Smoothing Variational Objectives with Sequential Monte Carlo for Nonlinear Dynamics

Anonymous Authors[1]

Abstract

Conductance based models of excitable cells are widely used in computational neuroscience to characterize the spiking activity of individual neurons. Recovering the multidimensional nonlinear dynamics that govern a cell from a single observation is a challenging problem motivating the development of novel techniques in time series analysis. Sequential Monte Carlo methods have been used to construct objective functions for variational inference on time series to perform simultaneous model inference and learning. We develop smoothed variational objectives analogous to forward-backwards message passing. We demonstrate that the use of information from the full time ordered sequence of observations improves both the state estimation and the dynamics learned. Experiments show that this method compares favorably against state of the art methods for variational inference in nonlinear dynamical systems. **TL;DR summary of paper**

1. Introduction

motivate the statistical problem with neuroscience question. introduce the method and outline the paper. Conductance based models of excitable cells are widely used in computational neuroscience to characterize the spiking activity of individual neurons. Recovering the multidimensional nonlinear dynamics that govern a cell from a single observation is a challenging problem motivating the development of novel techniques in time series analysis. Recently Sequential Monte Carlo methods have been used to construct objective functions for variational inference on time series to perform simultaneous model inference and learning. We develop smoothed variational objectives analogous to forward-backwards message passing. We demonstrate that the use of information from the full time ordered sequence of observations improves both the state estimation and the dynamics learned. Experiments show that this method compares favorably against state of the art methods for variational inference in nonlinear dynamical systems.

talk about dimensionality expansion, dimensionality reduction, prediction.

2. Related Work

Discuss SMC based methods including AESMC, VSMC, FIVO.

Contrast with pure VI methods including VRNN, DKF, LFADS, GfLDS, VIND.

3. Theory

3.1. Sequential Monte Carlo

Give a brief review of SMC and VI. Why is this approach promising as opposed to pure VI?

SMC methods factorize an ordered sequence of observations $\mathbf{X} \equiv \{\mathbf{x}_1, \dots \mathbf{x}_T\}$, $\mathbf{x}_t \in \mathbb{R}^{d_X}$ governed by an ordered a sequence of latent variables $\mathbf{Z} \equiv \{\mathbf{z}_1, \dots \mathbf{z}_T\}$, $\mathbf{z}_T \in \mathbb{R}^{d_Z}$ that evolve according to stochastic dynamics. The target distribution $\{p_\theta(\mathbf{z}_t|\mathbf{x}_t)\}_{t=1}^{T}$ is assumed intractable due to

SMC factorizes the target distribution $\{p_\theta(\mathbf{z}_t|\mathbf{x}_t)\}_{t=1}^{T}$ into a sequence of distributions of increasing spaces.

3.2. Auto-Encoding Sequential Monte Carlo

Recap main idea from AESMC and FIVO.

$$\log p(\mathbf{X}) = \log \int p(\mathbf{X}, \mathbf{Z}) d\mathbf{Z} = \log \frac{p(\mathbf{X}, \mathbf{Z})}{p(\mathbf{Z}|\mathbf{X})} \,. \tag{1}$$

$$\log p(\mathbf{X}) \geq \mathscr{L}_{\text{ELBO}}(\mathbf{X}) = \mathop{\mathbb{E}}_{q}[\log p(\mathbf{X}, \mathbf{Z})] - \mathop{\mathbb{E}}_{q}[\log q(\mathbf{Z}|\mathbf{X})] \,. \tag{2}$$

$$\hat{\mathcal{Z}}_{SMC} \equiv \prod_{t=1}^{T} \Big[\frac{1}{N} \sum_{n=1}^{N} w_t^{(n)} \Big] \tag{3}$$

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Importance Weighted Auto Encoders have been extended to the time series setting in the design of Sequential Monte Carlo (SMC) algorithms (Naesseth et al., 2018; Maddison et al., 2017; Le et al., 2018). SMC methods factorize the target distribution into a distribution of increasing spaces by performing importance sampling sequentially. Define importance weights $w_t^{(k)}$ at time t for sample k as follows:

$$w_t^{(k)} \equiv \frac{f(\mathbf{z}_t^{(k)}|\mathbf{z}_{t-1}^{(k)})g(\mathbf{x}_t|\mathbf{z}_t^{(k)})}{q(\mathbf{z}_t^{(k)}|\mathbf{z}_{t-1}^{(k)}, \mathbf{x}_t)} \tag{4}$$

where $\mathbf{z}_t^{(k)}$ is sampled:

$$\mathbf{z}_t^{(k)} \sim q(\mathbf{z}_t^{(k)}|\mathbf{z}_{t-1}^{(k)}, \mathbf{x}_t) \tag{5}$$

Various resampling schemes exist so that the samples which are referred to as particles are focused on promising regions of state space. This can be achieved by resampling particles according to their importance sampling weights:

$$\mathbf{a}_{t-1}^k \sim \text{Discrete}(\cdot|w_{t-1}^{(1)}, \cdots, w_{t-1}^{(K)}) \tag{6}$$

$$w_t^k \equiv \frac{f(\mathbf{z}_t^{(k)}|\mathbf{z}_{t-1}^{\mathbf{a}_{t-1}^k})g(\mathbf{x}_t|\mathbf{z}_t^{\mathbf{a}_{t-1}^k})}{q(\mathbf{z}_t^k|\mathbf{z}_{t-1}^{\mathbf{a}_{t-1}^k}, \mathbf{x}_t)} \tag{7}$$

At the last time step, we can evaluate the posterior as $\sum_{k=1}^{K} \bar{w}_T^k \delta_{\mathbf{Z}_{1:T}^k}(\mathbf{Z}_{1:T})$ averaging over paths to compute the functional integral. SMC also gives an unbiased estimate for the marginal likelihood:

$$\hat{\mathcal{Z}}_{SMC} \equiv \prod_{t=1}^{T} \Big[\frac{1}{K}\sum_{k=1}^{K} w_t^{(k)}\Big] \tag{8}$$

An important insight of (Maddison et al., 2017; Le et al., 2018) is that the SMC algorithm is deterministic conditioning on $(\mathbf{Z}_{1:T}^{(1:K)}, \mathbf{A}_{1:T-1}^{(1:K)})$. The proposal can thus be reparameterized to act as a variational distribution that can be encoded:

$$Q_{SMC}(\mathbf{Z}_{1:T}^{1:K}, \mathbf{A}_{1:T-1}^{1:K}) \equiv \left(\prod_{k=1}^{K} q_{1,\phi}(\mathbf{z}_1^k)\right) \times \left(\prod_{t=2}^{T}\prod_{k=1}^{K} q_{t,\phi}(\mathbf{z}_t^k|\mathbf{z}_{1:t-1}^{\mathbf{a}_{t-1}^k}) \cdot \text{Discrete}(\mathbf{a}_{t-1}^k|\mathbf{w}_{t-1}^{1:K})\right) \tag{9}$$

This gives a way of constructing a cost function for simultaneous model inference and learning. The cost is constructed by running SMC and performing stochastic gradient ascent on the importance weighted ELBO:

$$\mathscr{L}_{\text{ELBO SMC}}(\theta, \phi, \mathbf{Z}_{1:T}) \equiv \int Q_{SMC}(\mathbf{Z}_{1:T}^{1:K}, \mathbf{A}_{1:T-1}^{1:K}) \times \log \hat{\mathcal{Z}}_{SMC}(\mathbf{Z}_{1:T}^{1:K}, \mathbf{A}_{1:T-1}^{1:K}) d\mathbf{Z}_{1:T}^{1:N} d\mathbf{A}_{1:T}^{1:K} \tag{10}$$

Unlike pure variational methods, this does not involve inverting a block-tridiagonal matrix which mixes components of state space through the covariance. Information from the complete data $\mathbf{X}_{1:T}$ is not used to the future of $t < T$ to infer $\mathbf{z}_t$.

3.3. Smoothing Variational Objectives

Emphasize novelty of the proposed model, novelty of the experiments and the results. Discuss the implementation details.

The big question to answer is, how is this different from AESMC or FIVO in terms of theory? what is unique with respect to the applications and results?

nonlinear time invariant function learned on a smoothed SMC objective.

4. Results

Define quantitative evaluation metric. Motivate the use of k-step MSE_k and k-step R_k^2.

$$\text{MSE}_k = \sum_{t=0}^{T-k} (\mathbf{x}_{t+k} - \hat{\mathbf{x}}_{t+k})^2 \ , \quad R_k^2 = 1 - \frac{k\text{MSE}}{\sum_{t=0}^{T-k} (\mathbf{x}_{t+k} - \bar{\mathbf{x}})^2} \tag{11}$$

4.1. Fitzhugh Nagumo

$$\begin{aligned} \dot{V} &= f(V) - W + I_{ext} \,, \\ \dot{W} &= a(bV - cW) \end{aligned} \tag{12}$$

4.2. Lorenz Attractor

$$\begin{aligned} \dot{z}_1 &= \sigma(z_2 - z_1) \,, \\ \dot{z}_2 &= z_1(\rho - z_3) - z_2 \,, \\ \dot{z}_3 &= z_1 z_2 - \beta z_3 \,. \end{aligned} \tag{13}$$

4.3. Single Cell Recordings

5. Discussion

References

Anonymous. A novel variational family for hidden nonlinear markov models. In *Submitted to International Conference on Learning Representations*, 2019. under review.

Burda, Y., Grosse, R. B., and Salakhutdinov, R. Importance weighted autoencoders. *CoRR*, abs/1509.00519, 2015.

Figure 1. Summary of the Fitzhugh Nagumo results: (left) latent dynamics and paths for the original system (center) inferred 2D dynamics and paths from a noisy 1D observation (right) R_k^2 for various models.

Figure 2. Summary of the Lorenz results: (left) latent paths for the original system (center) inferred paths from a noisy 10D observation (right) R_k^2 for various models

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., and Bengio, Y. A recurrent latent variable model for sequential data. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 28*, pp. 2980–2988. Curran Associates, Inc., 2015.

Gao, Y., Archer, E. W., Paninski, L., and Cunningham, J. P. Linear dynamical neural population models through nonlinear embeddings. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 163–171. Curran Associates, Inc., 2016.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.

Le, T. A., Igl, M., Rainforth, T., Jin, T., and Wood, F. Auto-encoding sequential monte carlo. In *International Conference on Learning Representations*, 2018.

Maddison, C. J., Lawson, J., Tucker, G., Heess, N., Norouzi, M., Mnih, A., Doucet, A., and Teh, Y. Filtering variational objectives. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30*, pp. 6573–6583. Curran Associates, Inc., 2017.

Naesseth, C. A., Linderman, S. W., Ranganath, R., and Blei, D. M. Variational sequential monte carlo. In *AISTATS*, volume 84 of *Proceedings of Machine Learning Research*, pp. 968–977. PMLR, 2018.

Salimans, T., Kingma, D. P., and Welling, M. Markov chain monte carlo and variational inference: Bridging the gap. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pp. 1218–1226, 2015.

Sussillo, D., Józefowicz, R., Abbott, L. F., and Pandarinath, C. LFADS - latent factor analysis via dynamical systems. *CoRR*, abs/1608.06315, 2016.

Figure 3.