

Divide-and-Conquer Sequential Monte Carlo

Theoretical Properties and Limit Theorems

Adam M. Johansen

Joint work with Francesca Crucinio and Juan Kuntz

Applied Probability Seminar—May 15th, 2026

University of Warwick

a.m.johansen@warwick.ac.uk

<http://go.warwick.ac.uk/amjohansen/talks/>



- What is Sequential Monte Carlo (SMC)
- Some things we know about SMC
- What is Divide-and-Conquer SMC?
(D&C-SMC; Lindsten et al. (2017))
- What knowledge can we transfer from SMC to D&C-SMC?
- Some Theoretical Properties of D&C-SMC
(Kuntz et al., 2024)

Sequential Monte Carlo

The Essential Problem and SMC Solution

SMC Ingredients:

- Sequence of unnormalized (pathwise) targets $\hat{\rho}_t$ on $\mathbf{E}_t = \otimes_{s=0}^t E_s$.
- Normalizing constants $\hat{Z}_t = \hat{\rho}_t(\mathbf{E}_t)$
- Normalized counterparts $\hat{\mu}_t = \hat{\rho}_t / Z_t$.
- Proposals K_t : conditional laws over E_t given $\mathbf{x}_{t-1} \in \mathbf{E}_{t-1}$.
- Importance weights / potentials:

$$G_t = \frac{d\hat{\rho}_t}{d\hat{\rho}_{t-1} \otimes K_t}.$$

Goals

- Estimate $\hat{Z}_1, \dots, \hat{Z}_t, \dots$
- Approximate $\hat{\mu}_1, \dots, \hat{\mu}_t, \dots$

Algorithm

- Iterative importance sampling (IS) and resampling.

The Sequential Importance Resampling Algorithm

1: *Initialise*: for $n \leq N$, draw $\mathbf{X}_0^{n,N}$ independently from K_0 .

2: **for** $t = 1, \dots, T$ **do**

3: *Correct*: compute

$$\hat{\rho}_{t-1}^N = \frac{\hat{Z}_{t-1}^N}{N} \sum_{i=1}^N G_{t-1}(\mathbf{X}_{t-1}^{i,N}) \delta_{X_t^{i,N}} \quad \hat{\mu}_{t-1}^N = \frac{\hat{\rho}_{t-1}^N}{\hat{\rho}_{t-1}^N(\mathbf{E}_{t-1})}$$

4: *Resample*: e.g., for $n \leq N$, draw $\mathbf{X}_{t-}^{n,N} \stackrel{\text{ciid}}{\sim} \hat{\mu}_{t-1}^N$.

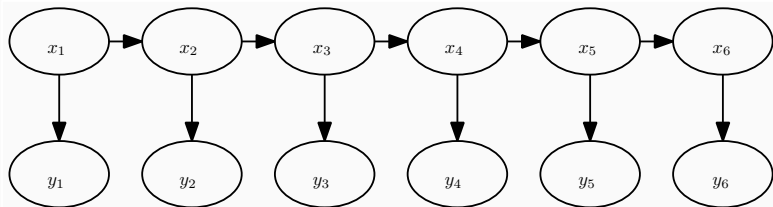
5: *Mutate*: for $n \leq N$, draw $X_t^{n,N} \sim K_t(\mathbf{X}_{t-}^{n,N}, dx_t)$ and set $\mathbf{X}_t^{n,N} := (X_t^{n,N}, \mathbf{X}_{t-}^{n,N})$.

6: *Estimate*:

$$\rho_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}, \quad Z_t^N = \rho_t^N(\mathbf{E}_t) \quad \text{and} \quad \mu_t^N := \rho_t^N / Z_t^N.$$

7: **end for**

SIR Example: Simple Particle Filters



- Unobserved Markov chain $\{X_n\}$ transition f .
- Observed process $\{Y_n\}$ conditional density g .
- The joint density is available:

$$p(x_{1:n}, y_{1:n} | \theta) = f_1^\theta(x_1) g^\theta(y_1 | x_1) \prod_{i=2}^n f^\theta(x_i | x_{i-1}) g^\theta(y_i | x_i).$$

- Natural SIR target distributions:

$$\mu_n^\theta(x_{1:n}) := p(x_{1:n} | y_{1:n}, \theta) \propto p(x_{1:n}, y_{1:n} | \theta) =: \rho_n^\theta(x_{1:n})$$

$$Z_n^\theta = \int p(x_{1:n}, y_{1:n} | \theta) dx_{1:n} = p(y_{1:n} | \theta)$$

- Choosing

$$\hat{\mu}_n^\theta(x_{1:n}) := p(x_{1:n}|y_{1:n}, \theta) \propto p(x_{1:n}, y_{1:n}|\theta) =: \hat{\rho}_n^\theta(x_{1:n})$$
$$\hat{Z}_n^\theta = \int p(x_{1:n}, y_{1:n}|\theta) dx_{1:n} = p(y_{1:n}|\theta)$$

- and $K_p(x_p|x_{1:p-1}) = f^\theta(x_p|x_{p-1})$ yields the bootstrap particle filter of Gordon et al. (1993),
- whereas $K_p(x_p|x_{1:p-1}) = p(x_p|x_{p-1}, y_p, \theta)$ yields the “locally optimal” particle filter.
- Note: Many alternative particle filters are SIR algorithms with other targets. Cf. Johansen and Doucet (2008); Doucet and Johansen (2011).

Sequential Monte Carlo Samplers: Another SIR Class

Given a sequence of targets $\bar{\mu}_1, \dots, \bar{\mu}_n$ on *arbitrary* spaces, Del Moral et al. (2006) extend the space:

$$\hat{\mu}_n(x_{1:n}) = \bar{\mu}_n(x_n) \prod_{p=n-1}^1 L_p(x_{p+1}, x_p)$$

$$\hat{\rho}_n(x_{1:n}) = \bar{\rho}_n(x_n) \prod_{p=n-1}^1 L_p(x_{p+1}, x_p)$$

$$\begin{aligned} \hat{Z}_n &= \int \rho_n(x_{1:n}) dx_{1:n} \\ &= \int \bar{\rho}_n(x_n) \prod_{p=n-1}^1 L_p(x_{p+1}, x_p) dx_{1:n} = \int \bar{\rho}_n(x_n) dx_n = \bar{Z}_n, \end{aligned}$$

and applies SMC to these extended targets.

Some dynamic semigroups

Linear $\rho_n = \rho_p Q_{p:n}$ where

$$Q_{n:n} := \text{Id}$$

$$Q_{p:n} := Q_{p+1} Q_{p+1:n}$$

$$Q_p(x, dy) := G_p(x) M_p(x, dy)$$

Non-linear $\mu_n = \Phi_{p:n}(\mu_p)$ where

$$\Phi_{n:n}(\mu) := \mu$$

$$\Phi_{p:n}(\mu) := \Phi_{p+1:n}(\Phi_{p+1}(\mu))$$

$$\Phi_p(\mu)(dy) := \int \frac{G(x) M(x, dy)}{\mu(G)} \mu(dx)$$

Correspondence $\Phi_{p:n}(\mu_p) = \frac{\mu_p Q_{p:n}}{\mu_p Q_{p:n}(1)}$.

On stability and semigroups

$$\begin{array}{ccccccc}
 \rho_1 & \rightarrow & \rho_2 = \rho_1 Q_2 & \rightarrow & \rho_3 = \rho_1 Q_{1:3} & \rightarrow & \dots \rightarrow \rho_1 Q_{1:n} \\
 \Downarrow & & & & & & \\
 \rho_1^N & \rightarrow & \rho_1^N Q_2 & \rightarrow & \rho_1^N Q_{1:3} & \rightarrow & \dots \rightarrow \rho_1^N Q_{1:n} \\
 & & \Downarrow & & & & \\
 & & \rho_2^N & \rightarrow & \rho_2^N Q_3 & \rightarrow & \dots \rightarrow \rho_2^N Q_{2:n} \\
 & & & & \Downarrow & & \\
 & & & & \rho_3^N & \rightarrow & \dots \rightarrow \rho_3^N Q_{3:n} \\
 & & & & & & \vdots \\
 & & & & & & \Downarrow \\
 & & & & & & \rho_{n-1}^N \rightarrow \rho_{n-1}^N Q_n \\
 & & & & & & \Downarrow \\
 & & & & & & \rho_n^N
 \end{array}$$

We have: $\rho_n^N - \rho_n = \sum_{p=1}^n (\rho_p^N - \rho_{p-1}^N Q_p) Q_{p:n}$

Theorem (Del Moral 2004: Theorem 7.4.2)

Under regularity conditions, for any $n \geq 1$, $\varphi \in \mathcal{C}_b E_n$:

$$\mathbb{E}[\rho^N(\varphi)] = \rho_n(\varphi).$$

Theorem (Del Moral 2004: Theorem 7.4.4)

Under regularity conditions, for any $n \geq 1$, $p \geq 1$, $\varphi_n \in \mathcal{C}_b(E_n)$:

$$\sqrt{N} \mathbb{E} [|\mu_n^N(\varphi_n) - \mu_n(\varphi_n)|^p]^{1/p} \leq c_{p,n} \|\varphi_n\|_\infty$$

By a Borel-Cantelli argument:

$$\lim_{N \rightarrow \infty} \mu_n^N(\varphi_n) \xrightarrow{\text{a.s.}} \mu_n(\varphi_n).$$

Central Limit Theorem

Proposition (Del Moral 2004: Proposition 9.4.2)

Under regularity conditions, for any $n \geq 1$:

$$\sqrt{N}(\mu_n^N(\varphi_n) - \mu_n(\varphi_n)) \xrightarrow{d} \mathcal{N}(0, \sigma_n^2(\varphi_n))$$

where

$$\sigma_n^2(\varphi_n) = \sum_{q=1}^n \mu_q [(\bar{Q}_{q,n}(\varphi_n - \mu_n(\varphi_n)))^2]$$

where

$$\bar{Q}_{q,n}(\varphi_n)(x_q) = Q_{q,n}(\varphi_n)(x_q) / \mu_q Q_{q,n}(\mathbf{1}).$$

SIR: Theoretical Justification — Some Of

In summary, under regularity conditions we have:

unbiasedness

$$\mathbb{E}[\widehat{Z}_n^N] = Z_n$$

slln

$$\lim_{N \rightarrow \infty} \widehat{\pi}_n^N(\varphi) \stackrel{\text{a.s.}}{=} \pi_n(\varphi)$$

clt For a normal random variable W_n^φ of appropriate variance:

$$\lim_{N \rightarrow \infty} \sqrt{N}[\widehat{\pi}_n^N(\varphi) - \pi_n(\varphi)] \stackrel{d}{=} W_n^\varphi$$

Auxiliary sequential importance resampling

Ingredients:

- Unnormalized targets ρ_t on $\mathbf{E}_t = \otimes_{s=0}^t E_s$.
- Normalized counterparts $\mu_t = \rho_t / Z_t$.
- Normalizing constants $Z_t = \rho_t(\mathbf{E}_t)$
- Sequences of auxiliary targets γ_{t-} and $\gamma_t := \gamma_{t-} \otimes K_t$.
- Auxiliary normalizing constants $\mathcal{Z}_t = \gamma_t(\mathbf{E}_t)$
- Normalized auxiliary targets $\pi_t = \gamma_t / \mathcal{Z}_t$.
- Proposal kernels K_t : conditional laws over E_t given \mathbf{E}_{t-1} .
- Importance weights / potential functions:

$$w_t = \frac{d\gamma_{t-}}{d\gamma_{t-1}}.$$

Algorithm: iterative importance sampling and resampling targeting auxiliary targets and an extra importance sampling step.

Auxiliary sequential importance resampling

- 1: *Propose*: for $n \leq N$, draw $\mathbf{X}_0^{n,N}$ independently from K_0 .
- 2: *Compute*: $\gamma_0^N := N^{-1} \sum_{n=1}^N \delta_{\mathbf{X}_0^{n,N}}$.
- 3: **for** $t = 1, \dots, T$ **do**
- 4: *Correct*: compute $\gamma_{t-}^N(d\mathbf{x}_{t-1}) := w_{t-}(\mathbf{x}_{t-1})\gamma_{t-1}^N(d\mathbf{x}_{t-1})$ and $\pi_{t-}^N := \gamma_{t-}^N / \gamma_{t-}^N(\mathbf{E}_{t-1})$.
- 5: *Resample*: for $n \leq N$, draw $\mathbf{X}_{t-}^{n,N}$ independently from π_{t-}^N .
- 6: *Mutate*: for $n \leq N$, draw $X_t^{n,N}$ independently from $K_t(\mathbf{X}_{t-}^{n,N}, d\mathbf{x}_t)$ and set $\mathbf{X}_t^{n,N} := (X_t^{n,N}, \mathbf{X}_{t-}^{n,N})$.
- 7: *Compute*: $\gamma_t^N := \frac{Z_t^N}{N} \sum_{n=1}^N \delta_{\mathbf{X}_t^{n,N}}$ where $Z_t^N := \gamma_{t-}^N(\mathbf{E}_{t-1})$.
- 8: **end for**

Note: At each step t , one obtains estimates of ρ_t , Z_t , and μ_t .

In the filtering setting, take:

- $\gamma_{t-}(d\mathbf{x}_{t-1}) = p(\mathbf{x}_{t-1}, \mathbf{y}_{t-1})\hat{p}(y_t|x_{t-1})d\mathbf{x}_{t-1}$
- $\pi_{t-} = \gamma_{t-}/\gamma_{t-}(\mathbf{E}_{t-1})$.

and one recovers the auxiliary particle filter of Pitt and Shephard (1999).

Divide-and-Conquer SMC
see Lindsten et al. (2017)

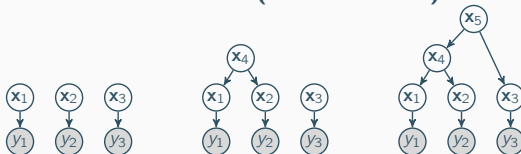
Divide-and-Conquer

Many models admit natural (or unnatural) decompositions:

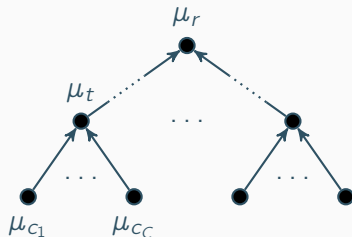
Level 0:

Level 1:

Level 2:

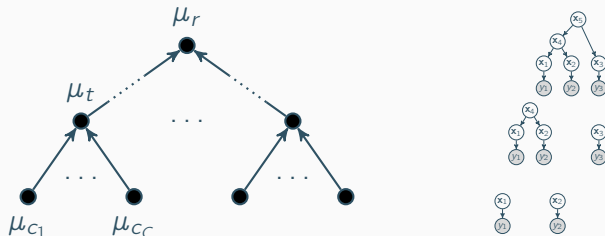


To which we can apply a divide-and-conquer strategy:



A few formalities...

- Use a tree, \mathbb{T} of models (with rootward variable inclusion):



- $t \in \mathbb{T}$ denotes a node; $r \in \mathbb{T}$ is the root.
- $\mathcal{C}_t = \{c_1, \dots, c_C\}$ denotes the children of t .
- E_t is the space of variables included in t but *not* its children.
- $\mathbf{E}_t = E_t \times \otimes_{c \in \mathcal{C}(t)} \mathbf{E}_c$ is the space of all variables included in \mathbb{T}_t : the subtree rooted at t .
- D&C-SMC can be viewed as a recursion over this tree.

The Divide-and-Conquer SMC Algorithm

dac_smc(u) for u in \mathbb{T} .

- 1: **if** u is a leaf (i.e. $u \in \mathbb{T}^\partial$) **then**
- 2: *Propose*: for $n \leq N$, draw $\mathbf{X}_u^{n,N}$ independently from K_u .
- 3: *Return*: $\gamma_u^N := N^{-1} \sum_{n=1}^N \delta_{\mathbf{X}_u^{n,N}}$.
- 4: **else**
- 5: **for** v in \mathcal{C}_u **do**
- 6: *Recurse*: set $\gamma_v^N := \text{dac_smc}(v)$.
- 7: **end for**
- 8: *Obtain*: $\gamma_{\mathcal{C}_u}^N = \prod_{v \in \mathcal{C}_u} \gamma_v^N$.
- 9: *Correct*: compute $\gamma_{u-}^N = w_{u-} \cdot \gamma_{\mathcal{C}_u}^N$ and $\pi_{u-}^N := \gamma_{u-}^N / \gamma_{u-}^N(\mathbf{E}_{\mathcal{C}_u})$.
- 10: *Resample*: for $n \leq N$, draw $\mathbf{X}_{u-}^{n,N} \sim \pi_{u-}^N$ independently.
- 11: *Mutate*: for $n \leq N$, draw $X_u^{n,N} \sim K_u(\mathbf{X}_{u-}^{n,N}, dx_u)$ and set $\mathbf{X}_u^{n,N} := (X_u^{n,N}, \mathbf{X}_{u-}^{n,N})$.
- 12: *Return*: $\gamma_u^N := N^{-1} \mathcal{Z}_u^N \sum_{n=1}^N \delta_{\mathbf{X}_u^{n,N}}$ where $\mathcal{Z}_u^N := \gamma_{u-}^N(\mathbf{E}_{\mathcal{C}_u})$.
- 13: **end if**

Theoretical Properties
see Kuntz et al. (2024)

Assumption (1. Absolute Continuity)

For all u in \mathbb{T} and v in \mathbb{T}^∂ , ρ_u is absolutely continuous w.r.t. γ_u , γ_{v_-} is absolutely continuous w.r.t. γ_{C_v} , and the Radon-Nikodym derivatives $w_u := d\rho_u/d\gamma_u$ and $w_{v_-} := d\gamma_{v_-}/d\gamma_{C_v}$ are positive everywhere.

Assumption (2. Boundedness)

For all u in \mathbb{T}^∂ and v in \mathbb{T} , $w_{u_-} = d\gamma_{u_-}/d\gamma_{C_u}$ and $w_v = d\rho_v/d\gamma_v$ are bounded: $\|w_{u_-}\|_\infty < \infty$ and $\|w_v\|_\infty < \infty$.

Unbiasedness of NC Estimates (Kuntz et al., 2024, Theorem 3)

If Assumptions 1–2 hold, then for all $u \in \mathbb{T}$:

$$\mathbb{E} [\rho_u^N(\varphi)] = \rho_u(\varphi), \quad \mathbb{E} [Z_u^N] = Z_u, \quad \forall N > 0, \quad \varphi \in \mathcal{B}_b(\mathbf{E}_u).$$

Dealing with product approximations: some motivation

- Given $\gamma_1^N \in \mathcal{P}(E_1)$ and $\gamma_2^N \in \mathcal{P}(E_1)$ which are empirical measure approximations of γ_1 and γ_2 . . . we need to control $\gamma_1^N \times \gamma_2^N - \gamma_1 \times \gamma_2$.
- Naïvely, we could try, e.g.

$$\begin{aligned} & \gamma_1^N \times \gamma_2^N - \gamma_1^N \times \gamma_2 + \gamma_1^N \times \gamma_2 - \gamma_1 \times \gamma_2 \\ &= \gamma_1^N \times (\gamma_2^N - \gamma_2) + (\gamma_1^N - \gamma_1) \times \gamma_2 \end{aligned}$$

but this is displeasing asymmetric, so:

$$\begin{aligned} & \gamma_1^N \times (\gamma_2^N - \gamma_2) + \gamma_1^N \times \gamma_2 - \gamma_1 \times \gamma_2 \\ &= (\gamma_1^N - \gamma_1) \times (\gamma_2^N - \gamma_2) + \gamma_1 \times (\gamma_2^N - \gamma_2) + (\gamma_1^N - \gamma_1) \times \gamma_2 \end{aligned}$$

is more appealing and suggests a general approach.

Hoeffding's Canonical Decomposition

Fix any u in $\mathbb{T}^{\mathcal{D}}$ and φ in $\mathcal{B}_b(\mathbf{E}_{\mathcal{C}_u})$. Note that,

$$\gamma_{\mathcal{C}_u}^N - \gamma_{\mathcal{C}_u} = \prod_{v \in \mathcal{C}_u} [\gamma_v^N - \gamma_v + \gamma_v] - \gamma_{\mathcal{C}_u} = \sum_{\emptyset \neq A \subseteq \mathcal{C}_u} \Delta_A^N \times \gamma_{\mathcal{C}_u}^A, \quad (1)$$

where $\Delta_A^N := \prod_{v \in A} (\gamma_v^N - \gamma_v)$ and $\gamma_{\mathcal{C}_u}^A := \gamma_{\mathcal{C}_u \setminus A}$ for all $A \subset \mathcal{C}_u$.

See Kuntz et al. (2022) for connections with U-statistics etc..

Product approximations: L^p error control

Lemma (L^p error control for product approximations)

Suppose that $(\eta_1^N)_{N=1}^\infty$ and $(\eta_2^N)_{N=1}^\infty$ are independent sequences of random probability measures on $(\Omega, \mathcal{F}, \mathbb{P})$, over (S_1, \mathcal{S}_1) and (S_2, \mathcal{S}_2) , respectively, and satisfying

$$\mathbb{E} \left[|\eta_k^N(\varphi) - \eta_k(\varphi)|^p \right]^{\frac{1}{p}} \leq \frac{C \|\varphi\|}{N^{1/2}} \quad \forall N > 0, \quad \varphi \in \mathcal{B}_b(S_k), \quad k \in \{1, 2\},$$

for a $p \geq 1$, with limits η_1 and η_2 that are also probabilities, then:

$$\mathbb{E} \left[|(\eta_1^N \times \eta_2^N)(\varphi) - (\eta_1 \times \eta_2)(\varphi)|^p \right]^{\frac{1}{p}} \leq \frac{C \|\varphi\|}{N^{1/2}} \quad \forall N > 0, \quad \varphi \in \mathcal{B}_b(S_1 \times S_2).$$

Proof Sketch.

Follows from the Hoeffding decomposition, Minkowski's inequality and judicious conditioning. □

Theorem (L^p Error Bounds (Kuntz et al., 2024, Theorem 5))

If Assumptions 1–2 hold, then, for each $p \geq 1$ and u in \mathbb{T} , there exist constants $C_u^\rho, C_u^\mu < \infty$ such that

$$\mathbb{E} \left[|\rho_u^N(\varphi) - \rho(\varphi)|^p \right]^{\frac{1}{p}} \leq \frac{C_u^\rho \|\varphi\|_\infty}{N^{1/2}},$$
$$\mathbb{E} \left[|\mu_u^N(\varphi) - \mu_u(\varphi)|^p \right]^{\frac{1}{p}} \leq \frac{C_u^\mu \|\varphi\|_\infty}{N^{1/2}},$$

for all $N > 0$ and φ in $\mathcal{B}_b(\mathbf{E}_u)$. In particular,

$$\mathbb{E}[|Z_u^N - Z_u|^p]^{1/p} \leq C_u^\rho / N^{1/2}$$

for all $N > 0$.

Laws of Large Numbers

Strong Law of Large Numbers (Kuntz et al., 2024, Theorem 1)

If Assumptions 1–2 are satisfied, u belongs to \mathbb{T} , and φ belongs to $\mathcal{B}_b(\mathbf{E}_u)$, then

$$\lim_{N \rightarrow \infty} \rho_u^N(\varphi) = \rho_u(\varphi), \quad \lim_{N \rightarrow \infty} \mu_u^N(\varphi) = \mu(\varphi), \quad \lim_{N \rightarrow \infty} Z_u^N = Z_u,$$

almost surely.

Strong Law of Large Numbers (Kuntz et al., 2024, Theorem 2)

If, in addition to Assumptions 1–2, the spaces $(E_u)_{u \in \mathbb{T}}$ are Polish and $(\mathcal{E}_u)_{u \in \mathbb{T}}$ are the corresponding Borel sigma algebras, then

$$\rho_u^N \rightharpoonup \rho_u, \quad \mu_u^N \rightharpoonup \mu_u, \quad \text{almost surely,}$$

for each u in \mathbb{T} , where \rightharpoonup denotes weak convergence as $N \rightarrow \infty$.

Towards a CLT: Local Error Control

Lemma (Local errors)

Fix any u in \mathbb{T} and a collection $(\varphi_v)_{v \in \mathbb{T}_u}$ with φ_v belonging to $\mathcal{B}_b(\mathbf{E}_v)$ for each v in \mathbb{T}_u . Let

$$V_v^N(\varphi_v) := \begin{cases} N^{1/2}(\pi_v^N(\varphi_v) - K_v(\varphi_v)) & \text{if } v \text{ is a leaf} \\ N^{1/2}(\pi_v^N(\varphi_v) - \pi_{v_-}^N(K_v\varphi_v)) & \text{otherwise} \end{cases} \quad \forall N > 0, v \in \mathbb{T}_u. \quad (2)$$

Under Assumptions 1–2, the collection $(V_v^N(\varphi_v))_{v \in \mathbb{T}_u}$ converges weakly as $N \rightarrow \infty$ to a collection $(V_v(\varphi_v))_{v \in \mathbb{T}_u}$ of independent zero mean random variables with Gaussian laws:

$$V_v(\varphi_v) \sim \mathcal{N}(0, \pi_v([\varphi_v - \pi_v(\varphi_v)]^2)) \quad \forall v \in \mathbb{T}_u. \quad (3)$$

Towards a CLT: A semigroup for the tree

In our context we have a 2 parameter semigroup indexed by the vertices of a tree; $\Gamma_{v,u}$ is defined when v is a descendent of u .

$$\Gamma_{u,u}(\mathbf{x}_u, d\mathbf{y}_u) := \delta_{\mathbf{x}_u}(d\mathbf{y}_u),$$

$$\Gamma_{v,u}(\mathbf{x}_v, d\mathbf{y}_u) := \delta_{\mathbf{x}_v}(d\mathbf{y}_v) \gamma_{\mathcal{C}_u^v}^v(d\mathbf{y}_{\mathcal{C}_u^v}) w_{u-}(\mathbf{y}_{\mathcal{C}_u}) K_u(\mathbf{y}_{\mathcal{C}_u}, d\mathbf{y}_u) \quad \forall v \in \mathcal{C}_u,$$

where $\gamma_{\mathcal{C}_u^v}^v := \prod_{r \in \mathcal{C}_u^v} \gamma_r$ with $\mathcal{C}_u^v := \mathcal{C}_u \setminus \{v\}$. For all other descendants $v \neq u$, we set

$$\Gamma_{v,u} = \Gamma_{v,r_1} \Gamma_{r_1,r_2} \cdots \Gamma_{r_l,u},$$

where v, r_1, \dots, r_l, u is the branch of \mathbb{T}_u connecting v and u .

Central Limit Theorem

Central Limit theorem (Kuntz et al., 2024, Theorem 6)

If Assumptions 1–2 hold, then, as $N \rightarrow \infty$,

$$N^{1/2} (\rho_u^N(\varphi) - \rho_u(\varphi)) \Rightarrow \mathcal{N}(0, \sigma_{\rho_u}^2(\varphi)),$$
$$N^{1/2} (\mu_u^N(\varphi) - \mu_u(\varphi)) \Rightarrow \mathcal{N}(0, \sigma_{\mu_u}^2(\varphi)),$$

for any given u in \mathbb{T} and φ in $\mathcal{B}_b(\mathbf{E}_u)$, where \Rightarrow denotes convergence in distribution,

$$\sigma_{\rho_u}^2(\varphi) := \sum_{v \in \mathbb{T}_u} \pi_v([\mathcal{Z}_v \Gamma_{v,u}[w_u \varphi] - \rho_u(\varphi)]^2),$$
$$\sigma_{\mu_u}^2(\varphi) := \sum_{v \in \mathbb{T}_u} \pi_v([\mathcal{Z}_v \Gamma_{v,u}[w_u Z_u^{-1}[\varphi - \mu_u(\varphi)]]]^2).$$

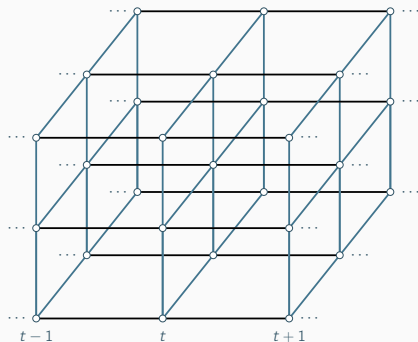
Some (Importance) Extensions

1. (Lightweight) Mixture Resampling [with Rejection Sampling]
2. Tempering (Del Moral et al., 2006).
3. Adaptation e.g., Zhou et al. (2016).

Skip to Conclusions

**Illustrative Application:
High-dimensional Filtering
See Crucinio and Johansen (2024)**

DaC-SMC for High-Dimensional Filtering i



Rough idea:

- Decompose space at each time.
- Implement marginal analogue of SMC (over time) — see Kück et al. (2006) and Crucinio and Johansen (2023).

Use, at node u at time t :

$$“\gamma_{t,u}(z_{t,u}) = g_{t,u}(z_{t,u}; (y_t(i)) : i \in \mathcal{V}_u) \frac{1}{N} \sum_{n=1}^N f_{t,u}(z_{t-1,\mathfrak{R}}^n, z_{t,u})”$$

where $f_{t,u}$ and $g_{t,u}$ approximate appropriate marginal quantities.

DaC-SMC for High-Dimensional Filtering ii

Leaf nodes: IS from with proposal $K_{t,u}$, weights are:

$$w_{t,u}(z_{t,u}, x_{1:t-1,u}) = \frac{g_{t,u}(z_{t,u}, (y_t(i))_{i \in \mathcal{V}_u}) \sum_{n=1}^N f_{t,u}(z_{t-1,\mathfrak{R}}^n, z_{t,u})}{\sum_{n=1}^N K_{t,u}(z_{t-1,\mathfrak{R}}^n, z_{t,u})}.$$

choosing $K_{t,u} \equiv f_{t,u}$ somewhat simplifies computation.

Intermediate nodes: IS using product of child nodes:

$$m_{t,u}(z_{t,c_u}) = \frac{g_{t,u}(z_{t,c_u}, (y_t(i))_{i \in \mathcal{V}_u})}{g_{t,\ell(u)}(z_{t,\ell(u)}, (y_t(i))_{i \in \mathcal{V}_{\ell(u)}}) g_{t,r(u)}(z_{t,r(u)}, (y_t(i))_{i \in \mathcal{V}_{r(u)}})} \times \frac{N^{-1} \sum_{n=1}^N f_{t,u}(z_{t-1,\mathfrak{R}}^n, z_{t,c_u})}{N^{-1} \sum_{n=1}^N f_{t,\ell(u)}(z_{t-1,\mathfrak{R}}^n, z_{t,\ell(u)}) N^{-1} \sum_{n=1}^N f_{t,r(u)}(z_{t-1,\mathfrak{R}}^n, z_{t,r(u)})},$$

with $O(N)$ computation cost. . . for each of N^2 particle pairs (strategy in paper has $O(N^{5/2})$ cost overall).

Lattice $V = \{1, \dots, d\}^2$. We take $d = 8$ and $d = 16$.

Dynamics $X_t(v) = X_{t-1}(v) + U_t(v)$, where $U_t(v) \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_x^2)$.

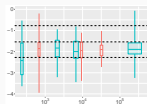
Observation $Y_t = X_t + V_t$; V_t to be multivariate t -distributed with $\nu = 10$ d.o.f., and precision structure $\Sigma_{v_j}^{-1} = \tau^{D(j,v)}$ if $D(j, v) \leq r_y$ and 0 otherwise. D denotes graph distance.

Data Simulated with $\sigma_x^2 = 1$, $\tau = -0.25$, $r_y = 1$ and $t = 10$.

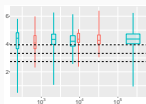
D&C Filtering: Spatial Example ii

$d = 8 \times 8$

(1, 1)

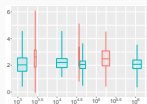


(8, 6)

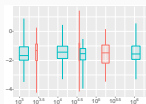


$d = 16 \times 16$

(1, 1)



(8, 8)



Runtime / s

Runtime / s

■ dac ■ dac-ada

Figure 1: Filtering mean estimates for two nodes for a 8×8 and a 16×16 lattice at time $t = 10$. 50 repetitions for $N = 100, 500, 1000$ and 5000 . The reference lines for the 8×8 grid show the average value of the filtering mean estimate and the interquartile range obtained with 50 repetitions of a bootstrap PF with $N = 10^5$ particles.

**Illustrative Application:
Hierarchical Monte Carlo Fusion
See Chan et al. (2023)**

Objective: combine approximations of “subposteriors”:

$$f(\mathbf{x}) \propto \prod_{c \in \mathcal{C}} f_c(\mathbf{x}), \quad (4)$$

Proposition (Dai et al. (2019))

If p_c is f_c^2 -invariant on \mathbb{R}^d then the density proportional to

$$g_{\mathcal{C}}(\bar{\mathbf{x}}^{(\mathcal{C})}, \mathbf{y}^{(\mathcal{C})}) := \prod_{c \in \mathcal{C}} \left[f_c^2(\mathbf{x}^{(c)}) \cdot p_c(\mathbf{y}^{(c)} | \mathbf{x}^{(c)}) \cdot \frac{1}{f_c(\mathbf{y}^{(c)})} \right], \quad (5)$$

admits marginal density $f^{(c)} \propto \prod_{c \in \mathcal{C}} f_c$ over $\mathbf{y}^{(c)} \in \mathbb{R}^d$.

This can be exploited by taking a proposal distribution proportional to:

$$h_{\mathcal{C}}(\tilde{\mathbf{x}}^{(\mathcal{C})}, \mathbf{y}^{(\mathcal{C})}) := \prod_{c \in \mathcal{C}} f_c(\mathbf{x}^{(c)}) \cdot \exp \left\{ -\frac{(\mathbf{y}^{(\mathcal{C})} - \tilde{\mathbf{x}}^{(\mathcal{C})})^{\top} \boldsymbol{\Lambda}_{\mathcal{C}}^{-1} (\mathbf{y}^{(\mathcal{C})} - \tilde{\mathbf{x}}^{(\mathcal{C})})}{2T} \right\}$$

where

$$\tilde{\mathbf{x}}^{(\mathcal{C})} := \left(\sum_{c \in \mathcal{C}} \boldsymbol{\Lambda}_c^{-1} \right)^{-1} \left(\sum_{c \in \mathcal{C}} \boldsymbol{\Lambda}_c^{-1} \mathbf{x}^{(c)} \right), \quad \boldsymbol{\Lambda}_{\mathcal{C}}^{-1} := \sum_{c \in \mathcal{C}} \boldsymbol{\Lambda}_c^{-1}.$$

Proposition

If $p_c(\mathbf{y}^{(c)}|\mathbf{x}^{(c)})$ is the transition density of a suitable Langevin diffusion

$$\frac{g_c(\bar{\mathbf{x}}^{(c)}, \mathbf{y}^{(c)})}{h_c(\bar{\mathbf{x}}^{(c)}, \mathbf{y}^{(c)})} \propto \rho_0(\bar{\mathbf{x}}^{(c)}) \cdot \rho_1(\bar{\mathbf{x}}^{(c)}, \mathbf{y}^{(c)}),$$

$$\rho_0(\bar{\mathbf{x}}^{(c)}) := \exp \left\{ - \sum_{c \in \mathcal{C}} \frac{(\tilde{\mathbf{x}}^{(c)} - \mathbf{x}^{(c)})^\top \boldsymbol{\Lambda}_c^{-1} (\tilde{\mathbf{x}}^{(c)} - \mathbf{x}^{(c)})}{2T} \right\},$$

$$\rho_1(\bar{\mathbf{x}}^{(c)}, \mathbf{y}^{(c)}) := \prod_{c \in \mathcal{C}} \mathbb{E}_{\mathbb{W}_{\Lambda_c}} \left[\exp \left\{ - \int_0^T \phi_c(\mathbf{X}_t^{(c)}) dt \right\} \right],$$

$$\phi_c(\mathbf{x}) := \frac{1}{2} \left(\nabla \log f_c(\mathbf{x})^\top \boldsymbol{\Lambda}_c \nabla \log f_c(\mathbf{x}) + \text{Tr}(\boldsymbol{\Lambda}_c \nabla^2 \log f_c(\mathbf{x})) \right),$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, and \mathbb{W}_{Λ_c} denotes the law of a Brownian bridge $\{\mathbf{X}_t^{(c)}, t \in [0, T]\}$ with $\mathbf{X}_0^{(c)} := \mathbf{x}^{(c)}$, $\mathbf{X}_T^{(c)} := \mathbf{y}^{(c)}$ and covariance matrix $\boldsymbol{\Lambda}_c$.

Merging Subposteriors

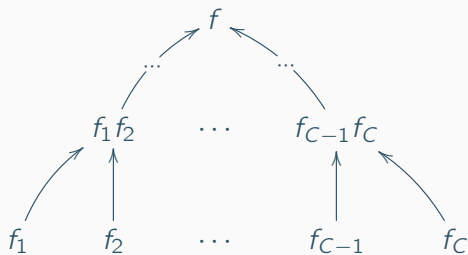
general.fusion($\mathcal{C}, \{\{\mathbf{x}_{0,i}^{(c)}, w_i^{(c)}\}_{i=1}^M, \mathbf{\Lambda}_c\}_{c \in \mathcal{C}}, N, T$)

Input: Samples $\{\mathbf{x}_{0,i}^{(c)}, w_i^{(c)}\}_{i=1}^M$ for $c \in \mathcal{C}$, matrices, $\{\mathbf{\Lambda}_c : c \in \mathcal{C}\}$, particle count, N , and time horizon, $T > 0$.

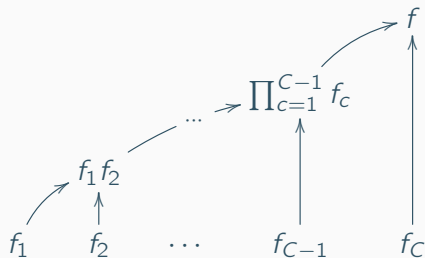
1. **Partial proposal:** Compose samples $\{\bar{\mathbf{x}}_{0,j}^{(c)}, \bar{w}_j\}_{j=1}^M$ where $\bar{w}_j := (\prod_{c \in \mathcal{C}} w_j^{(c)}) \cdot \rho_0(\bar{\mathbf{x}}_{0,j}^{(c)})$ for $j \in \{1, \dots, M\}$.
2. For i in 1 to N ,
 - 2.1 $\bar{\mathbf{x}}_{0,i}^{(c)}$: Sample $l \sim \text{categorical}(\bar{w}_{1:M})$ and set $\bar{\mathbf{x}}_{0,i}^{(c)} := \bar{\mathbf{x}}_{0,l}^{(c)}$.
 - 2.2 **Complete proposal:** Simulate $\mathbf{y}_i^{(c)} \sim \mathcal{N}_d(\bar{\mathbf{x}}_{0,i}^{(c)}, T\mathbf{\Lambda}_c)$.
 - 2.3 $\tilde{\rho}_{1,i}^{(c)}$: Compute importance weight $\tilde{\rho}_{1,i}^{(c)} := \tilde{\rho}_1^{(b)}(\bar{\mathbf{x}}_{0,i}^{(c)}, \mathbf{y}_i^{(c)})$.
3. For i in 1 to N compute $w_i^{(c)} = \tilde{\rho}_{1,i}^{(c)} / \sum_{k=1}^N \tilde{\rho}_{1,k}^{(c)}$.

Output: $\{\bar{\mathbf{x}}_{0,i}^{(c)}, \mathbf{y}_i^{(c)}, w_i^{(c)}\}_{i=1}^N$.

Some Decompositions Leading to an Algorithm



A balanced-binary tree.



A progressive tree.

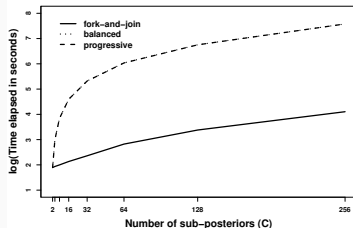
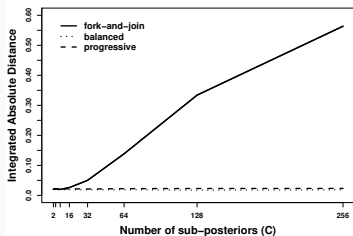
`d&c.fusion(v, N, T)`

Given: Sub-posteriors, $\{f_u\}_{u \in \text{Leaf}(\mathbb{T})}$, and preconditioning matrices $\{\mathbf{\Lambda}_u\}_{u \in \mathbb{T}}$.

Input: Node in tree, v , the number of particles N , and time horizon $T > 0$.

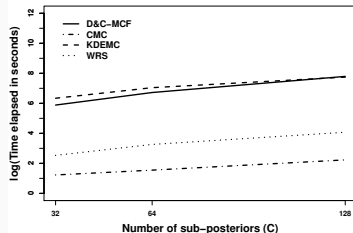
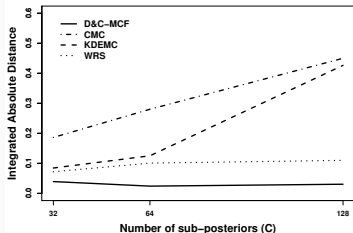
1. For $u \in \text{Ch}(v)$,
 - 1.1 $\{\mathbf{x}_i^{(u)}, \mathbf{y}_i^{(u)}, w_i^{(u)}\}_{i=1}^N \leftarrow \text{d\&c.fusion}(u, N, T)$.
2. If $v \in \text{Leaf}(\mathbb{T})$,
 - 2.1 For $i = 1, \dots, N$, sample $\mathbf{y}_i^{(v)} \sim f_v(\mathbf{y})$.
 - 2.2 **Output:** $\{\emptyset, \mathbf{y}_i^{(v)}, \frac{1}{N}\}_{i=1}^N$.
3. If $v \notin \text{Leaf}(\mathbb{T})$,
 - 3.1 **Output:** Call
`general.fusion($\text{Ch}(v), \{\{\mathbf{y}_i^{(u)}, w_i^{(u)}\}_{i=1}^N, \mathbf{\Lambda}_u\}_{u \in \text{Ch}(v)}, N, T)$` .

An Illustration of the Impact of the D&C Approach



Illustrative comparison of the effect of using different hierarchies, with $f \propto \prod_{c=1}^C f_c$, where $f_c \sim \mathcal{N}(0, C)$ for $c = 1, \dots, C$ (averaged over 50 runs).

Some Results for a Logistic Regression Example



[CMC=Consensus Monte Carlo; KDEMC=kernel density averaging approach of Neiswanger et al. (2014); WRS=Weierstrass Rejection Sampler]

* The '*Default of credit card clients*' data set available from <https://archive.ics.uci.edu/ml/datasets>. The data set comprised $m = 30000$ records of **response**: whether a default had occurred and binary covariates **Gender** and **Education**.

Conclusions

- $SMC \approx SIR$
- $D\&C-SMC \approx SIR + \text{Coalescence}$
- Distributed implementation is often straightforward
- D&C strategy can improve even serial performance
- D&C-SMC inherits many theoretical guarantees from SMC
- Some questions remain unanswered, e.g.:
 - How can we construct (near) optimal tree-decompositions?
- Some recent applications include:
 - Parallel (in time) Smoothing (Ding and Gandy, 2018; Corenflos et al., 2022)
 - High-dimensional Filtering (Crucinio and Johansen, 2024)
 - Fusing Bayesian analyses (Chan et al., 2023)

References

- R. Chan, M. Pollock, A. M. Johansen, and G. O. Roberts. Divide-and-conquer Monte Carlo fusion. *Journal of Machine Learning Research*, 24(193):1–82, June 2023. URL <http://jmlr.org/papers/v24/21-1274.html>.
- A. Corenflos, N. Chopin, and S. Särkkä. De-Sequentialized Monte Carlo: a parallel-in-time particle smoother. *Journal of Machine Learning Research*, 23(283):1–39, 2022. URL <http://jmlr.org/papers/v23/22-0140.html>.
- F. R. Crucinio and A. M. Johansen. Properties of marginal sequential Monte Carlo methods. *Statistics and Probability Letters*, 203(109914):1–8, December 2023. doi: 10.1016/j.spl.2023.109914.
- F. R. Crucinio and A. M. Johansen. A divide-and-conquer sequential Monte Carlo approach to high dimensional filtering. *Statistica Sinica*, 34:1093–1113, April 2024. doi: 10.5705/ss.202022.0243.
- H. Dai, M. Pollock, and G. O. Roberts. Monte Carlo Fusion. *Journal of Applied Probability*, 56(1):174–191, 2019.
- P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo methods for Bayesian Computation. In *Bayesian Statistics 8*. Oxford University Press, 2006.
- D. Ding and A. Gandy. Tree-based particle smoothing algorithms in a hidden Markov model. eprint 1808.08400, ArXiv Mathematics e-prints, 2018.
- A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. In D. Crisan and B. Rozovsky, editors, *The Oxford Handbook of Nonlinear Filtering*, pages 656–704. Oxford University Press, 2011.
- N. J. Gordon, S. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, April 1993. doi: 10.1049/ip-f-2.1993.0015.
- A. M. Johansen and A. Doucet. A note on the auxiliary particle filter. *Statistics and Probability Letters*, 78(12):1498–1504, September 2008. doi: 10.1016/j.spl.2008.01.032.
- H. Kück, N. de Freitas, and A. Doucet. SMC samplers for Bayesian optimal nonlinear design. In *Proceedings of the Nonlinear Statistical Signal Processing Workshop*, Cambridge, 2006.
- J. Kuntz, F. R. Crucinio, and A. M. Johansen. Product-form estimators: exploiting independence to scale up Monte Carlo. *Statistics and Computing*, 32(12):1–22, 2022. doi: 10.1007/s11222-021-10069-9.
- J. Kuntz, F. R. Crucinio, and A. M. Johansen. Divide-and-conquer sequential Monte Carlo: Properties and limit theorems. *Annals of Applied Probability*, 34(1B):1469–1523, 2024. doi: 10.1214/23-AAP1996.
- F. Lindsten, A. M. Johansen, C. A. Naesseth, B. Kirkpatrick, T. Schön, J. A. D. Aston, and A. Bouchard-Côté. Divide and conquer with sequential Monte Carlo samplers. *Journal of Computational and Graphical Statistics*, 26(2):445–458, 2017. doi: 10.1080/10618600.2016.1237363.
- M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.
- Y. Zhou, A. M. Johansen, and J. A. D. Aston. Towards automatic model comparison: An adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726, 2016. doi: 10.1080/10618600.2015.1060885.