Smooth interacting Monte Carlo approximation of contrast functions in HMM

François Le Gland
IRISA / INRIA Rennes

legland@irisa.fr
http://www.irisa.fr/aspi/

joint work with Natacha Caylus (Université de Rennes 1)
Arnaud Guyader (Université de Haute Bretagne, Rennes)
and Nadia Oudjane (Électricité de France R&D, Clamart)
Plan

• contrast functions in HMM
• interacting particle approximation
• global interacting particle approximation
• smooth interacting particle approximation
• linear tangent filter
• linear tangent filter: interacting particle approximation
• conclusion
statistical model, depending on a parameter

- Markov chain \( \{X_k, k \geq 0\} \) on \( E \), with Markov transition kernels

\[
P[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')
\]

and initial probability distribution

\[
P[X_0 \in dx] = \mu_0(dx)
\]

- observations \( \{Y_k, k \geq 0\} \) with values in \( F \), satisfying the memoryless channel assumption, with

\[
P[Y_k \in dy' \mid X_k = x'] = g_k(x', y') \lambda_k(dy')
\]

and likelihood function

\[
\Psi_k(x') = g_k(x', Y_k)
\]
conditional probability distributions

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$$
and
$$\mu_{k|k-1}(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k-1}]$$

going from $\mu_{k-1}$ to $\mu_k$ : decomposition into prediction / correction steps

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

where $\cdot$ denotes projective product, i.e.

$$\mu_k = \Psi_k \cdot \mu_{k|k-1} = \frac{\Psi_k \mu_{k|k-1}}{\mu_{k|k-1}(\Psi_k)}$$
or in a single step, introducing $R_k(x, dx') = Q_k(x, dx') \Psi_k(x')$

$$
\mu_{k-1} \rightarrow \mu_k = \frac{\mu_{k-1} R_k}{(\mu_{k-1} R_k)(E)}
$$

indeed

$$
(\mu_{k-1} R_k)(d x') = \int_E \mu_{k-1}(d x) Q_k(x, dx') \Psi_k(x')
$$

and

$$
(\mu_{k-1} R_k)(E) = \int_E \mu_{k|k-1}(d x') \Psi_k(x') = \mu_{k|k-1}(\Psi_k)
$$

log–likelihood function

$$
\ell_n = \sum_{k=1}^n \log \mu_{k|k-1}(\Psi_k) = \sum_{k=1}^n \log(\mu_{k-1} R_k)(E)
$$
Plan

- contrast functions in HMM
- **interacting particle approximation**
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter : interacting particle approximation
- conclusion
recall that

\[ \mu_{k-1} \rightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1} R_k}{(\mu_{k-1} R_k)(E)} \]

approximation : weighted empirical distribution

\[ \mu_k \approx \mu_k^N = \sum_{i=1}^{N} w_k^i \delta_{\xi_k^i} \quad \text{with} \quad \sum_{i=1}^{N} w_k^i = 1 \]

associated with particle system \( \Sigma_k = \{ (\xi_k^i, w_k^i) , i = 1 \cdots N \} \)

arbitrary importance decomposition of nonnegative kernel into

\[ R_k(x, dx') = P_k(x, dx') W_k(x, x') \]

with

- Markov mutation kernel \( P_k(x, dx') \), easy to simulate from

- selection function \( W_k(x, x') \), easy to evaluate
a natural importance decomposition is available

\[ R_k(x, dx') = Q_k(x, dx') \Psi_k(x') \]

example: sampling the solution of an SDE

\[ dX'_t = b(X'_t) \, dt + \sigma(X'_t) \, dW'_t \]

at discrete time instants \( t_0 < \cdots < t_k < \cdots \), yields a Markov chain \( X_k = X'_t \)

no explicit expression is available in general for the Markov kernel

\[ Q_k(x, dx') = \mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = \mathbb{P}[X'_{t_k} \in dx' \mid X'_{t_{k-1}} = x] \]

but it is easy to simulate the r.v. \( X_k \) given \( X_{k-1} = x \):

use any discretization scheme of the SDE between time instants \( t_{k-1} \) and \( t_k \), with initial condition \( X'_{t_{k-1}} = x \) \( \Box \)
starting from particle approximation

\[ \mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^{N} w_{k-1}^i \delta_{\xi_{k-1}^i} \]

and applying nonnegative kernel \( R_k(x, dx') \) exactly, yields

\[
(\mu_{k-1}^N R_k)(dx') = \sum_{i=1}^{N} w_{k-1}^i \ R_k(\xi_{k-1}^i, dx')
\]

\[= \sum_{i=1}^{N} \frac{w_{k-1}^i}{\pi_k^i} \ W_k(\xi_{k-1}^i, x') \ \pi_k^i \ P_k(\xi_{k-1}^i, dx') \]

with arbitrary importance decomposition of discrete weights into

\[ w_{k-1}^i = \frac{w_{k-1}^i}{\pi_k^i} \pi_k^i \]

auxiliary particle idea of Pitt and Shephard: sample product space \( E \times \{1 \cdots N\} \)
SIR (sampling / importance resampling) algorithm

going from $\Sigma_{k-1}$ to $\Sigma_k$

- **selection** of particles with higher weights: independently for $i = 1 \cdots N$
  \[
  \tau_k^i \sim (\pi_k^1 \cdots \pi_k^N) \text{ with values in index set } \{1 \cdots N\}
  \]

- **mutation** using importance Markov kernel: independently for $i = 1 \cdots N$
  \[
  \xi_k^i \sim P_k(\xi_k^{\tau_k^i}, dx')
  \]

- **weighting** according to importance weight: for $i = 1 \cdots N$
  \[
  w_k^i = \frac{1}{c_k^N} \frac{w_{k-1}^{\tau_k^i}}{\pi_k^i} \ W_k(\xi_k^{\tau_k^i}, \xi_k^i) \text{ s.t. } \sum_{k=1}^N w_k^i = 1
  \]

Discrete weights $(w_{k-1}^1 \cdots w_{k-1}^N)$ are used for selection and / or weighting.
particle approximation of log–likelihood function

\[ \ell_n = \sum_{k=0}^{n} \log(\mu_{k-1} R_k)(E) \]

using SIR algorithm

\[ \ell_n \approx \ell^N_n = \sum_{k=0}^{n} \log\left[ \frac{1}{N} \sum_{i=1}^{N} \frac{w_{k-1}^{\tau^i_k}}{\pi_k^{\tau^i_k}} W_k(\xi_{k-1}^{\tau^i_k}, \xi_k^{\tau^i_k}) \right] = \sum_{k=0}^{n} \log\left[ \frac{1}{N} c_k^N \right] \]

alternatively, using simple bootstrap algorithm

\[ \ell_n \approx \ell^N_n = \sum_{k=0}^{n} \log\left[ \frac{1}{N} \sum_{i=1}^{N} \Psi_k(\xi_k^i) \right] \]

unfortunately, particle systems \( \{\Sigma_k, 0 \leq k \leq n\} \) depend on the parameter pointwise, irregular, Monte Carlo approximation of log–likelihood function
Plan

• contrast functions in HMM
• interacting particle approximation
• global interacting particle approximation
• smooth interacting particle approximation
• linear tangent filter
• linear tangent filter : interacting particle approximation
• conclusion
define $R^0_k(x, dx')$, for some pivot (nominal) parameter value $\theta_0 \in \Theta$

**Assumption AC$_0$**

$$R_k(x, dx') = r_k(x, x') \ R^0_k(x, dx')$$

example: if

$$Q_k(x, dx') = q_k(x, x') \ Q^0_k(x, dx')$$

then

$$R_k(x, dx') = Q_k(x, dx') \ \Psi_k(x') = q_k(x, x') \ \frac{\Psi_k(x')}{\Psi^0_k(x')} \ Q^0_k(x, dx') \ \Psi^0_k(x')$$

\[\underbrace{r_k(x, x')} \quad \underbrace{R^0_k(x, dx')}\]

i.e. Assumption AC$_0$ holds
remark: assumption (★) is equivalent to

\[ \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \mathbb{E}^0[\phi(X_k) \Lambda_k \mid X_{k-1} = x] \]  

(★★)

for any test function \( \phi \), and for some (nonunique) r.v. \( \Lambda_k \).

indeed, if (★) holds, then

\[ \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \int_E Q_k(x, dx') \phi(x') \]

\[ = \int_E Q^0_k(x, dx') q_k(x, x') \phi(x') \]

\[ = \mathbb{E}^0[\phi(X_k) q_k(X_{k-1}, X_k) \mid X_{k-1} = x] \]

i.e. (★★) holds with

\( \Lambda_k = q_k(X_{k-1}, X_k) \)
conversely, if (⋆⋆) holds, then

\[ \int_{E} Q_k(x, dx') \phi(x') = \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] \]

\[ = \mathbb{E}^0[\phi(X_k) \Lambda_k \mid X_{k-1} = x] \]

\[ = \int_{E} \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x'] Q_k^0(x, dx') \phi(x') \]

i.e. (⋆) holds with

\[ q_k(x, x') = \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x'] \]

even though no explicit expression is available in general for

\[ q_k(x, x') = \mathbb{E}^0[\Lambda_k \mid X_{k-1} = x, X_k = x'] \]

it may be easy to simulate jointly under \( \mathbb{P}^0 \) the r.v.'s \( X_k \) and \( \Lambda_k \), given \( X_{k-1} = x \)
example (sampled SDE):

\[ dX'_t = b(X'_t) \, dt + \sigma(X'_t) \, dW'_t \]

if \((b(x) - b^0(x))\) is in the range of \(\sigma(x)\), and if \(\sigma(x)\) has full (column) rank, then \(\mathbb{P} \ll \mathbb{P}^0\), with Radon–Nikodym derivative

\[ \frac{d\mathbb{P}}{d\mathbb{P}^0}{\bigg|}_{\mathcal{F}_t} = \exp \left\{ \int_0^t \left[ \sigma^+(X'_s) (b(X'_s) - b^0(X'_s)) \right] \* dW'_s \right. \]

\[ - \frac{1}{2} \int_0^t \left| \sigma^+(X'_s) (b(X'_s) - b^0(X'_s)) \right|^2 ds \}

with pseudo–inverse \(\sigma^+(x) = [\sigma^*(x) \sigma(x)]^{-1} \sigma^*(x)\), hence under \(\mathbb{P}^0\)

\[ dX'_t = b_0(X'_t) \, dt + \sigma(X'_t) \, dW'_t \]

\(\mathbb{P}^0\)
in addition

\[ \Lambda_k = \exp \left\{ \int_{t_{k-1}}^{t_k} \left[ \sigma^+ (X'_s) (b(X'_s) - b^0(X'_s)) \right]^* dW'_s \\
- \frac{1}{2} \int_{t_{k-1}}^{t_k} |\sigma^+ (X'_s) (b(X'_s) - b^0(X'_s))|^2 ds \right\} \]

and it is easy to simulate jointly under \( \mathbb{P}^0 \) the r.v.'s \( X_k = X'_{t_k} \) and \( \Lambda_k \),
given \( X'_{t_{k-1}} = x \) \( \square \)
recall that

\[ \mu_{k-1} \rightarrow \mu_k = \bar{R}_k(\mu_{k-1}) = \frac{\mu_{k-1} R_k}{(\mu_{k-1} R_k)(E)} \]

smooth approximation: weighted empirical distribution

\[ \mu_k \approx \mu_k^N = \sum_{i=1}^{N} u_k^i w_k^{0,i} \delta_{\xi_k^0,i} \quad \text{with} \quad \sum_{i=1}^{N} w_k^{0,i} = 1 \quad \text{and} \quad \sum_{i=1}^{N} u_k^i w_k^{0,i} = 1 \]

associated with

- unique particle system \( \Sigma_k^0 = \{(\xi_k^{0,i}, w_k^{0,i}), i = 1 \cdots N\} \) defined using the pivot (nominal) value \( \theta_0 \) only, i.e. independent of \( \theta \)
- secondary weights \( S_k = \{u_k^i, i = 1 \cdots N\} \) depending on \( \theta \)

unique source of randomness (from Monte Carlo simulation, including interaction) can expect smooth approximation, if secondary weights are smooth (continuous, differentiable, etc.) w.r.t. parameter
how is this possible? starting from particle approximation

\[ \mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^{N} u_{k-1}^i w_{k-1}^{0,i} \delta_{\xi_{k-1}^{0,i}} \]

and applying nonnegative kernel \( R_k(x, dx') \) exactly, with decomposition

\[ R_k(x, dx') = r_k(x, x') R_k^0(x, dx') = r_k(x, x') W_k^0(x, x') P_k^0(x, dx') \]

under Assumption AC\(_0\), yields

\[ (\mu_{k-1}^N R_k)(dx') = \sum_{i=1}^{N} u_{k-1}^i w_{k-1}^{0,i} R_k(\xi_{k-1}^{0,i}, dx') \]

\[ = \sum_{i=1}^{N} u_{k-1}^i r_k(\xi_{k-1}^{0,i}, x') \frac{w_{k-1}^{0,i}}{\pi_{0,i}^k} W_k^0(\xi_{k-1}^{0,i}, x') \pi_{0,i}^k P_k^0(\xi_{k-1}^{0,i}, dx') \]

\[ \underbrace{r_k^i(x')} \quad \underbrace{w_k^0(x')} \quad \underbrace{m_k^0(dx')} \]

same auxiliary particle idea as above
smooth SIR (sampling / importance resampling) algorithm

going from \((\Sigma_{k-1}^0, S_{k-1})\) to \((\Sigma_k^0, S_k)\)

- **selection** of particles with higher weights: independently for \(i = 1 \cdots N\)

  \[\tau_{k}^{0,i} \sim (\pi_{k}^{0,1} \cdots \pi_{k}^{0,N})\]  \(\text{with values in index set } \{1 \cdots N\}\)

- **mutation** using importance Markov kernel: independently for \(i = 1 \cdots N\)

  \[\xi_{k}^{0,i} \sim P_{k}^{0}(\xi_{k-1}^{0,i}, dx')\]

- **weighting** according to importance weight: for \(i = 1 \cdots N\)

  \[w_{k}^{0,i} = \frac{1}{c_{k}^{0,N}} \frac{w_{k-1}^{0,\tau_{i}^{0,0}}}{\pi_{k}^{0,\tau_{i}^{0,0}}} W_{k}^{0}(\xi_{k-1}^{0,i}, \xi_{k}^{0,i}) \]  \(\text{s.t. } \sum_{k=1}^{N} w_{k}^{0,i} = 1\)

- **updating** secondary weights: for \(i = 1 \cdots N\)

  \[u_{k}^{i} = \frac{1}{b_{k}^{N}} u_{k-1}^{0,i} r_{k}(\xi_{k-1}^{0,i}, \xi_{k}^{0,i}) \]  \(\text{s.t. } \sum_{k=1}^{N} u_{k}^{i} w_{k}^{0,i} = 1\)
particle approximation of log–likelihood function

\[ \ell_n = \sum_{k=0}^{n} \log(\mu_{k-1} R_k)(E) \]

using SIR algorithm

\[ \ell_n \approx \ell_n^N = \sum_{k=0}^{n} \log \left( \frac{1}{N} \sum_{i=1}^{N} u_{k-1}^{0,i} R_k(\xi_{k-1}^{0,i}, \xi_k^{0,i}) \frac{w_{k-1}^{0,i}}{\pi_k^{0,i}} W_k^0(\xi_{k-1}^{0,i}, \xi_k^{0,i}) \right) \]

\[ = \sum_{k=0}^{n} \log \left( \frac{1}{N} \sum_{i=1}^{N} \frac{w_{k-1}^{0,i}}{\pi_k^{0,i}} W_k^0(\xi_{k-1}^{0,i}, \xi_k^{0,i}) \sum_{i=1}^{N} u_{k-1}^{0,i} R_k(\xi_{k-1}^{0,i}, \xi_k^{0,i}) w_k^{0,i} \right) \]

\[ = \sum_{k=0}^{n} \log \left( \frac{1}{N} c_k^{0,N} \right) + \sum_{k=0}^{n} \log b_k^N \]

particle systems \( \{\sum_k^0, 0 \leq k \leq n\} \) do not depend on the parameter

\[ \longrightarrow \text{global}, \text{possibly regular, Monte Carlo approximation of log–likelihood function} \]
convergence result: under technical assumptions

- on the model, including Lipschitz continuity of \( \theta \mapsto r_k(x, x') \) in some sense
- on the importance decomposition of kernels

it holds

\[
\sup_{\|\phi\|=1} \left\{ \mathbb{E} \left[ \sup_{\theta \in \Theta} |\mu^N_n(\phi) - \mu_n(\phi)|^p \right] \right\}^{1/p} \leq \frac{C_{p,n}}{\sqrt{N}}
\]

and

\[
\left\{ \mathbb{E} \left[ \sup_{\theta \in \Theta} |\ell^N_n - \ell_n|^p \right] \right\}^{1/p} \leq \frac{C_{p,n}}{\sqrt{N}}
\]

hence, almost surely as \( N \to \infty \)

\[
\sup_{\theta \in \Theta} |\ell^N_n - \ell_n| \longrightarrow 0
\]

and

\[
\hat{\theta}^N_n = \arg\max_{\theta \in \Theta} \ell^N_n \longrightarrow \arg\max_{\theta \in \Theta} \ell_n = \hat{\theta}_n
\]

interacting particle implementation of the MCML algorithm of Geyer (1994, 1996)
numerical example
comments

- for values of parameter too much away from pivot (nominal) value, essentially weighted Monte Carlo approximation, without interaction: poor approximation, already noticed by Cappé, Douc, Moulines and Robert (2002)

- local (in the vicinity of pivot (nominal) value, if not global) vs. pointwise particle approximation

- can expect smooth approximation, if secondary weights are smooth (continuous, differentiable, etc.) w.r.t. parameter
Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter: interacting particle approximation
- conclusion
nonnegative kernels \( R_k(x, dx') \) differentiable w.r.t. parameter \( \theta \)

Assumption AC

\[
\frac{\partial}{\partial \theta} R_k(x, dx') = s_k(x, x') R_k(x, dx')
\]

remark: if Assumption AC\(_0\) holds, i.e. if

\[
R_k(x, dx') = r_k(x, x') R_k^0(x, dx')
\]

and if functions \( r_k(x, x') \) are differentiable w.r.t. parameter \( \theta \)

then Assumption AC holds, with

\[
s_k(x, x') = \frac{\partial}{\partial \theta} \log r_k(x, x')
\]
remark: recall that, if

$$Q_k(x, dx') = q_k(x, x') Q^0_k(x, dx') \quad (\ast)$$

then Assumption AC$_0$ holds, with

$$r_k(x, x') = q_k(x, x') \frac{\Psi_k(x')}{\Psi^0_k(x')}$$

hence if functions $q_k(x, x')$ and $\Psi_k(x')$ are differentiable w.r.t. parameter $\theta$

then Assumption AC holds, with

$$s_k(x, x') = \frac{\partial}{\partial \theta} \log r_k(x, x') = \frac{\partial}{\partial \theta} \log q_k(x, x') + \frac{\partial}{\partial \theta} \log \Psi_k(x')$$
alternatively, if
\[
\int_{E} Q_k(x, dx') \phi(x') = \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \mathbb{E}^0[\phi(X_k) \Lambda_k \mid X_{k-1} = x] \quad (**) 
\]
for any test function \( \phi \), and for some (nonunique) r.v. \( \Lambda_k \), then
\[
\int_{E} \frac{\partial}{\partial \theta} Q_k(x, dx') \phi(x') = \mathbb{E}^0[\phi(X_k) \Lambda_k \frac{\partial}{\partial \theta} \log \Lambda_k \mid X_{k-1} = x] 
\]
\[
= \mathbb{E}[\phi(X_k) \frac{\partial}{\partial \theta} \log \Lambda_k \mid X_{k-1} = x] 
\]
\[
= \int_{E} \mathbb{E}\left[\frac{\partial}{\partial \theta} \log \Lambda_k \mid X_{k-1} = x, X_{k} = x'\right] Q_k(x, dx') \phi(x') 
\]
i.e.
\[
\frac{\partial}{\partial \theta} Q_k(x, dx') = I_k(x, x') Q_k(x, dx') 
\]
with
\[
I_k(x, x') = \mathbb{E}\left[\frac{\partial}{\partial \theta} \log \Lambda_k \mid X_{k-1} = x, X_{k} = x'\right] 
\]
and

\[
\frac{\partial}{\partial \theta} R_k(x, dx') = \left[ \frac{\partial}{\partial \theta} Q_k(x, dx') + Q_k(x, dx') \frac{\partial}{\partial \theta} \log \Psi_k(x') \right] \Psi_k(x')
\]

\[
= [I_k(x, x') + \frac{\partial}{\partial \theta} \log \Psi_k(x')] R_k(x, dx')
\]

i.e. Assumption AC holds, with

\[s_k(x, x') = I_k(x, x') + \frac{\partial}{\partial \theta} \log \Psi_k(x')\]

even though no explicit expression is available in general for

\[I_k(x, x') = \mathbb{E}\left[ \frac{\partial}{\partial \theta} \log \Lambda_k \mid X_{k-1} = x, X_k = x' \right]\]

it may be easy to simulate jointly under \(\mathbb{P}\) the r.v.’s \(X_k\) and \(\frac{\partial}{\partial \theta} \log \Lambda_k\),

given \(X_{k-1} = x\)
example (sampled SDE):

\[ dX'_t = b(X'_t) \, dt + \sigma(X'_t) \, dW'_t \]

recall that, if \((b(x) - b^0(x))\) is in the range of \(\sigma(x)\), and if \(\sigma(x)\) has full (column) rank, then

\[
\Lambda_k = \exp \left\{ \int_{t_{k-1}}^{t_k} \left[ \sigma^+(X'_s) (b(X'_s) - b^0(X'_s)) \right]^* dW'_s^0 \\
- \frac{1}{2} \int_{t_{k-1}}^{t_k} |\sigma^+(X'_s) (b(X'_s) - b^0(X'_s))|^2 ds \right\}
\]
therefore
\[
\frac{\partial}{\partial \theta} \log \Lambda_k = \int_{t_{k-1}}^{t_k} \left[ \sigma^+(X'_s) \frac{\partial}{\partial \theta} b(X'_s) \right]^* dW'_s
\]

\[
- \int_{t_{k-1}}^{t_k} \left[ \sigma^+(X'_s) \frac{\partial}{\partial \theta} b(X'_s) \right]^* [\sigma^+(X'_s) (b(X'_s) - b^0(X'_s))] ds
\]

\[
= \int_{t_{k-1}}^{t_k} \left[ \sigma^+(X'_s) \frac{\partial}{\partial \theta} b(X'_s) \right]^* [dW'_s - \sigma^+(X'_s) (b(X'_s) - b^0(X'_s))] ds
\]

\[
= \int_{t_{k-1}}^{t_k} \left[ \sigma^+(X'_s) \frac{\partial}{\partial \theta} b(X'_s) \right]^* dW'_s
\]

and it is easy to simulate jointly under \( \mathbb{P} \) the r.v.'s \( X_k = X'_{t_k} \) and \( \frac{\partial}{\partial \theta} \log \Lambda_k \),
given \( X'_{t_{k-1}} = x \)  \( \square \)
recall that
\[ \mu_k \approx \mu_k^N = \sum_{i=1}^{N} u_k^i w_k^{0,i} \delta_{\xi_k^{i}}, \]
with \[ \sum_{i=1}^{N} w_k^{0,i} = 1 \]
and
\[ u_k^i = \frac{1}{b_k^i} \frac{\tau_k^{0,i}}{u_{k-1}^i} r_k(\xi_k^{0}, \tau_k^{0,i}, \xi_k^{0,i}) \]
s.t. \[ \sum_{k=1}^{N} u_k^i w_k^{0,i} = 1 \]
under Assumption AC, secondary weights are differentiable w.r.t. parameter \( \theta \), hence particle approximation is differentiable w.r.t. parameter \( \theta \), with derivative
\[ \dot{\mu}_k^N = \frac{\partial}{\partial \theta} \mu_k^N = \sum_{i=1}^{N} \rho_k^i w_k^{0,i} \delta_{\xi_k^{i}}, \]
with \[ \sum_{k=1}^{N} \rho_k^i w_k^{0,i} = 0 \]
where (using logarithmic derivatives)
\[ \rho_k^i = \frac{\partial}{\partial \theta} u_k^i = \left[ \frac{\tau_k^{0,i}}{u_{k-1}^i} + s_k(\xi_k^{0}, \tau_k^{0,i}, \xi_k^{0,i}) - a_k^N \right] u_k^i \]
at pivot (nominal) value $\theta_0$, it holds

\[
\dot{\mu}_{k}^{0,N} = \left. \frac{\partial}{\partial \theta} \mu_k^N \right|_{\theta=\theta_0} = \sum_{i=1}^{N} \rho_{k}^{0,i} w_{k}^{0,i} \xi_{k}^{0,i}
\]

s.t. $\sum_{k=1}^{N} \rho_{k}^{0,i} w_{k}^{0,i} = 0$

where, taking $u_{k}^{0,i} \equiv 1$ into account

\[
\rho_{k}^{0,i} = \left. \frac{\partial}{\partial \theta} u_{k}^{i} \right|_{\theta=\theta_0} = \rho_{k-1}^{0,\tau_{k}^{0,i}} + s_{k}^{0}(\xi_{k-1}^{0,\tau_{k}^{0,i}}, \xi_{k}^{0,i}) - a_{k}^{0,N}
\]

questions

- is the optimal filter differentiable w.r.t. parameter (e.g. under Assumption AC)?

- does the derivative of the smooth particle approximation provide a reasonable approximation of the linear tangent filter?
Plan

• contrast functions in HMM
• interacting particle approximation
• global interacting particle approximation
• smooth interacting particle approximation
• linear tangent filter
• linear tangent filter: interacting particle approximation
• conclusion
recall that

\[ \mu_{k-1} \rightarrow \mu_k = R_k(\mu_{k-1}) = \frac{\mu_{k-1} R_k}{(\mu_{k-1} R_k)(E)} \]

if Assumption AC holds, i.e. if

\[ \frac{\partial}{\partial \theta} R_k(x, dx') = s_k(x, x') R_k(x, dx') \]

then

\[ (\mu_{k-1} \frac{\partial}{\partial \theta} R_k)(dx') = \int_E \mu_{k-1}(dx) s_k(x, x') R_k(x, dx') \]

and \( \{\mu_k, k \geq 0\} \) is differentiable w.r.t. parameter, with derivative \( \{\dot{\mu}_k, k \geq 0\} \) given by linear tangent equation

\[ (\mu_{k-1}, \dot{\mu}_{k-1}) \rightarrow \dot{\mu}_k = \frac{\dot{\mu}_{k-1} R_k + \mu_{k-1} \frac{\partial}{\partial \theta} R_k}{(\mu_{k-1} R_k)(E)} - a_k \mu_k \]

with constant \( a_k \) s.t. signed measure \( \dot{\mu}_k \) has zero total mass
Plan

• contrast functions in HMM
• interacting particle approximation
• global interacting particle approximation
• smooth interacting particle approximation
• linear tangent filter
• linear tangent filter : interacting particle approximation
• conclusion
by induction \( \hat{\mu}_k \ll \mu_k \)

joint approximation: weighted empirical distributions

\[
\mu_k \approx \mu_k^N = \sum_{i=1}^{N} w_k^i \delta_{\xi_k^i} \quad \text{with} \quad \sum_{i=1}^{N} w_k^i = 1
\]

\[
\hat{\mu}_k \approx \hat{\mu}_k^N = \sum_{i=1}^{N} \rho_k^i w_k^i \delta_{\xi_k^i} \quad \text{with} \quad \sum_{i=1}^{N} \rho_k^i w_k^i = 0
\]

associated with

- unique particle system \( \Sigma_k = \{(\xi_k^i, w_k^i), \ i = 1 \cdots N\} \)

- signed weights \( S'_k = \{\rho_k^i, \ i = 1 \cdots N\} \)
starting from particle approximations

\[
\mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^{N} w_{k-1}^i \delta_{\xi^i_{k-1}}
\]

\[
\dot{\mu}_{k-1} \approx \dot{\mu}_{k-1}^N = \sum_{i=1}^{N} \rho_{k-1}^i w_{k-1}^i \delta_{\xi^i_{k-1}}
\]

and applying nonnegative kernels \( R_k(x, dx') \) and \( \frac{\partial}{\partial \theta} R_k(x, dx') \) exactly, yields

\[
(\mu_{k-1}^N R_k)(dx') = \sum_{i=1}^{N} w_{k-1}^i R_k(\xi^i_{k-1}, dx')
\]

\[
= \sum_{i=1}^{N} \frac{w_{k-1}^i}{\pi_k^i} W_k(\xi^i_{k-1}, x') \pi_k^i P_k(\xi^i_{k-1}, dx')
\]

\[
\underbrace{w_k^i(x')} \quad \underbrace{m_k^i(dx')}
\]
and

\[
(\dot{\mu}_{k-1}^N R_k + \mu_{k-1}^N \frac{\partial}{\partial \theta} R_k)(dx')
\]

\[
= \sum_{i=1}^{N} w_{k-1}^i \left[ \rho_{k-1}^i + s_k(\xi_{k-1}^i, x') \right] R_k(\xi_{k-1}^i, dx')
\]

\[
= \sum_{i=1}^{N} \frac{w_{k-1}^i}{\pi_k^i} W_k(\xi_{k-1}^i, x') \left[ \rho_{k-1}^i + s_k(\xi_{k-1}^i, x') \right] \pi_k^i P_k(\xi_{k-1}^i, dx')
\]

under Assumption AC

same auxiliary particle idea as above
joint SIR (sampling / importance resampling) algorithm going from \((\Sigma_{k-1}, S'_{k-1})\) to \((\Sigma_k, S'_k)\)

- **selection** of particles with higher weights: independently for \(i = 1 \cdots N\)

\[ \tau^i_k \sim (\pi^i_k \cdots \pi^N_k) \] with values in index set \(\{1 \cdots N\}\)

- **mutation** using importance Markov kernel: independently for \(i = 1 \cdots N\)

\[ \xi^i_k \sim P_k(\xi_{k-1}^{\tau^i_k}, dx') \]

- **weighting** according to importance weight: for \(i = 1 \cdots N\)

\[ w^i_k = \frac{1}{c^N_k} \frac{w_{k-1}^{\tau^i_k}}{\pi^i_k} W_k(\xi_{k-1}^{\tau^i_k}, \xi^i_k) \mathrm{s.t.} \sum_{k=1}^{N} w^i_k = 1 \]

- **updating** signed weights: for \(i = 1 \cdots N\)

\[ \rho^i_k = \rho^i_{k-1} + s_k(\xi_{k-1}^{\tau^i_k}, \xi^i_k) - a^N_k \mathrm{s.t.} \sum_{k=1}^{N} \rho^i_k w^i_k = 0 \]
particle approximation of score function

recall expression for log–likelihood function

\[ \ell_n = \sum_{k=0}^{n} \log(\mu_{k-1} R_k)(E) \]

hence

\[ \frac{\partial}{\partial \theta} \ell_n = \sum_{k=0}^{n} \frac{(\mu_{k-1} R_k + \mu_{k-1} \frac{\partial}{\partial \theta} R_k)(E)}{(\mu_{k-1} R_k)(E)} \]

using SIR algorithm

\[ \ell_n \approx \ell_n^N = \sum_{k=0}^{n} \log\left[ \frac{1}{N} \sum_{i=1}^{N} \frac{w_{k-1}^i}{\pi_k} W_k(\xi_{k-1}^i, \xi_k^i) \right] = \sum_{k=0}^{n} \log\left[ \frac{1}{N} c_k^N \right] \]

and

\[ \frac{\partial}{\partial \theta} \ell_n \approx \frac{\partial}{\partial \theta} \ell_n^N = \sum_{k=0}^{n} \sum_{i=1}^{N} w_k^i \left[ \rho_{k-1}^i + s_k(\xi_{k-1}^i, \xi_k^i) \right] = \sum_{k=0}^{n} a_k^N \]
Plan

- contrast functions in HMM
- interacting particle approximation
- global interacting particle approximation
- smooth interacting particle approximation
- linear tangent filter
- linear tangent filter: interacting particle approximation
- conclusion
conclusion

• two different approaches to obtain interacting particle approximation of linear tangent filter

• more general version, under weaker form of Assumptions $AC_0$ and $AC$ (no explicit expression for $r_k(x, x')$ and $s_k(x, x')$)
references and some related work


- interacting particle (or MCMC) implementation of fully Bayesian approach: dominant approach, with many contributions, including recently Papavasiliou (PhD thesis 2002, 2003), Rossi (PhD thesis 2004)