High-dimensional inference using nested particle filters

— Nested sequential Monte Carlo methods

Christian A. Naesseth

Joint work with Fredrik Lindsten and Thomas B. Schön

LINKÖPING UNIVERSITY

Nonlinear filtering

The filtering problem for a nonlinear state space model,

$$x_{t+1} \mid x_t \sim f(x_{t+1} \mid x_t),$$
$$y_t \mid x_t \sim g(y_t \mid x_t),$$

amounts to computing

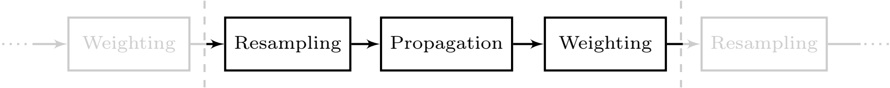
$$p(x_t \mid y_{1:t}) = \frac{g(y_t \mid x_t) \int f(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1}}{p(y_t \mid y_{1:t-1})}$$

for $t = 1, 2, \ldots$

The bootstrap filter

The *bootstrap particle filter* approximates $p(x_t \mid y_{1:t})$ by

$$\widehat{p}^N(x_t \mid y_{1:t}) := \sum_{i=1}^{N} \frac{W_t^i}{\sum_\ell W_t^\ell} \delta_{X_t^i}(x_t).$$

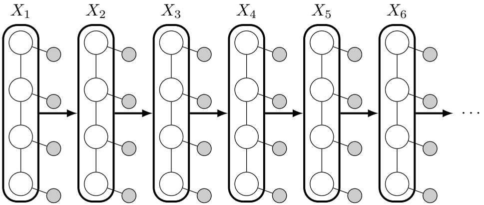


- *Resampling:* $\{(X_{t-1}^i, W_{t-1}^i)\}_{i=1}^N \to \{(\tilde{X}_{t-1}^i, 1/N)\}_{i=1}^N$.
- *Propagation:* $X_t^i \sim f(x_t \mid \tilde{X}_{t-1}^i)$.
- *Weighting:* $W_t^i = g(y_t \mid X_t^i)$.

$$\Rightarrow \{(X_t^i, W_t^i)\}_{i=1}^N$$

Particle filters in high dimension

- Known to perform poorly in high (say, $d \gtrsim 10$) dimensions.
- *ex)* Spatio-temporal model: $g(y_t \mid x_t) = \prod_{k=1}^{d} g(y_{t,k} \mid x_{t,k})$.



- $f(x_t \mid x_{t-1})$ is typically an *extremely* bad proposal distribution in HD.

Does a better proposal distribution improve our result?

Preview of the idea

- Optimal *proposals* and *resampling weights* — known conceptually but intractable to compute.
- Deterministic (e.g., Gaussian) approximations available, but often inadequate in high dimensions.

Idea behind Nested SMC:

- Use *SMC* to approximate the optimal proposals and resampling weights.
- Sampling distribution not available on closed form — still possible to obtain a valid algorithm!
- Nested SMC satisfies the conditions on the proposal approximation $\Rightarrow$ possible to use within itself (nesting to arbitrary degree).

Fully adapted auxiliary SMC sampler

Let $\bar{\pi}_t(x_{1:t}) = \mathcal{Z}_t^{-1}\pi_t(x_{1:t})$ for $t = 1, 2, \dots$ be a sequence of *target distributions*.

- *Optimal proposal:*
  $\bar{q}_t(x_t \mid x_{1:t-1}) = Z_t^{-1}(x_{1:t-1})q_t(x_t \mid x_{1:t-1})$, where

$$q_t(x_t \mid x_{1:t-1}) := \frac{\pi_t(x_{1:t})}{\pi_{t-1}(x_{1:t-1})} \qquad [= g(y_t \mid x_t) f(x_t \mid x_{t-1})]$$

- *Optimal resampling weights:*

$$\widetilde{W}_{t-1}^i := Z_t(X_{1:t-1}^i) \qquad [= p(y_t \mid X_{t-1}^i)]$$

Results in an *unweighted* set of particles $\{X_{1:t}^i\}_{i=1}^N$, such that $\bar{\pi}_t^N(x_{1:t}) = \frac{1}{N}\sum_{i=1}^N \delta_{X_{1:t}^i}(x_{1:t})$ approximates $\bar{\pi}_t$.

Nested SMC (I/II)

Some definitions...

Definition (Properly weighted sample). *Let* $\bar{\pi}(x) = \mathcal{Z}^{-1}\pi(x)$ *be a PDF. A (random) pair* $(X, W) \in \mathsf{X} \times \mathbb{R}_+$ *is properly weighted for* π *if* $\mathbb{E}[f(X)W] = \bar{\pi}(f)\mathcal{Z}$ *for all nonnegative measurable functions* f*.*

(A1) Let Q be a *class* and let $\mathrm{q} = \mathrm{Q}(q, M)$. Assume that:

1. At construction of q a (random) member variable is generated, accessible as $\widehat{Z} = \mathrm{q.GetZ()}$.
2. Q has a member function Simulate which returns a (random) variable $X = \mathrm{q.Simulate()}$ such that $(X, \widehat{Z})$ is properly weighted for q.

Fully adapted and nested SMC

Given $\{X^i_{1:t-1}\}_{i=1}^N$ targeting $\bar{\pi}_{t-1}(x_{1:t-1})$:

	Fully adapted SMC	Nested SMC
• *Initialisation:*	—	$\mathrm{q}^i = \mathrm{Q}(q_t(\cdot \mid X^i_{1:t-1}), M)$
• *Resampling weights:*	$Z_t(X^i_{1:t-1})$	$\widehat{Z}^i_t = \mathrm{q}^i.\mathrm{GetZ}()$
• *Propagation:*	$X^i_t \sim \bar{q}_t(x_t \mid X^{A^i_t}_{1:t-1})$	$X^i_t = \mathrm{q}^{A^i_t}.\mathrm{Simulate}()$

$\Rightarrow \{X^i_{1:t}\}_{i=1}^N$ targeting $\bar{\pi}_t(x_{1:t})$.

Nesting of NSMC

How can we construct the class Q with the desired properties?

One way is to use SMC or NSMC for this as well. (Hence the word "Nested" in "Nested SMC".)

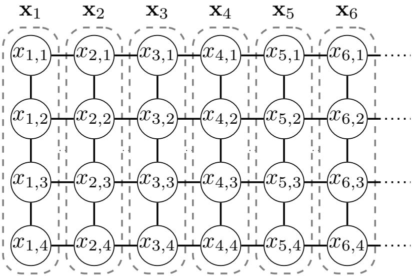
Theorem 1. *For a (properly weighted) NSMC algorithm with target $\pi_n(x_{1:n})$, let*

- $\widehat{\mathcal{Z}}_n := \prod_{t=1}^{n} \left\{ \frac{1}{N} \sum_{i=1}^{N} \widehat{Z}_t^i \right\}$,
- $X'_{1:n}$ *be generated by e.g. (standard) SMC-based backward simulation.*

Then, the pair $(X'_{1:n}, \widehat{\mathcal{Z}}_n)$ is properly weighted for π_n.

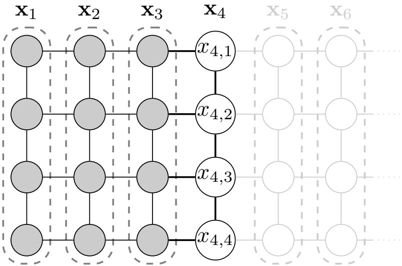
2D Markov Random Field – Nested SMC applied

1 spatial + 1 temporal dimension



$$\bar{\pi}_t(\mathbf{x}_{1:t}) = \frac{1}{\mathcal{Z}_t}\boldsymbol{\phi}_1(\mathbf{x}_1)\prod_{s=2}^{t}\left\{\boldsymbol{\phi}_s(\mathbf{x}_s)\boldsymbol{\psi}_s(\mathbf{x}_{s-1},\mathbf{x}_s)\right\}.$$

2D MRF – Nested SMC implementation (I/III)



Optimal proposals given by:

$$q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \boldsymbol{\phi}_t(\mathbf{x}_t)\boldsymbol{\psi}_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$$

$$= \left\{ \prod_{k=1}^{d} G_{t,k}(x_{t,k}) \prod_{k=2}^{d} m(x_{t,k-1}, x_{t,k}) \right\} \left\{ \prod_{k=1}^{d} \psi(x_{t-1,k}, x_{t,k}) \right\}$$

2D MRF – Nested SMC implementation (II/III)

Proposed algorithm:

Step 1: Initialisation

(Optimal weights $\widetilde{W}^i_{t-1} = Z_t(\mathbf{X}^i_{t-1})$ with $Z_t(\mathbf{x}_{t-1}) = \int q_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \mathrm{d}\mathbf{x}_t$.)

- For each particle $\{\mathbf{X}^i_{t-1}\}_{i=1}^N$:
  - Run PF with M particles for target $q_t(\mathbf{x}_t \mid \mathbf{X}^i_{t-1})$.
  - Estimate normalising constant: $\widehat{Z}^i_t = \prod_{k=1}^d \left\{ \frac{1}{N} \sum_{j=1}^M W^{i,j}_k \right\}$.

Step 2: Resampling

Resample $\{\mathbf{X}^i_{t-1}\}_{i=1}^N$ *and corresponding PFs* based on $\{\widehat{Z}^i_t\}_{i=1}^N$

2D MRF – Nested SMC implementation (III/III)

Step 3: Propagation

- Assume particle $\mathbf{X}_{t-1}^i$ resampled n_t^i times.
- For $i = 1, \ldots, N$, generate n_t^i descendants of $\mathbf{X}_{t-1}^i$ by *backward simulation*:
  - $\mathbb{P}(X'_{t,d} = X_{t,d}^{i,j}) = W_{t,d}^{i,j} \quad (j = 1, \ldots, M)$.
  - For $k = d - 1$ to 1,

$$\mathbb{P}(X'_{t,k} = X_{t,k}^{i,j}) = \frac{W_{t,k}^{i,j} m(X_{t,k}^{i,j}, X'_{t,k+1})}{\sum_{\ell=1}^{M} W_{t,k}^{i,\ell} m(X_{t,k}^{i,\ell}, X'_{t,k+1})} \quad (j = 1, \ldots, M).$$

$$\Rightarrow \mathbf{X}'_t = X'_{t,1:d} \overset{\text{approx.}}{\sim} \bar{q}_t(\cdot \mid \mathbf{X}_{t-1}^i).$$

Results in N *unweighted* particles: $\{\mathbf{X}_t^i\}_{i=1}^N$

Convergence of NSMC

Theorem 2. *Assume that* Q *satisfies the proper weighting condition* ***(A1)****. Then,*[1]

$$N^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N} f(X_{1:t}^{i}) - \bar{\pi}_t(f)\right) \xrightarrow{D} \mathcal{N}(0, \Sigma_t^M(f)),$$

where $\{X_{1:t}^{i}\}_{i=1}^{N}$ *are generated by the NSMC algorithm.*

- We obtain the standard $\sqrt{N}$ Monte Carlo rate.
- The convergence holds for *any value* of the precision parameter M.
- The asymptotic variance $\Sigma_t^M(f)$ does depend on M.

[1] Under certain regularity conditions on the test function f.

Related work

Monte Carlo approximation of SMC:

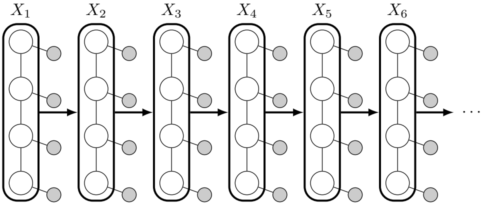
- Random-weight PF [2] – uses unbiased estimates of weights

NSMC uses unbiased estimates of weights + "approximate draws" from proposal (proper weighting condition)

SMC within SMC:

- SMC^2 [3] – joint $(\theta, x_{1:t})$, internal SMC for marginalisation
- EA-RBPF [4] – split state $x_t = (x_t^1, x_t^2)$, internal SMC for marginalisation
- Island PF [5] – PF where each particle is a PF.
- Space-time PF [6] – island PF for spatio-temporal models.

ex) Gaussian spatio-temporal model



Gaussian spatio-temporal model in the form of a 2D MRF, $d \times t$, i.e. $\dim x_t = d$.

$$p(x_{1:t}, y_{1:t}) \propto \prod_{s=1}^{t} \underbrace{\mathcal{N}(y_s; x_s, \tau^{-1} I)}_{G} \underbrace{\mathcal{N}(x_s; a x_{s-1}, I)}_{\psi} \underbrace{\mathcal{N}(x_s; 0, \Sigma)}_{m}$$

where Σ^{-1} is a banded matrix (reflecting local dependencies).

ex) Gaussian spatio-temporal model

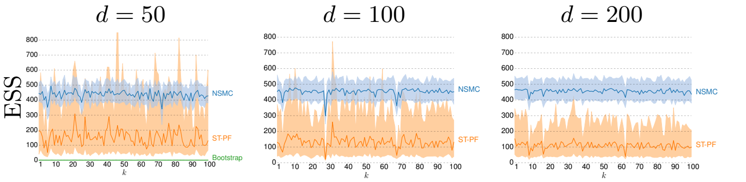
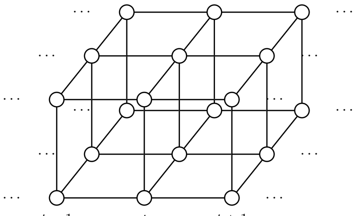


Figure: Median (over dimension) effective sample size (ESS) and 15–85% percentiles. $N = 500$ and $M = 2d$. (Results for 100 independent runs.)

$$\mathsf{ESS}_{t,k} := \left(\mathbb{E}\left[\frac{(\hat{x}_{t,k} - \mu_{t,k})^2}{\sigma^2_{t,k}} \right] \right)^{-1}$$

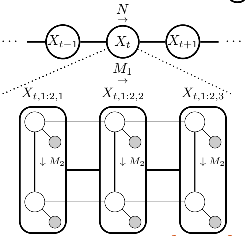
Beskos, A., Crisan, D., Jasra, A., Kamatani, K. and Zhou, Y. *A Stable Particle Filter in High-Dimensions*. arXiv:1412.3501, Dec. 2014.

ex) Spatio-temporal model for drought prediction



- System state $x_t = \{x_{t,k,\ell}\}_{k=1,\ell=1}^{K,L}$, i.e., dimension is $d = K \times L$.
- Binary variables: $x_{t,k,\ell} = 0$ (normal state) or $x_{t,k,\ell} = 1$ (drought).
- Yearly Gaussian observations of precipitation at each site.

ex) Spatio-temporal model for drought prediction



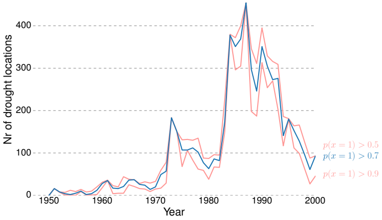
Exploit the rectangular structure in three levels:

Level 1: Instantiate a Nested SMC sampler targeting the full posterior filtering distribution.

Level 2: To sample x_t, we run a Nested SMC sampler, operating on the "columns" $x_{t,1:K,\ell}$, $\ell = 1, \ldots, L$.

Level 3: To sample each column $x_{t,1:K,\ell}$ we run a third level of SMC, operating on the individual components $x_{t,k,\ell}$, $k = 1, \ldots, K$.

ex) Spatio-temporal model for drought prediction



- Data from the Sahel region in Africa for years 1950–2000.
- $\{K, L\} = \{24, 44\}$ $(\Rightarrow d = 1\,056)$.
- $\{N, M_1, M_2\} = \{100, 40, 20\}$.

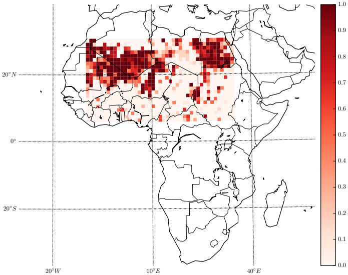


Figure: Sahel region in 1989.

Wrapping up

Summary:

- NSMC allows us to *"exactly approximate"* a fully adapted SMC sampler.
- Forward-backward strategy for lattice models.
- Provably correct for any number(s) M of particles in the "internal" filter(s).
- Modular to an arbitrary degree.
- Pushes the dimension-limit for SMC from "tens" to "hundreds" (?).

Worth to note:

- Can straightforwardly be used with, e.g., Particle MCMC for learning.
- Computational complexity of the method is $N \times M$ (for two layers). However, *much more efficient* than a bootstrap PF with $N \times M$ particles!
- NSMC *does not* beat the curse of dimensionality!
- Can we push the limit even further with blocking and localisation strategies?

References

C. A. Naesseth, F. Lindsten, and T. B. Schön.
Nested sequential Monte Carlo methods.
In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, Lille, France, July 2015.

P. Fearnhead, O. Papaspiliopoulos, G. O. Roberts, and A. Stuart.
Random-weight particle filtering of continuous time processes.
Journal of the Royal Statistical Society: Series B, 72(4):497–512, 2010.

N. Chopin, P. E. Jacob, and O. Papaspiliopoulos.
SMC^2: an efficient algorithm for sequential analysis of state-space models.
Journal of the Royal Statistical Society: Series B, 75(3):397–426, 2013.

A. Johansen, N. Whiteley, and A. Doucet.
Exact approximation of Rao-Blackwellised particle filters.
In *Proceedings of the 16th IFAC Symposium on System Identification (SYSID)*, Brussels, Belgium, July 2012.

C. Vergé, C. Dubarry, P. Del Moral, and E. Moulines.
On parallel implementation of sequential Monte Carlo methods: the island particle model.
Statistics and Computing, 25(2):243–260, 2015.

A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou.
A stable particle filter in high-dimensions.
arXiv.org, arXiv:1412.3501, December 2014.

NSMC and SMC Samplers

So far: Spatio-temporal models, where each x_t in itself has a chain-like or grid-like structure.

What about "general" high-dimensional problems?

SMC samplers can be used also for *fixed-dimensional* problems!

- $\bar{\pi}(x)$ distribution of interest.
- Construct a *bridging sequence* from some easy-to-sample $\bar{\pi}_0(x)$ to $\bar{\pi}(x)$.
- Run SMC for this sequence — keep only the particles at the final "time" step.

In our case:

- Want to sample from $\bar{q}_t(x_t \mid x_{1:t-1})$, the optimal proposal.
- *Bridging sequence*, e.g. from $\bar{\pi}_0(x_t) = f(x_t \mid x_{t-1})$ to $\bar{\pi}(x_t) = p(x_t \mid x_{t-1}, y_t)$.
- Run NSMC with "internal" SMC sampler for the bridging sequence.

Thank you!

www.liu.se

LINKÖPING UNIVERSITY