Make Stein's method great again for score-based generative models

Supervisor:
Yingzhen Li (yingzhen.li@imperial.ac.uk)

Short description:
Score-based generative models and diffusion models have become the state-of-the-art for generative modelling. Key to the success is the efficient estimation of noisy score function by denoising approaches. On the other hand, minimising Stein discrepancy as another line of score estimation technique has yet to see its success in the generative modelling field. This project has the ambition to challenge this status quo.

Long description:
Score-based generative models and diffusion models have become the state-of-the-art for generative modelling. Key to the success is the efficient estimation of noisy score function by denoising approaches. On the other hand, Stein's method as another line of score-based techniques has been successful for statistical and machine learning tasks such as goodness-of-fit tests, control variates and approximate Bayesian inference. But Stein's method has yet to succeed in the generative modelling field. Specifically, Stein discrepancy is a type of integral probability metrics which relies on finding the best test function to tell apart the difference between two distributions. A neural test function based approach would induce an adversarial training procedure like GANs for minimising Stein discrepancy, while kernel based Stein's method struggles to scale to high dimensions without the help of neural networks.

This project has the ambition to challenge this status quo and see to what extent we can push Stein's method to the extreme for generating realistic images and/or other types of static data. In particular, we will spend time on the following studies:

1. Investigate the Stein gradient estimator and scale it up for image data. This requires the usage of deep kernels (which are kernels build on top of neural network features), and a suitable algorithm for training such neural network features will be developed. This improved Stein gradient estimator will be tested using score-based generative models.
2. Compare and understand the properties of the estimated score functions from Stein's method vs other approaches such as denoising methods and minimising Fisher divergences. We will study the properties of the estimated vector fields, e.g., manifold behaviour and (local) conservativeness, and understand the impact of these properties for data generation.

Minimum viable thesis:
A straight-forward replacement of the score estimation method in Score-based generative models (Song et al. ICLR 2021) with the Stein gradient estimator (Li and Turner 2018). Documenting down the comparisons.

Required background & skills:
Student suitable for this project would have strong mathematical analysis skills. They should feel very comfortable in derivations with basic probability & statistics, linear algebra and calculus. They should also have experience with existing deep learning frameworks (e.g. Tensorflow or Pytorch). Having hands-on experience with generative models such as VAEs, GANs and diffusion models will be a plus factor.

Some References:

Score-based generative models:
Generative Modeling by Estimating Gradients of the Data Distribution
Improved Techniques for Training Score-Based Generative Models
Score-Based Generative Modeling through Stochastic Differential Equations
Score-based Generative Modeling in Latent Space
Maximum Likelihood Training of Score-Based Diffusion Models
Riemannian Score-Based Generative Modelling

Score estimation and score-matching:
Gradient Estimators for Implicit Models
A Spectral Approach to Gradient Estimation for Implicit Distributions
Sliced Score Matching: A Scalable Approach to Density and Score Estimation
Estimating High Order Gradients of the Data Distribution by Denoising
Estimation of Non-Normalized Statistical Models by Score Matching
A Connection Between Score Matching and Denoising Autoencoders

Stein's method in machine learning:
Measuring Sample Quality with Stein's Method
Measuring Sample Quality with Kernels
A Kernelized Stein Discrepancy for Goodness-of-fit Tests and Model Evaluation
A Kernel Test of Goodness of Fit
Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm
Learning the Stein Discrepancy for Training and Evaluating Energy-Based Models without Sampling
Sliced Kernelized Stein Discrepancy
Stein's Method Meets Computational Statistics: A Review of Some Recent Developments

Other related:
How to Train Your Energy-Based Models
Implicit Generation and Generalization in Energy-Based Models