\, u}{|u|^2}]\ |c(x)|^2 \leq \frac{\lambda_{\max}^{F_k}}{\alpha_k}\, |\partial f_k(x)|^2\ .$$

Notice also that

$$\int \mathbb{E}[\phi^2(X_{k+1}) \mid X_k = x]\, \mu(dx) = \langle Q_{k+1}\, \mu, \phi^2 \rangle\ .$$

Combining above estimates

$$|\, \langle\, [\partial\, Q_{k+1}]\, \mu, \phi\, \rangle\, |^2 \;\leq\; \frac{\lambda_{\max}^{F_k}}{\alpha_k}\, \langle\, Q_{k+1}\, \mu, \phi^2\, \rangle\, \int |\partial f_k(x)|^2\, \mu(dx)\ ,$$

for any test function ϕ defined in $\mathbb{R}^m$. Finally, taking $\phi = \mathbf{1}_A$ yields

$$|\, [\partial\, Q_{k+1}]\, \mu(A)\, |^2 \;\leq\; \frac{\lambda_{\max}^{F_k}}{\alpha_k}\, Q_{k+1}\, \mu(A)\, \int |\partial f_k(x)|^2\, \mu(dx)\ ,$$

for any Borel set A, hence

$$[\partial\, Q_{k+1}]\, \mu \ll Q_{k+1}\, \mu\ . \qquad \square$$

Proposition For any $k \geq 0$, $w_{k|k-1} \ll \mu_{k|k-1}$ and $w_k \ll \mu_k$.

PROOF

By induction w.r.t. time index.

Assume first that $w_{k-1} \ll \mu_{k-1}$.

Clearly, $Q_k \, w_{k-1} \ll Q_k \, \mu_{k-1} = \mu_{k|k-1}$.

If ∂f_k is square–integrable w.r.t. μ_{k-1}, then it follows from above Corollary that $[\partial \, Q_k] \, \mu_{k-1} \ll Q_k \, \mu_{k-1} = \mu_{k|k-1}$.

Recalling that

$$w_{k|k-1} = Q_k \, w_{k-1} + [\partial \, Q_k] \, \mu_{k-1} \ ,$$

yields $w_{k|k-1} \ll \mu_{k|k-1}$.

Assume now that $w_{k|k-1} \ll \mu_{k|k-1}$. Recalling that

$$w_k = \frac{\Psi_k \, w_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} - \frac{\Psi_k \, \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle} \, \frac{\langle w_{k|k-1}, \Psi_k \rangle}{\langle \mu_{k|k-1}, \Psi_k \rangle} \ ,$$

yields $w_k \ll \mu_k$. □

Representation for linear tangent prediction step, suitable for simulation : for any probability distribution μ on $\mathbb{R}^m$, any signed measure $w \ll \mu$ on $\mathbb{R}^m$, and any test function ϕ defined in $\mathbb{R}^m$

$$\langle Q_{k+1} \, w, \phi \rangle + \langle [\partial \, Q_{k+1}] \, \mu, \phi \rangle$$

$$= \langle w, Q_{k+1}^* \, \phi \rangle + \langle \mu, [\partial \, Q_{k+1}]^* \, \phi \rangle$$

$$= \int \mathbb{E}[\phi(X_{k+1}) \mid X_k = x] \, \frac{dw}{d\mu}(x) \, \mu(dx)$$

$$+ \int \mathbb{E}[\phi(X_{k+1}) \, \Xi_{k+1|k} \mid X_k = x] \, \mu(dx) \ .$$

Particle approximation of optimal filter

Idea : (i) generate a sample $\{\xi^i_{k|k-1} \, , \, i = 1, \cdots, N\}$ of i.i.d. random variables, called particle system, with common probability distribution $Q_k \, \mu^N_{k-1}$, where μ^N_{k-1} approximates μ_{k-1},

(ii) use corresponding empirical probability distribution

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

to approximate $\mu_{k|k-1} = Q_k \, \mu_{k-1}$.

Easy to implement, even in high dimensional problems, sufficient to simulate independent sample paths of hidden dynamical system.

Breakthrough : use sampling / importance resampling (SIR) techniques in correction step
automatically concentrates particles in regions of interest of the state space.

Gordon, Salmond and Smith '93

$$\mu^N_{k-1} \xrightarrow{\text{sampled prediction}} \mu^N_{k|k-1} = S^N(Q_k \, \mu^N_{k-1})$$

$$\xrightarrow{\text{correction}} \mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} \ ,$$

Algorithm : Particle approximation $\mu^N_{k|k-1}$ to optimal filter $\mu_{k|k-1}$, completely characterized by particle system $\{\xi^i_{k|k-1} \, , \, i = 1, \cdots, N\}$.

Transition from $\{\xi^i_{k|k-1} \, , \, i = 1, \cdots, N\}$ to $\{\xi^i_{k+1|k} \, , \, i = 1, \cdots, N\}$ decribed by following steps :

(i) Correction : for all $i = 1, \cdots, N$, compute weight

$$\omega^i_k = c_k^{-1} \, \Psi_k(\xi^i_{k|k-1}) \qquad \text{with} \qquad c_k = \sum_{i=1}^N \Psi_k(\xi^i_{k|k-1}) \ .$$

Then set

$$\mu^N_k = \Psi_k \cdot \mu^N_{k|k-1} = \sum_{i=1}^N \omega^i_k \, \delta_{\xi^i_{k|k-1}} \ .$$

(ii) Resampling : independently for all $i = 1, \cdots, N$, generate a r.v. ξ^i_k with discrete probability distribution μ^N_k (e.g. generate r.v.'s $\{\tau(i) \, , \, i = 1, \cdots, N\}$ with values in I according to some resampling rule, and set

$$\xi^i_k = \xi^{\tau(i)}_{k|k-1} \ ,$$

for all $i = 1, \cdots, N$).

(iii) Prediction : independently for all $i = 1, \cdots, N$, generate a r.v. W_k^i with density p_k, and set

$$\xi_{k+1|k}^i = f_k(\xi_k^i) + g_k(\xi_k^i)\, W_k^i \; .$$

Then set

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

Sequential particle method

Particle system with time–dependent number

$$N_k = \inf\{N \; : \; \delta_k \sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \geq \sup_{x \in \mathbb{R}^m} \Psi_k(x)\}$$

of particles in steps (ii) and (iii).

FG and Nadia Oudjane '01

Particle approximation of linear tangent optimal filter

Idea : because $w_{k|k-1} \ll \mu_{k|k-1}$ and $w_k \ll \mu_k$, approximate $w_{k|k-1}$ and w_k by weighted empirical probability distributions associated with same particle system $\{\xi_{k|k-1}^i \, , \, i = 1, \cdots, N\}$ already used for particle approximations of $\mu_{k|k-1}$ and μ_k, i.e. since

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

then set

$$w_{k|k-1}^N = \frac{1}{N} \sum_{i=1}^N \rho_{k|k-1}^i \; \delta_{\xi_{k|k-1}^i} \; ,$$

so that $w_{k|k-1}^N \ll \mu_{k|k-1}^N$.

Since $\langle w_{k|k-1}, 1 \rangle = 0$, ask that particle approximation satisfies $\langle w_{k|k-1}^N, 1 \rangle = 0$ as well, by requiring

$$\sum_{i=1}^N \rho_{k|k-1}^i = 0 \; .$$

Algorithm : Particle approximations $\mu_{k|k-1}^N$ and $w_{k|k-1}^N$ to optimal filter $\mu_{k|k-1}$ and its directional derivative $w_{k|k-1}$, completely characterized by particle and weight system $\{\xi_{k|k-1}^i, \rho_{k|k-1}^i \, , \, i = 1, \cdots, N\}$.

Transition from $\{\xi_{k|k-1}^i, \rho_{k|k-1}^i \, , \, i = 1, \cdots, N\}$ to $\{\xi_{k+1|k}^i, \rho_{k+1|k}^i \, , \, i = 1, \cdots, N\}$ described by following steps :

(i) Correction : for all $i = 1, \cdots, N$, compute weight

$$\omega_k^i = c_k^{-1} \, \Psi_k(\xi_{k|k-1}^i) \qquad \text{with} \qquad c_k = \sum_{i=1}^N \Psi_k(\xi_{k|k-1}^i) \; .$$

Then set

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

and

$$\begin{aligned} w_k^N &= F_k(\mu_{k|k-1}^N) \, w_{k|k-1}^N \\ &= \sum_{i=1}^N [\, \rho_{k|k-1}^i - \sum_{j=1}^N \rho_{k|k-1}^j \, \omega_k^j \,] \;\, \omega_k^i \, \delta_{\xi_{k|k-1}^i} \; . \end{aligned}$$

(ii) Resampling : independently for all $i = 1, \cdots, N$, generate a r.v. ξ_k^i with discrete probability distribution μ_k^N (e.g. generate r.v.'s $\{\tau(i) \, , \, i = 1, \cdots, N\}$ with values in I according to some resampling rule, and set

$$\xi_k^i = \xi_{k|k-1}^{\tau(i)} \qquad \text{and} \qquad \rho_k^i = \rho_{k|k-1}^{\tau(i)} - \frac{1}{N} \sum_{j=1}^N \rho_{k|k-1}^{\tau(j)} \; ,$$

for all $i = 1, \cdots, N$).

(iii) Prediction : independently for all $i = 1, \cdots, N$, generate a r.v. W_k^i with density p_k, and set

$$\begin{aligned} \xi_{k+1|k}^i &= f_k(\xi_k^i) + g_k(\xi_k^i) \, W_k^i \; , \\ \Xi_{k+1|k}^i &= \frac{-p'_k}{p_k}(W_k^i) \; g_k^*(\xi_k^i) \, a_k^{\ominus}(\xi_k^i) \, \partial f_k(\xi_k^i) \; , \\ \rho_{k+1|k}^i &= \rho_k^i + \Xi_{k+1|k}^i \; . \end{aligned}$$

Then set

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

and

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

Related work : particle approximation of signed measures, absolutely continuous w.r.t. optimal filter, and associated with additive functionals of hidden state sequence (generalization of auxiliary function in EM algorithm).

Olivier Cappé '01

Application to log–likelihood function ...

Recall that :

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

and

$$\langle \mu_{k|k-1}, \Psi_k \rangle \approx \langle \mu^N_{k|k-1}, \Psi_k \rangle = \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi^i_{k|k-1}) \ ,$$

hence

$$\ell_n \approx \ell^N_n = \sum_{k=0}^{n} \log[\, \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi^i_{k|k-1}) \,] \ .$$

... and to score function

Recall that :

$$\zeta_n = \sum_{k=0}^{n} \frac{\langle w_{k|k-1}, \Psi_k \rangle}{\langle \mu_{k|k-1}, \Psi_k \rangle} \ ,$$

and

$$\frac{\langle w_{k|k-1}, \Psi_k \rangle}{\langle \mu_{k|k-1}, \Psi_k \rangle} \approx \frac{\langle w^N_{k|k-1}, \Psi_k \rangle}{\langle \mu^N_{k|k-1}, \Psi_k \rangle} = \frac{\displaystyle\sum_{i=1}^{N} \rho^i_{k|k-1} \, \Psi_k(\xi^i_{k|k-1})}{\displaystyle\sum_{i=1}^{N} \Psi_k(\xi^i_{k|k-1})} \ ,$$

hence

$$\zeta_n \approx \zeta^N_n = \sum_{k=0}^{n} [\sum_{i=1}^{N} \rho^i_{k|k-1} \, \omega^i_k \,] \ .$$

Convergence analysis

Global error in particle approximation of linear tangent optimal filter results from

(i) accumulation of local error, essentially due to sampling,

(ii) transport by linear tangent filtering operators.

Two contributions to local error :
standard deviation $O(1/\sqrt{N})$, and bias $O(1/N)$.

Contraction property of linear tangent filtering operators, should hold under usual mixing assumption on unnormalized filtering kernels $R_n(x, dx') = Q_n(x, dx') \, \Psi_n(x')$.

FG and Nadia Oudjane '01
FG and Laurent Mevel '00

Perspectives

- Convergence analysis as $N \to \infty$.
- Large–time asymptotics as $n \to \infty$.
- Continuous time hidden state process, and effect of a time discretization scheme, e.g. Euler.
- Correlated (possibly non–Gaussian) noise sequences $\{W_k\}$ and $\{V_k\}$.
- Noise–free observations (ongoing work by Arnaud Doucet).
- Further numerical experiments.
- Application to industrial problems.