Towards Maximum Likelihood: Learning Undirected Graphical Models (UGMs) using Persistent Sequential Monte Carlo

Hanchen Xiong, Sandor Szedmak, Justus Piater
UNIVERSITY OF INNSBRUCK, AUSTRIA

BACKGROUND

In general, we can define undirected graphical models (UGMs):

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{\exp\left(-E(\mathbf{x};\boldsymbol{\theta})\right)}{\mathbf{Z}(\boldsymbol{\theta})} \quad (1)$$

$$where \quad E(\mathbf{x};\boldsymbol{\theta}) = -\boldsymbol{\theta}^{\top}\phi(\mathbf{x}) \quad (2)$$

Methods for learning undirected graphical models by *approximating the gradient of maximum log-likelihood*:

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}|\mathcal{D})}{\partial \boldsymbol{\theta}} = \underbrace{\mathbb{E}_{\mathcal{D}}(\phi(\mathbf{x}))}_{\psi^{+}} - \underbrace{\mathbb{E}_{\boldsymbol{\theta}}(\phi(\mathbf{x}))}_{\psi^{-}} \quad (3)$$

- contrastive divergence
- persistent contrastive divergence
- tempered transition
- parallel tempering
- Markov chain Monte Carlo (MCMC) maximum likelihood

All these method can be summarized as a **Robbins-Monro's Stochastic Approximation Procedure (SAP)**.

INTERPRETING LEARNING AS SQUENTIAL MONTE CARLO

Algorithm 1 Interpreting Learning as SMC

Input: training data $\mathcal{D} = \{\mathbf{x}^{(m)}\}_{m=1}^{M}$; learning rate η
Initialize $p(\mathbf{x};\boldsymbol{\theta}_0)$, $t \leftarrow 0$
Sample particles $\{\bar{\mathbf{x}}_0^{(s)}\}_{s=1}^{S} \sim p(\mathbf{x};\boldsymbol{\theta}_0)$
while ! stop criterion **do**
  // importance reweighting
  Assign $w^{(s)} \leftarrow \frac{1}{S}, \forall s \in S$
  // resampling is ignored because it has no effect
  // MCMC transition
  switch (algorithmic choice)
  case CD:
    generate a brand new particle set $\{\bar{\mathbf{x}}_{t+1}^{(s)}\}_{s=1}^{S}$ with one step Gibbs sampling from $\mathcal{D}$
  case PCD:
    evolve particle set $\{\bar{\mathbf{x}}_{t}^{(s)}\}_{s=1}^{S}$ to $\{\bar{\mathbf{x}}_{t+1}^{(s)}\}_{s=1}^{S}$ with one step Gibbs sampling
  case Tempered Transition**:**
    evolve particle set $\{\bar{\mathbf{x}}_{t}^{(s)}\}_{s=1}^{S}$ to $\{\bar{\mathbf{x}}_{t+1}^{(s)}\}_{s=1}^{S}$ with tempered transition
  case Parallel Tempering**:**
    evolve particle set $\{\bar{\mathbf{x}}_{t}^{(s)}\}_{s=1}^{S}$ to $\{\bar{\mathbf{x}}_{t+1}^{(s)}\}_{s=1}^{S}$ with parallel tempering
  end switch
  // update distribution
  Compute the gradient $\Delta\boldsymbol{\theta}_t$ according to (3)
  $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \eta\Delta\boldsymbol{\theta}_t$
  $t \leftarrow t+1$
end while
Output: estimated parameters $\boldsymbol{\theta}^{*} = \boldsymbol{\theta}_t$

ROBBBINS-MONRO'S SAP

SAP Algorithm:

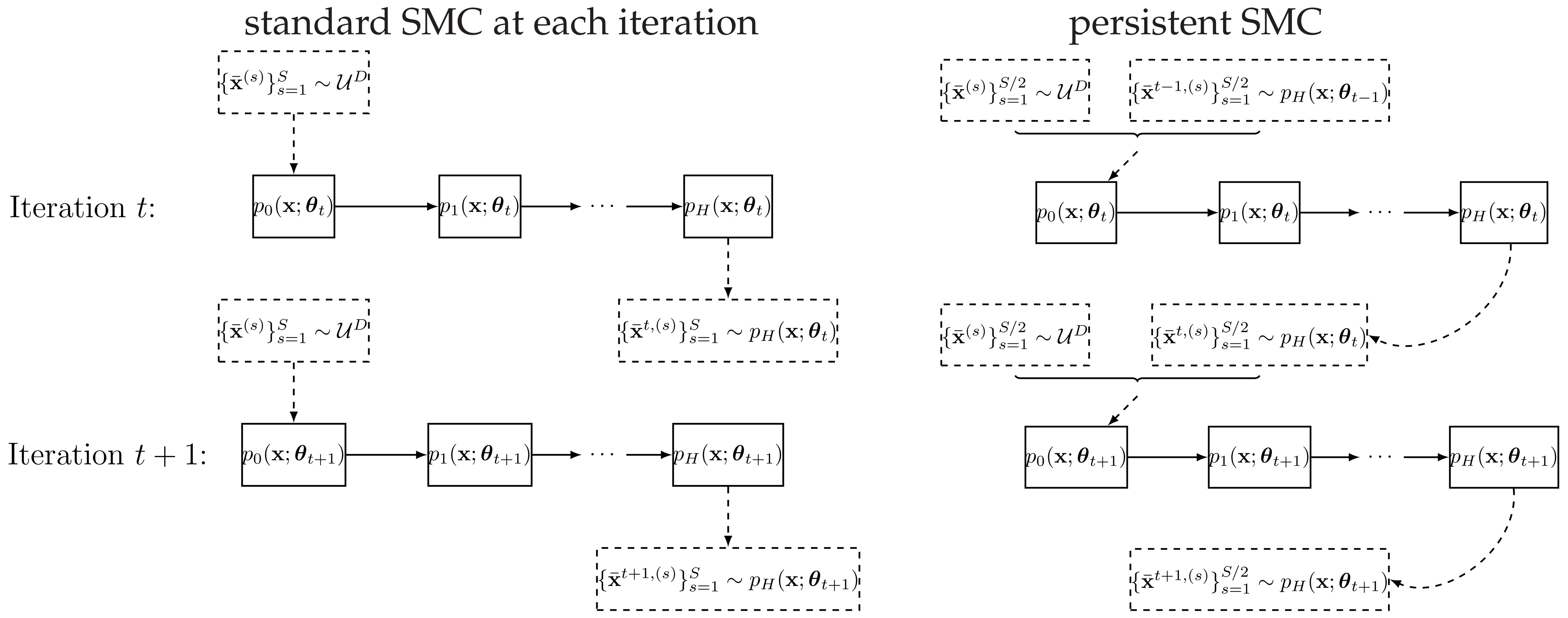
1. Training data set $\mathcal{D} = \{\mathbf{x}^1, \cdots, \mathbf{x}^M\}$. Randomly initialize model parameters $\boldsymbol{\theta}^0$ and N particles $\{\mathbf{s}^{0,1}, \cdots, \mathbf{s}^{0,N}\}$.
2. **for** $t = 0 : T$ **do**
3. **for** $n = 1 : N$ **do**
4. Sample $\mathbf{s}^{t+1,n}$ from $\mathbf{s}^{t,n}$ using transition operator $T_{\boldsymbol{\theta}}$.
5. **end for**
6. Update: $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t + \eta\Big[\frac{1}{M}\sum_{m=1}^{M}\phi(\mathbf{x}^m) - \frac{1}{N}\sum_{n=1}^{N}\phi(\mathbf{s}^{t+1,n})\Big]$
7. Decrease η.
8. **end for**

ACKNOWLEDGEMENT

This research has received funding from the EU FP7 Programme under grant agreement no. 270273, Xperience.

XPERIENCE.ORG
SEVENTH FRAMEWORK PROGRAMME
CogSys
Cognitive Systems

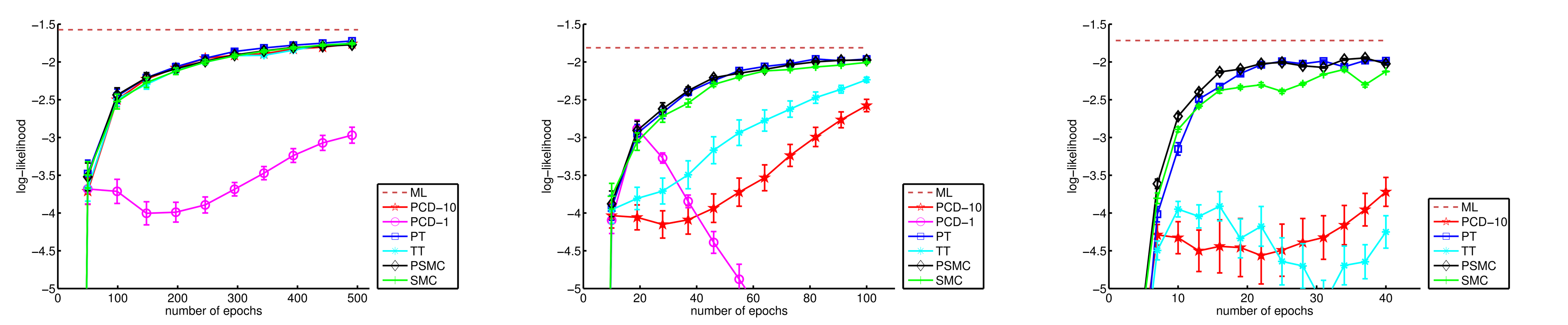
PERSISTENT SEQUENTIAL MONTE CARLO



where $\mathcal{U}^D$ is a uniform distribution on $\mathbf{x}$, and intermediate sequential distributions are: $p_h(\mathbf{x};\boldsymbol{\theta}_{t+1}) \propto p(\mathbf{x};\boldsymbol{\theta}_t)^{1-\beta_h} p(\mathbf{x};\boldsymbol{\theta}_{t+1})^{\beta_h}$ where $0 \le \beta_H \le \beta_{H-1} \le \cdots \beta_0 = 1$.

EXPERIMENTAL RESULTS AND COMPARISON

1. large learning rates:



2. high dimensional: