Optimal scaling for conditional sequential Monte Carlo methods in high dimensions

Axel Finke   Alexandre H. Thiery

Department of Statistics and Applied Probability
National University of Singapore

10th July 2018

Problem formulation

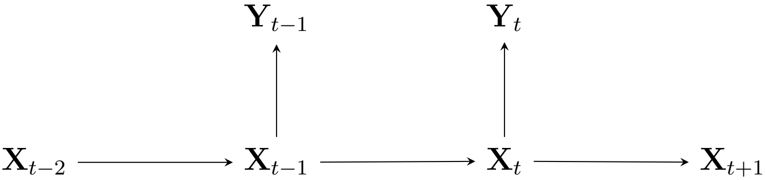
Background: CSMC algorithms (D fixed)

Breakdown of CSMC as $D \to \infty$

Novel 'random-walk' CSMC algorithm

Motivation: High-dimensional state-space model

- D-dimensional latent states: $\mathbf{X}_t = \begin{bmatrix} X_{t,1} \\ \vdots \\ X_{t,D} \end{bmatrix}$,
- T observations: $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$.



- want to approximate $\boldsymbol{\pi}_{T,D}(\mathbf{x}_{1:T}) = p(\mathbf{x}_{1:T} | \mathbf{y}_{1:T})$,
- needs MCMC updates on $(T \times D)$-dimensional space.

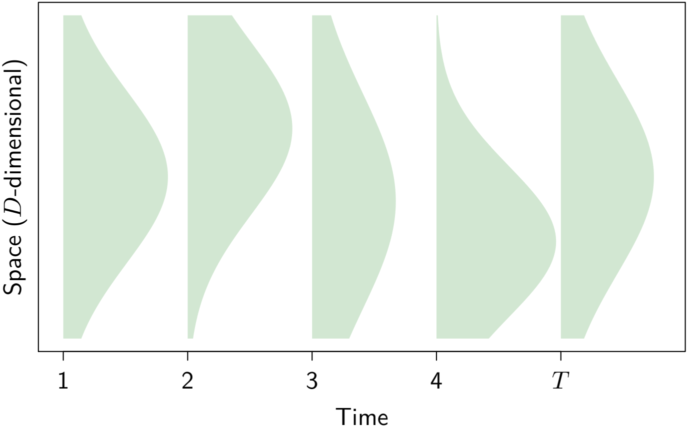
Problem formulation

Background: CSMC algorithms (D fixed)

Breakdown of CSMC as $D \to \infty$

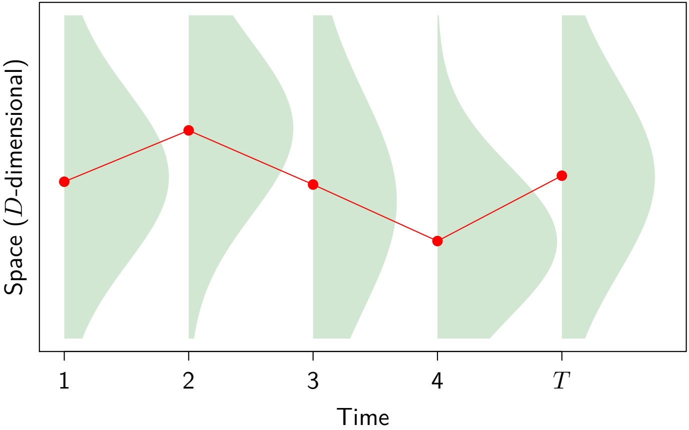
Novel 'random-walk' CSMC algorithm

CSMC algorithm



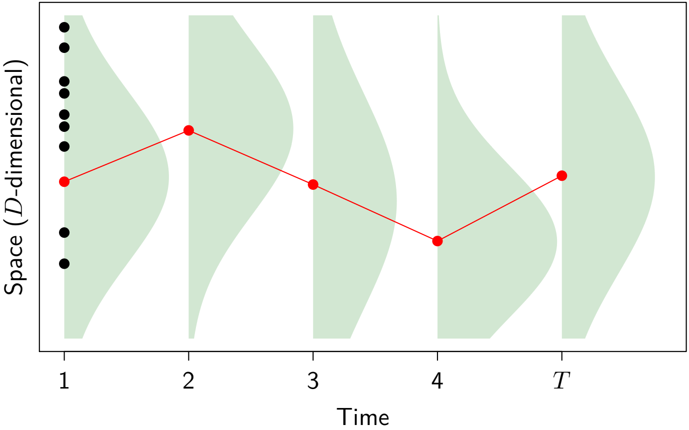
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



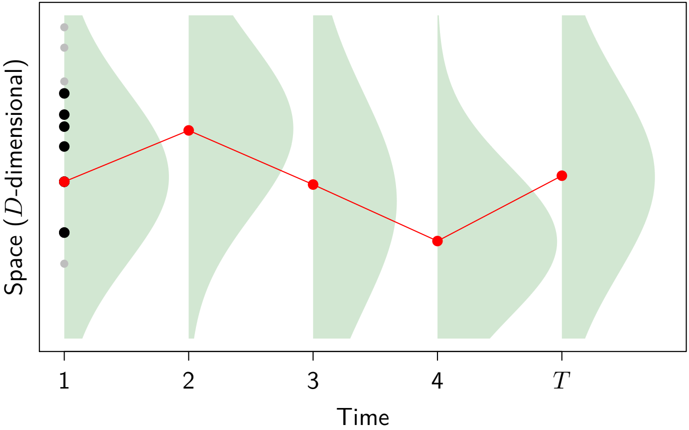
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



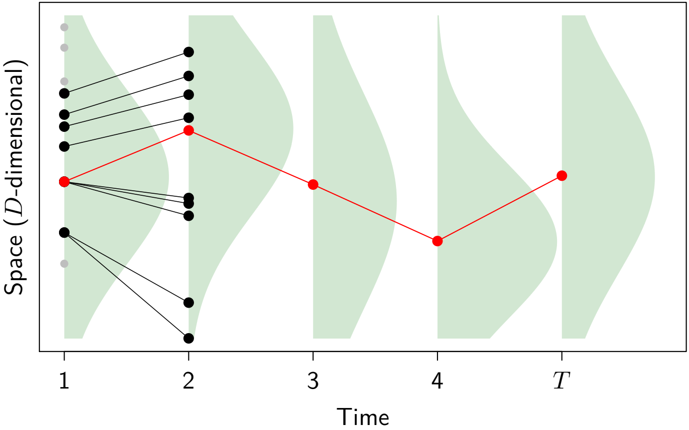
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



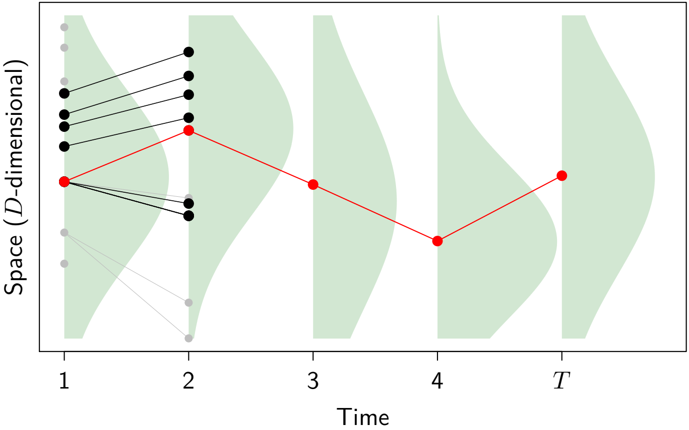
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



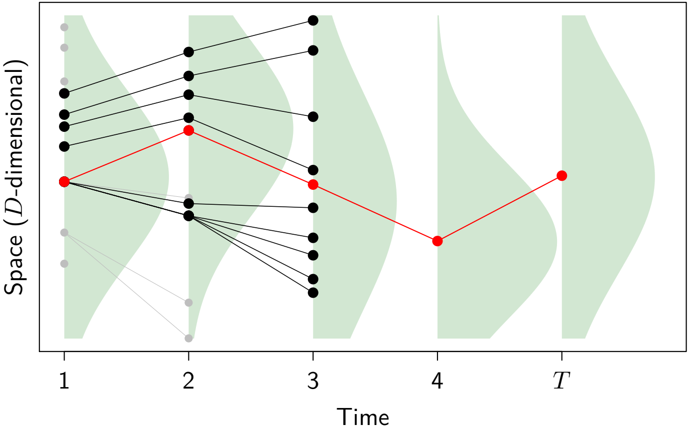
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



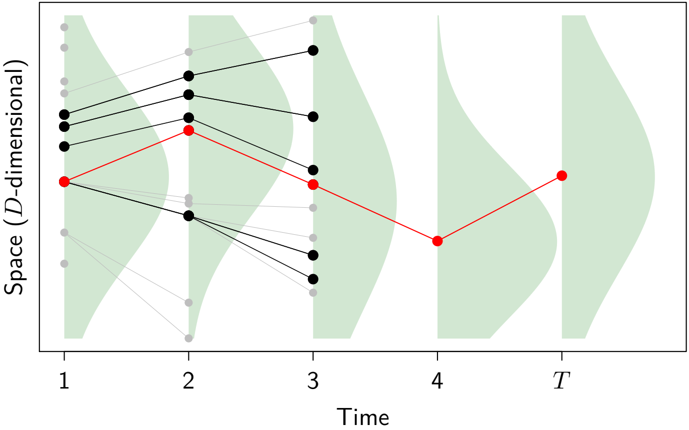
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



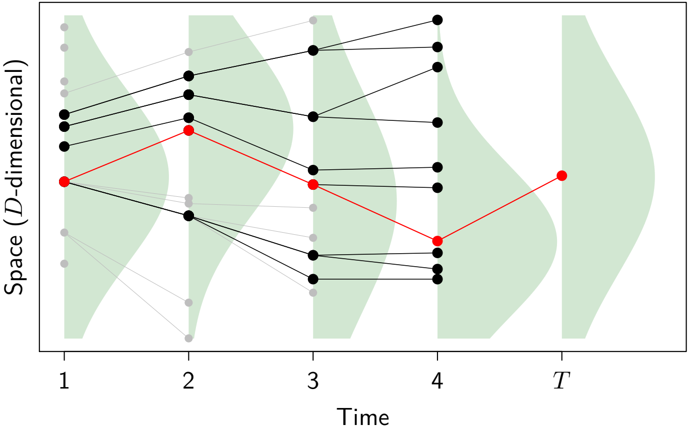
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



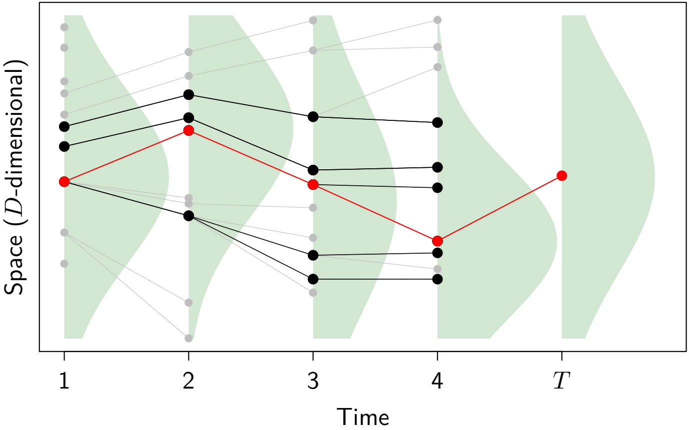
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



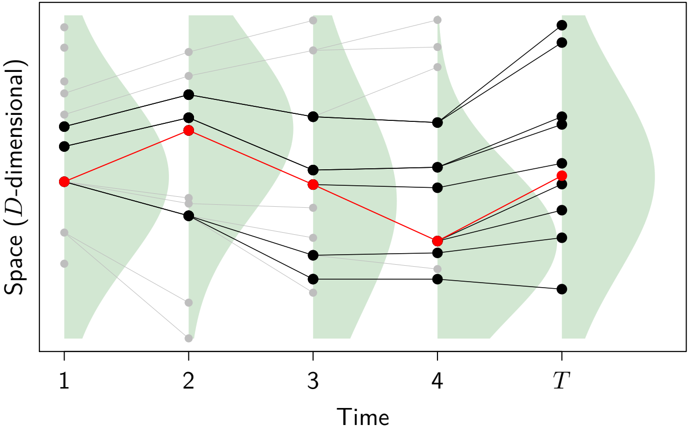
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



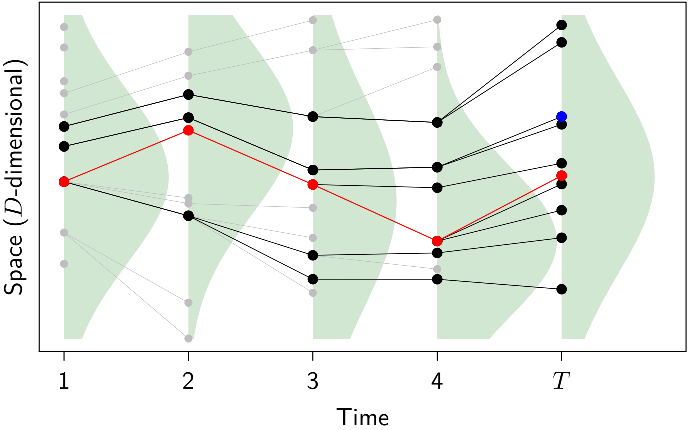
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm



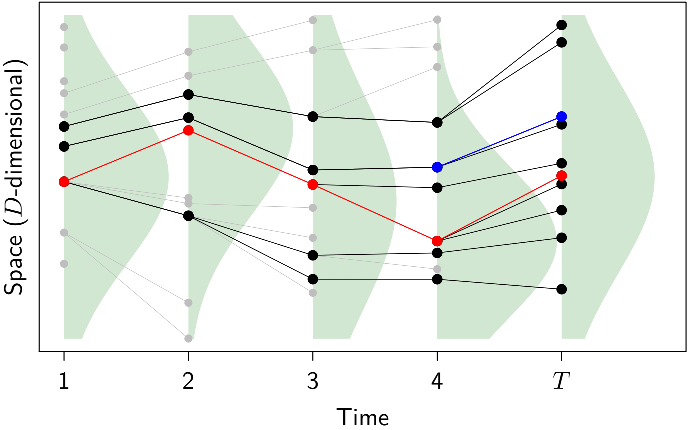
- Given reference path $\mathbf{X}_{1:T} \sim \boldsymbol{\pi}_{T,D}$ (current state of chain)
- propagate N particles/particle lineages via
  - sampling ('mutation') from the model dynamics,
  - resampling ('selection') according to importance weights.

CSMC algorithm: Selecting new reference path



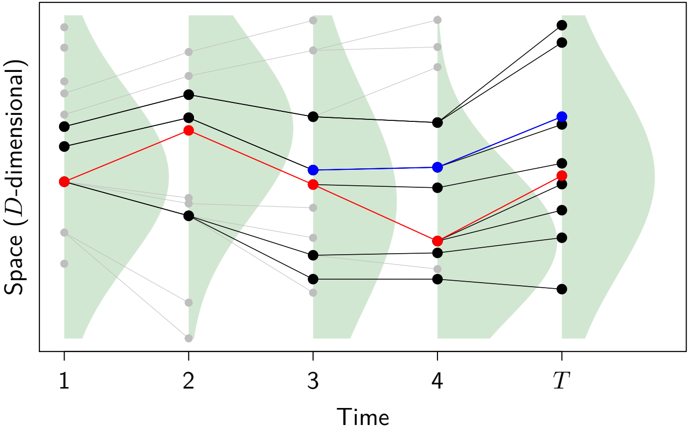
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



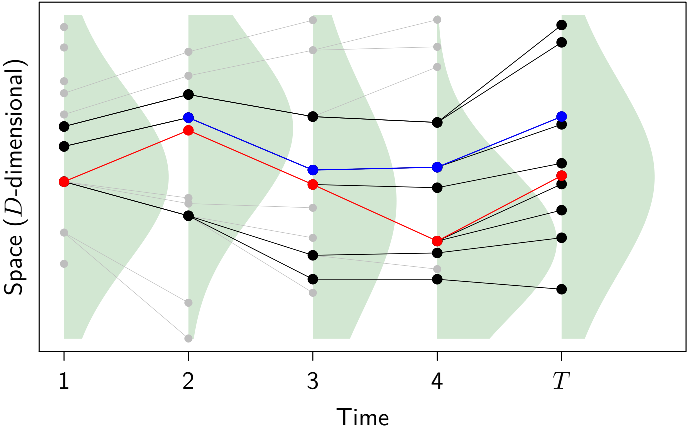
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



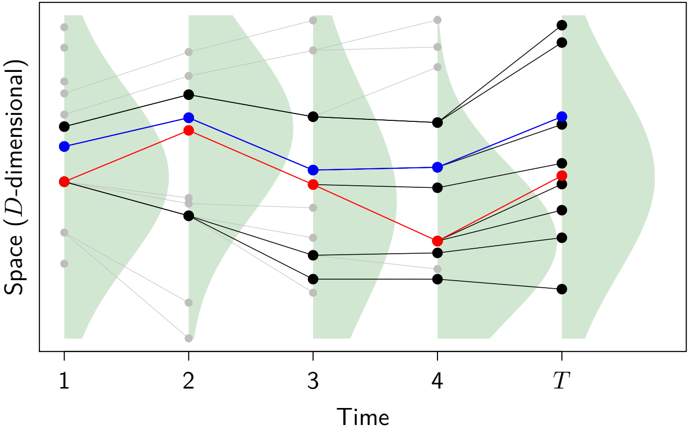
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



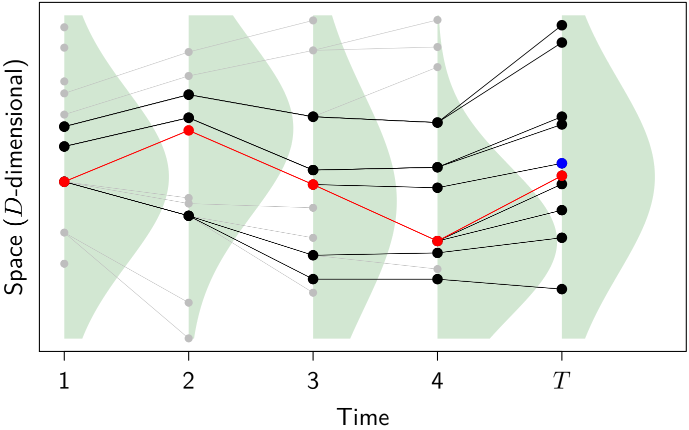
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



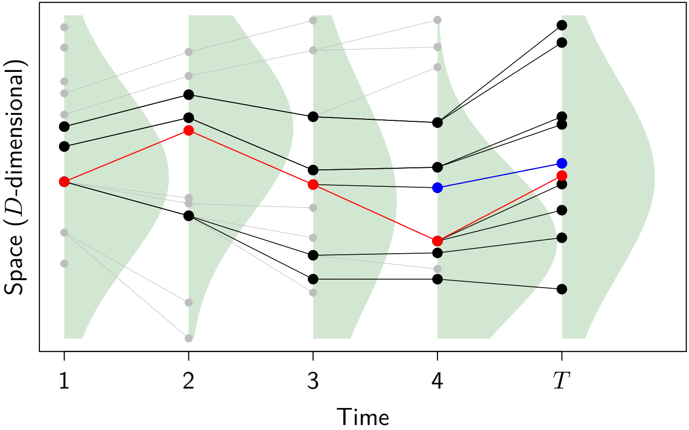
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



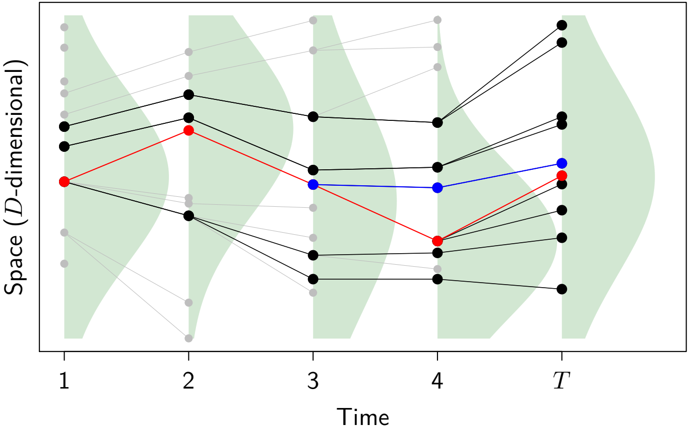
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



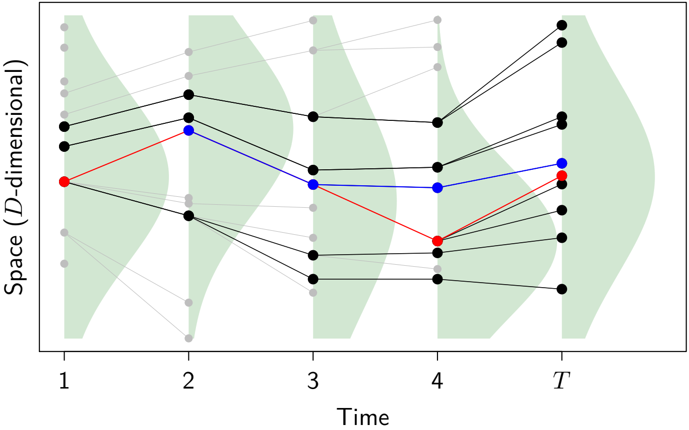
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the ‘acceptance rates’ requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



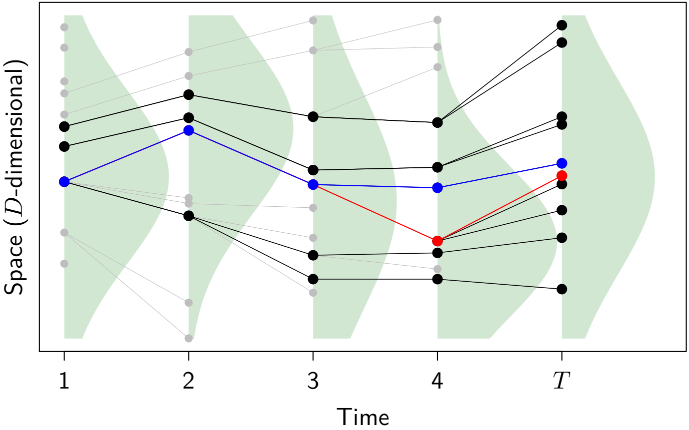
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the ‘acceptance rates’ requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



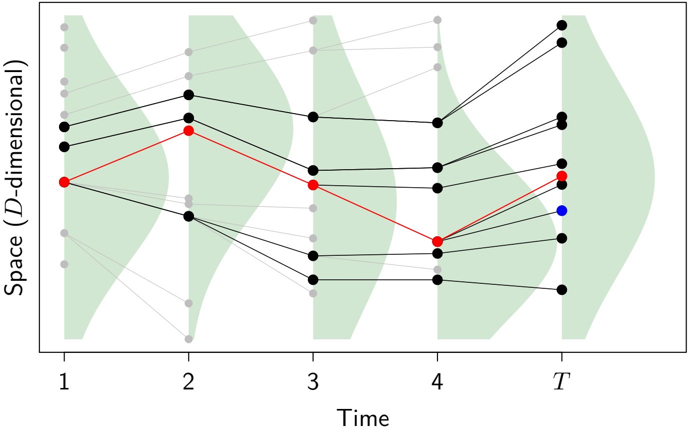
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Selecting new reference path



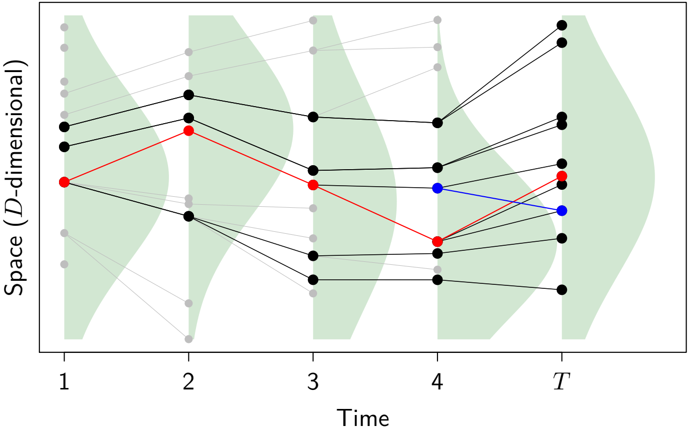
- induces $\boldsymbol{\pi}_{T,D}$-invariant Markov kernel $P^N_{T,D}(\mathbf{x}_{1:T}, \mathrm{d}\mathbf{x}_{1:T})$,
- **Problem:** $\mathbf{x}_{1:T}$ coalescing with $\mathbf{x}_{1:T}$
  - controlling the 'acceptance rates' requires $N = \mathrm{O}(T)$ (Andrieu et al., 2018; Koskela et al., 2018)

CSMC algorithm: Backward-sampling extension



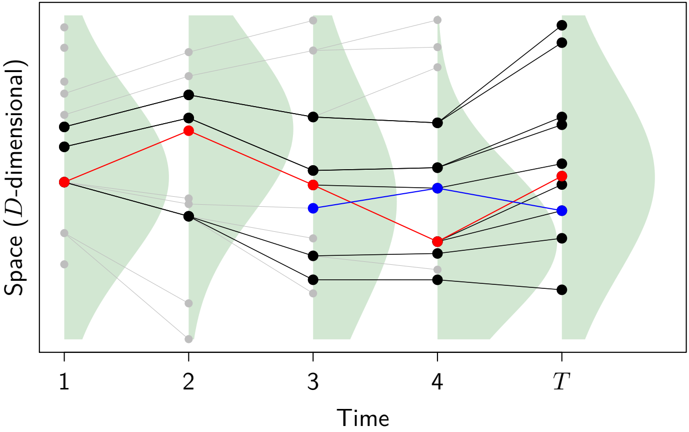
- forms new lineage;
- no longer need to grow N with T (Lee et al., 2018)

CSMC algorithm: Backward-sampling extension



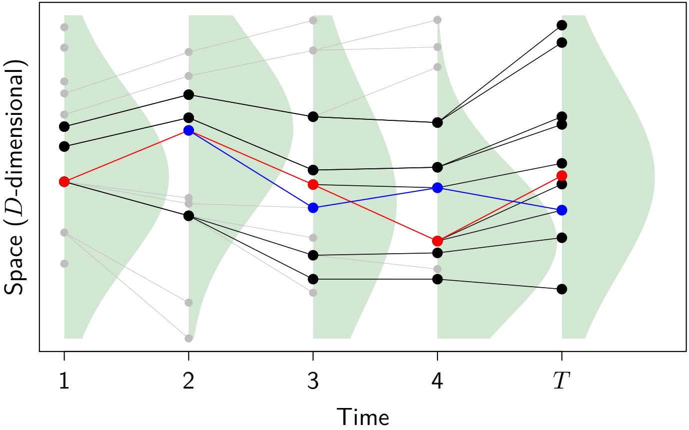
- forms new lineage;
- no longer need to grow N with T (Lee et al., 2018)

CSMC algorithm: Backward-sampling extension



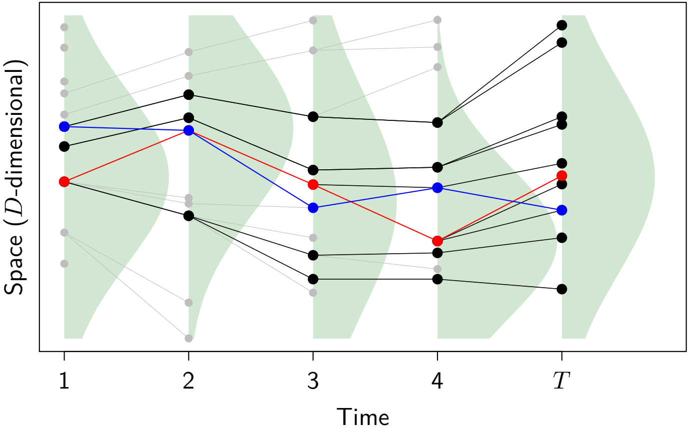
- forms new lineage;
- no longer need to grow N with T (Lee et al., 2018)

CSMC algorithm: Backward-sampling extension



- forms new lineage;
- no longer need to grow N with T (Lee et al., 2018)

CSMC algorithm: Backward-sampling extension



- forms new lineage;
- no longer need to grow N with T (Lee et al., 2018)

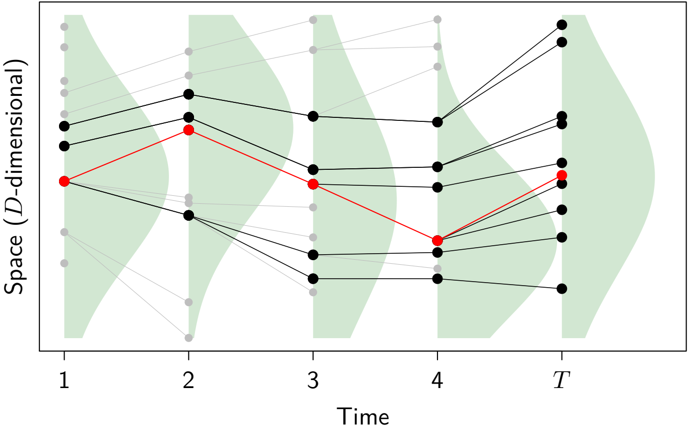
Problem formulation

Background: CSMC algorithms (D fixed)

Breakdown of CSMC as $D \to \infty$

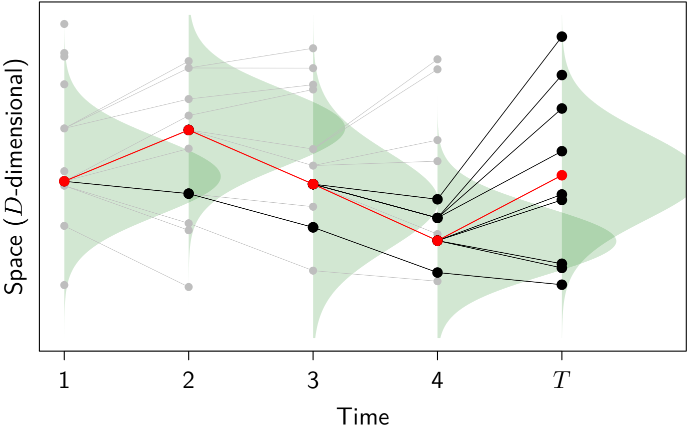
Novel 'random-walk' CSMC algorithm

CSMC in high dimensions ($D \to \infty$)



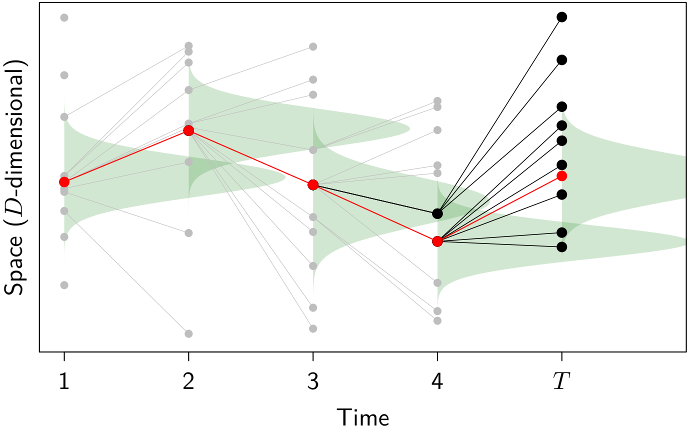
- all acceptance rates go to 0
- even with backward sampling.

CSMC in high dimensions ($D \to \infty$)



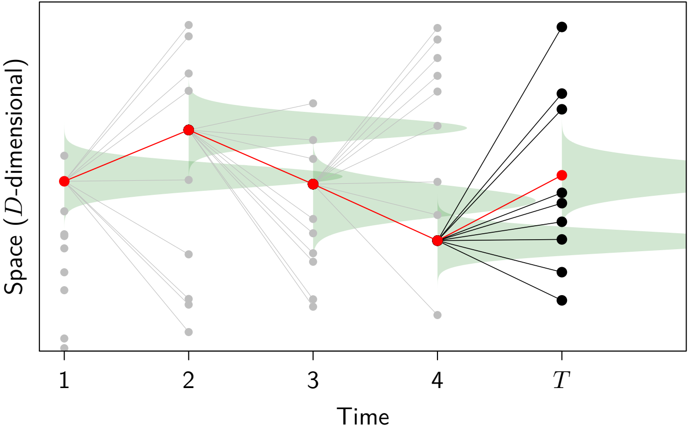
- all acceptance rates go to 0
- even with backward sampling.

CSMC in high dimensions ($D \to \infty$)



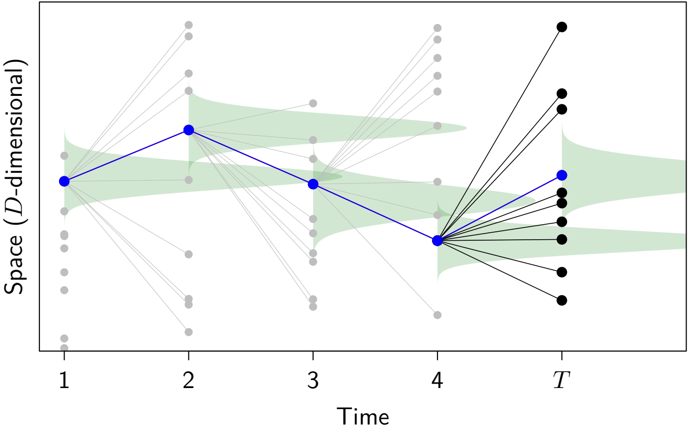
- all acceptance rates go to 0
- even with backward sampling.

CSMC in high dimensions ($D \to \infty$)



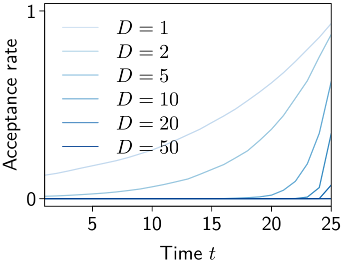
- all acceptance rates go to 0
- even with backward sampling.

CSMC in high dimensions ($D \to \infty$)

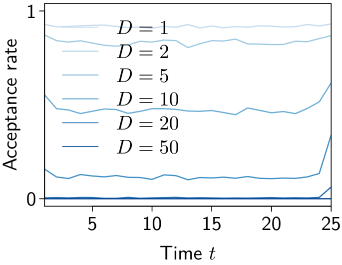


- all acceptance rates go to 0
- even with backward sampling.

Numerical illustration (state-space model)



CSMC



CSMC + backward sampling

Problem formulation

Background: CSMC algorithms (D fixed)

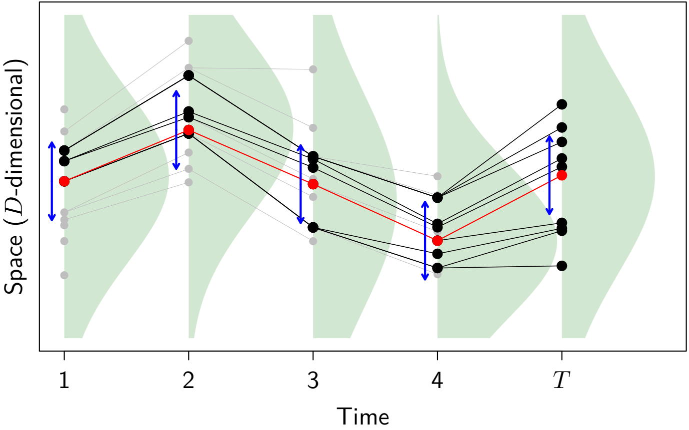
Breakdown of CSMC as $D \to \infty$

Novel 'random-walk' CSMC algorithm

Simple case: $T = N = 1$

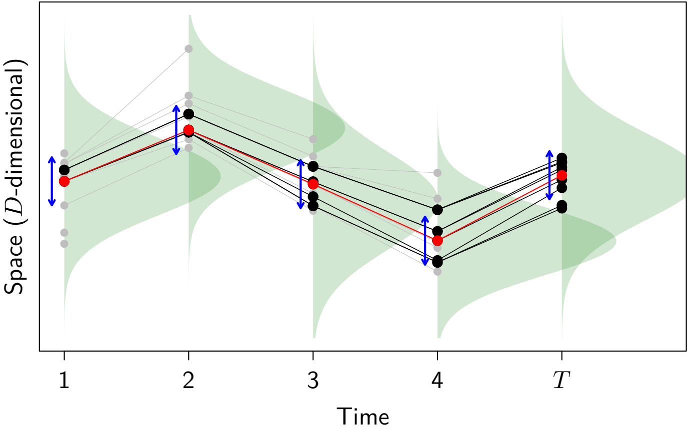
- CSMC reduces to Independent Metropolis–Hastings,
- **Problem:** 'global' proposals are difficult to design;
  - acceptance rate is typically $\mathrm{O}(\mathrm{e}^{-D})$.
- **Remedy:** suitably scaled 'local' proposals
  - e.g. random walk with variance σ^2/D,
  - stabilises acceptance rate as $D \to \infty$.
- **'No free lunch':** need $\mathrm{O}(D)$ iterations
  - non-trivial (diffusion) limit (Roberts et al., 1997).
- Extension to $N > 1$ proposals in Bédard et al. (2012).

Random-walk CSMC algorithm ($D \to \infty$)



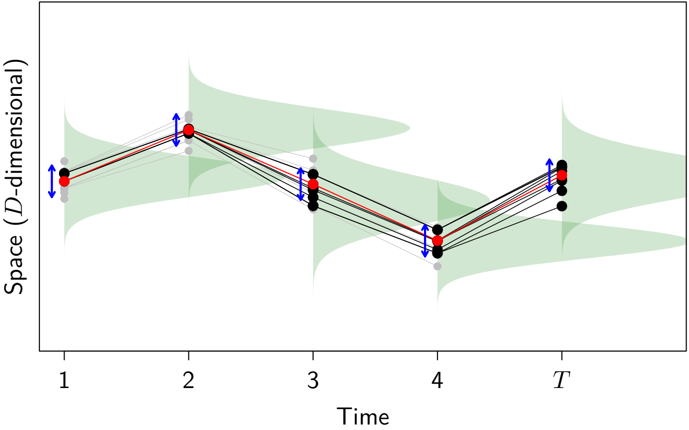
- (marginal) proposal : $\mathrm{Normal}(\mathbf{x}_t, \sigma_t^2/D)^{\otimes D}$,
- acceptance rates converge to non-trivial limits

Random-walk CSMC algorithm ($D \to \infty$)



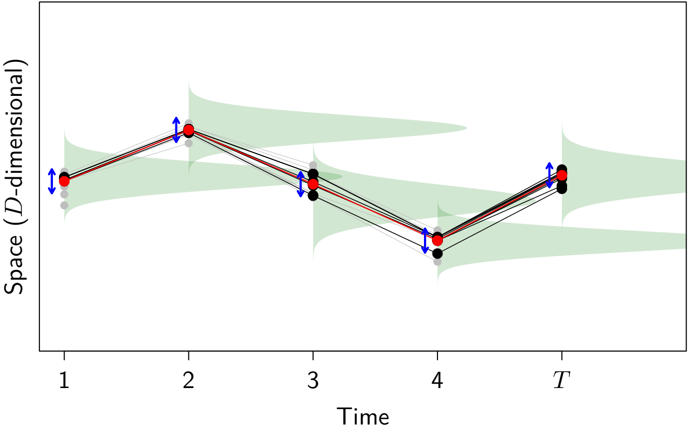
- (marginal) proposal : $\mathrm{Normal}(\mathbf{x}_t, \sigma_t^2/D)^{\otimes D}$,
- acceptance rates converge to non-trivial limits

Random-walk CSMC algorithm ($D \to \infty$)



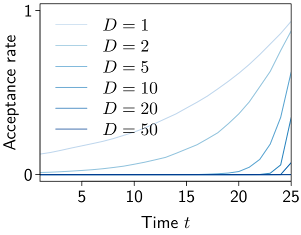
- (marginal) proposal : $\mathrm{Normal}(\mathbf{x}_t, \sigma_t^2/D)^{\otimes D}$,
- acceptance rates converge to non-trivial limits

Random-walk CSMC algorithm ($D \to \infty$)

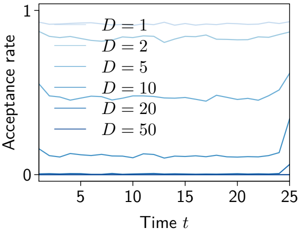


- (marginal) proposal : $\mathrm{Normal}(\mathbf{x}_t, \sigma_t^2/D)^{\otimes D}$,
- acceptance rates converge to non-trivial limits

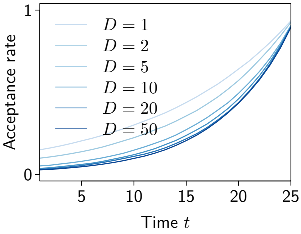
Numerical illustration (state-space model), ctd



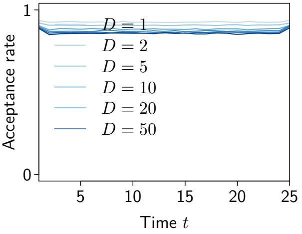
CSMC



CSMC + backward sampling

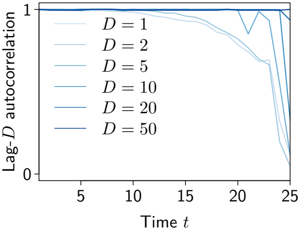


Random-walk CSMC

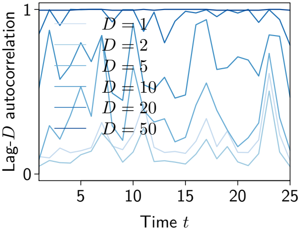


Random-walk CSMC
+ backward sampling

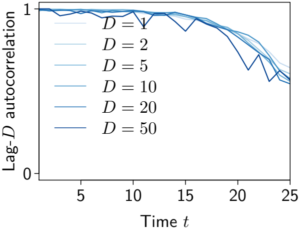
Numerical illustration (state-space model), ctd



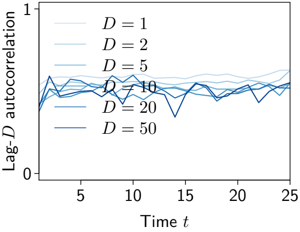
CSMC



CSMC + backward sampling



Random-walk CSMC



Random-walk CSMC
+ backward sampling

Literature I

Andrieu, C., Lee, A., and Vihola, M. (2018). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli*, 24(2):842–872.

Bédard, M., Douc, R., and Moulines, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Processes and their Applications*, 122(3):758–786.

Koskela, J., Jenkins, P. A., Johansen, A. M., and Spano, D. (2018). Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo. *ArXiv e-prints*, 1804.01811.

Lee, A., Singh, S. S., and Vihola, M. (2018). Coupled conditional backward sampling particle filter. *ArXiv e-prints*, 1806.05852.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.