Stochastic Continuous Normalizing Flows
Training SDEs as ODEs

Liam Hodgkinson, Chris van der Heide, Fred Roosta, Michael W. Mahoney

Neural ODEs

Ordinary differential equations (ODEs) formed from single-(or multi-)layer neural networks are called ***neural ODEs***:

$$\frac{\mathrm{d}}{\mathrm{d}t}Z(t) = f(Z(t), t, \theta). \quad (1)$$

They can be trained with the ***adjoint method*** [1].

Stratonovich Correction

STEP 1

Because the chain rule does not hold in Itô calculus, *naive approximations do not work.*

Solution: Use a different calculus

The Itô SDE (3) can be ***converted to a Stratonovich SDE***:

$$\mathrm{d}Z_t = \tilde{\mu}(Z_t, t, \theta)\mathrm{d}t + \sigma(Z_t, t, \theta) \circ \mathrm{d}B_t, \quad (4)$$

using a ***Stratonovich drift correction***.

Stratonovich SDEs satisfy the chain rule, and ***behave like ODEs*** (reversible, approximable).

Regularization

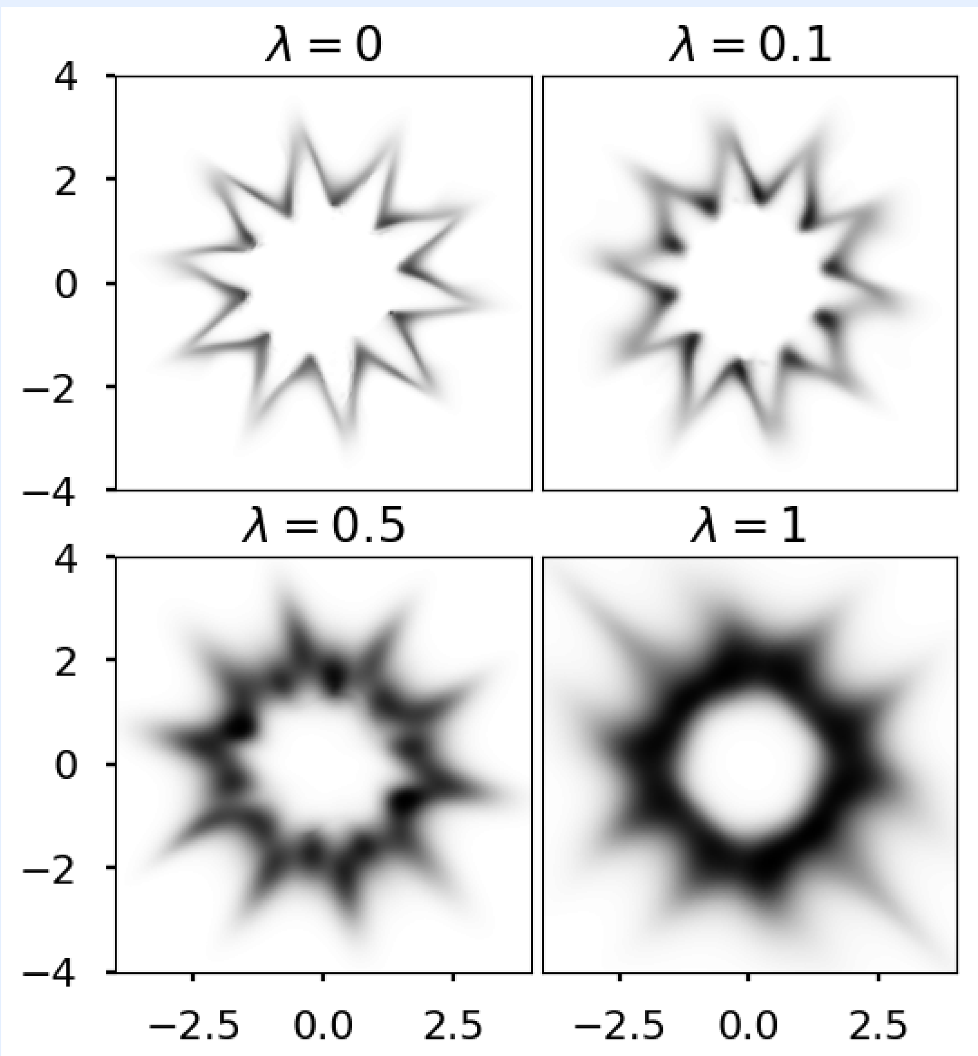

Figure 2: Fitting a star-shaped density with ***stochastic continuous normalizing flows*** with drift parameterized by a 3-layer NN, and diffusion coefficient λI. Increasing the diffusion adds a ***regularization*** effect to the fit.

Acknowledgements

This work has been supported by the Australian Research Council Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS), under grant number CE140100049. We would also like to acknowledge DARPA, NSF, and ONR for providing partial support.

Continuous Normalizing Flow

Under the assumption that $Z(0) \sim p_0(\theta)$, (1) becomes a ***generative model*** — a continuous normalizing flow [1, 2].

It can be trained by ***maximum likelihood estimation*** or ***variational inference*** using

Theorem (Chen et al., 2018). *If $Z(t)$ satisfies (1), the probability density p_t of $Z(t)$ satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t}\log p_t(Z(t)) = -\nabla_z \cdot f(Z(t), t, \theta) \quad (2)$$

Wong–Zakai Approximation

STEP 2

Let $B(t) \approx B_t$ be a pathwise ***approximation*** of Brownian motion.

Approximations include:

- ***Piecewise linear:*** Linear interpolation of Gaussian random walk—recovers Euler scheme.
- ***Karhunen-Loève expansion:*** A smooth Fourier series for Brownian motion.

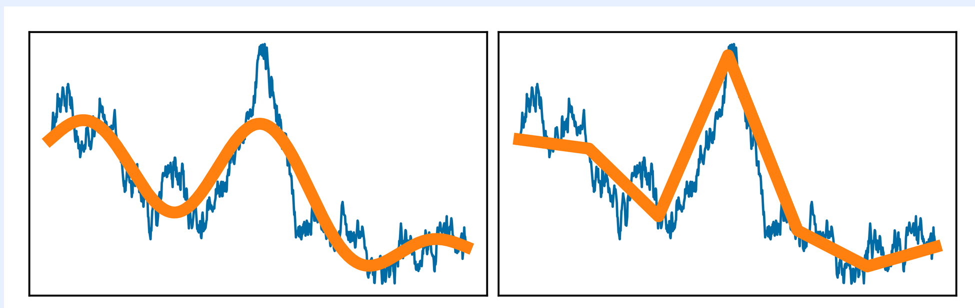

Figure 1: Approximations (orange) of Brownian motion (blue): (left) Karhunen-Loève; (right) linear spline/Euler

A ***Wong–Zakai approximation*** of (4) is the random ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}Z(t) = \tilde{\mu}(Z(t), t, \theta) + \sigma(Z(t), t, \theta)\frac{\mathrm{d}B(t)}{\mathrm{d}t} \quad (5)$$

Samplers

Langevin diffusions are commonly used as ***sampling algorithms***. Can we learn them?

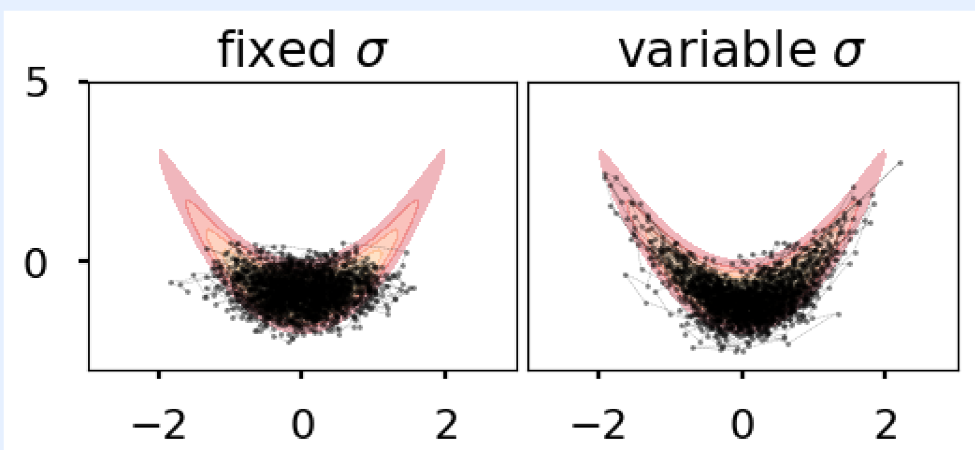

Figure 3: Simulating SDEs trained to a banana target distribution with drift parameterized by a 3-layer NN. (left) Diffusion coefficient is fixed $\sigma \equiv I$; (right) Diffusion coefficient parameterized by 3-layer NN.

Neural SDEs

An Itô stochastic differential equation (SDE) is of the form

$$\mathrm{d}Z_t = \mu(Z_t, t, \theta)\mathrm{d}t + \sigma(Z_t, t, \theta)\mathrm{d}B_t, \quad (3)$$

where B_t is **Brownian motion**, μ is the drift, and σ the diffusion coefficient.

(Neural) SDEs are (neural) ODEs with injected noise.

Problem: Training SDEs is often complex.

Objective: Develop a general procedure for training SDEs using ODE training procedures.

Three Steps:

1. Perform a Stratonovich drift correction
2. Approximate the Brownian motion
3. Train the resulting random ODE

Training a Random ODE

STEP 3

The approximation (5) is an ***ODE with latent variable*** $B(t)$. It can be trained ***using any ODE training procedure***.

Metatheorem. *The following procedures are* **consistent** *as $B(t) \to B_t$.*

- ***Stochastic adjoint method:*** Adjoint method applied to (5)—equivalent to [3].
- ***Density estimation:*** Monte Carlo estimation for density using (2).
- ***Stochastic continuous normalizing flow (SCNF):*** Perform semi-implicit variational inference using (2).

References

[1] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*, 2018.

[2] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

[3] Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. In *International Conference on Artificial Intelligence and Statistics*, pages 3870–3882. PMLR, 2020.