Graphical model inference: Sequential Monte Carlo meets deterministic approximations

Fredrik Lindsten (LiU/UU), Jouni Helske (LiU), and Matti Vihola (JYU)

UPPSALA UNIVERSITET · LINKÖPING UNIVERSITY · JYVÄSKYLÄN YLIOPISTO / UNIVERSITY OF JYVÄSKYLÄ

Summary and contributions

Sequential Monte Carlo can be used for graphical model inference via **sequential graph decompositions** – we sample from a sequence of increasing subgraphs until we recover the full graph.

Can have poor performance if the implicit posteriors over subgraphs differ from the corresponding marginals of the full graph.

We propose to sample from a sequence of **twisted intermediate distributions:**

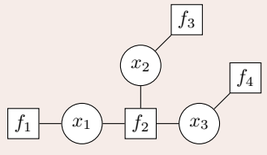
- Twisting can serve as a **look-ahead** – optimal twisting imply that we sample from the correct marginals at each iteration.
- Twisting functions can be computed using your favorite approximate inference method!
- Post-correction of the biases associated with this deterministic inference method.
- **Unbiased** normalizing constant estimates.

Probabilistic graphical models

Generic models for encoding dependencies in complex and high-dimensional statistical models.

Applications in spatial statistics, topic modeling, computational biology, etc.

Factor graph formulation



We consider **factor graphs** with joint distribution

$$\pi(x_{1:T}) = \frac{1}{Z} \prod_{j \in \mathcal{F}} f_j(x_{\mathcal{I}_j}).$$

Here,

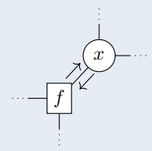
- $\mathcal{F}$ denotes the set of factors,
- $\mathcal{I} := \{1, \ldots, T\}$ is the set of variables,
- $\mathcal{I}_j = \mathrm{Ne}(j)$ is the index set of variables on which factor f_j depends.

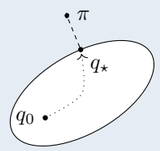
Task at hand

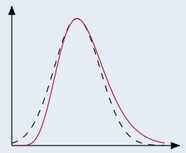
- Compute expectations w.r.t. $\pi(x_{1:T})$,
- Estimate the **normalizing constant** Z.

Deterministic inference methods

- Message passing
- Variational inference
- Laplace approximations
- …







Sequential Monte Carlo (SMC, [1])

SMC can be used to sample π based on a **sequential graph decomposition.**

Unnormalized target at iteration t is

$$\gamma_t(x_{1:t}) = \prod_{j \in \mathcal{F}_t} f_j(x_{\mathcal{I}_j})$$

for $t = 1, \ldots, T$. Here $\mathcal{F}_t$ is the set of factors depending only on $x_{1:t}$.

Algorithm. For $t = 1, \ldots, T$,

- Simulate particles, $x_t^i \sim q_t(x_t \mid x_{1:t-1}^i)$
- Compute weights $w_t^i \propto \omega_t(x_{1:t}^i)$ with,

$$\omega_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{\gamma_{t-1}(x_{1:t-1}) q_t(x_t \mid x_{1:t-1})}.$$

- Resample (if ESS is too low).

Twisted SMC

Introduce **twisting functions** ψ_t and define twisted intermediate targets as,

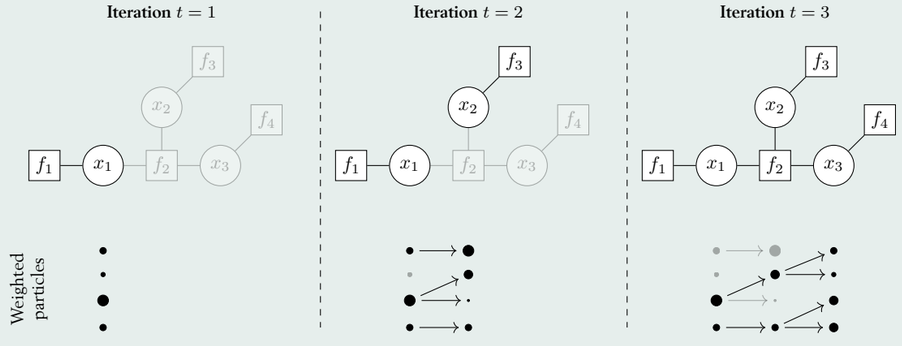
$$\gamma_t^\psi(x_{1:t}) := \psi_t(x_{1:t}) \gamma_t(x_{1:t}) = \psi_t(x_{1:t}) \prod_{j \in \mathcal{F}_t} f_j(x_{\mathcal{I}_j}).$$

Allows for **looking ahead** at future variables and dependencies.

Proposition (Optimal twisting). With

$$\psi_t^*(x_{1:t}) := \int \prod_{j \in \mathcal{F} \setminus \mathcal{F}_t} f_j(x_{\mathcal{I}_j}) dx_{t+1:T},$$

the SMC algorithm outputs i.i.d. draws from π and the normalizing constant estimate is exact; $\widehat{Z} = Z$ w.p.1.



How do we select the twisting functions?

Twisting functions via deterministic approximations

Finding optimal twisting functions boils down to solving the original inference problem!
Instead: use your favorite **deterministic inference method** to approximate them!

- Sub-optimality only affects efficiency (not consistency or unbiasedness) of SMC.
- The consecutive SMC stage can be seen as a post-correction of the biases.

Loopy belief propagation

Proposition. For a tree-structured graphical model:

$$\psi_t^*(x_{1:t}) = \prod_{j \in \mathcal{F} \setminus \mathcal{F}_t} \prod_{s \in \{1:t\} \cap \mathcal{I}_j} \mu_{j \to s}(x_s),$$

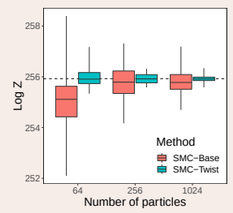
where $\mu_{j \to s}$ is the message from factor j to variable s.

Optimal for trees, but we can use the same expression for the twisting functions for arbitrary graphs, akin to loopy belief propagation.

ex) Ising model

Compute normalizing constant of 16×16 square lattice Ising model

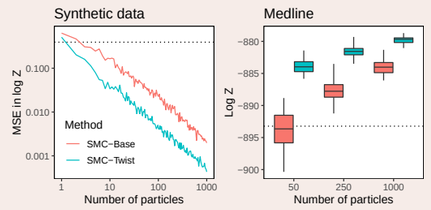
- $J_{ij} \equiv 0.44$,
- $H_i \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-1, 1)$,



$$\pi(x_{1:T}) = \frac{1}{Z} \exp\left(\sum_{(i,j) \in \mathcal{E}} J_{ij} x_i x_j + \sum_{i \in \mathcal{V}} H_i x_i \right),$$

ex) Topic model evaluation

Computing likelihood of held out documents for latent Dirichlet allocation model. Special purpose EP algorithm by [2] used to approximate the twisting functions. (Dotted line is plain EP estimate.)



Expectation propagation

Approximate the factors of the model:

$$\pi(x_{1:T}) \approx \frac{\prod_{j \in \mathcal{F}} \widetilde{f}_j(x_{\mathcal{I}_j})}{\int \prod_{j \in \mathcal{F}} \widetilde{f}_j(x_{\mathcal{I}_j}) dx_{1:T}}.$$

Iteratively update the factors so that

$$\widetilde{f}_j(x_{\mathcal{I}_j}) \widetilde{\pi}^{-j}(x_{1:T}) \approx f_j(x_{\mathcal{I}_j}) \widetilde{\pi}^{-j}(x_{1:T}).$$

Natural approximation of twisting functions:

$$\psi_t(x_{1:t}) = \int \prod_{j \in \mathcal{F} \setminus \mathcal{F}_t} \widetilde{f}_j(x_{\mathcal{I}_j}) dx_{t+1:T}.$$

Laplace approximation

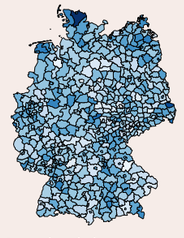
For a latent Gaussian Markov Random Field, with prior $p(x_{1:T}) = \mathcal{N}(x_{1:T} \mid \mu, Q^{-1})$ where $Q_{ij} \neq 0$ if and only if x_i and x_j share a factor, a Gaussian approximation of the posterior is obtained by Laplace's method.

Approximate twisting functions given by:

$$\psi_t(x_{1:t}) = \int \prod_{s=t+1}^{T} \{\widetilde{p}(y_s \mid x_s)\} p(x_{t+1:T} \mid x_{1:t}) dx_{t+1:T}.$$

ex) CAR–Binomial model

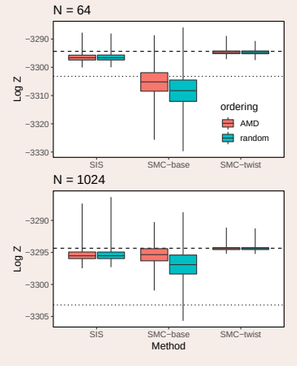
How to order the variables? We compare random ordering and approximate minimum degree (AMD) reordering.



Model:

- Precision: $Q_{tt} = 0.1 \times \{|\mathrm{Ne}(t)| + 1\}$ and $Q_{tt'} = -0.1$ if $t \sim t'$
- $y_t \sim \mathrm{Binomial}(10, \mathrm{logit}^{-1}(x_t))$
- Spatial structure $\sim$ regions in Germany
- $T = 544$

(Dashed line is true log-likelihood, dotted line is estimate from Laplace approximation.)



References

[1] C. A. Naesseth, F. Lindsten and T. B. Schön. Sequential Monte Carlo methods for graphical models. *Advances in Neural Information Processing Systems 27*, 2014, 1862-1870.

[2] T. Minka and J. Lafferty. Expectation-propagation for the Generative Aspect Model. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002.