Conditional particle filters with bridge backward sampling

Matti Vihola (Jyväskylä)

MCM @ Paris
28 June 2023

FINNISH CENTRE OF EXCELLENCE IN RANDOMNESS AND STRUCTURES 2022–2029

ACADEMY OF FINLAND

Joint work with: Santeri Karppinen (Jyväskylä)
Sumeetpal S. Singh (Wollongong)
Nicolas Chopin (ENSAE)
Tomás Soto (LUT)

UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS AND STATISTICS

Outline

Introduction: Smoothing & conditional particle filters

Problematic case: Time-discretised path integral model

Solution: CPF with bridge backward sampling

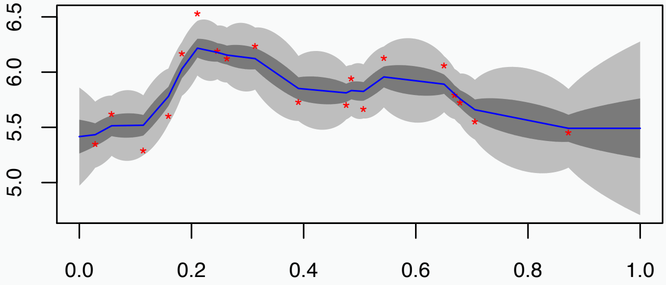
Introduction: Smoothing & conditional particle filters

The smoothing problem

Hidden Markov model (a.k.a. general state space model):

- Latent Markov chain X_k with initial law $m_1(\,\cdot\,)$ and transitions $m_k(x_{k-1},\,\cdot\,)$
- Conditionally independent observations y_k with densities $g_k(\,\cdot\mid X_k)$

We aim at inferring $\pi_T = \mathcal{L}(X_{1:T} \mid Y_{1:T} = y_{1:T})$ (smoothing).

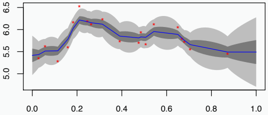


The smoothing problem

Hidden Markov model (a.k.a. general state space model):

- Latent Markov chain X_k with initial law $m_1(\,\cdot\,)$ and transitions $m_k(x_{k-1},\,\cdot\,)$
- Conditionally independent observations y_k with densities $g_k(\,\cdot \mid X_k)$

We aim at inferring $\pi_T = \mathcal{L}(X_{1:T} \mid Y_{1:T} = y_{1:T})$ (smoothing).



The Feynman–Kac notation/generalisation:

- Proposal distributions: $M_1(\,\cdot\,)$, $M_k(x_{k-1},\,\cdot\,)$ — e.g. $M_k \equiv m_k$
- Potential (or weight) functions: $G_k(\,\cdot\,) \geq 0$ — e.g. $G_k(\,\cdot\,) \equiv g_k(y_k \mid \,\cdot\,)$

$\rightsquigarrow$ different 'proposal distributions', 'lookaheads' and more general models...

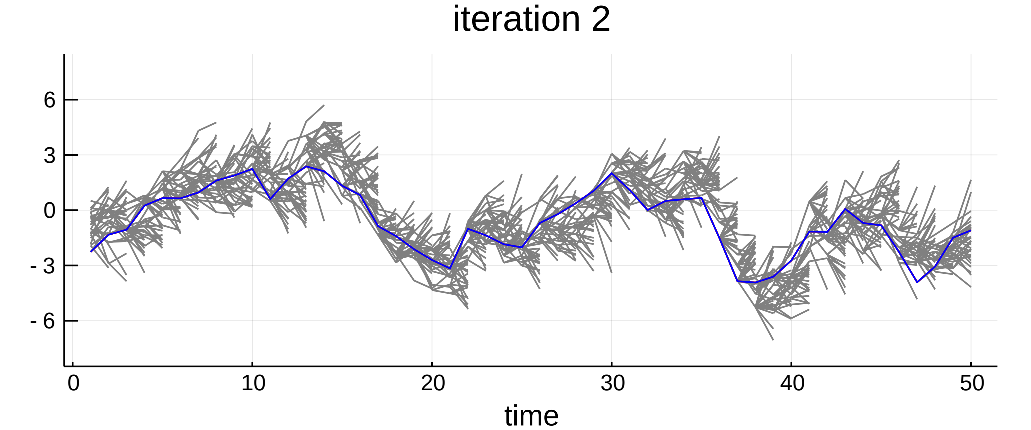
Conditional particle filter (CPF)

Markov transition $x^*_{1:T} \to (X_1^{B_1}, \ldots, X_T^{B_T})$ leaving π_T invariant $\forall N \geq 2!$ [1]

1. Reference path $x^*_{1:T}$
2. Sample remaining $N-1$ particles sequentially from M_1, M_2, ..., M_T;
   conditional resample proportional to $G_k(\,\cdot\,)$
3. Pick particle B_T at last time T;
   Ancestor trace $B_{T-1}, \ldots, B_1$ and output $(X_1^{B_1}, \ldots, X_T^{B_T})$

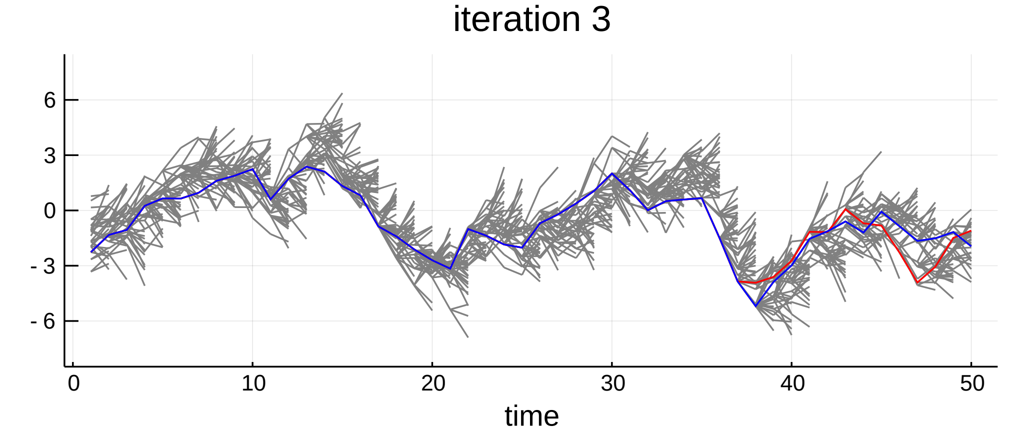
[1] Andrieu, Doucet & Holenstein (*J. Roy. Statist. Soc. Ser. B.*, 2010)

Iterated CPF on noisy AR(1)



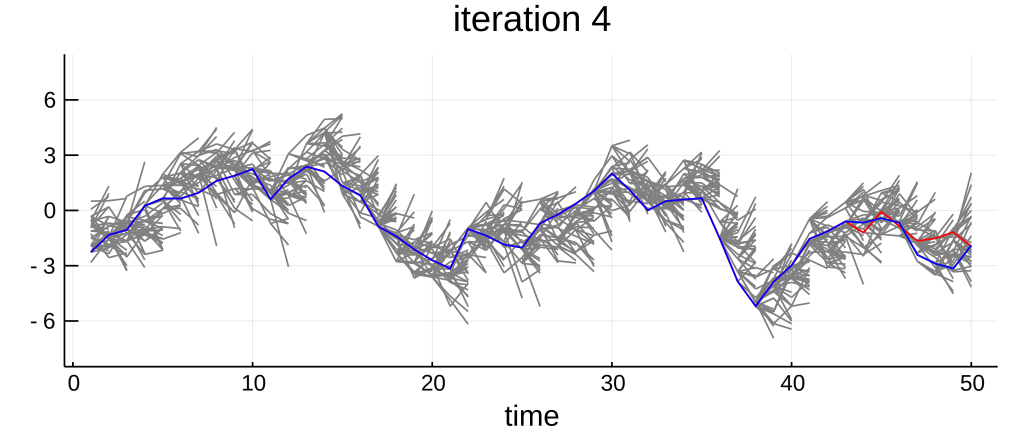
- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

Iterated CPF on noisy AR(1)



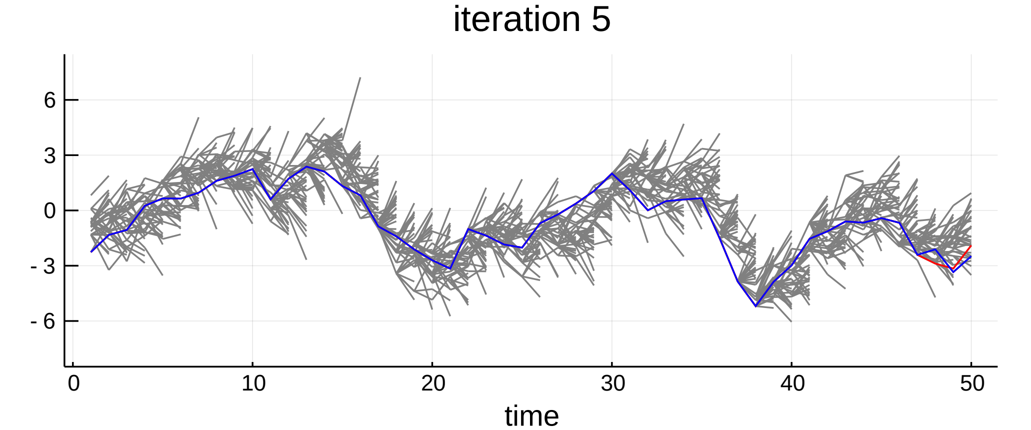
- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

Iterated CPF on noisy AR(1)



- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

Iterated CPF on noisy AR(1)



- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

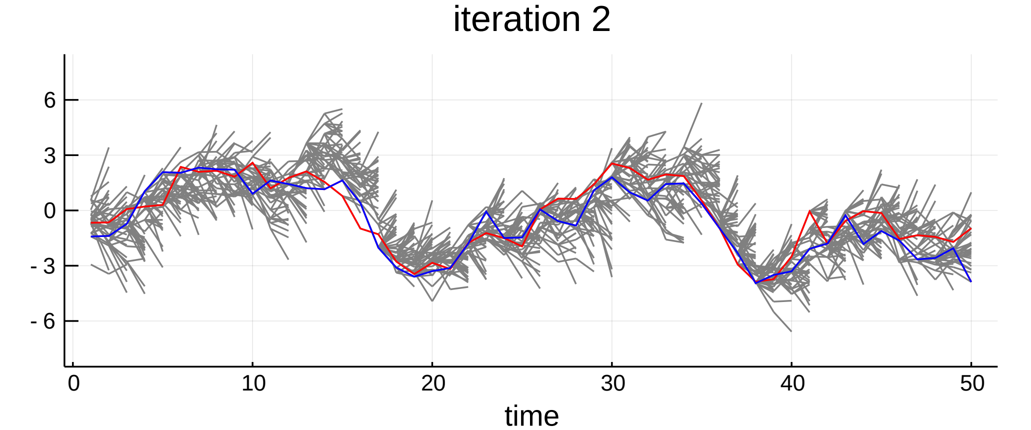
Conditional particle filter with backward sampling (CPF-BS)

Efficient MCMC smoothing: π_T-invariant transition $x^*_{1:T} \to (X_1^{B_1}, \ldots, X_T^{B_T})$[2]

1. Reference path $x^*_{1:T}$
2. Sample remaining $N-1$ particles sequentially from M_1, M_2, ..., M_T; conditional resample proportional to $G_k(\,\cdot\,)$
3. Pick particle B_T at last time T; Backward sample $B_{T-1}, \ldots, B_1$ and output $(X_1^{B_1}, \ldots, X_T^{B_T})$

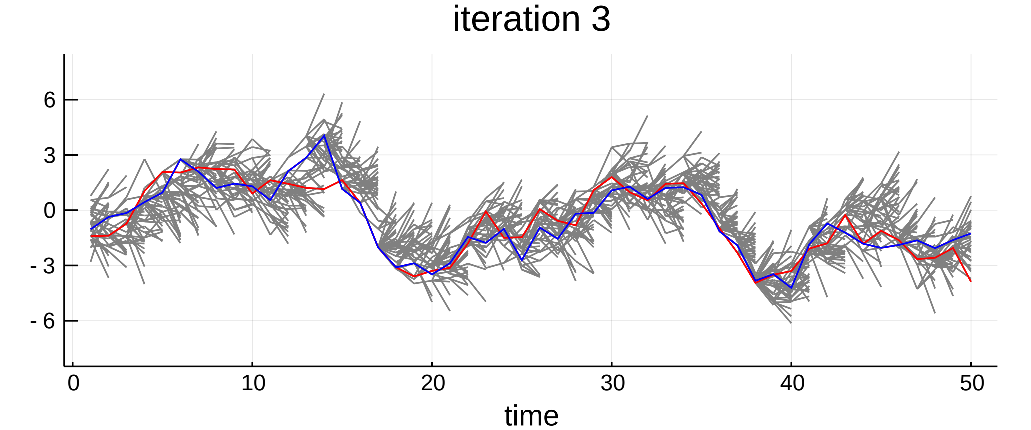
[2]Whiteley (*J. Roy. Statist. Soc. Ser. B.*, 2010); Algorithmic variant: Lindsten, Jordan & Schön (*J. Mach. Learn. Res.*, 2014) ancestor sampling CPF

Iterated CPF-BS on noisy AR(1)



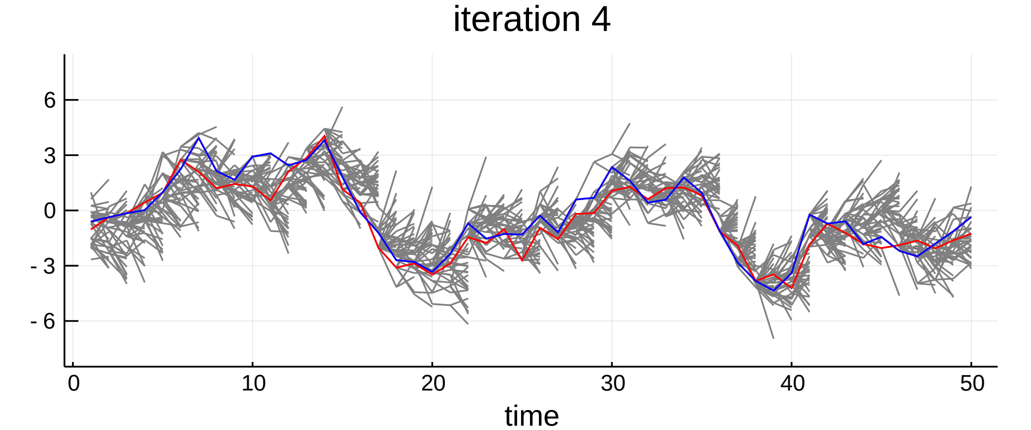
- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

Iterated CPF-BS on noisy AR(1)



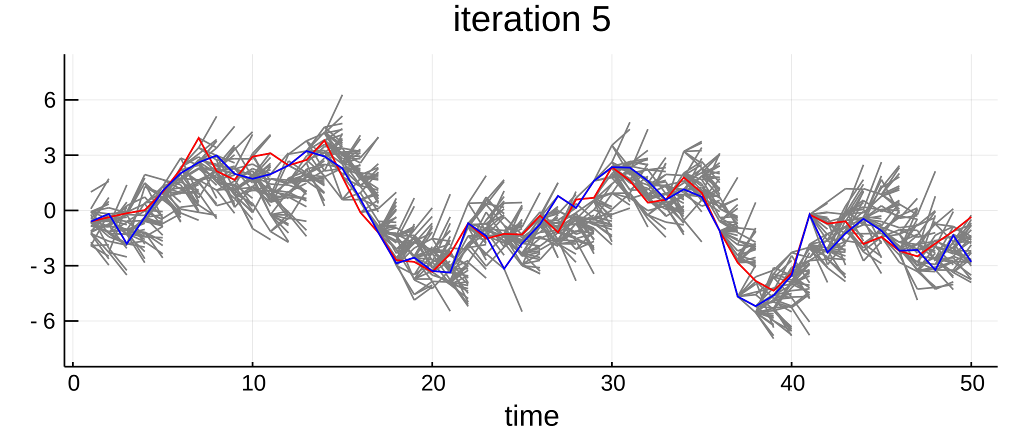
- Reference $x^*_{1:T}$, Output $X^{B_{1:T}}_{1:T}$

Iterated CPF-BS on noisy AR(1)



- Reference $x^*_{1:T}$, Output $X_{1:T}^{B_{1:T}}$

Iterated CPF-BS on noisy AR(1)



- Reference $x^*_{1:T}$, Output $X^{B_{1:T}}_{1:T}$

Why to use CPF-BS?

- Hard high-dimensional inference problem: $X_k \in \mathbb{R}^d \implies X_{1:T} \in \mathbb{R}^{dT}$.
- Impressive empirical performance even with very large T
- Theoretical guarantees under general conditions[3]:
  - Number of particles N sufficiently large (but constant in T!)
  - Mixing time at most linear in T

- CPF-BS can be plugged into particle Gibbs, where the model has unknown parameters θ:
  - Update $X_{1:T} \mid \theta$ by CPF-BS (using $(M^\theta_{1:T}, G^\theta_{1:T})$)
  - Update $\theta \mid X_{1:T}$ by your favourite MCMC...

[3] Lee, Singh & V (*Ann. Statist.*, 2020)

Problematic case: Time-discretised path integral model

The path integral model

- A continuous-time analogue of the HMM:

$$\Pi_\tau(A) = \frac{1}{\mathcal{Z}} \mathbb{E}\Big[\overbrace{1(Z_{[0,\tau]} \in A)}^{\text{Markov process law}} \overbrace{\exp\Big(-\int_0^\tau V(Z_u)\mathrm{d}u\Big)}^{\text{Weight}}\Big],$$

  where $(Z_t)_{t\geq 0} \sim \mathbb{M}$ is a Markov process (typically diffusion, from SDE) and $\mathcal{Z} > 0$ is a normalising constant.
- Potential function $V \geq 0$ modulates the law $\mathbb{M}$ — the 'environment'
  - $V(x)$ has small 'penalty' value $\rightsquigarrow$ preferred state x
  - $V(x)$ has high 'penalty' value $\rightsquigarrow$ unlikely state x

Time-discretised path integral model

- Time-discretisation $0, \Delta, 2\Delta, \ldots, \Delta T = \tau$ and Riemann sum approximation:

$$\pi_T(x_{1:T}) = \frac{1}{\mathcal{Z}}\Bigg(M_1(x_1)\prod_{k=2}^{T} M_k^{\Delta}(x_k \mid x_{k-1})\Bigg)\Bigg(\prod_{k=1}^{T-1} \overbrace{e^{-\Delta V(x_k)}}^{G_k^{\Delta}(x_k)}\Bigg),$$

  where
  - $X_k = Z_{\Delta(k-1)}$
  - M_k^{Δ} are the (approximate) transition probabilities from $Z_{\Delta(k-1)}$ to $Z_{\Delta k}$
- When $\Delta \to 0$:
  - $M_k^{\Delta}(x_k \mid x_{k-1}) \to \delta_{x_{k-1}}(x_k)$
  - $G_k^{\Delta}(x) \approx 1 - \Delta V(x) \to 1$
  - $T = \tau/\Delta \to \infty$
- Problem should not get 'harder' as $\Delta \to 0$, but 'standard' CPF-BS degenerates...

CPF resampling for $\Delta \to 0$

- Original CPF & CPF-BS used conditional multinomial resampling[4]
  - Degenerates $\times$
- Two other conditional resamplings for CPF[5]:
  - residual degenerates $\times$
  - systematic seems stable $\checkmark$ (empirically...)
- Can we say something more about resamplings in the limit?

[4] Andrieu, Doucet & Holenstein (*J. Roy. Statist. Soc. Ser. B.*, 2010)
[5] Chopin & Singh (*Bernoulli*, 2015)
[6] Chopin, Singh, Soto & V, *Ann. Statist.* (2022)

CPF resampling for $\Delta \to 0$

- Original CPF & CPF-BS used conditional multinomial resampling[4]
  - Degenerates $\times$
- Two other conditional resamplings for CPF[5]:
  - residual degenerates $\times$
  - systematic seems stable $\checkmark$ (empirically...)
- Can we say something more about resamplings in the limit?
- At least for ~~condiitonal~~ particle filters[6]:
  - 'Killing' $\checkmark$
  - 'Stratified' (with mean partition) $\checkmark$
  - 'Systematic' (w.m.p) $\checkmark$
  - 'Srinivasan Sampling process' (w.m.p.) (SSP) $\checkmark$

[4] Andrieu, Doucet & Holenstein (*J. Roy. Statist. Soc. Ser. B.*, 2010)
[5] Chopin & Singh (*Bernoulli*, 2015)
[6] Chopin, Singh, Soto & V, *Ann. Statist.* (2022)

Interlude (for experts): Particle filter limit as $\Delta \to 0$

Assumption: Instantaneous resampling rate

For all $v^{1:N} \geq 0$ and all $a^{1:N} \in \{1{:}N\}^N$, $a^{1:N} \neq 1{:}N$ the limit:

$$\iota(a^{1:N}, v^{1:N}) = \lim_{\Delta \to 0+} \frac{1}{\Delta} r\big(a^{1:N} \mid \exp(-\Delta v^1), \ldots, \exp(-\Delta v^N)\big)$$

exists and the term inside the limit is uniformly bounded for bounded potentials $v^{1:N} \leq v^* < \infty$ and $\Delta \in (0,1)$.

Theorem (Chopin, Singh, Soto, V, 2022)

Under Assumption & technical conditions, the discrete-time particle filter converges to Markov process with generator

$$\mathcal{L}f(x^{1:N}) = \mathcal{L}^{\mathbb{M}} f(x^{1:N}) + \sum_{a^{1:N} \neq 1:N} \iota\big(a^{1:N}, (V(x^1), \ldots, V(x^N))\big)\big(f(x^{a^{1:N}}) - f(x^{1:N})\big)$$

Interlude (cont): Limiting overall death rates

- Simple expressions for the overall death rate:

$$\iota^*(v^{1:N}) = \textstyle\sum_{a^{1:N}\neq 1:N} \iota(a^{1:N}, v^{1:N}),$$

which captures $\mathbb{P}(A_k^{1:N} \neq 1{:}N) \approx \Delta\iota^*\big(V(X_k^1), \ldots, V(X_k^N)\big)$

Proposition

$$\iota^*_{\text{killing}}(v^{1:N}) = \frac{N-1}{N}\sum_{i=1}^{N}(v^i - v_{\min}) = (N-1)(\bar{v} - v_{\min}) \qquad v_{\min} = \min_j v^j$$

$$\iota^*_{\text{stratified}}(v^{1:N}) = \sum_{j=1}^{N} j(\bar{v} - v^{\varpi(j)}) \qquad \bar{v} = \frac{1}{N}\sum_{i=1}^{N} v^i$$

$$\iota^*_{\text{systematic}}(v^{1:N}) = \frac{1}{2}\sum_{i=1}^{N}|\bar{v} - v^i| = \iota^*_{\text{ssp}}(v^{1:N})$$

Interlude (cont.): Ordering the overall death rates

Theorem (Chopin, Singh, Soto, V, 2022)

The overall resampling intensities of killing, stratified and systematic/SSP resampling satisfy

$$\iota^*_{\text{killing}}(v^{1:N}) \geq \iota^*_{\text{systematic}}(v^{1:N}), \qquad \text{and} \qquad \iota^*_{\text{stratified}}(v^{1:N}) \geq \iota^*_{\text{systematic}}(v^{1:N}),$$

for all potential values $v^{1:N}$. However, ι^*_{killing} and $\iota^*_{\text{stratified}}$ do not satisfy such order in general.

⇝ Systematic/SSP with mean partition preferable

- Karppinen, Singh & V (to appear):
  - Conditional killing resampling
  - Conditional systematic *with mean partition* ✓✓

Backward sampling with stiff dynamics M_k

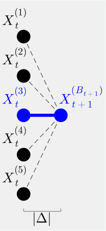
As $\Delta \to 0$:

- Continuous paths $\implies M_k^{\Delta}(\,\cdot \mid x) \to \delta_x(\,\cdot\,)$.
- Backward sampling step:

$$\mathbb{P}(B_k = i \mid B_{k+1}) \propto G_k^{\Delta}(X_k^i)\, \overbrace{M_{k+1}^{\Delta}(X_{k+1}^{B_{k+1}} \mid X_k^i)}^{\approx 0 \text{ for } i \neq A_{k+1}^{B_{k+1}}}$$

$$\implies B_k = A_{k+1}^{B_{k+1}} \text{ with high probability.}$$

$\therefore$ CPF-BS degenerates to CPF with ancestor tracing;
the benefits of backward sampling are lost...

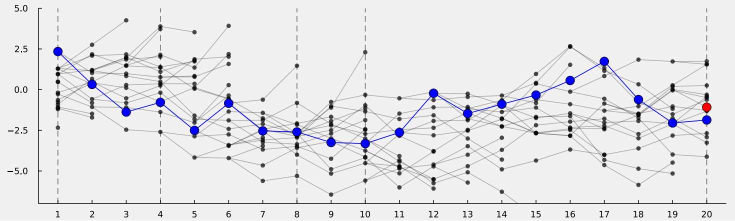


Solution: CPF with bridge backward sampling

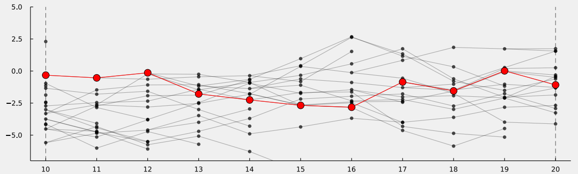
CPF with bridge backward sampling (CPF-BBS)

Generalisation of backward sampling to blocks of length > 1

- We use auxiliary 'bridge CPF' steps
- Requires tractable M_k (can calculate conditionals, such as linear-Gaussian)
- Initialisation: run CPF with reference $x^*_{1:T}$ and choose a particle $X_T^{(B_T)}$

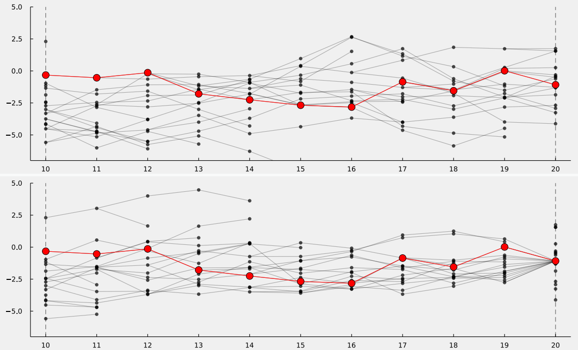


Bridge CPF: block from $\ell = 10$ to $u = 20$



1. Trace back from $X_u^{B_u}$
   $\rightarrow$ 'block reference' $\hat{X}_{\ell:u}$

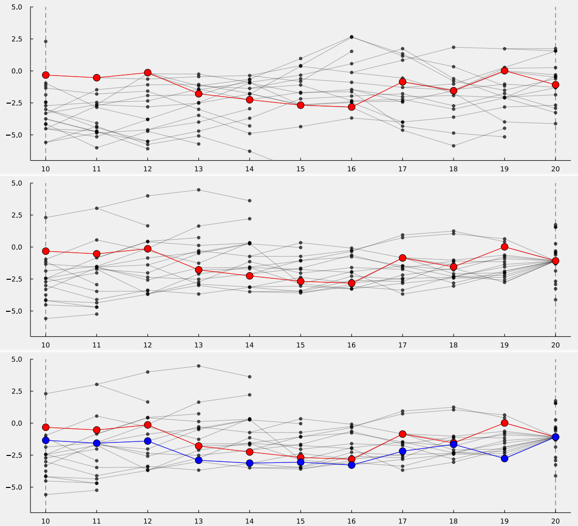
Bridge CPF: block from $\ell = 10$ to $u = 20$



1. Trace back from $X_u^{B_u}$
   $\to$ 'block reference' $\hat{X}_{\ell:u}$

2. Simulate particles $\tilde{X}_{\ell+1:u-1}^{1:N}$ with auxiliary CPF:
   - reference $\hat{X}_{\ell:u}$
   - conditional transitions $M_{k|k-1,u}(\,\cdot \mid x_{k-1}, \hat{X}_u)$
   - modified potentials $G_k(x_{k-1}, x_k) M_{u|\ell}(\hat{X}_u \mid x_\ell)^{\frac{1}{u-\ell}}$,

Bridge CPF: block from $\ell = 10$ to $u = 20$



1. Trace back from $X_u^{B_u}$
   $\to$ 'block reference' $\hat{X}_{\ell:u}$

2. Simulate particles $\tilde{X}^{1:N}_{\ell+1:u-1}$ with auxiliary CPF:
   - reference $\hat{X}_{\ell:u}$
   - conditional transitions $M_{k|k-1,u}(\,\cdot\mid x_{k-1}, \hat{X}_u)$
   - modified potentials $G_k(x_{k-1}, x_k) M_{u|\ell}(\hat{X}_u \mid x_\ell)^{\frac{1}{u-\ell}}$,

3. Output trajectory $\tilde{X}^{\tilde{B}_{\ell:u}}_{\ell:u}$

Choosing the blocks

- CPF-BBS efficient if blocks chosen appropriately
- Natural rule: maximise 'probability of local update' (PLU):

$$\mathbb{P}(\tilde{X}_\ell^{\tilde{B}_\ell} \neq \hat{X}_\ell)$$

Choosing the blocks

- CPF-BBS efficient if blocks chosen appropriately
- Natural rule: maximise 'probability of local update' (PLU):

$$\mathbb{P}(\tilde{X}_\ell^{\tilde{B}_\ell} \neq \hat{X}_\ell) \approx \mathrm{PLU}_M(\ell,u)\mathrm{PLU}_G(\ell,u)\Big(1-\tfrac{1}{N}\Big)^{-1}$$

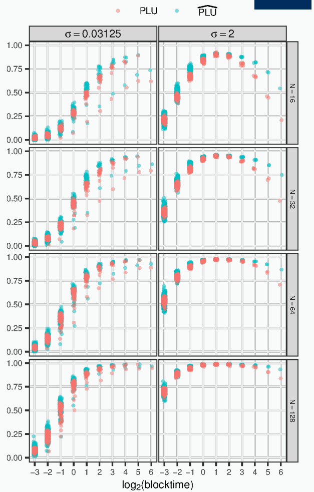
- Based on two approximate estimators for PLU:

$$\mathrm{PLU}_M(\ell,u) = 1 - \frac{M_{u|\ell}(x_u^* \mid x_\ell^*)}{\sum_{j=1}^N M_{u|\ell}(x_u^* \mid x_\ell^{(j)})} \text{ (short blocks)}$$

$$\mathrm{PLU}_G(\ell,u) = \left(1-\frac{1}{N}\right)\prod_{k=\ell}^{u-1}\left(1-\frac{p_k N}{(N-1)^2}\right) \text{ (long blocks),}$$

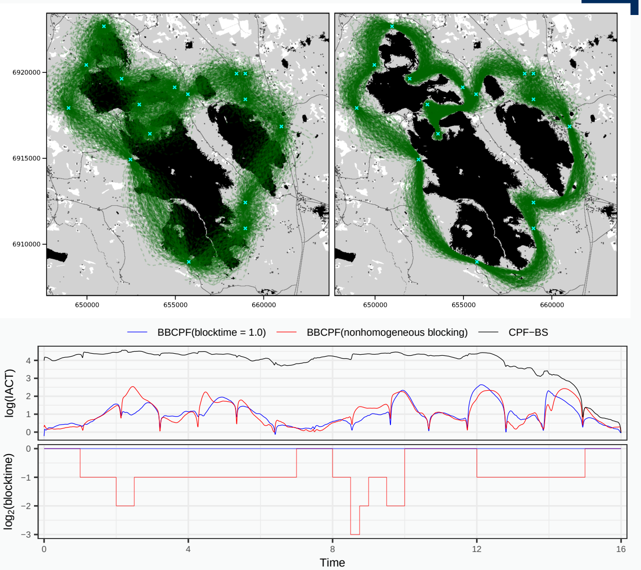
where $p_k = \frac{1}{2}\sum_{i=1}^N \left|W_k^i - \frac{1}{N}\right|$ systematic 'resampling rate'

- One trial run, can evaluate with multiple candidate ℓ, u...



Example: Interpolation with terrain preferences

- Ornstein–Uhlenbeck velocity conditioned on linear Gaussian observations × (left)
- Path integral model: avoid dark water bodies (right)[a]
- Hand-tuned constant size blocking
- PLU-based automatic 'optimised' blocking



[a] Animation: @MattiVihola

Concluding remarks

CPF can work with weakly informative G_k:

- Use conditional systematic resampling with mean partition with CPF

CPF-BS can be extended to work with stiff M_k (if tractable):

- Use bridge backward sampling CPF
- Good heuristics to choose appropriate blocking

Discussion:

- Weak potentials and stiff dynamics arise with HMMs, too!
- The method requires tractable $M_{1:T}$, which can be from a Gaussian approximation...
- Blocked particle Gibbs (Singh, Lindsten & Moulines, *Biometrika*, 2015) can also be used, but requires block size & *block overlap* calibration...

References

- **C. Andrieu, A. Doucet and R. Holenstein.**
  Particle Markov chain Monte Carlo methods.
  J. R. Stat. Soc. Ser. B. Stat. Methodol., 2010.
- **N. Whiteley.**
  Discussion on “PMCMC methods”.
  J. R. Stat. Soc. Ser. B. Stat. Methodol., 2010.
- **F. Lindsten, M. I. Jordan and T. B. Schön.**
  Particle Gibbs with ancestor sampling.
  J. Mach. Learn. Res., 2014.
- **N. Chopin and S. S. Singh**
  On particle Gibbs sampling.
  Bernoulli, 2015.
- **A. Lee, S. S. Singh and M. Vihola.**
  Coupled conditional backward sampling particle filter.
  Ann. Statist., 2020.
- **S. S. Singh, F. Lindsten and E. Moulines.**
  Blocking strategies and stability of particle Gibbs samplers.
  Biometrika, 2017.
- **N. Chopin, S. S. Singh, T. Soto and M. Vihola**
  On resampling schemes for particle filters with weakly informative observations.
  Ann. Statist., 2022.
- **S. Karppinen, S. S. Singh and M. Vihola.**
  Conditional particle filters with bridge backward sampling
  J. Comput. Graph. Statist., to appear. (*arXiv:2205.13898*)

Thank you!

Julia package `https://github.com/skarppinen/Resamplings.jl`
& codes `https://github.com/skarppinen/cpf-bbs`

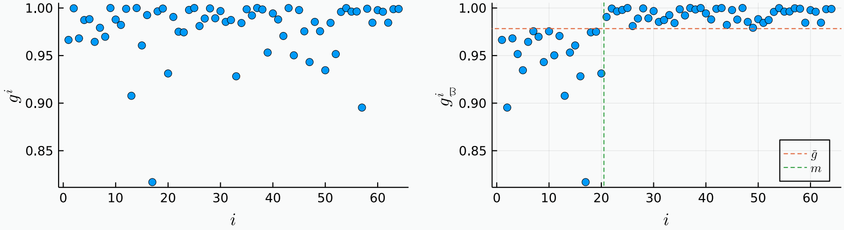
Mean partition

Definition

Suppose that $g^{1:N} \in \mathbb{R}^N$. A permutation ϖ of $1{:}N$ is a mean partition (order) for $g^{1:N}$ if the re-indexed $g_\varpi^i = g^{\varpi(i)}$ satisfy

$$g_\varpi^1, \dots, g_\varpi^m \le \bar{g} \qquad \text{and} \qquad g_\varpi^{m+1}, \dots, g_\varpi^N > \bar{g},$$

where the pivot $\bar{g}$ is the mean $\bar{g} = \frac{1}{N}\sum_{i=1}^N g^i$.

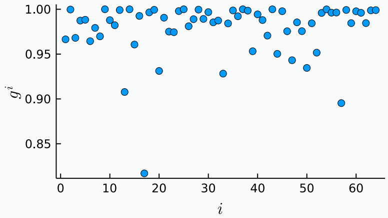


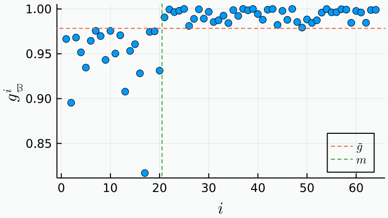
Stratified/Systematic/SSP with mean partition

Slightly modified resampling schemes:

1. Mean partition order the $g^{1:N} \to g_{\varpi}^{1:N}$
   - costs $O(N)$, like the first stage of Quicksort...
2. Apply the Strat./Syst./SSP resampling: $\tilde{A}^{1:N} \sim r(\,\cdot \mid g_{\varpi}^{1:N})$
3. Map the $\tilde{A}^{1:N}$ back to 'original indexing' $A^{1:N}$:

$$A^{\varpi(i)} = \varpi(\tilde{A}^i)$$

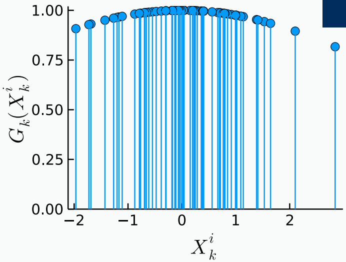




Weak potentials

The potentials G_k are 'weakly informative' if the weights $G_k(X_k^1), \ldots, G_k(X_k^N)$ are 'nearly constant'

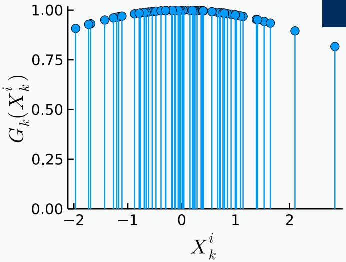
$\rightsquigarrow$ Normalised weights $W_k^i = \dfrac{G_k(X_k^i)}{\sum_{j=1}^N G_k(X_k^j)} \approx \dfrac{1}{N}$



Weak potentials

The potentials G_k are 'weakly informative' if the weights $G_k(X_k^1), \ldots, G_k(X_k^N)$ are 'nearly constant'

$\rightsquigarrow$ Normalised weights $W_k^i = \dfrac{G_k(X_k^i)}{\sum_{j=1}^N G_k(X_k^j)} \approx \dfrac{1}{N}$



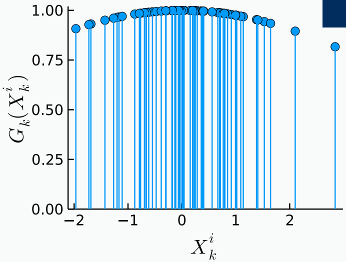
Examples of weakly informative scenarios:

1. If $G_k(x_k) = g_k(y_k \mid x_k)$ in the HMM context, Y_k have high variability wrt. M_k

Weak potentials

The potentials G_k are 'weakly informative' if the weights $G_k(X_k^1), \ldots, G_k(X_k^N)$ are 'nearly constant'

$\rightsquigarrow$ Normalised weights $W_k^i = \dfrac{G_k(X_k^i)}{\sum_{j=1}^N G_k(X_k^j)} \approx \dfrac{1}{N}$



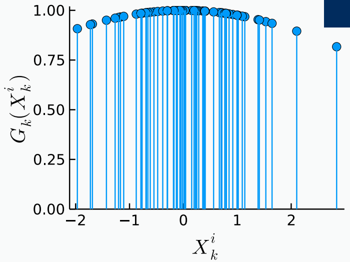
Examples of weakly informative scenarios:

1. If $G_k(x_k) = g_k(y_k \mid x_k)$ in the HMM context, Y_k have high variability wrt. M_k
2. When M_k form a good approximation of the smoothing distribution
   (e.g. Laplace approximation with linear-Gaussian m_k and nonlinear g_k...)

Weak potentials

The potentials G_k are 'weakly informative' if the weights $G_k(X_k^1), \dots, G_k(X_k^N)$ are 'nearly constant'

$\rightsquigarrow$ Normalised weights $W_k^i = \dfrac{G_k(X_k^i)}{\sum_{j=1}^N G_k(X_k^j)} \approx \dfrac{1}{N}$



Examples of weakly informative scenarios:

1. If $G_k(x_k) = g_k(y_k \mid x_k)$ in the HMM context, Y_k have high variability wrt. M_k
2. When M_k form a good approximation of the smoothing distribution (e.g. Laplace approximation with linear-Gaussian m_k and nonlinear g_k...)
3. When M_k, G_k come from a continuous-time path integral model

BridgeCPF$(\underline{\boldsymbol{x}}_\ell, b^*_{\ell:u-1}, x^*_{\ell+1:u})$

$W_\ell^{(1:N)} \leftarrow M_{u|\ell}(x_u^* \mid \underline{x}_\ell)^{\frac{1}{u-\ell}}$; $\underline{\tilde{\boldsymbol{X}}}_\ell \leftarrow \underline{\boldsymbol{x}}_\ell$

for $v = \ell + 1 : u - 1$ **do**

$\quad \tilde{A}_{v-1}^{(1:N)} \leftarrow r^{(b^*_{v-1}, b^*_v)}\left(\ \cdot \ \mid G_{v-1}(\tilde{\boldsymbol{X}}_{v-1}^{(1:N)}) W_{v-1}^{(1:N)}\right)$

$\quad$ Draw $\tilde{X}_v^{(i)} \sim \bar{M}_v(\ \cdot \ \mid \tilde{X}_{v-1}^{(\tilde{A}_{v-1}^{(i)})}, \tilde{X}_u^*)$ for $i \neq b_v^*$ and set $\tilde{X}_v^{(b_v^*)} = x_v^*$

$\quad$ Set $\tilde{\boldsymbol{X}}_v^{(i)} \leftarrow (X_{v-1}^{(\tilde{A}_{v-1}^{(i)})}, \tilde{X}_v^{(i)})$ for $i \in \{1{:}N\}$.

$\quad W_v^{(1:N)} \leftarrow W_{v-1}^{(\tilde{A}_{v-1}^{(1:N)})}$

end for

Draw $\tilde{B}_{u-1} \sim \mathrm{Categorical}(\tilde{\omega}_{u-1}^{(1:N)})$ where $\tilde{\omega}_{u-1}^{(j)} = G_{u-1}(\tilde{\boldsymbol{X}}_{u-1}^{(j)}) G_u(\tilde{X}_{u-1}^{(j)}, x_u^*) W_{u-1}^{(j)}$

$\tilde{B}_{\ell:u-2} \leftarrow$ AncestorTrace$(\underline{\tilde{A}}_{\ell:u-2}, \tilde{B}_{u-1})$

output $\left((x_\ell^{(\tilde{B}_\ell)}, \tilde{X}_{\ell+1:u-1}^{(\tilde{B}_{\ell+1:u-1})}), \tilde{B}_{\ell:u-1}\right)$

Conditional particle filter (CPF)

Markov transition $x^*_{1:T} \to (X_1^{B_1}, \ldots, X_T^{B_T})$ leaving π_T invariant $\forall N \geq 2$!

ConditionalParticleFilter$(M_{1:T}, G_{1:T}, r, N, x^*_{1:T}, b^*_{1:T})$

$X_1^{b_1^*} \leftarrow x_1^*$ and $X_1^i \sim M_1(\,\cdot\,)$ for $i \neq b_1$
for $k = 2, \ldots, T$ **do**
  $A_k^{1:N} \sim r^{(b^*_{k-1}, b^*_k)}\big(G_{k-1}(X^1_{k-1}), \ldots, G_{k-1}(X^N_{k-1})\big)$ Conditional resample
  $X_k^{b_k^*} \leftarrow x_k^*$ and $X_k^i \sim M_k(\,\cdot \mid X_{k-1}^{A_k^i})$ for $i \neq b_k^*$
end for
Draw $B_T \sim \mathrm{Categorical}\big(G_T(X_T^1), \ldots, G_T(X_T^N)\big)$ Pick particle at end time T
for $k = T-1, \ldots, 1$ **do**
  $B_k \leftarrow A_{k+1}^{B_{k+1}}$ Ancestor trace
end for
output $(X_1^{B_1}, \ldots, X_T^{B_T})$

[6] Andrieu, Doucet & Holenstein (*J. Roy. Statist. Soc. Ser. B.*, 2010)

Conditional particle filter with backward sampling (CPF-BS)

Efficient MCMC smoothing: π_T-invariant transition $x^*_{1:T} \to (X_1^{B_1}, \ldots, X_T^{B_T})$[7]

CPF-BS($M_{1:T}, G_{1:T}, r, N, x^*_{1:T}, b^*_{1:T}$)

```
X_1^{b_1^*} ← x_1^* and X_1^i ~ M_1( · )                                  for i ≠ b_1
for k = 2, ..., T do
    A_k^{1:N} ~ r^{(b_{k-1}^*, b_k^*)}(G_{k-1}(X_{k-1}^1), ..., G_{k-1}(X_{k-1}^N))    Conditional resample
    X_k^{b_k^*} ← x_k^* and X_k^i ~ M_k( · | X_{k-1}^{A_k^i})                 for i ≠ b_k^*
end for
Draw B_T ~ Categorical(G_T(X_T^1), ..., G_T(X_T^N))                         Pick index at end
for k = T-1, ..., 1 do
    B_k ~ Categorical(w_k^{1:N}), where w_k^i = G_k(X_k^i) M_{k+1}(X_{k+1}^{(B_{k+1})} | X_k^{(i)})
end for
output (X_1^{B_1}, ..., X_T^{B_T})
```

[7] Whiteley (*J. Roy. Statist. Soc. Ser. B.*, 2010);
Algorithmic variant: Lindsten, Jordan & Schön (*J. Mach. Learn. Res.*, 2014) ancestor sampling CPF