Variational Marginal Particle Filters

Manning College of Information and Computer Sciences,
University of Massachusetts Amherst

Jinlin Lai, Justin Domke, Daniel Sheldon
{jinlinlai,domke,sheldon}@cs.umass.edu

Background

State space models (SSMs):

$$p(x_{1:T}, y_{1:T}) = f(x_1) \prod_{t=2}^{T} f(x_t \mid x_{t-1}) \prod_{t=1}^{T} g(y_t \mid x_t)$$

Sequential Monte Carlo (SMC):

SMC is an algorithm that uses weighted particle systems to approximate the posterior $p(x_{1:T} \mid y_{1:T})$ and estimate the likelihood $p(y_{1:T})$ of a SSM.

Variational SMC (VSMC):

VSMC learns the parameters of proposal distributions by optimizing a lower bound of log-likelihood. It can also be used for learning the parameters of the SSM.

Introduction

Problems of existing methods:

- Does SMC give the tightest lower bound?
- Can we make the objective fully differentiable?

Contributions of this work:

- We give a novel proof of the correctness of SMC in terms of transformations of estimator-coupling pairs (Domke and Sheldon, 2019)
- We show that the marginal particle filter (MPF) is SMC with Rao-Blackwellization operations, implying that MPF gives a tighter bound than SMC
- We propose to optimize a variational bound based on MPF, which can be made differentiable through implicit reparameterization of mixtures (Graves, 2016; Figurnov et al., 2018)

Operations on Estimator-Coupling Pairs

- **Marginalize (Lemma 1)**: marginalize a subset of ω to get a lower variance estimator
  - Major difference between SMC and MPF
- **Replicate (Lemma 2)**: replicate estimator-coupling pairs N times
  - The sampling of N particles at each time-step
- **ExtendTarget (Lemma 3)**: extend the target distribution to $\gamma(x, x')$
  - Augmenting the target distribution via proposal distribution
- **ChangeTarget (Lemma 3)**: change the target distribution to $\gamma'(x')$
  - Changing the target distribution via proposal distribution

Marginal Particle Filters

By mechanically applying **ExtendTarget** and **Replicate** operations multiple times to importance weighted variational inference (IWVI) for $p(x_1, y_1)$, we can get SMC for $p(x_{1:T}, y_{1:T})$.

Replacing **ExtendTarget** with **ChangeTarget** and **Marginalize** operations, we get MPF for $p(x_{1:T}, y_{1:T})$.

Algorithm 2 Marginal Particle Filter (Klaas et al., 2005)

Require: $p(x_{1:T}, y_{1:T})$, $y_{1:T}$, $\{r_t(x_t|x_{t-1})\}$, N

Sample $x_1^i \sim r_1(x_1)$ for all i

Set $v_1^i = \frac{f_1(x_1^i)g(y_1|x_1^i)}{r_1(x_1^i)}$ for all i

for $t = 2, \ldots, T$ **do**

  for $i = 1, \ldots, N$ **do**

    Set $\overline{v}_{t-1}^j = v_{t-1}^j / \sum_{k=1}^N v_{t-1}^k$ for all j

    Sample $j \sim \text{Categorical}\left(\overline{v}_{t-1}^{1:N}\right)$

    Sample $x_t^i \sim r_t(x_t|x_{t-1}^j)$

    Set $v_t^i = \frac{\sum_{j=1}^N v_{t-1}^j f(x_t^i|x_{t-1}^j)g(y_t|x_t^i)}{\sum_{j=1}^N v_{t-1}^j r_t(x_t^i|x_{t-1}^j)}$

Training the Variational Bound

Let $\hat{p}_{\text{MPF}}(y_{1:T}) = \prod_{t=1}^{N} \frac{1}{N} \sum_{i=1}^{N} v_t^i$ be the unbiased estimator of $p(y_{1:T})$. We propose the variational objective

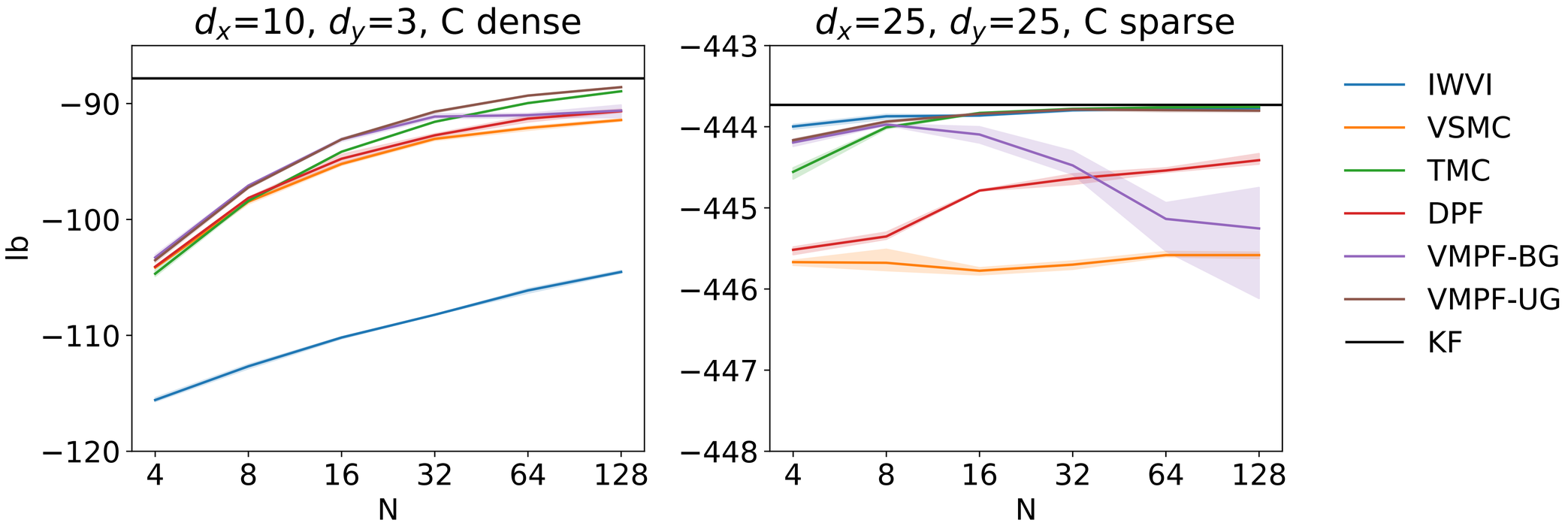
$$\mathscr{L}_{\text{VMPF}}(\phi, \theta) = \mathbb{E}\left[\log \hat{p}_{\text{MPF}}(y_{1:T}; \phi, \theta)\right]$$

Biased Gradients (VMPF-BG): ignore the gradient paths for probabilities of categorical variables

Unbiased Gradients (VMPF-UG): use implicit reparameterization gradients (Graves, 2016; Figurnov et al., 2018) for mixtures

Experiments

Linear Gaussian State Space Models



Final $\mathbb{E}[\log \hat{p}(y_{1:T})]$ as a function of N for IWVI, VSMC, TMC, DPF, VMPF-BG and VMPF-UG under two settings for the linear Gaussian SSM. Black line: true log-likelihood.

Experiments

Deep Markov Models

N	Method	Nottingham	JSB	MuseData	Piano-midi.de
4	IWVI	-3.86 (0.04)	-7.40 (0.01)	-8.19 (0.04)	-8.78 (0.01)
	VSMC	-3.38 (0.03)	-7.16 (0.01)	-7.68 (0.02)	-8.39 (0.03)
	TMC	-3.65 (0.02)	-7.51 (0.01)	-8.41 (0.02)	-9.00 (0.01)
	DPF	-3.33 (0.02)	-7.14 (0.01)	-7.75 (0.01)	-8.45 (0.01)
	VMPF-BG	**-3.29** (0.03)	**-7.04** (0.01)	**-7.67** (0.00)	**-8.35** (0.01)
16	IWVI	-3.84 (0.03)	-7.33 (0.02)	-8.16 (0.02)	-8.74 (0.02)
	VSMC	-3.06 (0.02)	-6.81 (0.00)	-7.22 (0.02)	-8.03 (0.00)
	TMC	-3.23 (0.02)	-6.98 (0.02)	-7.72 (0.00)	-8.37 (0.01)
	DPF	-3.08 (0.02)	-6.78 (0.00)	-7.22 (0.01)	-8.18 (0.02)
	VMPF-BG	**-2.96** (0.02)	**-6.64** (0.01)	**-7.14** (0.01)	**-7.92** (0.01)

Test set nats per time-step for DMM trained with IWVI, VSMC, DPF, VMPF-BG (higher is better) on four polyphonic music datasets. Mean and standard deviation of 3 runs are reported.

Thank you for watching!