Smoothing Estimates of Diffusion Processes
PICE Smoother (PICES)

Hans-Christian Ruiz[1], Hilbert J. Kappen[1]

[1] Donders Institute, Radboud University, Nijmegen, The Netherlands

DONDERS INSTITUTE

INTRODUCTION

Access to complex dynamic phenomena is usually limited by indirect, **noisy measurements**. Given a time sequence of observations, we have to **infer the underlying process** to learn more about the system.
For instance in Neuroscience, the brain activity of humans during experiments has to be measured by non-invasive methods like fMRI or EEG.

$$p(X_{[0,T]}|y_{0:T}) \propto p(X_{[0,T]})\exp\left(\sum_{j=1}^{J}\log\left[g(y_{t_j}|X_{t_j})\right]\right) \qquad dX_t = F(X_t,t)dt + \sigma_{dyn}(X_t,t)dW_t$$

Mathematically, this problem is formulated as a Bayesian inference problem, where the aim is to find the **posterior/smoothing distribution** over processes given the time series. The exact solution to this problem is in general intractable when the dynamics is non-linear or the observation is a non-linear function of the hidden process. Hence, approximate methods are required to overcome this difficulty.

One can resort to sampling methods such as Sequential Monte Carlo methods. However, a major challenge for sampling from the posterior distribution is the **weight degeneracy** of importance sampling schemes, which gives a small effective sampling size (ESS). The degeneracy problem is especially acute when the prior deviates from the posterior due to model mismatch.

We present PICE Smoother, a **novel method to sample efficiently** from the smoothing distribution over hidden diffusion processes in continuous time. We use an adaptive importance sampler based on the path integral cross-entropy (PICE) control theory.

METHOD

PICES significantly **reduces the weight degeneracy** by adapting the process with a feedback controller. This is done by **minimizing the Kullback-Leibler divergence** between the posterior and the proposal distribution.

Step -1: Input

$$u(x,t|\hat{\theta}_0) \equiv 0$$
$$q_0(X_0|\phi_0) = p(X_0)$$

Initialize controller at zero
Set prior $p(X_0)$ as first proposal distribution

Step 0: Initialization

$$X_0^i \sim q_l(X_0|\phi_l)$$
$$\hat{S}_i = -\log\left(\frac{p(X_0^i)}{q_l(X_0^i|\phi_l)}\right)$$

Draw N particles from proposal distribution $q_l(X_0|\phi_l)$
Compute importance sampling (IS) correction $\hat{S}_i$

Step 1: Generate Trajectories from the Adapted Stochastic Differential Equation

$$dX_t = F(X_t,t)dt + u(X_t,t)dt + \sigma_{dyn}(X_t,t)dW_t \quad dW_t \sim \mathcal{N}(0,dt)$$

The feedback control $u(X_t,t)$ adapts the prior/uncontrolled dynamics
Propagate the N particles via numerical integration of the adapted process

Step 2: Calculate Cost

$$S_i := -\sum_{k=1}^{J}\log\left[g(y_{t_k}|X_{t_k}^i)\right] + \frac{1}{2}\int_0^T u^\dagger u ds + \int u^\dagger dW_s + \hat{S}_i \qquad \text{Cost}$$

$$w_i = \exp\left(-S_i\right) / \sum_{j=1}^{N}\exp\left(-S_j\right) \qquad \text{Importance Weights}$$

The cost S_i contains the log-likelihood $\log\left[g(y|x)\right]$ and the IS correction term for diffusion processes. The cost defines the IS weights w_i, that normalized give a point-mass representation of the posterior

Step 3: Update

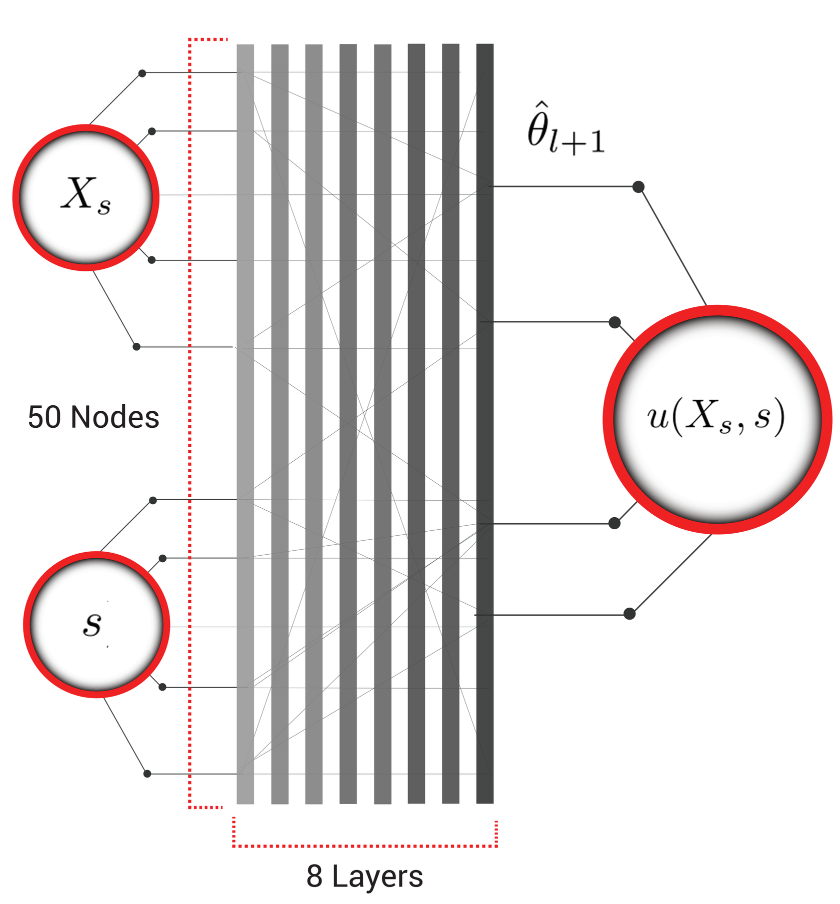
$$\hat{\theta}_{l+1} = \hat{\theta}_l + \eta\left\langle \int_0^T dW_s \frac{\partial \hat{u}(X_s,s|\hat{\theta}_l)}{\partial \hat{\theta}_l}\right\rangle_{u_l}$$

$$\text{Update } \phi_l \text{ e.g. } \mu_l = \langle X\rangle_{u_l},\ \Sigma_l^2 = \langle X \otimes X\rangle_{u_l}$$

$$\langle h(X)\rangle := \sum_{i=1}^{N} w_i h(X^i)$$

Update rule for the control parameters $\hat{\theta}_l$ comes from the cross-entropy method for diffusion processes
The weighted statistics approximate statistics with respect to the posterior
Use statistics of the posterior marginal at $t = 0$ to update the proposal distribution $q_l(X_0|\phi_l)$ in step 0

CONTROL



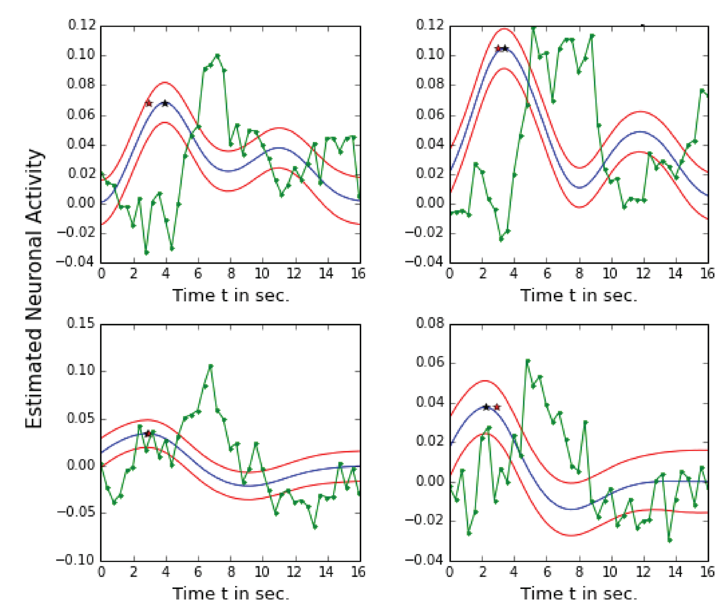
- A **good parametrization** of the controller is crucial. Better approximations to the optimal control function result in **more efficient IS**.
- The choice of the parametrization of the optimal control is an open question.
- PICE places **no restriction** on the parametric form of the importance control function.
- Deep Neural Networks (DNN) posses the flexibility needed to **approximate complex functions**.
- The choice of DNN alleviates the burden of constructing a "good" parametrization.
- Pitfall: over parametrization has a negative impact on the ESS.

RESULTS

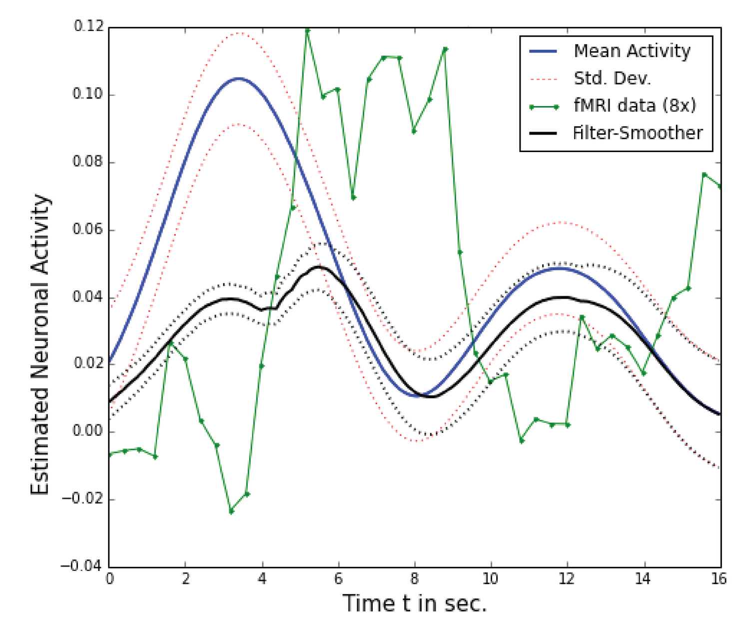
Estimate the neuronal activity of a subject during an experiment, where subjects had to respond to stimuli by pressing a button.
The motor cortex is modeled as a 1-D process: $dz = -zdt + \sigma_z dW_z$

The BOLD response to the neuronal activity is a non-linear 4D coupled system.
Observation noise is Gaussian.

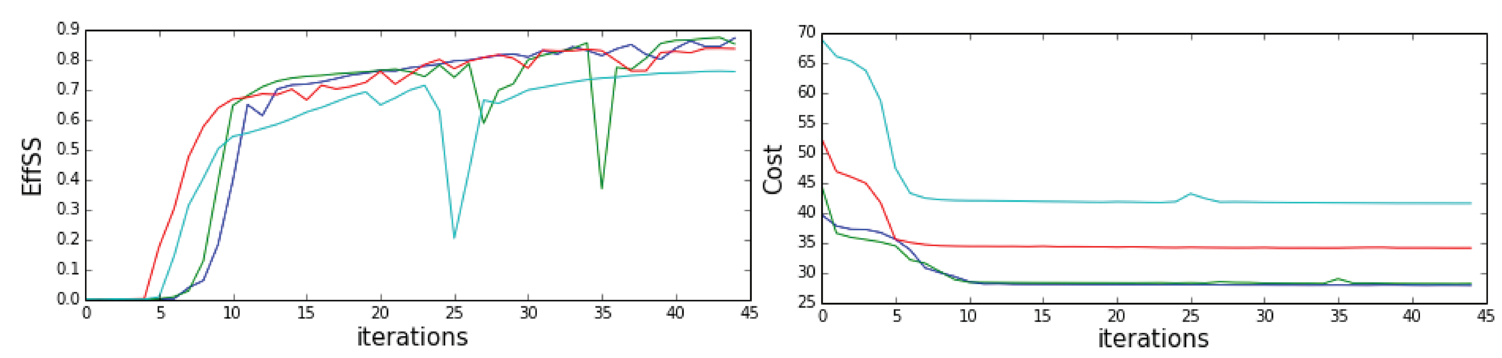
To reconstruct the signal in the motor cortex we iterate PICES 45 times. Per iteration used 500 CPUs with 600 particles each.



Note: No Input assumed
Smoothing estimates of the hidden neuronal activity in the motor cortex (**blue**: mean; **red**: standard deviation). **Black star**: maximum of the mean estimate. **Red star**: measured response time.
The values of the fMRI time-series are scaled with a factor of 8 for comparison **(green)**.



Note: No Input assumed
Comparison of estimates: standard Bootstrap Filter-Smoother **(black)** and PICES **(blue)**
The CPU time for both methods is about 40 minutes.
For BFS we used 500 CPUs with $N = 10^4$ particles per CPU.



Effective Sampling Size vs iterations (left) for four fMRI time series: The initial fraction of ESS is around 3×10^{-6}, i.e. only one particle out of the $N = 3 \times 10^5$ has a significant contribution to the estimations.
Right: The mean cost $\mathbb{E}_u\left[S\right]$ under the control at each iteration.

Note: A shallow network or a simple linear feedback controller performed significantly worse for fMRI data.

CONCLUSIONS

- The prior process is **adapted with a control drift**, the importance controller.
- This adaptation reduces the total cost incurred by the entire time series.
  This minimization is accompanied by an **increase in the sampling efficiency**.
- The parametrization of the importance controller is arbitrary. Here, we use a deep neural network with 8 hidden layers and 50 nodes width each.
- Our results on fMRI show that adaptive importance sampling via a controlled diffusion process improves the efficiency of the sampler **five orders** of magnitude.
- The feedback control accounts in a certain degree for **model mismatch** due to the lack of inputs.
- There is **no restriction** on the drift and diffusion terms of the prior process, nor on the observation model.
- The sampling and the gradient computations are **easily parallelizable** and can be implemented efficiently in a distributed manner.
- **PICES is an alternative to particle smoothing methods**.

OUTLOOK

In some cases, the optimal control shows **discontinuities in time**. The correct approximation of these discontinuities has a major **impact on the scalability** of the method to very large time series. How can we ensure that the neural network learns efficiently with enough time resolution to scale the method?

We need a better understanding of the impact of the network architecture on the efficiency of the sampler.
What type of architectures and transfer functions facilitate high temporal resolution?
Other parametrizations of the controllers may prove more useful, especially when the time scale of the optimal control function is very short.

The introduction of an **adaptive annealing procedure** is beneficial when the variance of the estimates are too large to learn accurate parameters. However, this introduces a bias in the gradient estimation.
How does this affect the performance of the controller after convergence?
How to choose the annealing threshold?

ACKNOWLEDGES

This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.
The work of H.-Ch. Ruiz is supported by the European Commission through the FP7 Marie Curie Initial Training Network 289146, NETT: Neural Engineering Transformative Technologies.

REFERENCES

H. J. Kappen and H.-C. Ruiz. Adaptive importance sampling for control and inference. Journal of Statistical Physics, 2016.

H.-C. Ruiz and H. J. Kappen. Particle Smoothing for Hidden Diffusion Processes: Adaptive Path Integral Smoother. Under review at IEEE TSP; arXiv preprint arXiv:1605.00278, 2016.

Hans Ruiz: hruiz@science.ru.nl

SNN ADAPTIVE INTELLIGENCE NETT SURF SARA Radboud University Radboudumc