Communication Efficient Sequential Monte Carlo

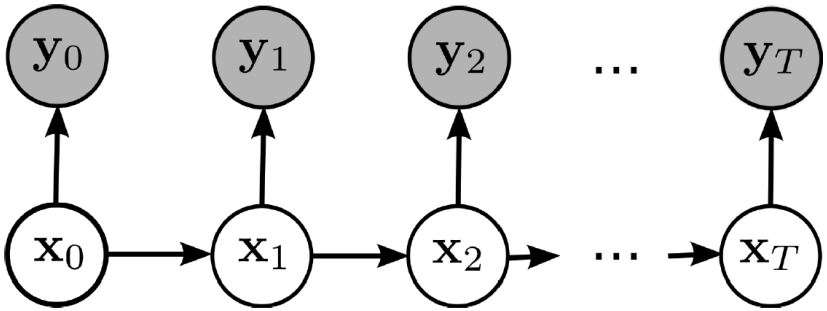
Deborshee Sen
Department of Statistics and Applied Probability,
National University of Singapore

deborshee.sen@u.nus.edu

SMC Workshop, Uppsala (2017)

Joint work with Alexandre Thiery of NUS

Hidden Markov model



- *Latent* process $\{X_t\}_{t\geq 0}$ on $(\mathsf{X}, \mathcal{X})$;
  – A Markov chain.
- *Observed* process $\{Y_t\}_{t\geq 0}$ on $(\mathsf{Y}, \mathcal{Y})$.

Interested in:

1. $P(x_T \mid y_{0:T-1}) \equiv \pi_T(x_T)$
   – known as *predictive density*.
2. $\gamma_T = \int_{\mathsf{X}^T} p(y_{0:T-1} \mid x_{0:T-1})\, p(x_{0:T-1})\, \mathsf{d}x_{0:T-1}$, $\gamma_0 \equiv 1$
   – *likelihood* of first T observations $y_{0:T-1}$.

Bootstrap particle filter

Hidden Markov model dynamics:

$$X_0 \sim \pi_0(\cdot)$$
$$X_t|\{X_{t-1} = x_{t-1}\} \sim K_t(x_{t-1}, \cdot), \quad t \geq 1,$$
$$Y_t|\{X_t = x_t\} \sim g_t(x_t, \cdot), \quad t \geq 0.$$

- Bootstrap particle filter generates samples $\{\widetilde{X}_{0,i}; 1 \leq i \leq N\}$ from π_0;
  – **Local** operation.
- Assigns *weights* $w_{t,i} \propto g_t(\widetilde{X}_{t,i}, y_t)$, $\sum_{i=1}^N w_{t,i} = 1$;
  – **Local** operation.
- *Resamples* $\{\widetilde{X}_{t,i}; 1 \leq i \leq N\}$ according to $\{w_{t,i}; 1 \leq i \leq N\}$ to get $\{X_{t,i}; 1 \leq i \leq N\}$;
  – **Global** operation.
- *Mutates* $X_{t,i}$ according to $K_{t+1}(X_{t,i}, \cdot)$: $\widetilde{X}_{t+1,i} \sim K_{t+1}(X_{t,i}, \cdot)$;
  – **Local** operation.

Resampling

- Resampling consists of first choosing *ancestors* $\{a_{t,i}; 1 \leq i \leq N\}$ and then setting $X_{t,i} = \widetilde{X}_{t,a_{t,i}}$ such that

$$\mathrm{E}\left\{\frac{1}{N}\sum_{i=1}^{N}\varphi(X_{t,i})\right\} = \sum_{i=1}^{N} w_{t,i}\,\varphi(\widetilde{X}_{t,i}) \tag{1}$$

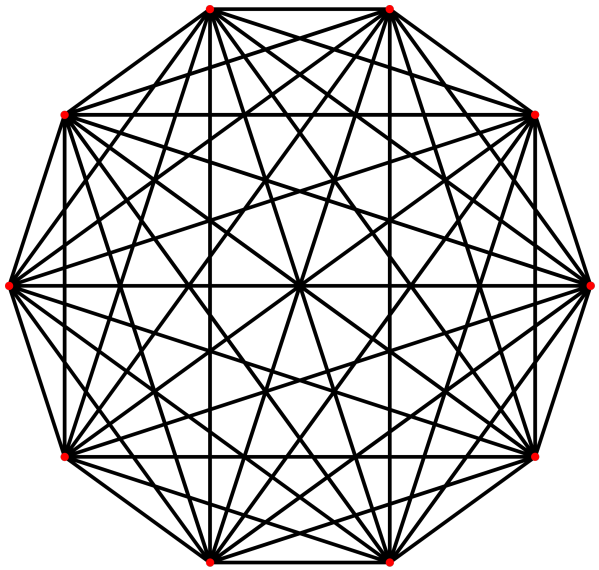
  for any function φ for which the expectation is finite.
- There are many ways of choosing $\{a_{t,i}; 1 \leq i \leq N\}$ such that condition (1) is satisfied, e.g., multinomial resampling, residual resampling, systematic resampling.

Bottleneck in parallelising: resampling

Question How to parallelise the resampling step?

Solution Reduce interactions between particles.

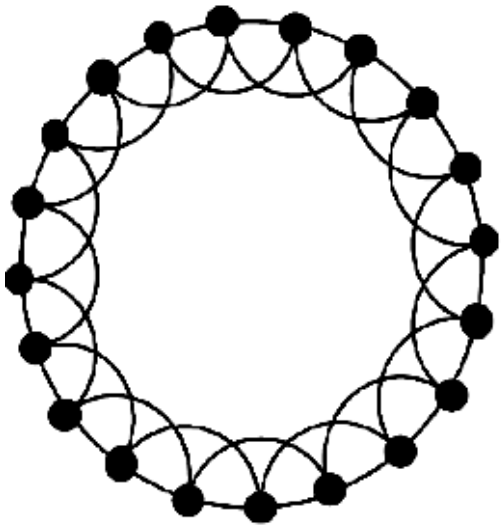
- Bootstrap particle filter has interactions between all particles.



Important consideration Bootstrap particle filter is **stable**.

Local exchange particle filter

- *Heine & Whiteley (2016)* considers a particle filter with a "local exchange" mechanism.



- They call it the *local exchange particle filter*.
- They prove a central limit theorem as the network size $\to \infty$.
- They provide counter-examples for stability.

αSMC

- Proposed by *Whiteley, Lee & Heine (2016).*
- Interactions between particles is controlled by a sequence of Markov transition matrices $\{\alpha_t\}_{t \geq 0}$.
- Each α_t can be interpreted as a (weighted) network of connections.
- Particles resample locally from among neighbours they are connected to in the network.
- αSMC provides an estimate $\widehat{\pi}_T^N$ of π_T and $\widehat{\gamma}_T^N$ of γ_T.
  – N denotes the number of particles used.

αSMC

- Let α_t^{ij} denote the (i,j)-th element of the matrix α_t.

Algorithm

<u>At $t = 0$:</u>

Set $W_{0,i} = 1$ and sample $X_{0,i} \sim \pi_0$.

<u>At $t \geq 1$:</u>

Set $W_{t,i} = \sum_{j=1}^{N} \alpha_{t-1}^{ij} W_{t-1,j}\, g_{t-1}(X_{t-1,j})$.

Sample

$$X_{t,i} \sim \frac{1}{W_{t,i}} \sum_{j=1}^{N} \alpha_{t-1}^{ij} W_{t-1,j}\, g_{t-1}(X_{t-1,j})\, K_t(X_{t-1,j}, \cdot).$$

end

$$\widehat{\pi}_T^N(\varphi) = \sum_{i=1}^{N} \frac{W_{T,i}}{\sum_{j=1}^{N} W_{T,j}} \varphi(X_{T,i}) \quad \text{and} \quad \widehat{\gamma}_T^N = \frac{1}{N} \sum_{i=1}^{N} W_{T,i}.$$

Instances of αSMC

Sequential importance sampling

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Local exchange particle filter

$$\begin{pmatrix} 1/3 & 1/3 & 0 & 0 & \cdots & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1/3 & 0 & 0 & 0 & \cdots & 1/3 & 1/3 \end{pmatrix}$$

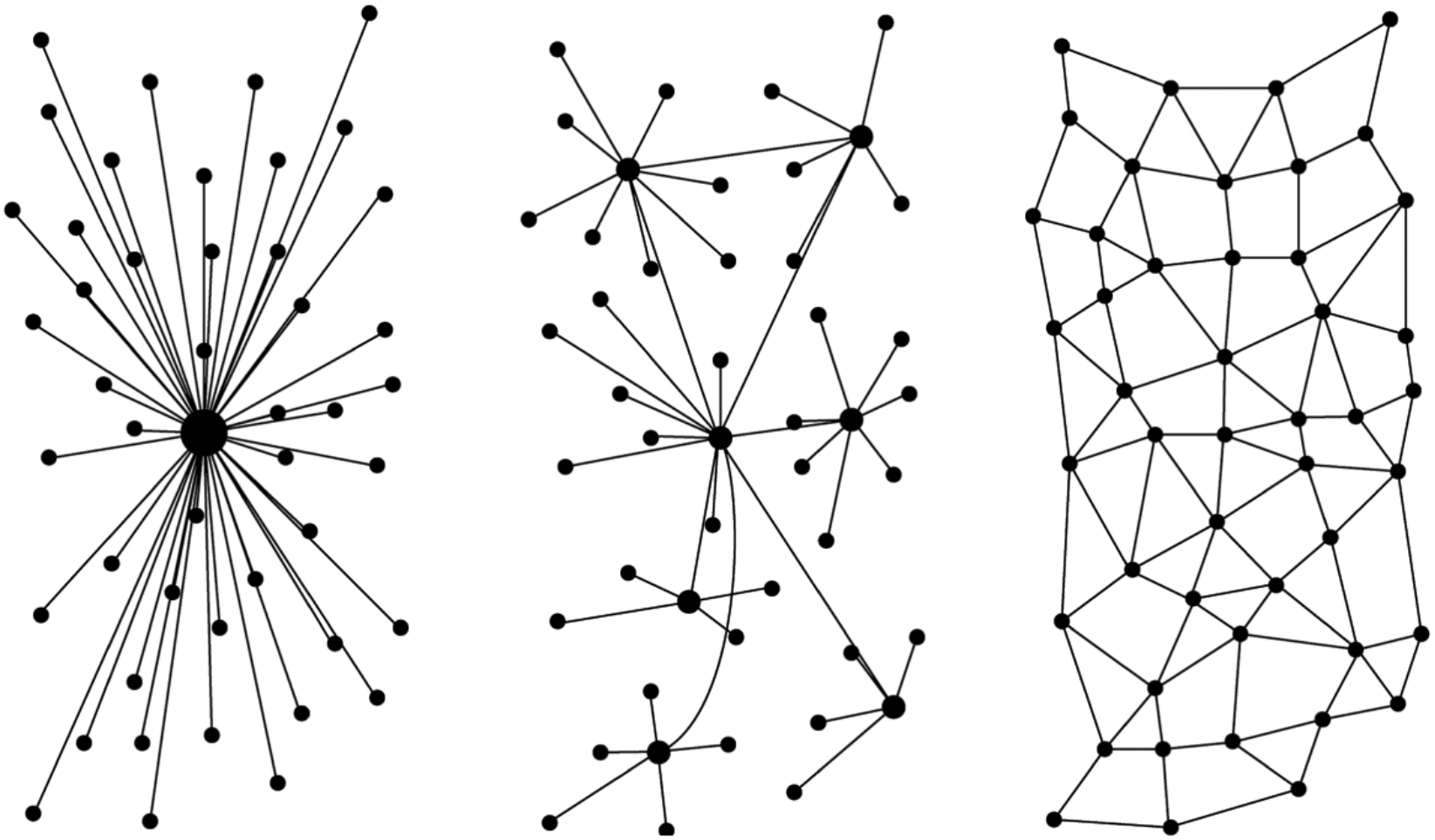
Bootstrap particle filter

$$\begin{pmatrix} 1/N & 1/N & \cdots & 1/N \\ 1/N & 1/N & \cdots & 1/N \\ 1/N & 1/N & \cdots & 1/N \\ \vdots & \vdots & \ddots & \vdots \\ 1/N & 1/N & \cdots & 1/N \end{pmatrix}$$

In general

$$\begin{pmatrix} 0 & \square & 0 & \square & \cdots & 0 \\ \square & 0 & \square & 0 & \cdots & \square \\ 0 & 0 & \square & \square & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \square & 0 & 0 & \square & \cdots & \square \end{pmatrix}$$

Network architectures



Centralised (bootstrap particle filter)

Decentralised (island particle filter)

Distributed (αSMC)

Choosing networks

- *Whiteley, Lee & Heine (2016)* recognises that algorithmically controlling the effective sample size (ESS) leads to time-uniform convergence.
- They introduce schemes for adaptively choosing $\{\alpha_t\}_{t\geq 0}$ such that the resulting αSMC is stable.
- *Lee & Whiteley (2016)* extends the work of *Whiteley, Lee & Heine (2016)*.
- They show how graphs that model the interaction among particles can be induced by tree data structures which model the network topology of a distributed computing environment.
- They present efficient distributed algorithms that are stable and operate on forests associated with these trees.

Random connections

Assume that there exist constants $\kappa_K < \infty$ and $\kappa_g < \infty$ such that

$$\kappa_K^{-1} \leq K_t \leq \kappa_K \quad \text{and} \quad \kappa_g^{-1} \leq g_t \leq \kappa_g \quad \forall\, t.$$

- Consider αSMC with N particles where each particle connects **randomly** to C_N other particles at each time.

Theorem (Sen-T (2017))

If $C_N \to \infty$ as $N \to \infty$ and connections are chosen randomly,

$$\lim_{N\to\infty} \text{var}\left\{\sqrt{N}\,\widehat{\gamma}_T^N\right\} = \lim_{N\to\infty} \text{var}\left\{\sqrt{N}\,\widehat{\gamma}_{T,\text{bootstrap}}^N\right\},$$

$$\lim_{N\to\infty} \text{var}\left\{\sqrt{N}\,\widehat{\pi}_T^N(\varphi)\right\} = \lim_{N\to\infty} \text{var}\left\{\sqrt{N}\,\widehat{\pi}_{T,\text{bootstrap}}^N(\varphi)\right\}.$$

- If $C_N \to \infty$, there is **no asymptotic cost** of αSMC over the bootstrap particle filter!

Random connections

Theorem (Sen-T (2017))

If $C_N \equiv C$ and connections are chosen randomly,

$$\lim_{N\to\infty} \mathrm{var}\left\{\sqrt{N}\,\widehat{\gamma}_T^N\right\} = \lim_{N\to\infty} \mathrm{var}\left\{\sqrt{N}\,\widehat{\gamma}_{T,\mathrm{bootstrap}}^N\right\} + \frac{Cst_T^{\gamma}}{C} + \mathcal{O}\left(\frac{1}{C^2}\right),$$

$$\lim_{N\to\infty} \mathrm{var}\left\{\sqrt{N}\,\widehat{\pi}_T^N(\varphi)\right\} = \lim_{N\to\infty} \mathrm{var}\left\{\sqrt{N}\,\widehat{\pi}_{T,\mathrm{bootstrap}}^N(\varphi)\right\} + \frac{Cst_T^{\pi}}{C} + \mathcal{O}\left(\frac{1}{C^2}\right).$$

- If $C_N \equiv$ C, the **asymptotic cost** of αSMC over the bootstrap particle filter is $\mathcal{O}(1/C)$.

Random connections

Theorem (Sen-T (2017))

The asymptotic variance of αSMC is **stable** *over time if* $C_N > \kappa_g^4 \, \kappa_K$

$$\sup_{T \geq 0} \lim_{N \to \infty} \operatorname{var}\left\{ \sqrt{N} \, \widehat{\pi}_T^N(\varphi) \right\} < \infty.$$

Spectral gap

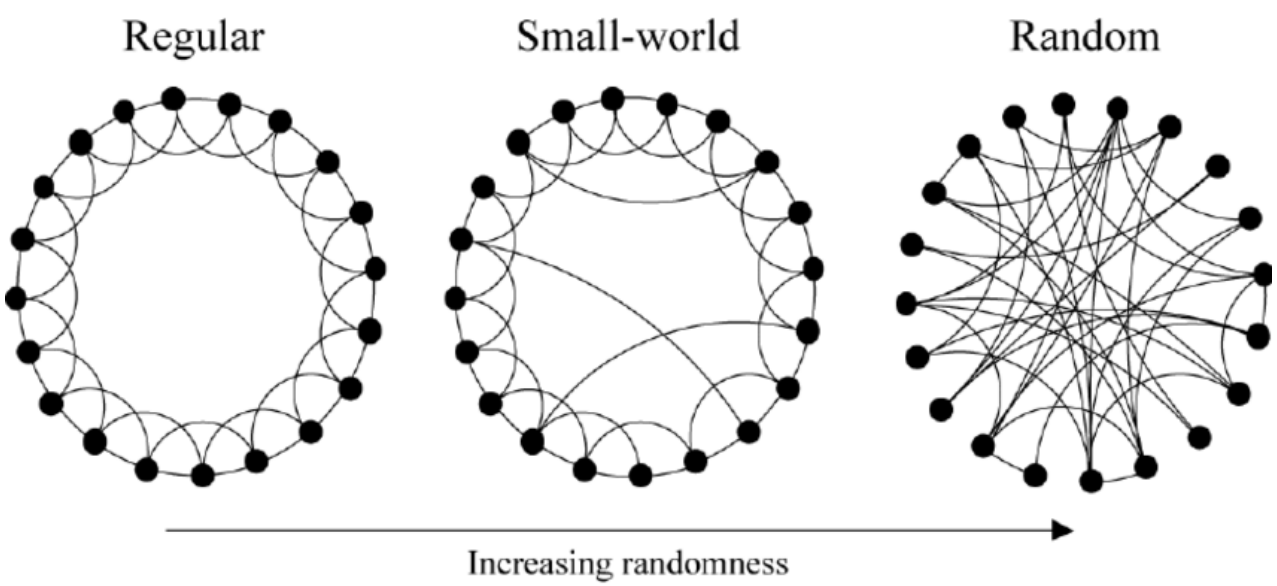
- Eigenvalues of Markov transition matrix $\alpha \in \mathbb{R}^{N \times N}$

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > -1.$$

- The absolute spectral gap of α is

$$\mathrm{Gap}(\alpha) := 1 - \max_{i=2}^{N} |\lambda_i|.$$

- This quantifies how rapidly information can flow in the network.



Examples of spectral gaps

- For sequential importance sampling, $\text{Gap}(\alpha) \equiv 0$.
  – Not stable.

- For the local exchange particle filter, $\text{Gap}(\alpha) \to 0$ as $N \to \infty$.
  – Not necessarily stable.

- For the bootstrap particle filter, $\text{Gap}(\alpha) \equiv 1$.
  – Stable.

Question Is there a threshold for the spectral gap beyond which αSMC is stable?

Stability of αSMC

- Consider a fixed doubly-stochastic matrix α.

Theorem (Sen-T (2017))

The αSMC algorithm is stable in time as soon as the absolute spectral gap is large enough

$$Gap(\alpha) > 1 - \frac{1}{\kappa_g^2}.$$

- As long as the absolute spectral gap is large enough,

$$\sup \left\{ \sqrt{N} \times \left|\left|\left| \widehat{\pi}_T^N - \pi_T \right|\right|\right| \, : \, T \geq 0 \right\} < \infty,$$

where

$$\left|\left|\left| \mu - \nu \right|\right|\right|^2 \equiv \sup \left\{ \mathrm{E} \left[\left(\mu(\varphi) - \nu(\varphi) \right)^2 \right] \, : \, \varphi \in \mathcal{B}(\mathsf{X}) \right\}.$$

- $\mathcal{B}(\mathsf{X})$ denotes the set of all bounded functions on $(\mathsf{X}, \mathcal{X})$.

Graphs with large spectral gaps

- Ramanujan graphs are C-regular graphs on N vertices for which

$$\mathrm{Gap}(\alpha) \geq 1 - \frac{2\sqrt{C-1}}{C}.$$

- Alon-Friedman: with high probability, the absolute spectral gap of a random walk on a C-regular graph with N vertices is large,

$$\mathrm{Gap}(\alpha) \geq_{\mathrm{P}} 1 - \frac{2}{\sqrt{C}} - \frac{\varepsilon}{C}$$

  for any $\varepsilon > 0$.

- This gives an algorithmic way of creating large C-regular graphs with large spectral gaps (i.e., expander graphs).

Spectral gaps of random graphs

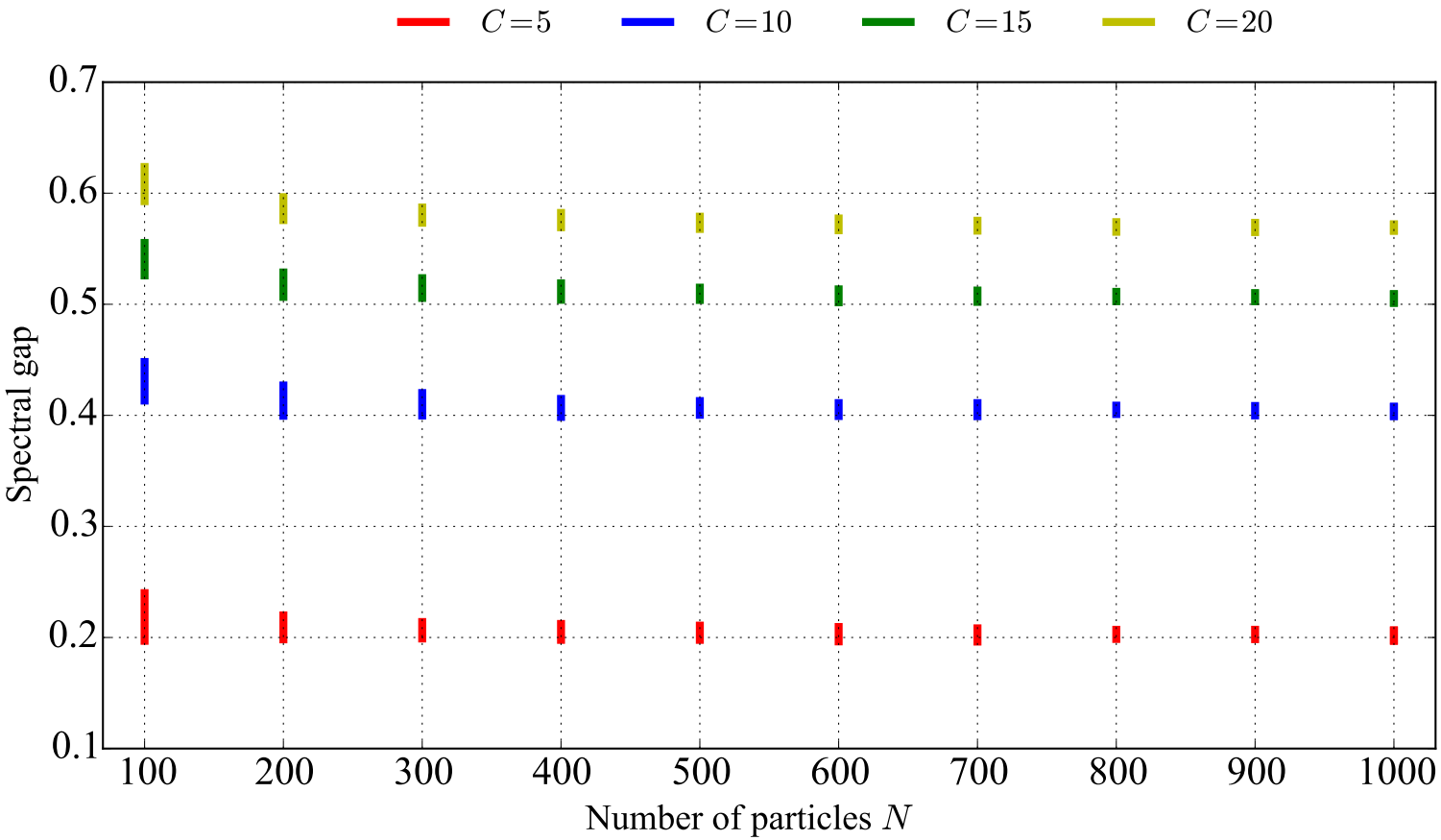


Figure: 90% confidence interval for adjacency matrices of spectral gaps of random C-regular graphs on N vertices.

Numerical example

- Consider the hidden Markov model given by

$$\begin{aligned} X_{t+1} &= -\frac{1}{2}(X_t - 1) + \mathcal{N}(0,1), \quad X_0 = 0, \\ Y_t &= X_t + \mathcal{N}(0, 0.2^2). \end{aligned}$$

- We compare the following:

(a) Local exchange particle filter.

(a) αSMC with a random C-regular graph.

– A random α matrix is chosen at time 0 and then fixed thereafter.

(b) αSMC when connections are chosen randomly at each time.

- The bootstrap particle filter is used as a benchmark.

Performance w.r.t. connectivity strength

- $N = 2 \times 10^3$, $T = 2 \times 10^2$.

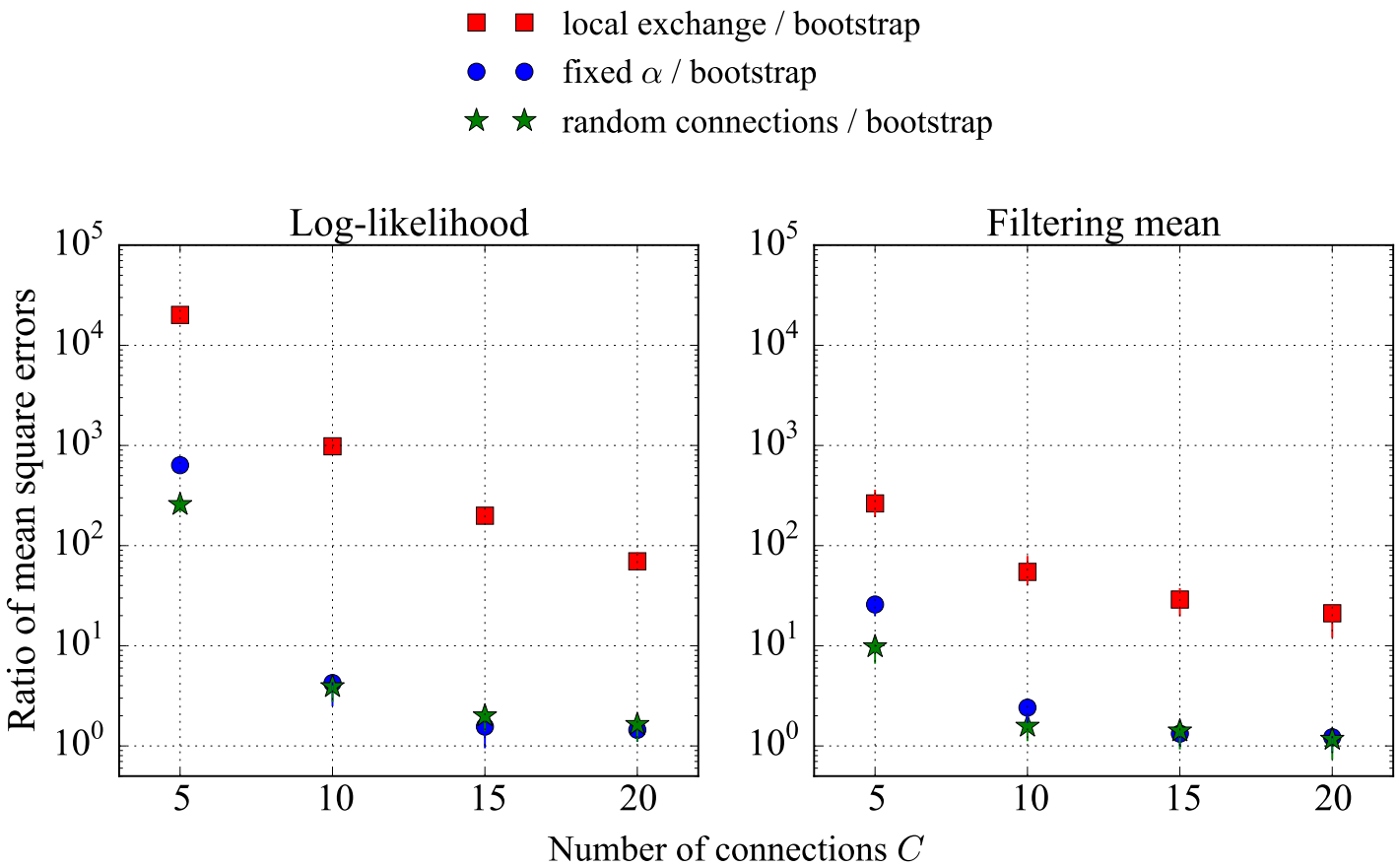


Figure: Relative performances with respect to the bootstrap particle filter of using (a) a local exchange mechanism, (b) a random C-regular graph, and (c) random connections at each time.

Performance w.r.t. network size

- $T = 2 \times 10^2$.

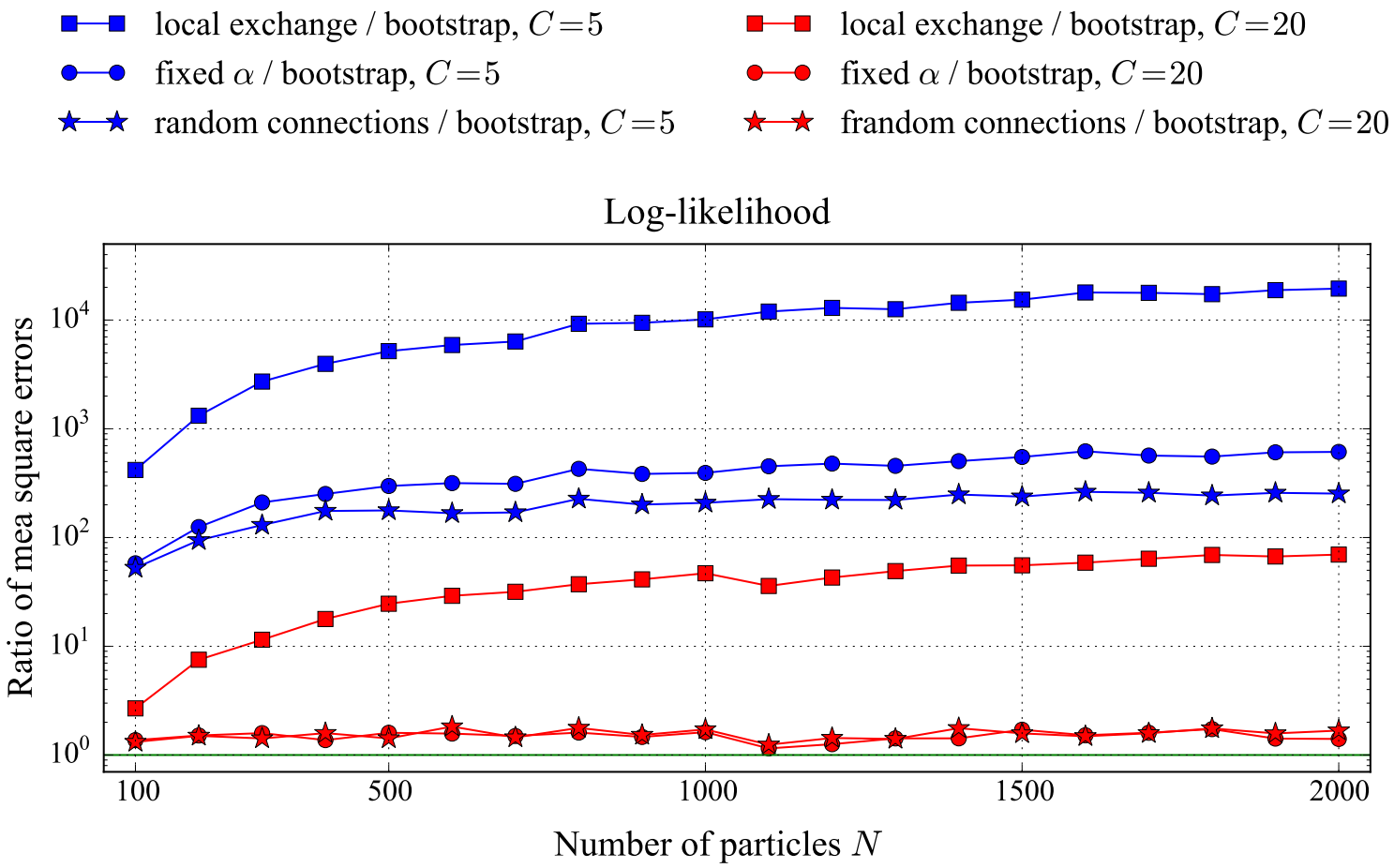


Figure: Relative performances of αSMC with respect to the bootstrap particle filter.

Performance w.r.t. network size

- $T = 2 \times 10^2$.

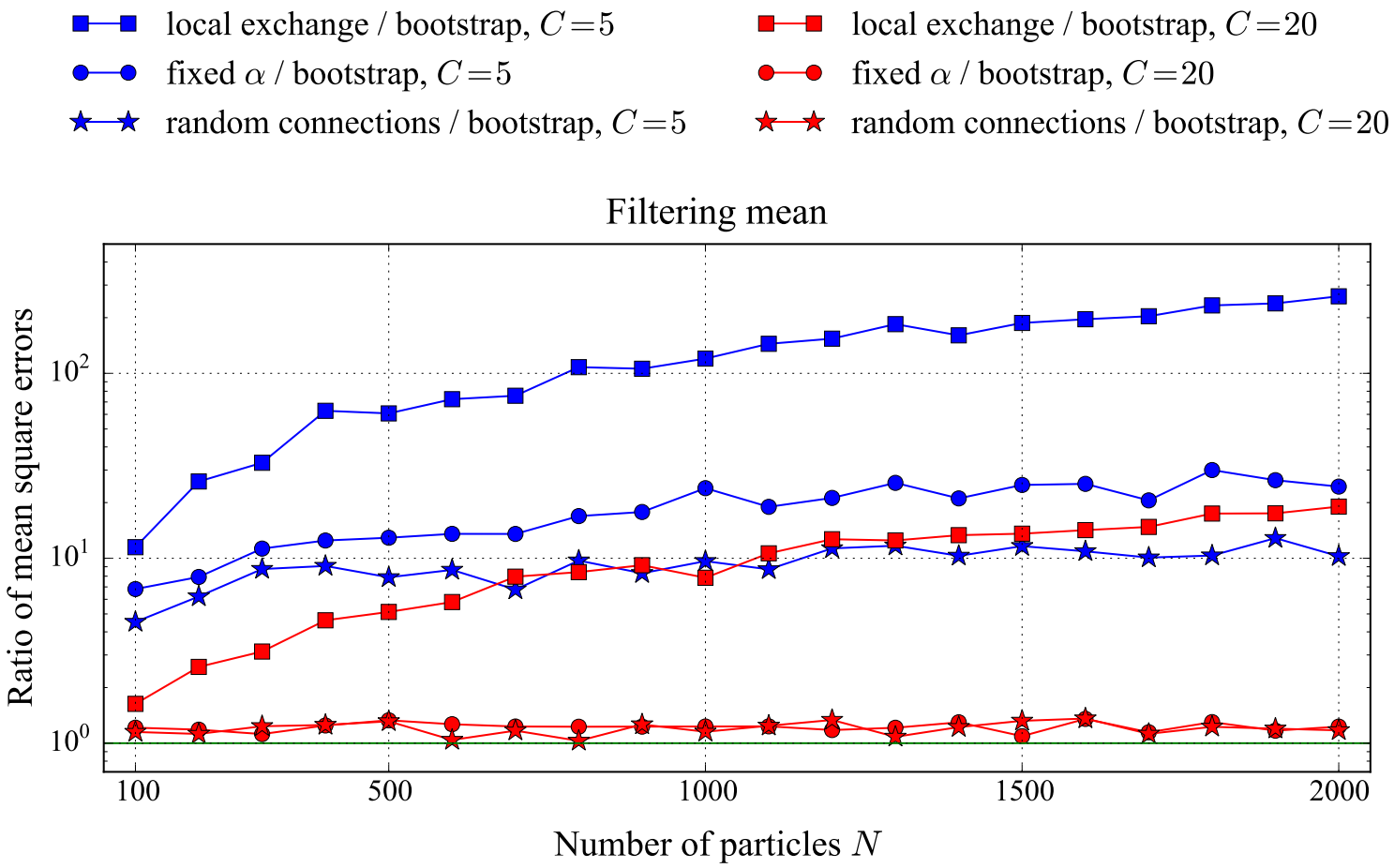


Figure: Relative performances of αSMC with respect to the bootstrap particle filter.

Performance over time

- $N = 10^3$, $C = 10$.

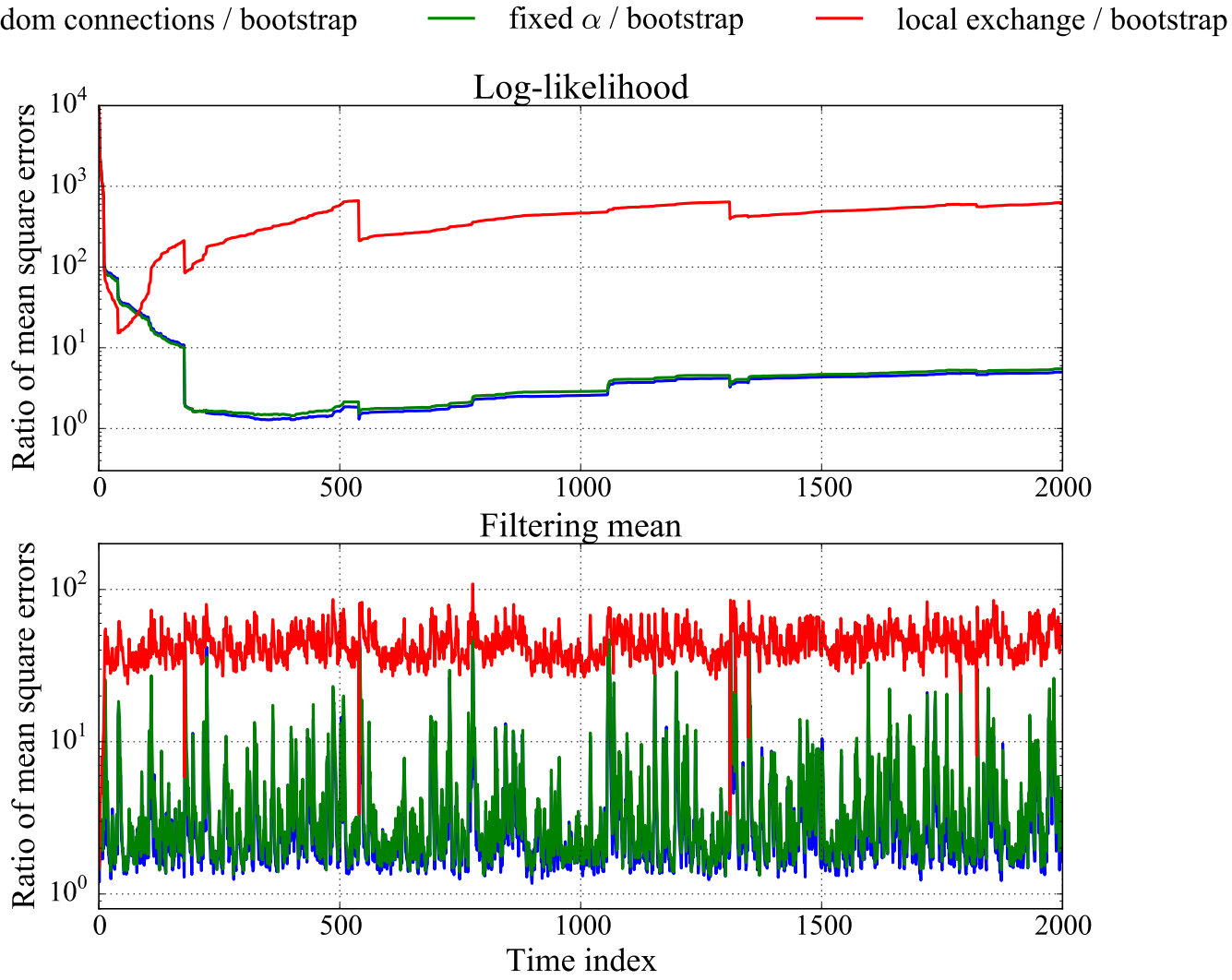


Figure: Relative performances of αSMC with respect to the bootstrap particle filter.

Performance over time

- $N = 10^3$, $C = 30$.

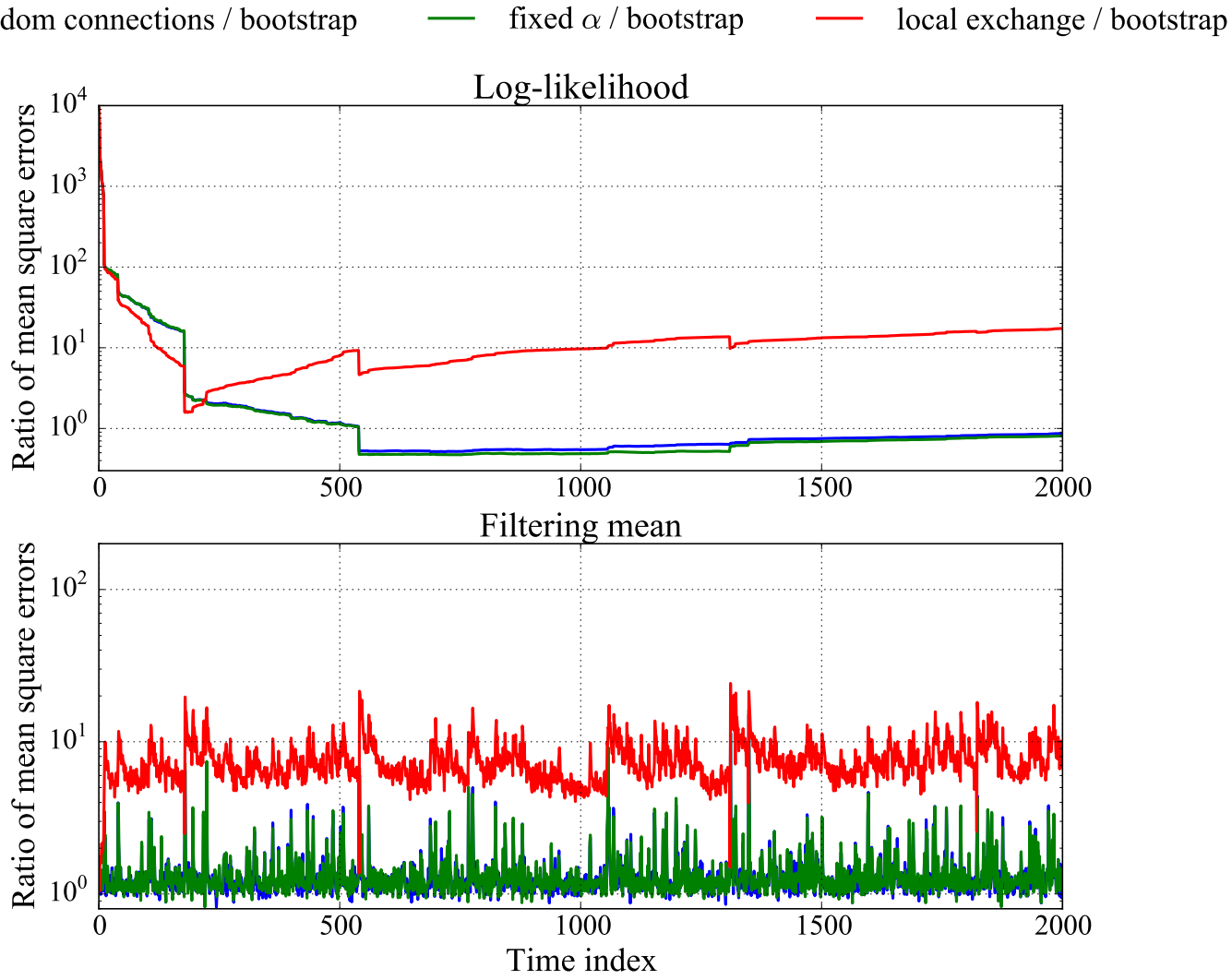


Figure: Relative performances of αSMC with respect to the bootstrap particle filter.

Conclusions

- The resampling step is the main bottleneck in parallelising particle filters.
- Reducing interactions between particles is one way to speed things up.
- Controlling the absolute spectral gap of the communication network is enough to guarantee time-uniformly stable algorithms.
- This can be done using Ramanujan graphs and expander graphs.
- If connections between particles are random, the resulting algorithm is asymptotically equivalent to the bootstrap particle filter if the number of connections increases to infinity with the number of particles.
- If the number of random connections is fixed at C, the extra asymptotic variance over the bootstrap particle filter is $\mathcal{O}(1/C)$.

References

αSMC:

1. Heine, K., & Whiteley, N. (2016). "Fluctuations, stability and instability of a distributed particle filter with local exchange." *Stochastic Processes and their Applications.*
2. Lee, A., & Whiteley, N. (2016). "Forest resampling for distributed sequential Monte Carlo." *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(4), 230-248.
3. Whiteley, N., Lee, A., & Heine, K. (2016). "On the role of interaction in sequential Monte Carlo algorithms." *Bernoulli*, 22(1), 494-529.

Spectral gaps:

1. Alon, N. "Eigenvalues and expanders." *Combinatorica* 6.2 (1986): 83-96.
2. Friedman, J. "A proof of Alon's second eigenvalue conjecture." *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing.* ACM, 2003.
3. Lubotzky, Alexander, Ralph Phillips, and Peter Sarnak. "Ramanujan graphs." *Combinatorica* 8.3 (1988): 261-277