Integrable Nonparametric Flows

David Pfau and Danilo Rezende

{pfau,danilor}@google.com

DeepMind

Background and Motivation

Normalizing flows [1]: represent a probability distribution with an invertible transformation of a base distribution to a target distribution

$$\log p_T(\mathbf{x}) = \log p_0\left(f^{-1}(\mathbf{x})\right) - \log\left|\mathbf{J}_f\left(f^{-1}(\mathbf{x})\right)\right|$$

Neural ODEs [2]: represent the invertible transformation as an ODE

$$f(\mathbf{x}(0)) = \mathbf{x}(T) \qquad \frac{\partial \mathbf{x}}{\partial t} = \mathbf{v}(\mathbf{x})$$

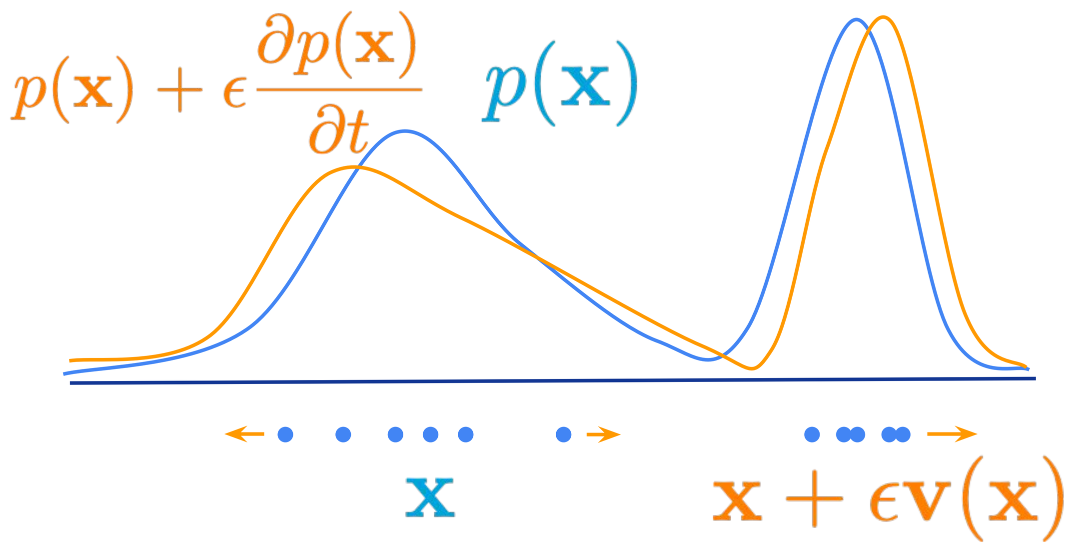
$$\frac{\partial \log p(\mathbf{x})}{\partial t} = -\mathbf{v}(\mathbf{x})^T \nabla \log p(\mathbf{x}) - \nabla \cdot \mathbf{v}(\mathbf{x})$$

$$\frac{\partial p(\mathbf{x})}{\partial t} = -\nabla \cdot (p(\mathbf{x})\mathbf{v}(\mathbf{x}))$$

(also appears in Stein operator, Fokker-Planck equation, and continuity equation for conservation of mass)

Typically, we know initial $p_0(\mathbf{x})$ and final samples $\mathbf{x}(T)$, and solve for $\mathbf{v}$ by maximum likelihood. *What if instead, we knew p(**x**) and infinitesimal change dp(**x**)/dt? How would we solve for **v** then?*

Method



v is underdetermined - restrict p**v** to be *integrable*

$$\nabla u(\mathbf{x}) = p(\mathbf{x})\mathbf{v}(\mathbf{x})$$

Then can solve *Poisson equation*

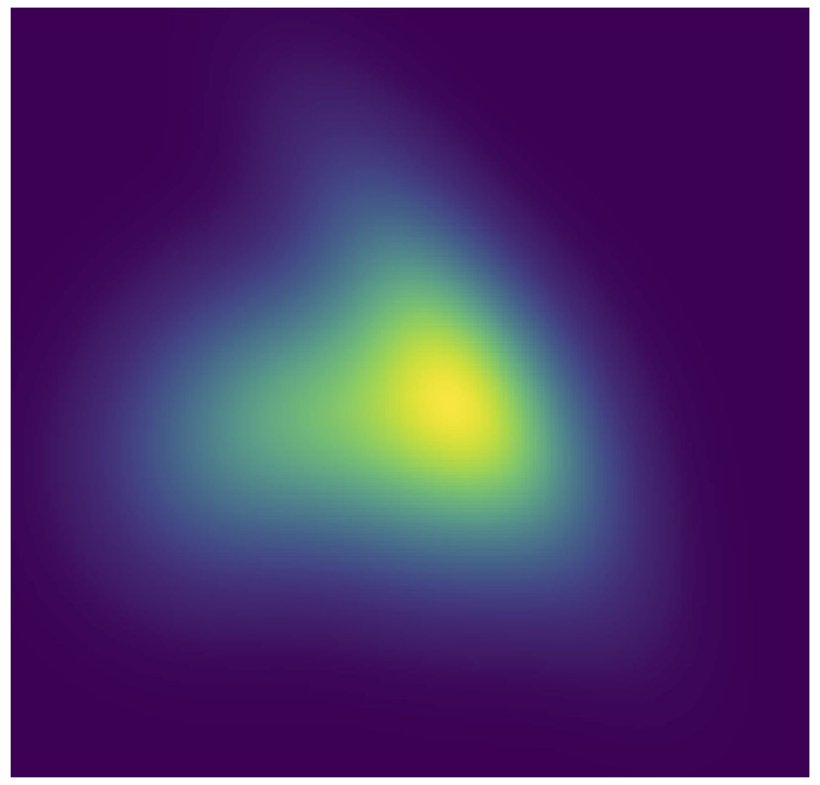
$$\nabla^2 u(\mathbf{x}) = -\frac{\partial p(\mathbf{x})}{\partial t}$$

Equivalent to treating p**v** as *electric field* created by particles with charge dp(x)/dt. Can be solved empirically by Coulomb kernel:

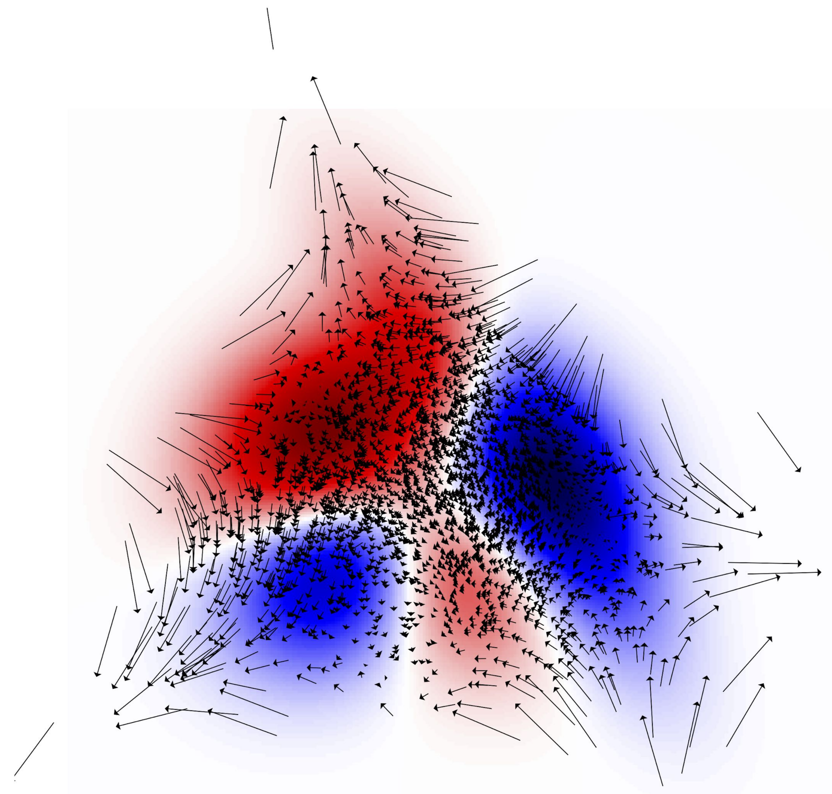
$$p(\mathbf{x}_i)\mathbf{v}(\mathbf{x}_i) = \frac{1}{N-1}\sum_{j\neq i}\frac{\partial \log p(\mathbf{x}_j)}{\partial t}\frac{\Gamma\left(\frac{n}{2}\right)(\mathbf{x}_i - \mathbf{x}_j)}{2\pi^{n/2}|\mathbf{x}_i - \mathbf{x}_j|^n}$$

Unlike Stein variational gradient descent [3] there is only one possible kernel, which is known to be *optimal* for related applications [4].

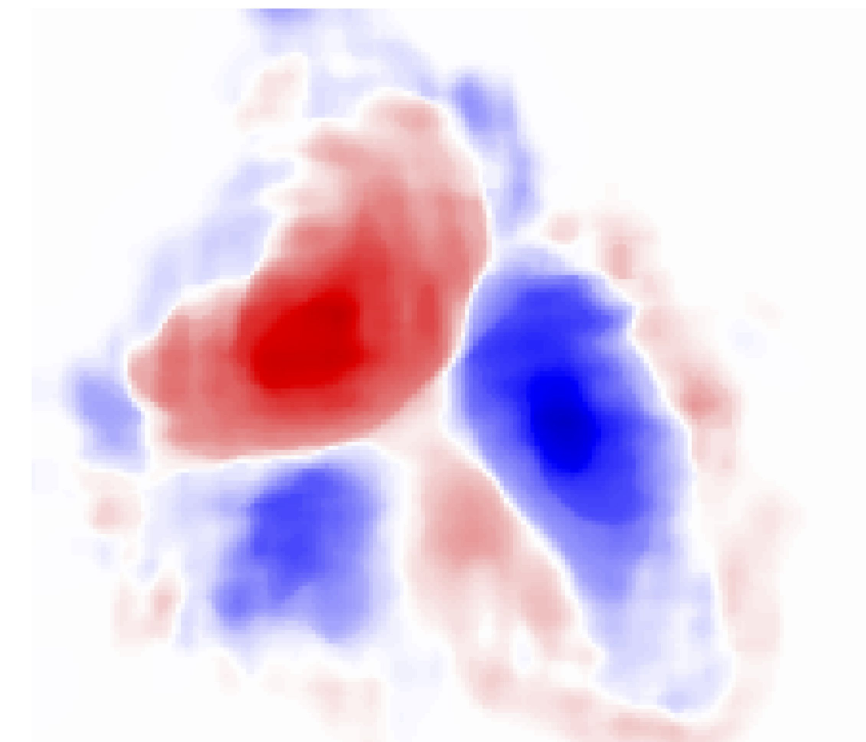
Results



Density



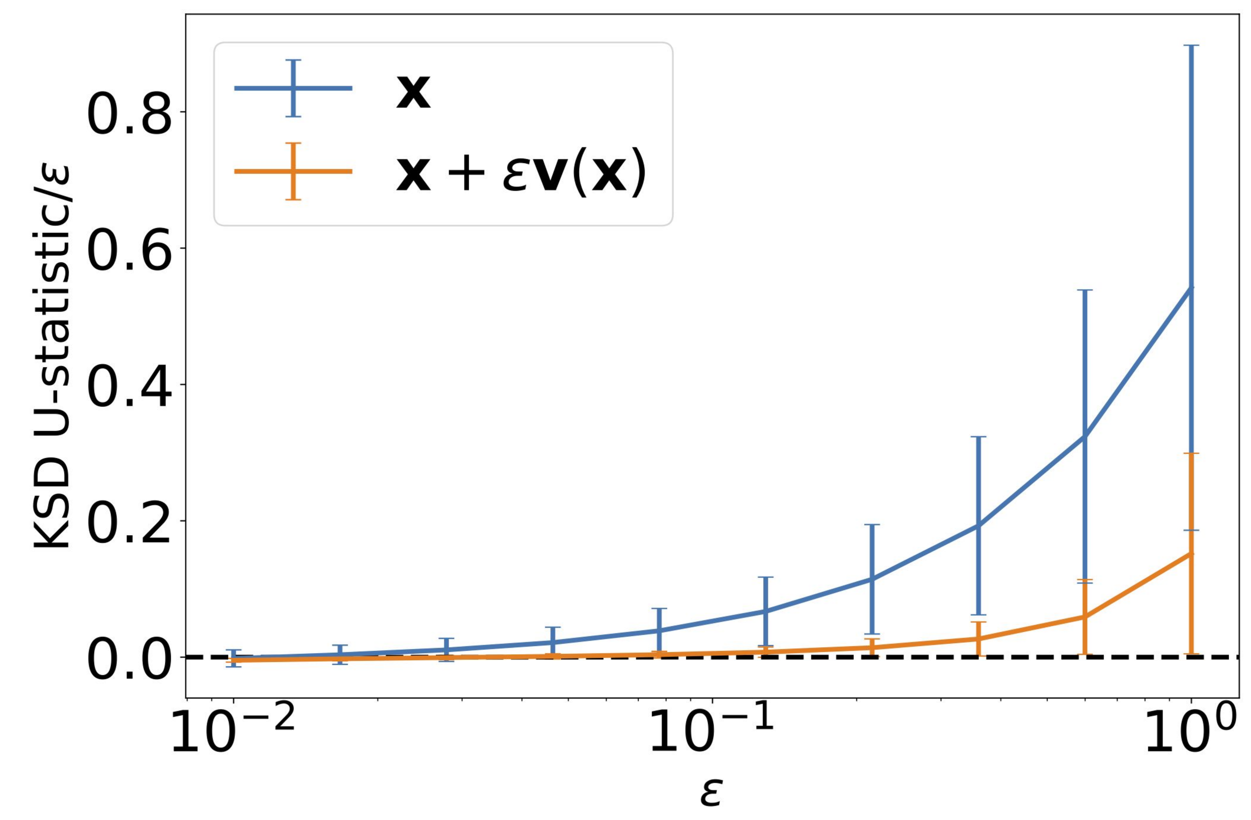
Perturbation to density and Integrable flow field



Kernel density estimate of perturbation from particles

Mixture of Gaussians in 2D:

- 10k samples, perturb the means
- Evaluate with kernelized Stein discrepancy [5, 6]
- Flow-perturbed samples closely match perturbed density across a range of scales



Kernelized Stein Discrepancy between particles and perturbed density

Future Directions

Applications: Accelerating convergence when sampling and optimization are coupled:

$$\min_\theta \mathbb{E}_\mathbf{x}\left[f_\theta(\mathbf{x})\right] \rightleftarrows \mathbf{x} \sim p_\theta(\mathbf{x})$$

- Variational Quantum Monte Carlo
  - *f*=energy, *p*=wavefunction
- Policy Gradient Methods
  - *f*=value, *p*=policy
- Variational Inference
  - *f*=ELBO, *p*=variational posterior

Scaling: How to beat the curse of dimensionality

- How do we get around the need for the partition function when we only have an unnormalized distribution?
- How do we get around the r^{n-1} dropoff in field strength in high dimensions?

References

[1] D. Rezende and S. Mohamed, "Variational Inference with Normalizing Flows" (2015). ICML

[2] R. T. Q. Chen, Y. Rubanova, J. Bettencourt and D. Duvenaud, "Neural Ordinary Differential Equations" (2018). NeurIPS

[3] Q. Liu and D. Wang, "Stein Variational Gradient Descent" (2016). NeurIPS

[4] S. Hochreiter and K. Obermayer, "Optimal Kernels for Unsupervised Learning" (2005). IJCNN

[5] Q. Liu, J. Lee and M. I. Jordan, "A Kernelized Stein Discrepancy for Goodness-of-Fit Tests and Model Evaluation" (2016). ICML

[6] K. Chwialkowski, H. Strathmann and A. Gretton, "A Kernel Test of Goodness of Fit" (2016). ICML

`https://tinyurl.com/integrable-flows`