

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe¹ and Kyle Cranmer¹

¹New York University

In this note, ... [GL: todo.]

I. INTRODUCTION

[GL: Prescribed vs. implicit. See case of non-diff models in Balaji et al.]

II. PROBLEM STATEMENT

We consider a family of parameterized densities $p_\theta(\mathbf{x})$ defined implicitly through the simulation of a stochastic generative process, where $\mathbf{x} \in \mathbb{R}^d$ is the data and θ are the parameters of interest. The simulation may involve some complicated latent process, such that

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \quad (1)$$

where $\mathbf{z} \in \mathbb{R}^m$ is a latent variable providing an external source of randomness.

We assume that we already have an accurate simulation of the stochastic generative process that defines $p_\theta(\mathbf{x}|\mathbf{z})$, as specified through a deterministic function $g(\cdot; \theta) : \mathbb{R}^m \rightarrow \mathbb{R}^d$. That is,

$$p_\theta(\mathbf{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_d} \int_{\{\mathbf{z}: g(\mathbf{z}; \theta) \leq \mathbf{x}\}} p(\mathbf{z})d\mathbf{z}. \quad (2)$$

The simulator g is assumed to be a non-invertible function, that can only be used to generate data in forward mode. For this reason, evaluating the integral in Eqn. 2 is intractable. Importantly, and as increasingly found in science, we consider the additional constraint that g is a non-differentiable model, e.g. as when specified as a computer program.

Given some observed data $\{\mathbf{x}_i | i = 1, \dots, N\}$ drawn from the (unknown) true distribution p_r , our goal is the inference of the parameters of interest θ^* that minimize the divergence between p_r and the modeled data distribution p_θ induced by $g(\cdot; \theta)$. That is,

$$\theta^* = \arg \min_{\theta} \rho(p_r, p_\theta), \quad (3)$$

where ρ is some divergence.

III. BACKGROUND

A. Generative adversarial networks

Generative adversarial networks (GAN) were first proposed by [2] as a way to build an implicit generative model capable of producing samples from random noise \mathbf{z} . More specifically, a generative model $g(\cdot; \theta)$ is pit against an adversarial network d whose antagonistic objective is to recognize real data \mathbf{x} from generated data $g(\mathbf{z}; \theta)$. Both models g and d are trained simultaneously, in such a way that g learns to maximally confuse its adversary d (which happens when g produces samples comparable to the observed data), while d continuously adapts to changes in g . At the equilibrium and assuming enough capacity in the networks, it can be shown that g induces a distribution that is indistinguishable from the distribution of the observed data \mathbf{x} .

More recently, ...

[GL: Explain WGAN, loss and optimum.]

B. Variational optimization

IV. ADVERSARIAL VARIATIONAL OPTIMIZATION

V. EXPERIMENTS

A. Toy problem

B. Physics example

VI. RELATED WORKS

[GL: Implicit generative models.] [GL: ABC.] [GL: carl [1].] [GL: Wood's papers.] [GL: CMA-ES.]

VII. SUMMARY

ACKNOWLEDGMENTS

KC and GL are both supported through NSF ACI-1450310, additionally KC is supported through PHY-1505463 and PHY-1205376.

-
- [1] CRANMER, K., PAVEZ, J., AND LOUPPE, G. Approximating likelihood ratios with calibrated discriminative classifiers. *arXiv preprint arXiv:1506.02169* (2015).
 - [2] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A., AND BENGIO, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems* (2014), pp. 2672–2680.