SMC Methods for Graphical Models

Christian A. Naesseth

Joint work with F. Lindsten, A. M. Johansen, T. B. Schön, B. Kirkpatrick, J. Aston, A. Bouchard-Côté

LINKÖPINGS UNIVERSITET TEKNISKA HÖGSKOLAN

UPPSALA UNIVERSITET

Summary

- Probabilistic Graphical Models (PGM) – used in computer vision, localisation, error-correcting coding, speech recognition, computational biology, machine learning, etc.
- Cyclic and/or continuous PGMs in general makes inference non-trivial and approximate methods are needed.
- **Proposed Sequential Monte Carlo (SMC)-based methods:**
  - *"Standard" SMC* Exploits a sequential decomposition to construct the joint probability of the PGM using "standard" SMC methods.
  - *D&C SMC* New SMC algorithm based on a divide-and-conquer strategy.

Graphical Models

We consider models on the form,

$$p(X_{\mathcal{V}}) = \frac{1}{Z}\prod_{C\in\mathcal{C}} \psi_c(X_C),$$

where the graph $G = (\mathcal{V}, \mathcal{E})$ has vertex set $\mathcal{V} = \{x_1, \dots, x_{|\mathcal{V}|}\}$, edge set $\mathcal{E}$, cliques C and $Z = \int \prod_{C\in\mathcal{C}} \psi_c(X_C) dX_{\mathcal{V}}$ is the *partition function* (normalisation constant).

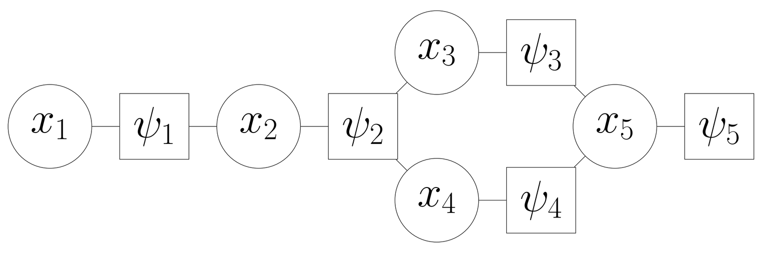

Figure 1: Example factor graph describing dependencies between random variables x in a PDF.

The construction applies to a variety of interesting problems! Factor graphs can in fact be used to represent both Bayesian networks (e.g. state-space models, hierarchical models) and Markov networks (e.g. Restricted Boltzmann Machines, Ising models, ...)

Sequential Decomposition

Sequential decomposition of a PGM is all about using structure encoded by factors in the graph to construct a valid sequence of target distributions for an SMC sampler.

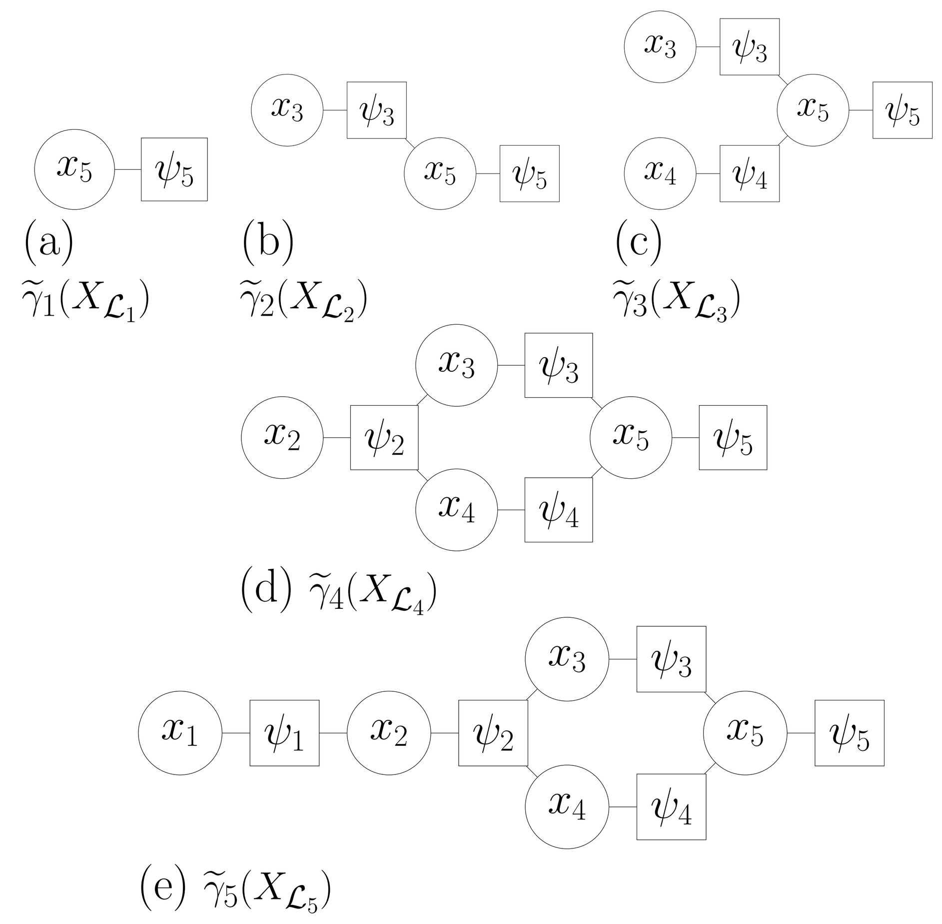

Figure 4: An example of one possible sequential decomposition of the factor graph in Figure 1.

Sequential Monte Carlo

A standard approach to approximate a sequence of distributions. Denote the newly added random variables ξ_k^i and all the added random variables at iteration k, $X_{\mathcal{L}_k} \subset \{x_1, \dots, x_{\mathcal{V}}\}$.

Algorithm 1 SMC for PGM

Perform each step for $i = 1, \dots, N$.

Sample $X_{\mathcal{L}_1}^i \sim r_1(\cdot)$. Set $w_1^i = W_1(X_{\mathcal{L}_1}^i)$.

$\widehat{Z}_1^N := \left(\frac{1}{N}\sum_{i=1}^N w_k^i\right)$.

for $k = 2$ **to** K **do**

Sample a_k^i according to resampling weights:

$\mathbb{P}(a_k^i = j) = \frac{\nu_{k-1}^j w_{k-1}^j}{\sum_l \nu_{k-1}^l w_{k-1}^l}, \quad j \in \{1, \dots, N\}$

Sample $\xi_k^i \sim r_k(\cdot | X_{\mathcal{L}_{k-1}}^{a_k^i})$ and set $X_{\mathcal{L}_k}^i = X_{\mathcal{L}_{k-1}}^{a_k^i} \cup \xi_k^i$. Set $w_k^i = W_k(X_{\mathcal{L}_k}^i)$.

$\widehat{Z}_k^N = \widehat{Z}_{k-1}^N \cdot \frac{1}{N}\sum_{i=1}^N \nu_k^i w_k^i$.

end for

Divide-and-Conquer SMC

New SMC method that utilises a divide-and-conquer strategy to approximate a distribution of interest. Multiple independent particle populations are resampled, merged and propagated as the method progresses on an auxiliary tree-structured decomposition of the PGM.

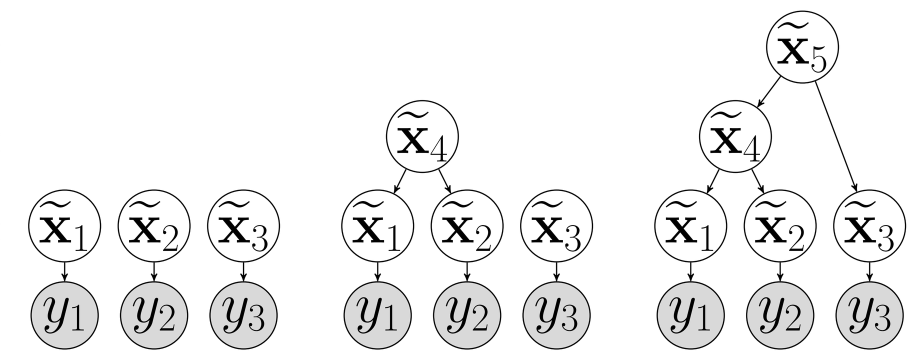

Figure 5: Decomposition of a hierarchical Bayesian model.

Algorithm 2 dc_smc(t)

1. For $c \in \mathcal{C}(t)$:
   (a) $((\mathbf{x}_c^i, \mathbf{w}_c^i)_{i=1}^N, \widehat{Z}_c^N) \leftarrow$ dc_smc(c).
   (b) Resample $(\mathbf{x}_c^i, \mathbf{w}_c^i)_{i=1}^N$ to obtain the equally weighted particle system $(\check{\mathbf{x}}_c^i, 1)_{i=1}^N$.
2. For particle $i = 1, \dots, N$:
   (a) Simulate $\widetilde{\mathbf{x}}_t^i \sim q_t(\cdot \mid \check{\mathbf{x}}_{c_1}^i, \dots, \check{\mathbf{x}}_{c_C}^i)$ from some proposal kernel on $\widetilde{\mathsf{X}}_t$, and where $(c_1, c_2, \dots, c_C) = \mathcal{C}(t)$.
   (b) Set $\mathbf{x}_t^i = (\check{\mathbf{x}}_{c_1}^i, \dots, \check{\mathbf{x}}_{c_C}^i, \widetilde{\mathbf{x}}_t^i)$.
   (c) Compute $\mathbf{w}_t^i = \frac{\gamma_t(\mathbf{x}_t^i)}{\prod_{c\in\mathcal{C}(t)} \gamma_c(\check{\mathbf{x}}_c^i)} \frac{1}{q_t(\widetilde{\mathbf{x}}_t^i \mid \check{\mathbf{x}}_{c_1}^i, \dots, \check{\mathbf{x}}_{c_C}^i)}$.
3. Compute $\widehat{Z}_t^N = \left\{\frac{1}{N}\sum_{i=1}^N \mathbf{w}_t^i\right\} \prod_{c\in\mathcal{C}(t)} \widehat{Z}_c^N$.
4. Return $((\mathbf{x}_t^i, \mathbf{w}_t^i)_{i=1}^N, \widehat{Z}_t^N)$.

Example: 64×64 Classical XY model ($\beta = 1.0$)

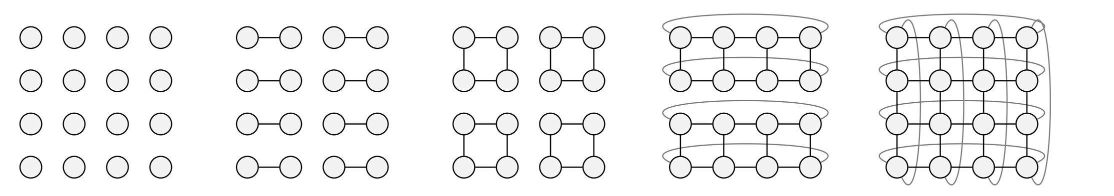

Figure 6: The disconnected components correspond to the groups of variables that are targeted by the different populations of the D&C SMC algorithm. At the final iteration, corresponding to the rightmost figure, we recover the original, connected model.

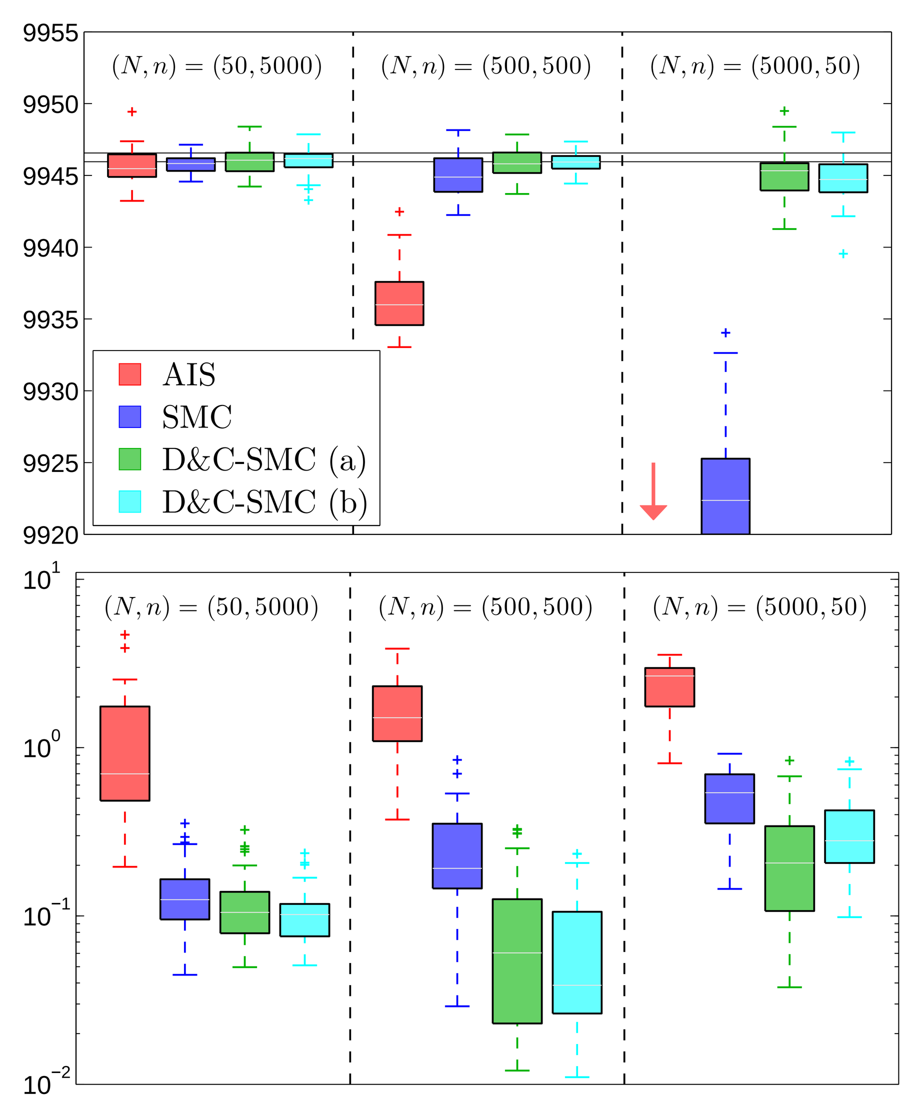

Figure 7: Box-plots over 50 runs of each sampler, using N particles and a total of n annealing steps. D&C SMC (b) uses fewer particles, $\bar{N} = 50$, 309, and 475 in the three columns (left to right), respectively, to match its computational cost to the other methods. *Top:* Estimates of $\log(Z)$. The horizontal lines correspond to min and max of 10 runs of AIS with $(N, n) = (10, 100\,000)$. *Bottom:* MSE (log-scale) for the estimates of $\mathbb{E}[x_k]$ (averaged over the grid).

Example: 16×16 ($\beta = 1.1$) and 64×64 Classical XY model ($\beta = \{0.5, 1.1\}$)

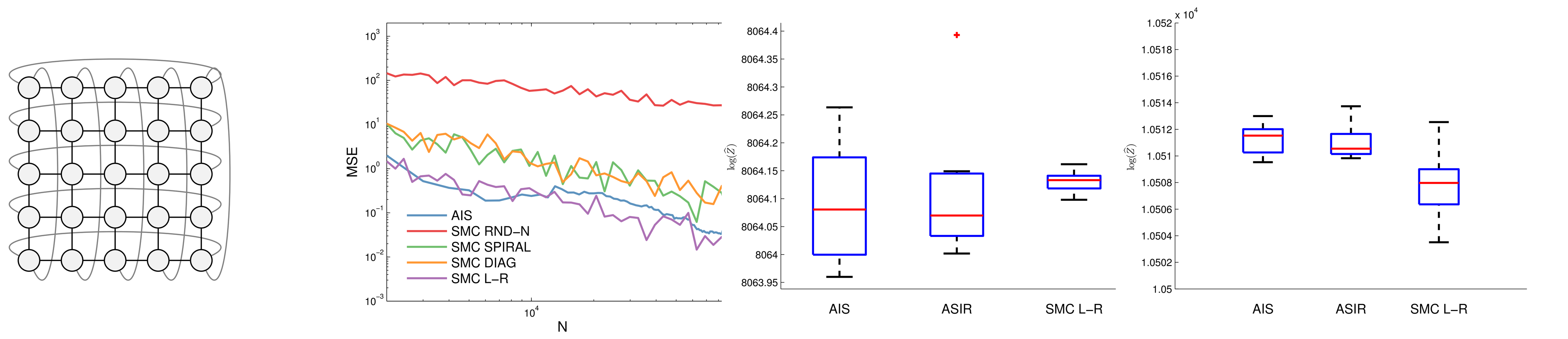

Figure 2: *Left:* XY model *Middle Left:* MSE in the estimates of $\log Z$ for AIS and four different orderings in SMC for PGM. *Middle Right, Right:* The logarithm of the estimated partition function for the 64×64 XY model with inverse temperature 0.5 (middle right) and 1.1 (right).

Example: Evaluation of LDA topic models (likelihood estimation)

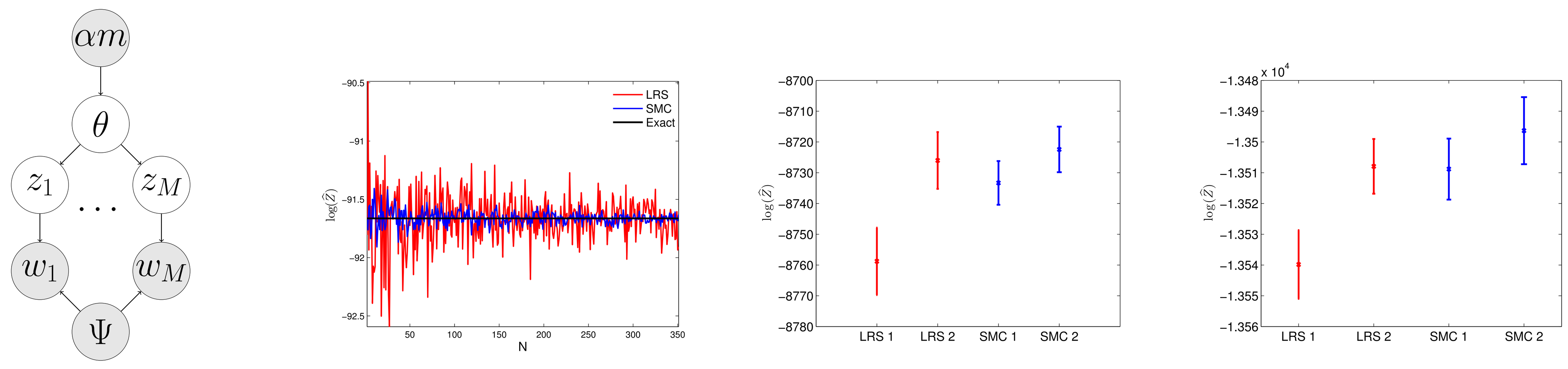

Figure 3: *Left:* LDA topic model. *Middle - Right:* Estimates of the log-likelihood of heldout documents for various datasets using SMC for PGM and Left-Right Sequential sampler.

[1] C. A. Naesseth, F. Lindsten and T. B. Schön, *Sequential Monte Carlo for Graphical Models*, arXiv.org, arXiv:1402.0330.

[2] F. Lindsten, A. M. Johansen, C. A. Naesseth, B. Kirkpatrick, T. B. Schön, J. Aston, A. Bouchard-Côté. *Divide-and-Conquer with Sequential Monte Carlo*, arXiv.org, arXiv:1406.4993.