Natural gradient enables fast sampling in spiking neural networks

Paul Masset*, Jacob A. Zavatone-Veth*
J. Patrick Connor, Venkatesh N. Murthy# and Cengiz Pehlevan#

NEURAL INFORMATION PROCESSING SYSTEMS

@paul_masset

@jzavatoneveth

HARVARD Department of Physics
Harvard John A. Paulson School of Engineering and Applied Sciences
MCB Department of Molecular and Cellular Biology
Center for Brain Science

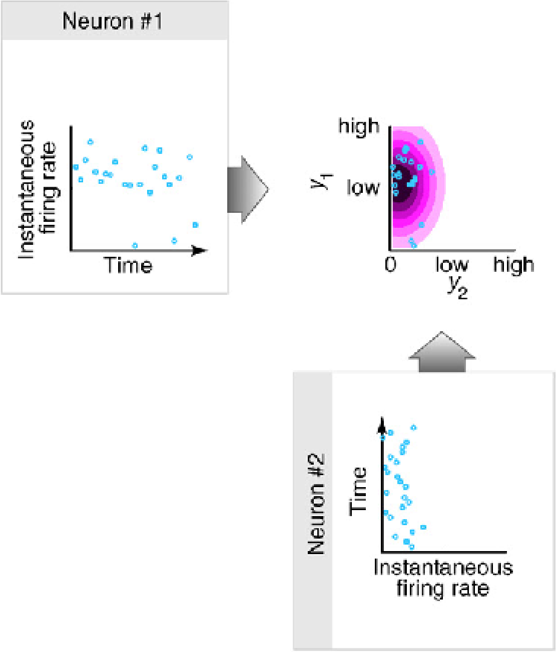
Ambiguity in perception demands probabilistic processing

Ambiguity in perception demands probabilistic processing

Ambiguity in perception demands probabilistic processing

Probabilistic computations need to be efficient in high-dimension

Sampling-based probabilistic inference in neural circuits

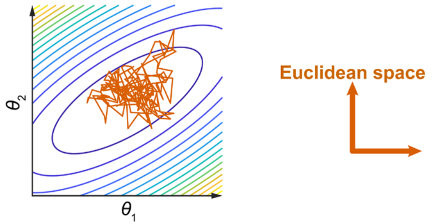


A simple method to sample a distribution $p(\mathbf{z}) \propto \exp[-U(\mathbf{z})]$ in a rate network is using Langevin dynamics

$$d\mathbf{z}(t) = -\nabla U(\mathbf{z})\, dt + \sqrt{2}\, d\mathbf{W}(t)$$

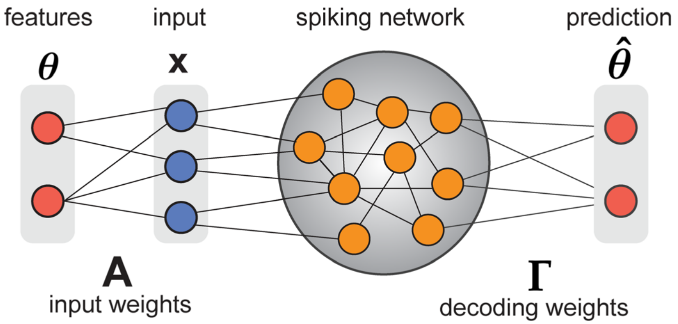
For a Gaussian $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, this yields the linear dynamics

$$d\mathbf{z}(t) = -\mathbf{\Sigma}^{-1}\mathbf{z}\, dt + \sqrt{2}\, d\mathbf{W}(t)$$



Fiser et al., TICS, 2010; Hennequin et al., NeurIPS 2014

Sampling-based inference in spiking neural networks



However, this approach has only been used to sample 2-dimensional Gaussians and has not been shown to scale to higher-dimensional distributions.

Boerlin et al., PLoS Comp. Bio., 2013, Savin & Denève, NeurIPS, 2014

The "complete recipe" for stochastic gradient MCMC

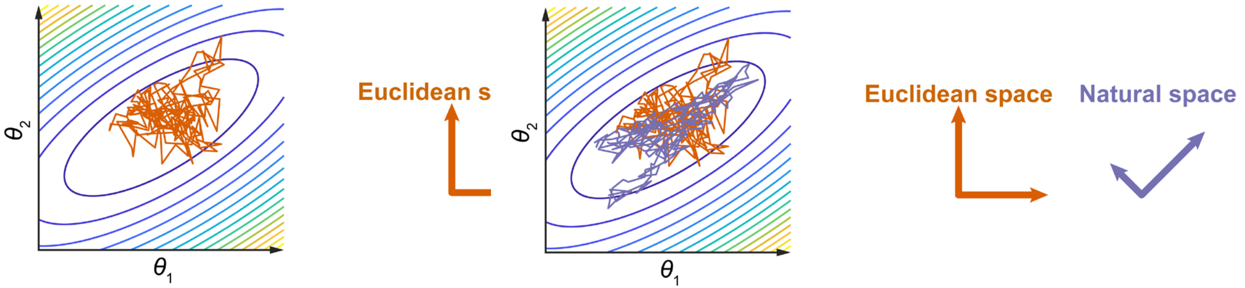
Naïve Langevin sampling becomes slow in high dimensions

Ma et al. proposed the 'complete recipe' for stochastic gradient MCMC, which shows that all diffusions that sample $p(\mathbf{z}) \propto \exp[-U(\mathbf{z})]$ can be written as

$$d\mathbf{z}(t) = \{-[\mathbf{D}(\mathbf{z}) + \mathbf{S}(\mathbf{z})]\nabla U(\mathbf{z}) + \mathrm{div}[\mathbf{D}(\mathbf{z}) + \mathbf{S}(\mathbf{z})]\}\, dt + \sqrt{2\mathbf{D}(\mathbf{z})}\, d\mathbf{W}(t)$$

for a symmetric matrix field $\mathbf{D}$ and skew-symmetric matrix field $\mathbf{S}$.

This captures many accelerated algorithms, including Riemannian and Hamiltonian dynamics



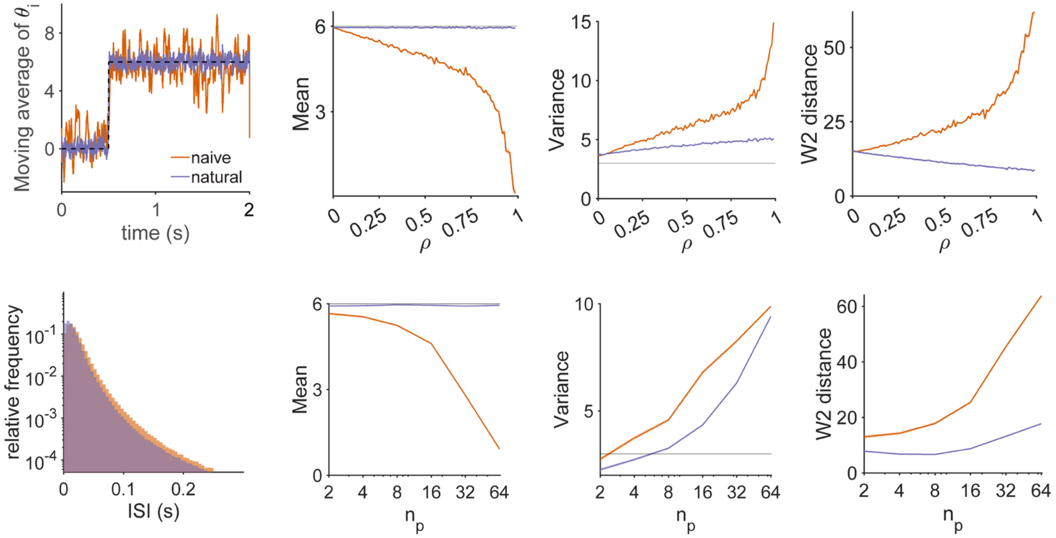
We can leverage these principles to achieve fast sampling in spiking networks

Ma et al., NeurIPS, 2015

Fast sampling in a network approximating Langevin sampling

With naïve geometry, the Savin-Denève network cannot sample from strongly-correlated distributions

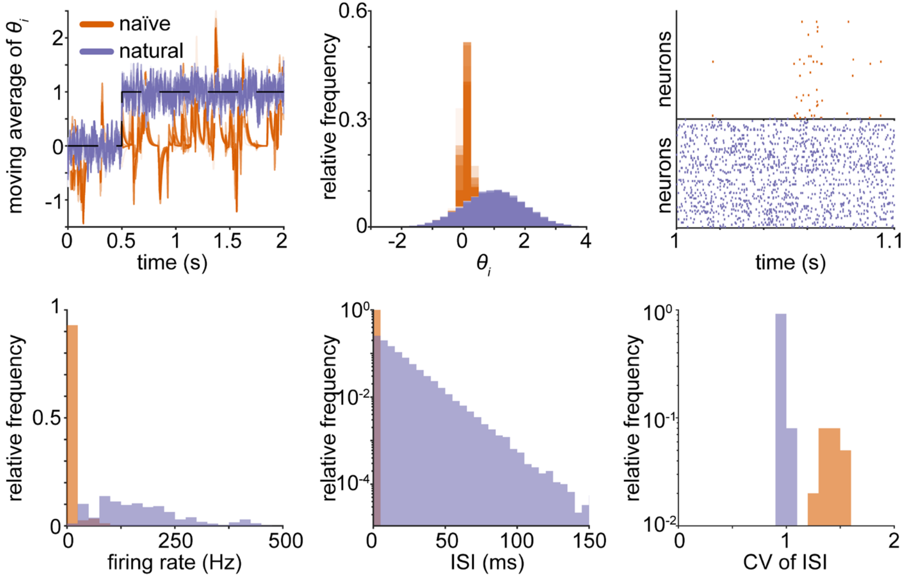
But, with natural gradients, it can sample accurately even in high dimensions



Here, we sample from a 10-dimensional equicorrelated Gaussian with $\rho = 0.75$

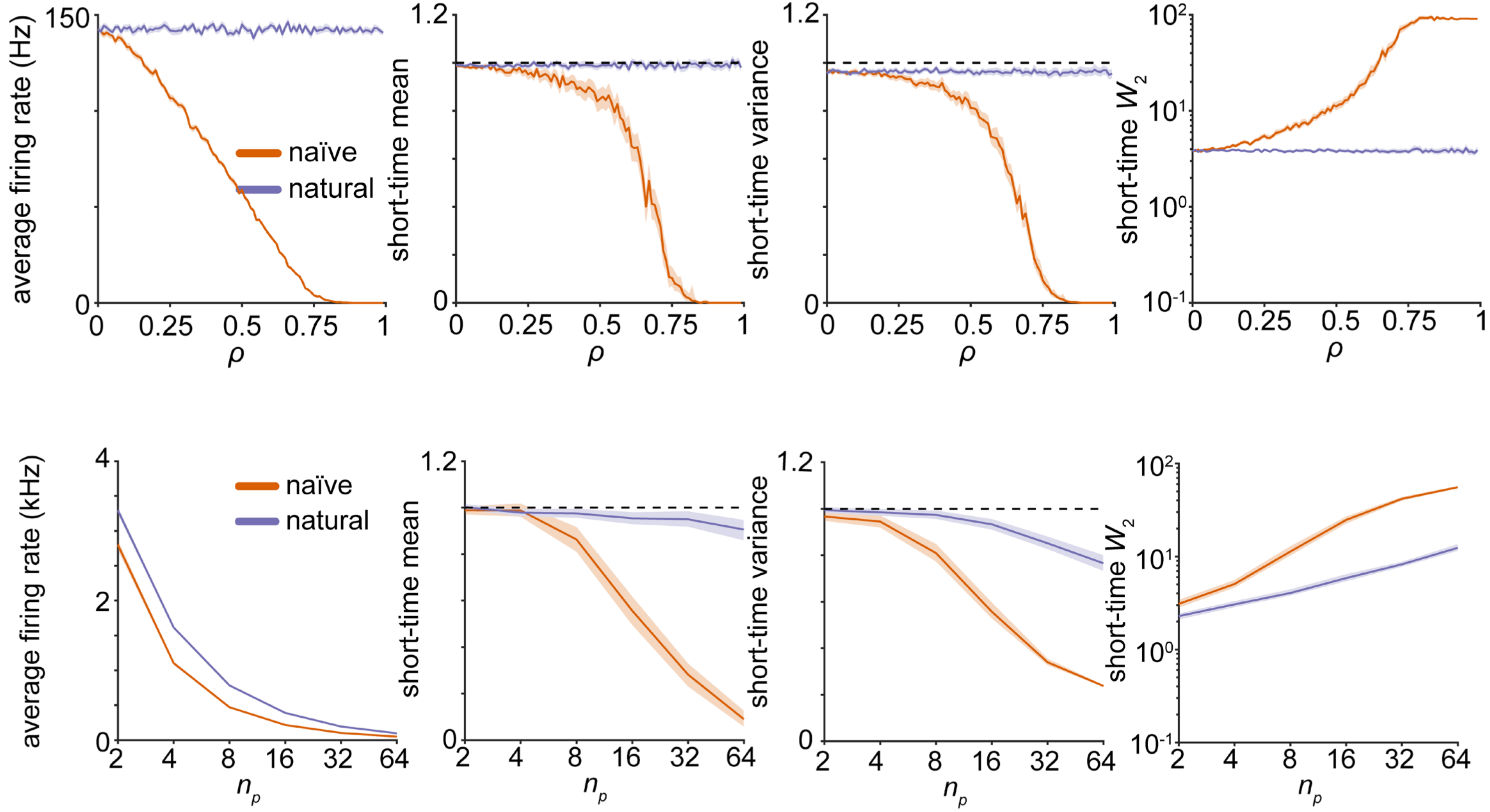
Fast sampling in a network with probabilistic spiking

We can also leverage these principles to achieve fast sampling in a network with probabilistic spiking, derived from Metropolis-Hastings sampling.



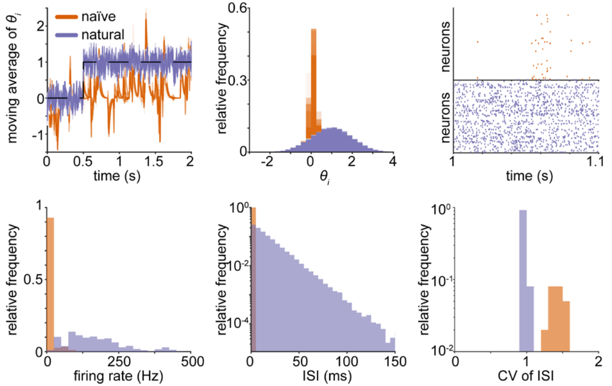
Here, spiking is suppressed entirely with naïve geometry

Fast sampling in a network with probabilistic spiking



Conclusion

We show that the “complete recipe” enables fast sampling in spiking neural networks simulating Langevin sampling with noisy voltage dynamics and sampling using probabilistic spike rules



In our paper, we provide a unifying derivation of spiking networks designed to sample from Gaussian distributions using probabilistic rules, of which the efficient balanced network is a limit.

$$\dot{\mathbf{V}}(t) = -\frac{1}{\tau_m}\mathbf{V}(t) - \boldsymbol{\Gamma}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\mathbf{o}(t) + \boldsymbol{\Gamma}^\top\boldsymbol{\Sigma}^{-1}\left(\dot{\boldsymbol{\mu}}(t) + \frac{1}{\tau_m}\boldsymbol{\mu}(t)\right)$$

Acknowledgements



Patrick Connor | Venki Murthy | Cengiz Pehlevan

Members of the Pehlevan & Murthy Groups

Funding:

HARVARD Mind Brain Behavior

NIH National Institutes of Health *Turning Discovery Into Health*

Google Research

HARVARD *Brain Science Initiative*

NSF

Link to paper: