Divide-and-Conquer Sequential Monte Carlo

Some Properties and Applications

Adam M. Johansen

Collaborators include:
John Aston, Alexandre Bouchard-Côté, Ryan Chan, Francesca Crucinio, Brent Kirkpatrick, Juan Kuntz, Fredrik Lindsten, Christian Næsseth, Murray Pollock, Gareth Roberts and Thomas Schön

Uppsala Universitet Division of Systems and Control—June 13th, 2024

University of Warwick
a.m.johansen@warwick.ac.uk
http://go.warwick.ac.uk/amjohansen/talks/

Outline

- Sequential Monte Carlo (SMC)
- Divide-and-Conquer SMC

  (D&C-SMC; Lindsten et al. (2017))
- Some Theoretical Properties of D&C-SMC

  (Kuntz et al., 2024)
- Some Illustrative Applications:
  - High-dimensional Filtering

    (Crucinio and Johansen, 2024)
  - Hierarchical Fusion

    (Chan et al., 2023)
- Conclusions

Sequential Monte Carlo

The Essential Problem and SMC Solution

SMC Ingredients:

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

$$w_t = \frac{d\rho_t}{d\rho_{t-1} \otimes K_t}.$$

Goals

- Estimate $Z_1, \ldots, Z_t, \ldots$
- Approximate $\mu_1, \ldots, \mu_t, \ldots$

Algorithm

- Iterative importance sampling (IS) and resampling.

The Sequential Importance Resampling Algorithm

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

2: *Correct:* compute
$$\rho_0^N := \frac{1}{N} \sum_{n=1}^{N} w_0(\mathbf{X}_0^{n,N}) \delta_{\mathbf{X}_0^{n,N}},$$
where $w_0 := d\rho_0/dK_0$, $Z_0^N = \rho_0^N(\mathbf{E}_0)$ and $\mu_0^N := \rho_0^N / Z_0^N$.

3: **for** $t = 1, \ldots, T$ **do**

4: *Resample:* for $n \leq N$, draw $\mathbf{X}_{t-}^{n,N} \sim \mu_{t-1}^N$ independently[1].

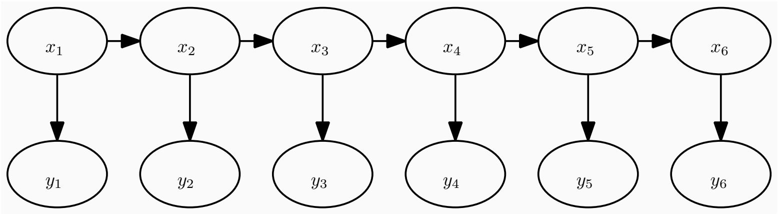
5: *Mutate:* for $n \leq N$, draw $X_t^{n,N} \sim K_t(\mathbf{X}_{t-}^{n,N}, dx_t)$ and set $\mathbf{X}_t^{n,N} := (X_t^{n,N}, \mathbf{X}_{t-}^{n,N})$.

6: *Correct:* compute
$$\rho_t^N = \frac{Z_{t-1}^N}{N} \sum_{i=1}^{N} w_t(\mathbf{X}_t^{n,N}) \delta_{\mathbf{X}_t^{n,N}},$$
$Z_t^N = \rho_t^N(\mathbf{E}_t)$ and $\mu_t^N := \rho_t^N / Z_t^N$.

7: **end for**

[1] Or something better...

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:

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

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

Bootstrap PFs and Similar

- Choosing

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

- 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 $\bar{\mu}_1, \ldots, \bar{\mu}_n$ on *arbitrary* spaces, Del Moral et al. (2006) extend the space:

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

SIR: Theoretical Justification — Some Of

Under regularity conditions we have:

unbiasedness

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

slln

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

clt For a normal random variable W_n of appropriate variance:

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

although establishing this requires a little work (cf., e.g. Del Moral (2004).

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

Propose: for $n \le N$, draw $\mathbf{X}_0^{n,N}$ independently from K_0.

Compute: $\gamma_0^N := N^{-1} \sum_{n=1}^N \delta_{\mathbf{X}_0^{n,N}}$.

for $t = 1, \ldots, T$ **do**

    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 $\pi_{t_-}^N := \gamma_{t_-}^N / \gamma_{t_-}^N(\mathbf{E}_{t-1})$.

    Resample: for $n \le N$, draw $\mathbf{X}_{t_-}^{n,N}$ independently from $\pi_{t_-}^N$.

    Mutate: for $n \le N$, draw $X_t^{n,N}$ independently from $K_t(\mathbf{X}_{t_-}^{n,N}, dx_t)$ and set $\mathbf{X}_t^{n,N} := (X_t^{n,N}, \mathbf{X}_{t_-}^{n,N})$.

    Compute: $\gamma_t^N := \frac{\mathcal{Z}_t^N}{N} \sum_{n=1}^N \delta_{\mathbf{X}_t^{n,N}}$ where $\mathcal{Z}_t^N := \gamma_{t_-}^N(\mathbf{E}_{t-1})$.

end for

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

Auxiliary Particle Filters

In the filtering setting, take:

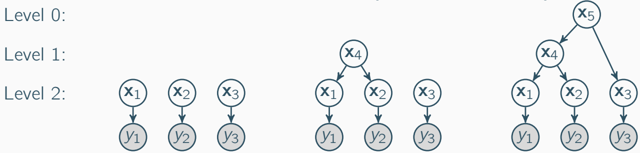
- $\gamma_{t_-}(d\mathbf{x}_{t-1}) = p(\mathbf{x}_{t-1}, \mathbf{y}_{t-1})\hat{p}(y_t|x_{t-1})d\mathbf{x}_{t-1}$
- $\pi_{t_-} = \gamma_{t_-}/\gamma_{t_-}(\mathbf{E}_{t-1})$.

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

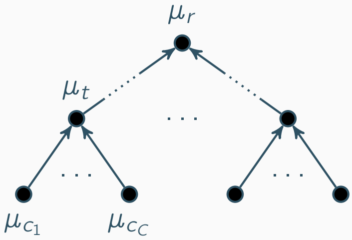
Divide-and-Conquer SMC
see Lindsten et al. (2017)

Divide-and-Conquer

Many models admit natural (or unnatural) decompositions:

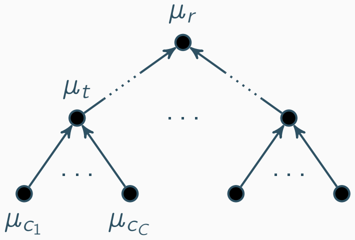


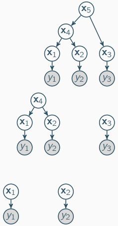
To which we can apply a divide-and-conquer strategy:



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.
- $\mathcal{C}_t = \{c_1, \ldots, c_C\}$ denotes the children of t.
- E_t is the space of variables included in t but *not* its children.
- $\mathbf{E}_t = E_t \times \otimes_{c \in C(t)} \mathbf{E}_c$ 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** $\mathbb{T}$.

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

Theoretical Properties
see Kuntz et al. (2024)

Theoretical Properties: Regularity Assumptions

Assumption (1. Absolute Continuity)

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

Assumption (2. Boundedness)

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

Theoretical Properties i

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

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

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 $N > 0$.

Theoretical Properties ii

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

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

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 $N \to \infty$.

Theoretical Properties iii

Central Limit theorem (Kuntz et al., 2024, Theorem 6)
If Assumptions 1–2 hold, then, as $N \to \infty$,

$$N^{1/2}\left(\rho_u^N(\varphi) - \rho_u(\varphi)\right) \Rightarrow \mathcal{N}(0, \sigma^2_{\rho_u}(\varphi)),$$
$$N^{1/2}\left(\mu_u^N(\varphi) - \mu_u(\varphi)\right) \Rightarrow \mathcal{N}(0, \sigma^2_{\mu_u}(\varphi)),$$

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

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

Theoretical Properties iv

More on the CLT

In particular, $N^{1/2}\left(Z_u^N - Z_u\right) \Rightarrow \mathcal{N}(0, \sigma^2_{Z_u})$ as $N \to \infty$ with

$$\sigma^2_{Z_u} := Z_u^2 \sum_{v \in \mathbb{T}_u} \pi_v \left(\left[\frac{d\mu_u^v}{d\pi_v} - 1 \right]^2 \right), \tag{1}$$

where μ_u^v denotes the $\mathbf{E}_v$-marginal of μ_u (i.e. $\mu_u^v(A) := \mu_u(A \times E_{\mathbb{T}_u \setminus \mathbb{T}_v})$ for all A in $\mathcal{E}_v$).

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

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

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

One Key Ingredient: Multinomial Expansion
Fix any u in $\mathbb{T}^{\partial}$ and φ in $\mathcal{B}_b(\mathbf{E}_{\mathcal{C}_u})$. Note that,

$$\gamma^N_{\mathcal{C}_u} - \gamma_{\mathcal{C}_u} = \prod_{v \in \mathcal{C}_u} [\gamma^N_v - \gamma_v + \gamma_v] - \gamma_{\mathcal{C}_u} = \sum_{\emptyset \neq A \subseteq \mathcal{C}_u} \Delta^N_A \times \gamma^{\not A}_{\mathcal{C}_u}, \quad (2)$$

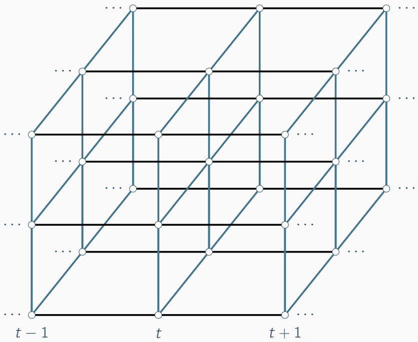
where $\Delta^N_A := \prod_{v \in A}(\gamma^N_v - \gamma_v)$ and $\gamma^{\not A}_{\mathcal{C}_u} := \gamma_{\mathcal{C}_u \setminus A}$ for all $A \subset \mathcal{C}_u$.

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).

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:

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

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:

$$m_{t,u}(z_{t,\mathcal{C}_u}) = \frac{g_{t,u}(z_{t,\mathcal{C}_u}, (y_t(i))_{i \in \mathcal{V}_u})}{g_{t,\ell(u)}(z_{t,\ell(u)}, (y_t(i))_{i \in \mathcal{V}_{\ell(u)}}) g_{t,r(u)}(z_{t,r(u)}, (y_t(i))_{i \in \mathcal{V}_{r(u)}})} \times$$

$$\frac{N^{-1} \sum_{n=1}^N f_{t,u}(z_{t-1,\mathfrak{R}}^n, z_{t,\mathcal{C}_u})}{N^{-1} \sum_{n=1}^N f_{t,\ell(u)}(z_{t-1,\mathfrak{R}}^n, z_{t,\ell(u)}) N^{-1} \sum_{n=1}^N f_{t,r(u)}(z_{t-1,\mathfrak{R}}^n, z_{t,r(u)})},$$

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

D&C Filtering: Spatial Example i

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 $U_t(v) \overset{\text{i.i.d}}{\sim} N(0, \sigma_x^2)$.

Observation $Y_t = X_t + V_t$; V_t to be multivariate t-distributed with $\nu = 10$ d.o.f., and precision structure $\Sigma_{vj}^{-1} = \tau^{D(j,v)}$ 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

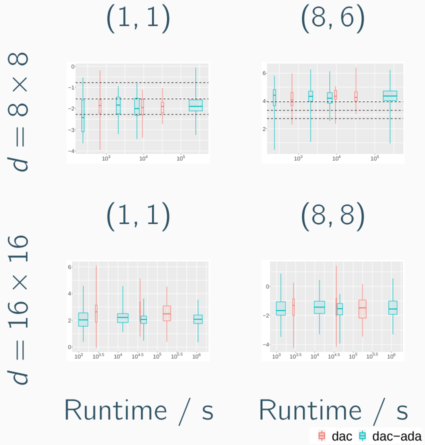


Figure 1: Filtering mean estimates for two nodes for a 8×8 and a 16×16 lattice at time $t = 10$. 50 repetitions for $N = 100, 500, 1000$ and 5000. The reference lines for the 8×8 grid show the average value of the filtering mean estimate and the interquartile range obtained with 50 repetitions of a bootstrap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

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

Monte Carlo Fusion i

Objective: combine approximations of "subposteriors":

$$f(\mathbf{x}) \propto \prod_{c \in \mathcal{C}} f_c(\mathbf{x}), \tag{3}$$

Proposition (Dai et al. (2019))

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

$$g_{\mathcal{C}}\left(\vec{\mathbf{x}}^{(\mathcal{C})}, \mathbf{y}^{(\mathcal{C})}\right) := \prod_{c \in \mathcal{C}} \left[f_c^2\left(\mathbf{x}^{(c)}\right) \cdot p_c\left(\mathbf{y}^{(\mathcal{C})} \middle| \mathbf{x}^{(c)}\right) \cdot \frac{1}{f_c\left(\mathbf{y}^{(\mathcal{C})}\right)} \right], \tag{4}$$

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

Monte Carlo Fusion ii

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

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

where

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

Monte Carlo Fusion iii

Proposition

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

$$\frac{g_{\mathcal{C}}(\vec{\mathbf{x}}^{(\mathcal{C})},\mathbf{y}^{(\mathcal{C})})}{h_{\mathcal{C}}(\vec{\mathbf{x}}^{(\mathcal{C})},\mathbf{y}^{(\mathcal{C})})} \propto \rho_0(\vec{\mathbf{x}}^{(\mathcal{C})})\cdot\rho_1(\vec{\mathbf{x}}^{(\mathcal{C})},\mathbf{y}^{(\mathcal{C})}),$$

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

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

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

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

Merging Subposteriors

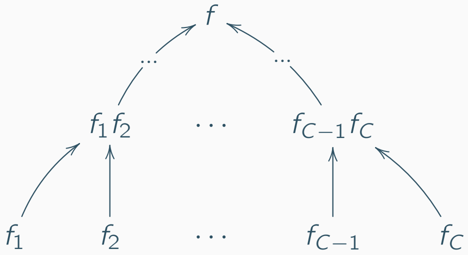
`general.fusion`$(\mathcal{C}, \{\{\mathbf{x}_{0,i}^{(c)}, w_i^{(c)}\}_{i=1}^M, \mathbf{\Lambda}_c\}_{c\in\mathcal{C}}, N, T)$

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, $T > 0$.

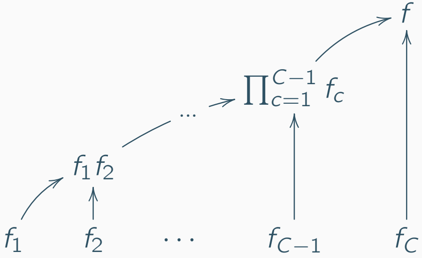
1. **Partial proposal:** Compose samples $\{\vec{\mathbf{x}}_{0,j}^{(\mathcal{C})}, \vec{w}_j\}_{j=1}^M$ where $\vec{w}_j := \left(\prod_{c\in\mathcal{C}} w_j^{(c)}\right) \cdot \rho_0(\vec{\mathbf{x}}_{0,j}^{(\mathcal{C})})$ for $j \in \{1, \dots, M\}$.
2. For i in 1 to N,
   - 2.1 $\vec{\mathbf{x}}_{0,i}^{(\mathcal{C})}$**:**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

Output: $\left\{\vec{\mathbf{x}}_{0,i}^{(\mathcal{C})}, \mathbf{y}_i^{(\mathcal{C})}, w_i^{(\mathcal{C})}\right\}_{i=1}^N$.

Some Decompositions Leading to an Algorithm



A balanced-binary tree.



A progressive tree.

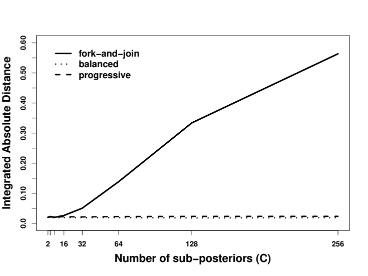
`d&c.fusion`(v, N, T)

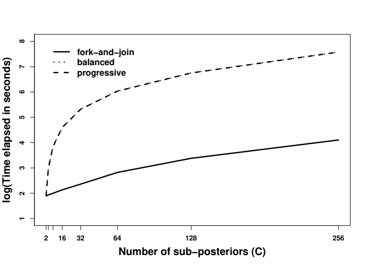
Given: Sub-posteriors, $\{f_u\}_{u \in \mathrm{Leaf}(\mathbb{T})}$, and preconditioning matrices $\{\mathbf{\Lambda}_u\}_{u \in \mathbb{T}}$.

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

1. For $u \in \mathrm{Ch}(v)$,
    1.1 $\left\{\mathbf{x}_i^{(u)}, \mathbf{y}_i^{(u)}, w_i^{(u)}\right\}_{i=1}^{N} \leftarrow$ `d&c.fusion`(u, N, T).
2. If $v \in \mathrm{Leaf}(\mathbb{T})$,
    2.1 For $i = 1, \ldots, N$, 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 $v \notin \mathrm{Leaf}(\mathbb{T})$,
    3.1 **Output:** Call `general.fusion`$(\mathrm{Ch}(v), \{\{\mathbf{y}_i^{(u)}, w_i^{(u)}\}_{i=1}^{N}, \mathbf{\Lambda}_u\}_{u \in \mathrm{Ch}(v)}, N, T)$.

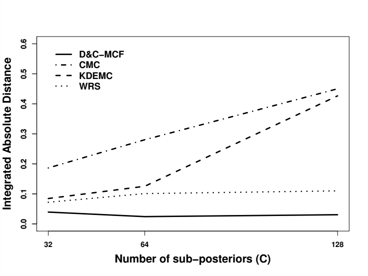
An Illustration of the Impact of the D&C Approach

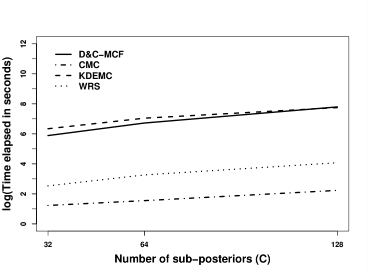




Illustrative comparison of the effect of using different hierarchies, with $f \propto \prod_{c=1}^{C} f_c$, where $f_c \sim \mathcal{N}(0, C)$ for $c = 1, \dots, C$ (averaged over 50 runs).

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 + 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 other recent applications include:
  - Parallel (in time) Smoothing (Ding and Gandy, 2018; Corenflos et al., 2022)

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. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Probability and Its Applications. Springer Verlag, New York, 2004.

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: Fiteen 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. 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.