Variational Objectives for Markovian Dynamics with Backward Simulation

Antonio Moretti* Zizhao Wang* Luhuan Wu*

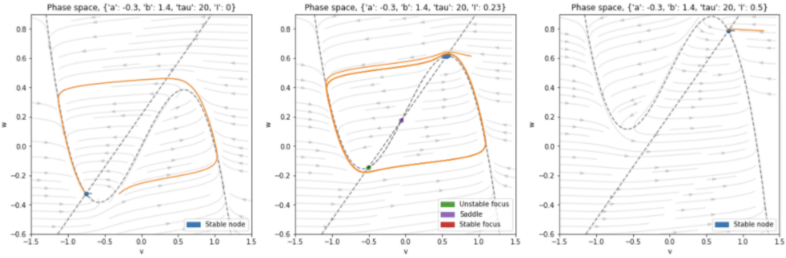
Iddo Drori Itsik Pe'er

Columbia University

8/5/2020

Overview

- HMMs formalize measurement on a latent **dynamical system**.



Overview

- HMMs formalize measurement on a latent **dynamical system**.

$$
\begin{array}{ccccccccc}
Z_0 & \xrightarrow[f]{} & Z_1 & \xrightarrow[f]{} & Z_2 & \xrightarrow[f]{} & \cdots & \xrightarrow[f]{} & Z_{T-1} \\
\downarrow g & & \downarrow g & & \downarrow g & & & & \downarrow g \\
X_0 & & X_1 & & X_2 & & \cdots & & X_{T-1}
\end{array}
$$

Overview

- Two distinct tasks:

$$\begin{tikzcd}
Z_0 \arrow[r, "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "f"'] & Z_{T-1} \arrow[d, "g"'] \\
X_0 & X_1 & X_2 & \cdots & X_{T-1}
\end{tikzcd}$$

Overview

- Two distinct tasks:
  - Infer **latent dynamics** and **latent trajectories** $Z_{1:T}$

$$\begin{tikzcd}
Z_0 \arrow[r, "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "f"'] & Z_{T-1} \arrow[d, "g"'] \\
X_0 & X_1 & X_2 & \cdots & X_{T-1}
\end{tikzcd}$$

Overview

- Two distinct tasks:
  - Infer **latent dynamics** and **latent trajectories** $Z_{1:T}$
  - Recover **transition** $f(\cdot)$ and **emission** $g(\cdot)$ functions from $X_{1:T}$

$$\begin{tikzcd} Z_0 \arrow[r, "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "f"'] & Z_{T-1} \arrow[d, "g"'] \\ X_0 & X_1 & X_2 & \cdots & X_{T-1} \end{tikzcd}$$

Overview

- Two distinct tasks:
    - Infer **latent dynamics** and **latent trajectories** $Z_{1:T}$
    - Recover **transition** $f(\cdot)$ and **emission** $g(\cdot)$ functions from $X_{1:T}$

$$\begin{tikzcd}
Z_0 \arrow[r, "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "f"'] & Z_{T-1} \arrow[d, "g"'] \\
X_0 & X_1 & X_2 & \cdots & X_{T-1}
\end{tikzcd}$$

$\Rightarrow$ Recent connections between VI and SMC: **filtered objectives**.

Overview

- **Filtering SMC** operates on **sequence of probability spaces**

$$
\begin{tikzcd}
X_0 \arrow[d, "q_1"] & X_1 \arrow[d, "q_1"] & X_2 \arrow[d, "q_1"] & \cdots & X_{T-1} \arrow[d, "q_1"] \\
Z_0 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "q_2", "f"'] & Z_{T-1} \arrow[d, "g"'] \\
\hat{X}_0 & \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_{T-1}
\end{tikzcd}
$$

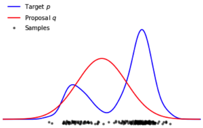
Overview

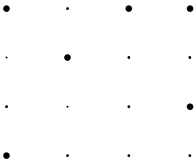
- **Filtering SMC** operates on **sequence of probability spaces**
  - Decompose the problem across time steps.

$$
\begin{tikzcd}
X_0 \arrow[d, "q_1"] & X_1 \arrow[d, "q_1"] & X_2 \arrow[d, "q_1"] & \cdots & X_{T-1} \arrow[d, "q_1"] \\
Z_0 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "q_2", "f"'] & Z_{T-1} \arrow[d, "g"'] \\
\hat{X}_0 & \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_{T-1}
\end{tikzcd}
$$

Overview

- **Filtering SMC** operates on **sequence of probability spaces**
  - Decompose the problem across time steps.
  - Define **tractable** proposal q, draw k samples $\{z_0^k\}_{k=1}^K \sim q(\cdot)$

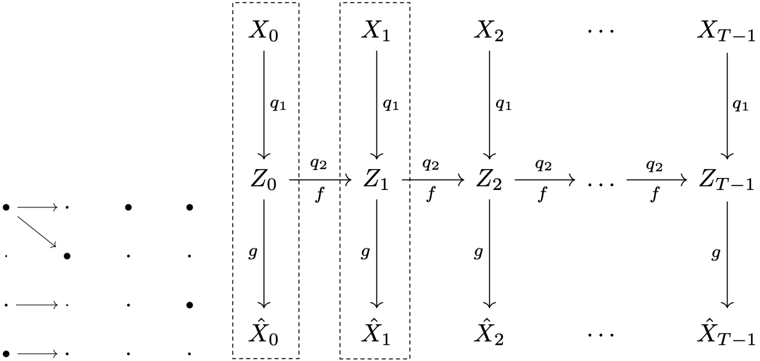




$$
\begin{tikzcd}
X_0 \arrow[d, "q_1"] & X_1 \arrow[d, "q_1"] & X_2 \arrow[d, "q_1"] & \cdots & X_{T-1} \arrow[d, "q_1"] \\
Z_0 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "q_2", "f"'] & Z_{T-1} \arrow[d, "g"'] \\
\hat{X}_0 & \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_{T-1}
\end{tikzcd}
$$

Overview

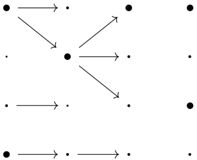
- **Filtering SMC** operates on **sequence of probability spaces**
  - Decompose the problem across time steps.
  - **Tractable** proposal q, draw k samples $\{z_1^k\}_{k=1}^K \sim q(\cdot|z_0^{a_0^k})$



  - **Resample particles** to focus on areas of high probability mass

Overview

- **Filtering SMC** operates on **sequence of probability spaces**
  - Decompose the problem across time steps.
  - **Tractable** proposal q, draw k samples $\{z_2^k\}_{k=1}^K \sim q(\cdot|z_1^{a_1^k})$

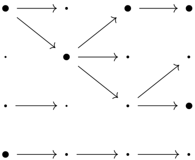


$$
\begin{tikzcd}
X_0 \arrow[d, "q_1"] & X_1 \arrow[d, "q_1"] & X_2 \arrow[d, "q_1"] & \cdots & X_{T-1} \arrow[d, "q_1"] \\
Z_0 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "q_2", "f"'] & Z_{T-1} \arrow[d, "g"'] \\
\hat{X}_0 & \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_{T-1}
\end{tikzcd}
$$

  - **Resample particles** to focus on areas of high probability mass

Overview

- **Filtering SMC** operates on **sequence of probability spaces**
  - Decompose the problem across time steps.
  - **Tractable** proposal q, draw k samples $\{z^k_{T-1}\}^K_{k=1} \sim q(\cdot|z^{a^k_{T-2}}_{T-2})$



$$\begin{tikzcd}
X_0 \arrow[d, "q_1"] & X_1 \arrow[d, "q_1"] & X_2 \arrow[d, "q_1"] & \cdots & X_{T-1} \arrow[d, "q_1"] \\
Z_0 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_1 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & Z_2 \arrow[r, "q_2", "f"'] \arrow[d, "g"'] & \cdots \arrow[r, "q_2", "f"'] & Z_{T-1} \arrow[d, "g"'] \\
\hat{X}_0 & \hat{X}_1 & \hat{X}_2 & \cdots & \hat{X}_{T-1}
\end{tikzcd}$$

  - **Resample particles** to focus on areas of high probability mass

Variational Inference in a Nutshell

- Approx *intractable* $p_\theta(z, x)$ w/ *tractable* $q_\phi(z|x)$

Variational Inference in a Nutshell

- Approx *intractable* $p_\theta(z, x)$ w/ *tractable* $q_\phi(z|x)$

  $\Rightarrow$ Trade integration for optimization

Variational Inference in a Nutshell

- Approx *intractable* $p_\theta(z, x)$ w/ *tractable* $q_\phi(z|x)$

  $\Rightarrow$ Trade integration for optimization

- Exploit hidden symmetry

Variational Inference in a Nutshell

- Approx *intractable* $p_\theta(z, x)$ w/ *tractable* $q_\phi(z|x)$

  $\Rightarrow$ Trade integration for optimization

- Exploit hidden symmetry

  $\Rightarrow$ Duality: $\underset{\theta,\phi}{\text{argmin}}\ \mathcal{D}_{KL}\left(q_\phi(z|x)||p_\theta(z|x)\right) \equiv \underset{\theta,\phi}{argmax}\ \mathcal{L}_{ELBO}$

Variational Objectives

- Weighting samples by ratio p/q **corrects for approximation**

Variational Objectives

- Weighting samples by ratio p/q **corrects for approximation**
  - $\Rightarrow$ Bias proposal towards true posterior

$$w_t^k := \frac{f(z_t^k | z_{t-1}^{a_{t-1}^k}) g(x_t | z_t^k)}{q(z_t^k | z_{t-1}^{a_{t-1}^k}, x_t)}$$

Variational Objectives

- Weighting samples by ratio p/q **corrects for approximation**
  - $\Rightarrow$ Bias proposal towards true posterior

$$w_t^k := \frac{f(z_t^k | z_{t-1}^{a_{t-1}^k}) g(x_t | z_t^k)}{q(z_t^k | z_{t-1}^{a_{t-1}^k}, x_t)}$$

- SMC constructs **filtered estimate** of $p_\theta(x)$

$$\hat{\mathcal{Z}}_{SMC} := \prod_{t=1}^{T} \left[\frac{1}{K} \sum_{k=1}^{K} w_t^k \right],$$

Variational Objectives

- Weighting samples by ratio p/q **corrects for approximation**
  - $\Rightarrow$ Bias proposal towards true posterior

$$w_t^k := \frac{f(z_t^k | z_{t-1}^{a_{t-1}^k}) g(x_t | z_t^k)}{q(z_t^k | z_{t-1}^{a_{t-1}^k}, x_t)}$$

- SMC constructs **filtered estimate** of $p_\theta(x)$ and objective:

$$\hat{\mathcal{Z}}_{SMC} := \prod_{t=1}^{T} \left[\frac{1}{K} \sum_{k=1}^{K} w_t^k \right], \qquad \mathcal{L}_{SMC} := \mathop{\mathbb{E}}_{q} \left[\log \hat{\mathcal{Z}}_{SMC} \right]$$

Variational Objectives

- Weighting samples by ratio p/q **corrects for approximation**
  - $\Rightarrow$ Bias proposal towards true posterior

$$w_t^k := \frac{f(z_t^k | z_{t-1}^{a_{t-1}^k}) g(x_t | z_t^k)}{q(z_t^k | z_{t-1}^{a_{t-1}^k}, x_t)}$$

- SMC constructs **filtered estimate** of $p_\theta(x)$ and objective:

$$\hat{\mathcal{Z}}_{SMC} := \prod_{t=1}^{T} \left[\frac{1}{K} \sum_{k=1}^{K} w_t^k \right], \qquad \mathcal{L}_{SMC} := \mathop{\mathbb{E}}_{q} \left[\log \hat{\mathcal{Z}}_{SMC} \right]$$

- $\mathcal{L}_{SMC}$ is a **biased estimator** for $\log p_\theta(x)$, bias is $\mathcal{O}(K^{-1})$

Particle Smoothing

Forward Filtering Backward Simulation (Godsill, 2004):

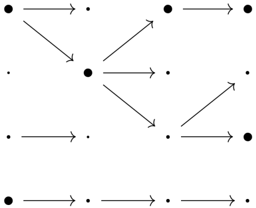
$$p(\mathrm{z}_{1:T}|\mathrm{x}_{1:T}) = p(\mathrm{z}_T|\mathrm{x}_{1:T}) \prod_{t=1}^{T-1} p(\mathrm{z}_t|\mathrm{z}_{t+1:T}, \mathrm{x}_{1:T}) \ ,$$

- Samples drawn from continuous q in **forward pass**

Particle Smoothing

Forward Filtering Backward Simulation (Godsill, 2004):

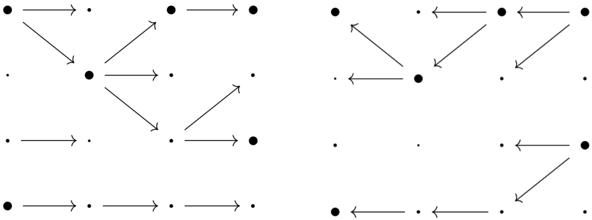
- Samples drawn from continuous q in **forward pass**



Particle Smoothing

Forward Filtering Backward Simulation (Godsill, 2004):

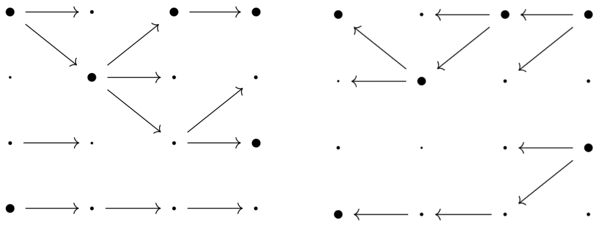
- Continuous samples drawn from q in forward pass.
- Backward pass **only performs resampling**.



Particle Smoothing

Forward Filtering Backward Simulation (Godsill, 2004):

- Continuous samples drawn from q in forward pass.
- Backward pass **only performs resampling**.



$\Rightarrow$ **Limits expressiveness of variational family**.

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

$$p(z_{1:T}|x_{1:T}) = p(z_T|x_{1:T}) \prod_{t=1}^{T-1} p(z_t|z_{t+1:T}, x_{1:T})$$

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

$$p(z_{1:T}|x_{1:T}) = p(z_T|x_{1:T}) \prod_{t=1}^{T-1} p(z_t|z_{t+1:T}, x_{1:T})$$

- Design q to approx p w/ *self-normalized importance sampling*

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

$$p(z_{1:T}|x_{1:T}) = p(z_T|x_{1:T}) \prod_{t=1}^{T-1} p(z_t|z_{t+1:T}, x_{1:T})$$

- Design q to approx p w/ *self-normalized importance sampling*

  $\Rightarrow$ Subsample *subparticles* and compute *subweights*.

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

$$p(z_{1:T}|x_{1:T}) = p(z_T|x_{1:T}) \prod_{t=1}^{T-1} p(z_t|z_{t+1:T}, x_{1:T})$$

- Design q to approx p w/ *self-normalized importance sampling*
  - $\Rightarrow$ Subsample *subparticles* and compute *subweights*.
- Select subparticle index w/ prob proportional to subweight

Particle Smoothing Variational Objectives

Can we use smoothing posterior to design a **continuous-domain backward proposal?**

$$p(z_{1:T}|x_{1:T}) = p(z_T|x_{1:T}) \prod_{t=1}^{T-1} p(z_t|z_{t+1:T}, x_{1:T})$$

- Design q to approx p w/ *self-normalized importance sampling*

  $\Rightarrow$ Subsample *subparticles* and compute *subweights*.

- Select subparticle index w/ prob proportional to subweight

- Yields i.i.d. sample trajectories and smooth objective.

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(z_t | z_{t+1}, x_{1:T})$$

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(\mathrm{z}_t | \mathrm{z}_{t+1}, \mathrm{x}_{1:T})$$

- Sample M **subparticles** for each $k \in \{1, \cdots, K\}$

$$\tilde{\mathrm{z}}_t^{k,1:M} \sim q(\mathrm{z}_t | \tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})$$

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(\mathrm{z}_t|\mathrm{z}_{t+1}, \mathrm{x}_{1:T})$$

- Sample M **subparticles** for each $k \in \{1, \cdots, K\}$

$$\tilde{\mathrm{z}}_t^{k,1:M} \sim q(\mathrm{z}_t|\tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})$$

- Use samples to define **subweights**:

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(\mathrm{z}_t|\mathrm{z}_{t+1}, \mathrm{x}_{1:T})$$

- Sample M **subparticles** for each $k \in \{1, \cdots, K\}$

$$\tilde{\mathrm{z}}_t^{k,1:M} \sim q(\mathrm{z}_t|\tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})$$

- Use samples to define **subweights**:

$$\omega_{t|T}^{k,m} = p(\tilde{\mathrm{z}}_t^{k,m}|\tilde{\mathrm{z}}_{t+1}^{k,m}, \mathrm{x}_{1:T})$$

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(z_t | z_{t+1}, x_{1:T})$$

- Sample M **subparticles** for each $k \in \{1, \cdots, K\}$

$$\tilde{z}_t^{k,1:M} \sim q(z_t | \tilde{z}_{t+1}^k, x_{1:T})$$

- Use samples to define **subweights**:

$$\begin{aligned}\omega_{t|T}^{k,m} &= p(\tilde{z}_t^{k,m} | \tilde{z}_{t+1}^{k,m}, x_{1:T}) \\ &\propto \int p(z_{t-1}, \tilde{z}_t^{k,m} | x_{1:t-1}) dz_{t-1} \frac{f(\tilde{z}_{t+1}^k | \tilde{z}_t^{k,m}) g(x_t | \tilde{z}_t^{k,m})}{q(\tilde{z}_t^{k,m} | \tilde{z}_{t+1}^k, x_{1:T})}\end{aligned}$$

Particle Smoothing Variational Objectives

- Define **continuous-domain backward proposal**

$$q(\mathrm{z}_t|\mathrm{z}_{t+1}, \mathrm{x}_{1:T})$$

- Sample M **subparticles** for each $k \in \{1, \cdots, K\}$

$$\tilde{\mathrm{z}}_t^{k,1:M} \sim q(\mathrm{z}_t|\tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})$$

- Use samples to define **subweights**:

$$\begin{aligned}
\omega_{t|T}^{k,m} &= p(\tilde{\mathrm{z}}_t^{k,m}|\tilde{\mathrm{z}}_{t+1}^{k,m}, \mathrm{x}_{1:T}) \\
&\propto \int p(\mathrm{z}_{t-1}, \tilde{\mathrm{z}}_t^{k,m}|\mathrm{x}_{1:t-1}) d\mathrm{z}_{t-1} \frac{f(\tilde{\mathrm{z}}_{t+1}^k|\tilde{\mathrm{z}}_t^{k,m}) g(\mathrm{x}_t|\tilde{\mathrm{z}}_t^{k,m})}{q(\tilde{\mathrm{z}}_t^{k,m}|\tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})} \\
&\approx \left[\sum_{j=1}^{K} \bar{w}_{t-1}^j f(\tilde{\mathrm{z}}_t^{k,m}|z_{t-1}^j)\right] \frac{f(\tilde{\mathrm{z}}_{t+1}^k|\tilde{\mathrm{z}}_t^{k,m}) g(\mathrm{x}_t|\tilde{\mathrm{z}}_t^{k,m})}{q(\tilde{\mathrm{z}}_t^{k,m}|\tilde{\mathrm{z}}_{t+1}^k, \mathrm{x}_{1:T})}.
\end{aligned}$$

Particle Smoothing Variational Objectives

Construct estimator $\hat{\mathcal{Z}}_{SVO}$

$$\hat{\mathcal{Z}}_{SVO} := \frac{1}{K} \sum_{k=1}^{K} \frac{p(\tilde{z}_{1:T}^{k}, x_{1:T})}{q(\tilde{z}_{1:T}^{k} | x_{1:T})} \,,$$

Particle Smoothing Variational Objectives

Construct estimator $\hat{\mathcal{Z}}_{SVO}$ and objective $\mathcal{L}_{SVO}$:

$$\hat{\mathcal{Z}}_{SVO} := \frac{1}{K}\sum_{k=1}^{K} \frac{p(\tilde{z}^k_{1:T}, x_{1:T})}{q(\tilde{z}^k_{1:T} | x_{1:T})}, \quad \mathcal{L}_{SVO} := \mathop{\mathbb{E}}_{q}\left[\log \hat{\mathcal{Z}}_{SVO}\right]$$

Particle Smoothing Variational Objectives

Construct estimator $\hat{\mathcal{Z}}_{SVO}$ and objective $\mathcal{L}_{SVO}$:

$$\hat{\mathcal{Z}}_{SVO} \coloneqq \frac{1}{K} \sum_{k=1}^{K} \frac{p(\tilde{z}_{1:T}^k, x_{1:T})}{q(\tilde{z}_{1:T}^k | x_{1:T})}, \quad \mathcal{L}_{SVO} \coloneqq \mathbb{E}_q \left[\log \hat{\mathcal{Z}}_{SVO} \right]$$

where proposal factorizes,

$$q(\tilde{z}_{1:T}^k | x_{1:T}) \coloneqq M^T \cdot \omega_{T|T}^k \cdot q(\tilde{z}_T^k | x_{1:T}) \prod_{t=1}^{T-1} \left[\omega_{t|T}^k \cdot q(\tilde{z}_t^k | \tilde{z}_{t+1}^k, x_{1:T}) \right].$$

Particle Smoothing Variational Objectives

Construct estimator $\hat{\mathcal{Z}}_{SVO}$ and *smoothing variational objective*:

$$\hat{\mathcal{Z}}_{SVO} := \frac{1}{K}\sum_{k=1}^{K}\frac{p(\tilde{\mathrm{z}}^k_{1:T}, \mathrm{x}_{1:T})}{q(\tilde{\mathrm{z}}^k_{1:T}|\mathrm{x}_{1:T})}, \quad \mathcal{L}_{SVO} := \mathop{\mathbb{E}}_{q}\left[\log \hat{\mathcal{Z}}_{SVO}\right]$$

where proposal factorizes

$$q(\tilde{\mathrm{z}}^k_{1:T}|\mathrm{x}_{1:T}) := M^T \cdot \omega^k_{T|T} \cdot q(\tilde{\mathrm{z}}^k_T|\mathrm{x}_{1:T}) \prod_{t=1}^{T-1}\left[\omega^k_{t|T} \cdot q(\tilde{\mathrm{z}}^k_t|\tilde{\mathrm{z}}^k_{t+1}, \mathrm{x}_{1:T})\right].$$

- SVO generates **unbiased estimate** of marginal likelihood.

$$\hat{\mathcal{Z}}_{SVO} \to p_\theta(x_{1:T})$$

Evaluation Metric

How to evaluate model performance?

- PSVO should be *more* than a good **autoencoder**.

Evaluation Metric

How to evaluate model performance?

- PSVO should be *more* than a good **autoencoder**.
- $\mathcal{L}_{ELBO}$ is **not comparable** across models.

Evaluation Metric

How to evaluate model performance?

- PSVO should be *more* than a good **autoencoder**.
- $\mathcal{L}_{ELBO}$ is **not comparable** across models.
- **Latent trajectories** connected by *smooth transformations* can result in **equivalent representation** of $\hat{x}_{1:T}$.

Evaluation Metric

- Repetedly apply **transition function** $f(\cdot)$ and **emission function** $g(\cdot)$ to form predictions.

$$\mathrm{MSE}_k = \sum_{t=0}^{T-k} \left(\mathsf{x}_{t+k} - \hat{\mathsf{x}}_{t+k}\right)^2 , \quad R_k^2 = 1 - \frac{\mathrm{MSE}_k}{\sum_{t=0}^{T-k} \left(\mathsf{x}_{t+k} - \bar{\mathsf{x}}\right)^2}$$

Evaluation Metric

- Repetedly apply **transition function** $f(\cdot)$ and **emission function** $g(\cdot)$ to form predictions.
- Compare *predicted observations* $\hat{x}_{1:T-k}$ and data $x_{1:T-k}$

$$\mathrm{MSE}_k = \sum_{t=0}^{T-k} \left(\mathsf{x}_{t+k} - \hat{\mathsf{x}}_{t+k}\right)^2 \ , \quad R_k^2 = 1 - \frac{\mathrm{MSE}_k}{\sum_{t=0}^{T-k} \left(\mathsf{x}_{t+k} - \bar{\mathsf{x}}\right)^2}$$

Evaluation Metric

- Repetedly apply **transition function** $f(\cdot)$ and **emission function** $g(\cdot)$ to form predictions.

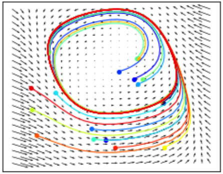
$$Z_0 \xrightarrow{f} Z_1 \xrightarrow{f} Z_K \xrightarrow{f} \dots \xrightarrow{f} Z_{T-K}$$

Evaluation Metric

- Repetedly apply **transition function** $f(\cdot)$ and **emission function** $g(\cdot)$ to form predictions.
- Compare *predicted observations* $\hat{x}_{1:T-k}$ and data $x_{1:T-k}$

$$
\begin{tikzcd}
Z_0 \arrow[r, "f"] & Z_1 \arrow[r, "f"] & Z_K \arrow[r, "f"] \arrow[d, "g"'] & \dots \arrow[r, "f"] & Z_{T-K} \arrow[d, "g"'] \\
 & & \hat{X}_K & \dots & \hat{X}_{T-K}
\end{tikzcd}
$$

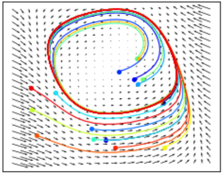
Particle Smoothing Variational Objectives: FHN



- Integrate over 200 time bins w/ random initial points

$$\dot{z}_1 = z_1 - z_1^3/3 - z_2$$
$$\dot{z}_2 = a(bz_1 - cz_2)$$

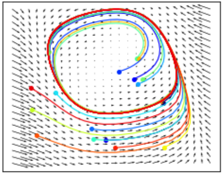
Particle Smoothing Variational Objectives: FHN



- Take 1D observation forming **partially observable system**

$$x_t = \mathcal{N}(z_1, \sigma^2)$$

Particle Smoothing Variational Objectives: FHN

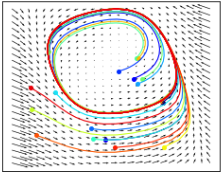


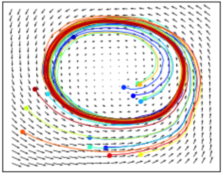
- Take 1D observation forming **partially observable system**

$$x_t = \mathcal{N}(z_1, \sigma^2)$$

$\Rightarrow$ Filtering **cannot infer** *initial point* from 1D observation.

Particle Smoothing Variational Objectives: FHN

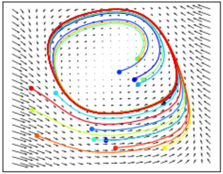


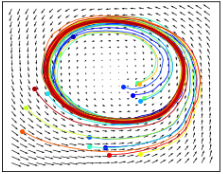


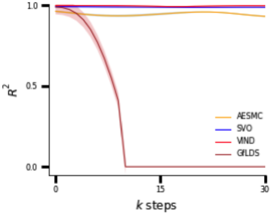
- Inferred hidden trajectories and dynamics topologically similar with ground truth

$$\dot{z}_1 = z_1 - z_1^3/3 - z_2$$
$$\dot{z}_2 = a(bz_1 - cz_2)$$

Particle Smoothing Variational Objectives: FHN



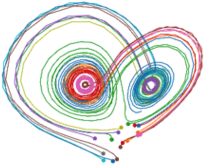




- Outperforms filtered nonlinear objectives and linear dynamics.

$$\dot{z}_1 = z_1 - z_1^3/3 - z_2$$
$$\dot{z}_2 = a(bz_1 - cz_2)$$

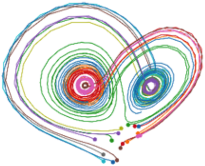
Particle Smoothing Variational Objectives: Lorenz



- Integrate over 250 time bins w/ random initial points

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

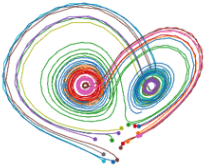
Particle Smoothing Variational Objectives: Lorenz

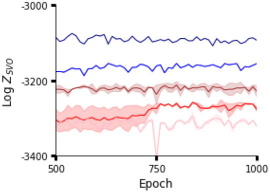


- Define 10D **nonlinear** observations for dimension reduction

$$\dot{z}_1 = \sigma(z_2 - z_1)$$
$$\dot{z}_2 = z_1(\rho - z_3) - z_2$$
$$\dot{z}_3 = z_1 z_2 - \beta z_3$$

Particle Smoothing Variational Objectives: Lorenz

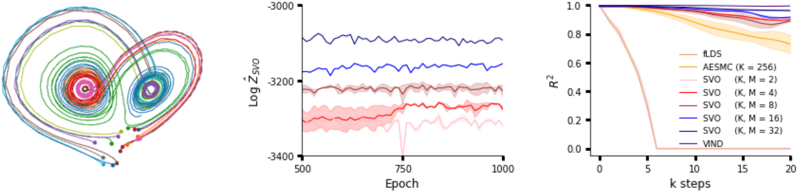




- **Tighter bounds** $\log \mathcal{Z}_{SVO} \rightarrow \log p_\theta(x_{1:T})$ as K, M increase

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

Particle Smoothing Variational Objectives: Lorenz



- $\mathcal{L}_{SVO}$ **consistently outperforms** $\mathcal{L}_{SMC}$ with fewer particles

$$\dot{z}_1 = \sigma(z_2 - z_1)$$
$$\dot{z}_2 = z_1(\rho - z_3) - z_2$$
$$\dot{z}_3 = z_1 z_2 - \beta z_3$$

Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$

- VI is a tradeoff b/t **tractability** of $q_\phi(z_{1:T}|x_{1:T})$ vs **expressiveness** of $p_\theta(z_{1:T}, x_{1:T})$

Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$

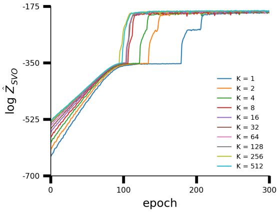
- VI is a tradeoff b/t **tractability** of $q_\phi(z_{1:T}|x_{1:T})$ vs **expressiveness** of $p_\theta(z_{1:T}, x_{1:T})$
- If **variational family** is *limited*, training both θ and ϕ can pull $p_\theta(z_{1:T}, x_{1:T}) \to q_\phi(z_{1:T}|x_{1:T})$.

Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$

- VI is a tradeoff b/t **tractability** of $q_\phi(z_{1:T}|x_{1:T})$ vs **expressiveness** of $p_\theta(z_{1:T}, x_{1:T})$
- If **variational family** is *limited*, training both θ and ϕ can pull $p_\theta(z_{1:T}, x_{1:T}) \to q_\phi(z_{1:T}|x_{1:T})$

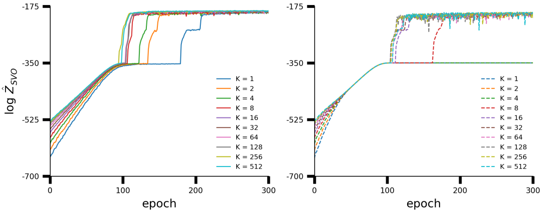
$\Rightarrow$ Share transition function b/t $q_\phi(z_{1:T}|x_{1:T})$ and $p_\theta(z_{1:T}, x_{1:T})$

Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$



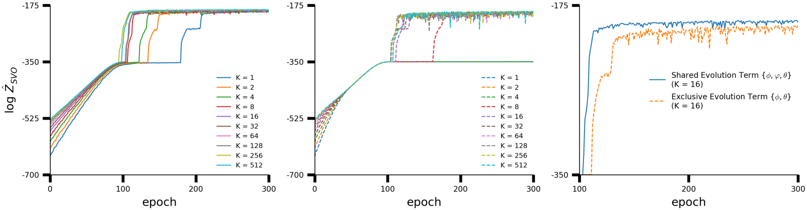
- $\mathcal{L}_{SVO}$ convergence when sharing $f(z_t|\psi(z_{t-1}))$ b/t q_ϕ and p_θ

Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$



- **Slower convergence** of $\mathcal{L}_{SVO}$ and lower values for *separate transition parameters*.

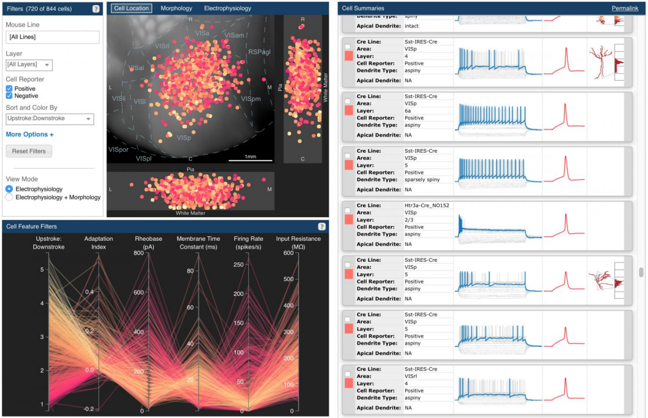
Particle Smoothing Variational Objectives: $\mathcal{L}_{SVO}$



- A closer look at $\mathcal{L}_{SVO}$ convergence when $K = 16$.

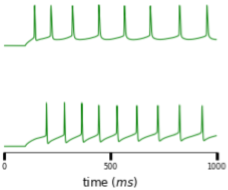
$\Rightarrow$ **Faster convergence** w/ shared parameters

Particle Smoothing Variational Objectives: Allen



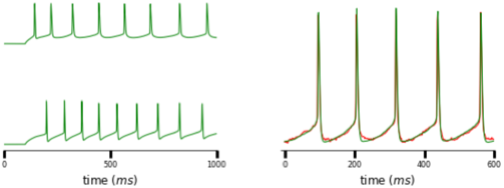
- Electrophysiology data of individual neurons from mouse visual cortex downloaded from Allen Brain Atlas.

Particle Smoothing Variational Objectives: Allen



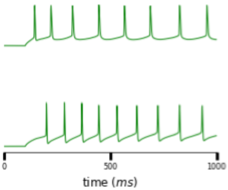
- Download 30 trials of neuronal spiking from input current.

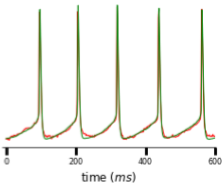
Particle Smoothing Variational Objectives: Allen

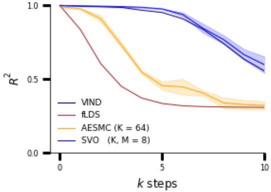


- 10-millisecond prediction captures **depolarization** and **hyperpolarization nonlinearities**.

Particle Smoothing Variational Objectives: Allen







- SVO **outperforms** filtered objectives and linear systems.

Takeaways

PSVO:

- *Consistently outperforms* **filtered objectives**.

Takeaways

PSVO:

- *Consistently outperforms* **filtered objectives**.
- *Learns* **nonlinear transition and emission functions** from *partially observable systems*.

Takeaways

PSVO:

- *Consistently outperforms* **filtered objectives**.
- *Learns* **nonlinear transition and emission functions** from *partially observable systems*.
- *Augments* **backward proposal support** and *boosts* **particle diversity**.

Takeaways

PSVO:

- *Consistently outperforms* **filtered objectives**.
- *Learns* **nonlinear transition and emission functions** from *partially observable systems*.
- *Augments* **backward proposal support** and *boosts* **particle diversity**.
- Well-motivated *variational objective* $\mathcal{L}_{SVO}$ from a **consistent** and **unbiased likelihood estimate**.

Thank You

- Implementation and datasets for experiments online:

  $\Rightarrow$ https://github.com/amoretti86/PSVO

- Thanks to Christian Naesseth and Daniel Hernandez for helpful discussions

References I

- Yuri Burda, Roger B. Grosse, and Ruslan Salakhutdinov, *Importance weighted autoencoders.*, CoRR **abs/1509.00519** (2015).
- Yuanjun Gao, Evan Archer, Liam Paninski, and John P. Cunningham, *Linear dynamical neural population models through nonlinear embeddings*, 2016.
- Simon J Godsill, Arnaud Doucet, and Mike West, *Monte carlo smoothing for nonlinear time series*, Journal of the American Statistical Association **99** (2004), no. 465, 156–168.
- Alan L Hodgkin and Andrew F Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, The Journal of physiology **117** (1952), no. 4, 500–544.

References II

- Daniel Hernandez, Antonio Khalil Moretti, Ziqiang Wei, Shreya Saxena, John Cunningham, and Liam Paninski, *Nonlinear evolution via spatially-dependent linear dynamics for electrophysiology and calcium data*, 2018.
- Allan R. Jones, Caroline C. Overly, and Susan M. Sunkin, *The allen brain atlas: 5 years and beyond*, Nature Reviews Neuroscience **10** (2009), 821 EP –.
- Diederik P. Kingma and Max Welling, *Auto-encoding variational bayes*, CoRR **abs/1312.6114** (2013).
- Tuan Anh Le, Maximilian Igl, Tom Rainforth, Tom Jin, and Frank Wood, *Auto-encoding sequential monte carlo*, International Conference on Learning Representations, 2018.

References III

Chris J Maddison, John Lawson, George Tucker, Nicolas Heess, Mohammad Norouzi, Andriy Mnih, Arnaud Doucet, and Yee Teh, *Filtering variational objectives*, Advances in Neural Information Processing Systems 30 (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), Curran Associates, Inc., 2017, pp. 6573–6583.

Antonio Khalil Moretti, Zizhao Wang, Luhuan Wu, Iddo Drori, and Itsik Pe'er, *Particle smoothing variational objectives*, 2019.

Antonio K Moretti, Zizhao Wang, Luhuan Wu, Iddo Drori, and Itsik Pe'er, *Variational objectives for markovian dynamics with backward simulation*, European Conference on Artificial Intelligence (2020).

References IV

Christian A. Naesseth, Scott W. Linderman, Rajesh Ranganath, and David M. Blei, *Variational sequential monte carlo*, 2017.