Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe, Joeri Hermans and Kyle Cranmer

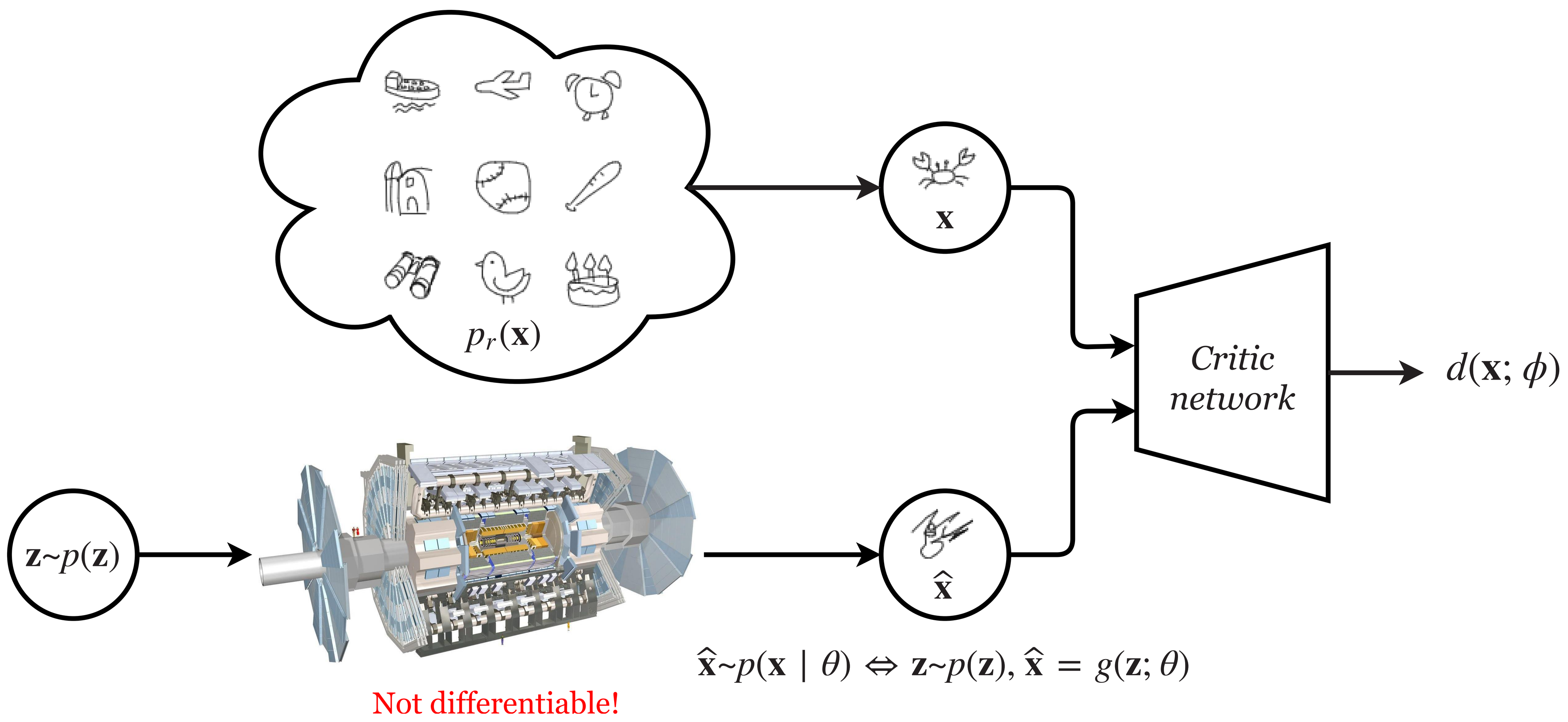
Presented at AISTATS 2019, Osaka, Japan ✉ g.louppe@uliege.be 🐦 glouppe

Abstract

Complex computer simulators are increasingly used across fields of science as generative models tying parameters of an underlying theory to experimental observations. Inference in this setup is often difficult, as simulators rarely admit a tractable density or likelihood function. **We introduce Adversarial Variational Optimization (AVO), a likelihood-free inference algorithm for fitting a non-differentiable generative model**. We adapt the training procedure of generative adversarial networks by replacing the differentiable generative network with a domain-specific simulator. We solve the resulting non-differentiable minimax problem by minimizing variational upper bounds of the two adversarial objectives. Effectively, **the procedure results in learning a proposal distribution over simulator parameters, such that the JS divergence between the marginal distribution of the synthetic data and the empirical distribution is minimized**. We evaluate and compare the method with simulators producing both discrete and continuous data.

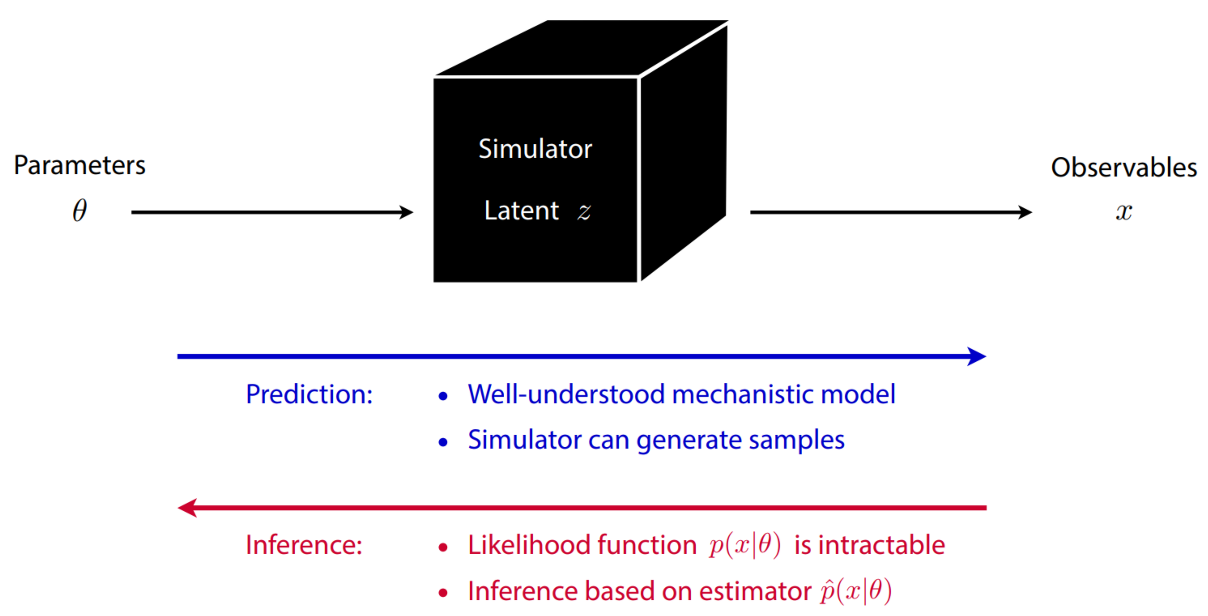
tl ;dr.

1. Take the adversarial training setup of GANs.
2. Replace the generator network with a scientific simulator.
3. Bypass the non-differentiability with REINFORCE.



$$\mathcal{L}_d(\phi) = \mathbb{E}_{\mathbf{x}\sim p_r(\mathbf{x})}[-\log(d(\mathbf{x};\phi))] + \mathbb{E}_{\hat{\mathbf{x}}\sim p(\mathbf{x}|\theta)}[-\log(1-d(\hat{\mathbf{x}};\phi))]$$
$$\mathcal{L}_g(\theta) = \mathbb{E}_{\hat{\mathbf{x}}\sim p(\mathbf{x}|\theta)}[\log(1-d(\hat{\mathbf{x}};\phi))]$$

Likelihood-free inference



In scientific simulators, the likelihood of observations $\mathbf{x}$ given model parameters $\boldsymbol{\theta}$ is implicitly defined as

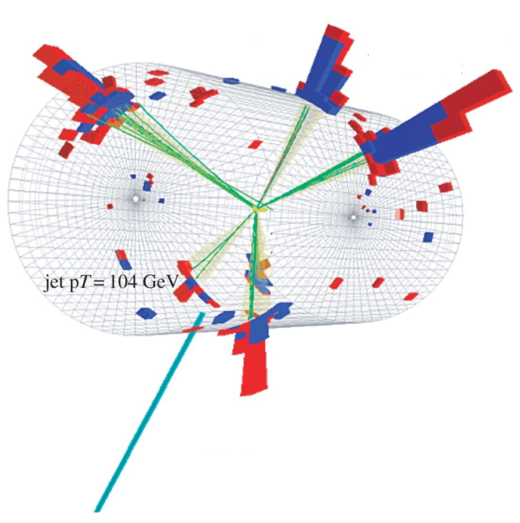
$$p(\mathbf{x}|\boldsymbol{\theta}) = \int p(\mathbf{x}|\mathbf{z},\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta})d\mathbf{z}.$$

This makes it **intractable to evaluate**.
Our goal is to estimate the parameters $\boldsymbol{\theta}^*$ that minimize the JSD divergence between the (empirical) data distribution $p_r(\mathbf{x})$ and the implicit model $p(\mathbf{x}|\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \text{JSD}(p_r(\mathbf{x}), p(\mathbf{x}|\boldsymbol{\theta})).$$

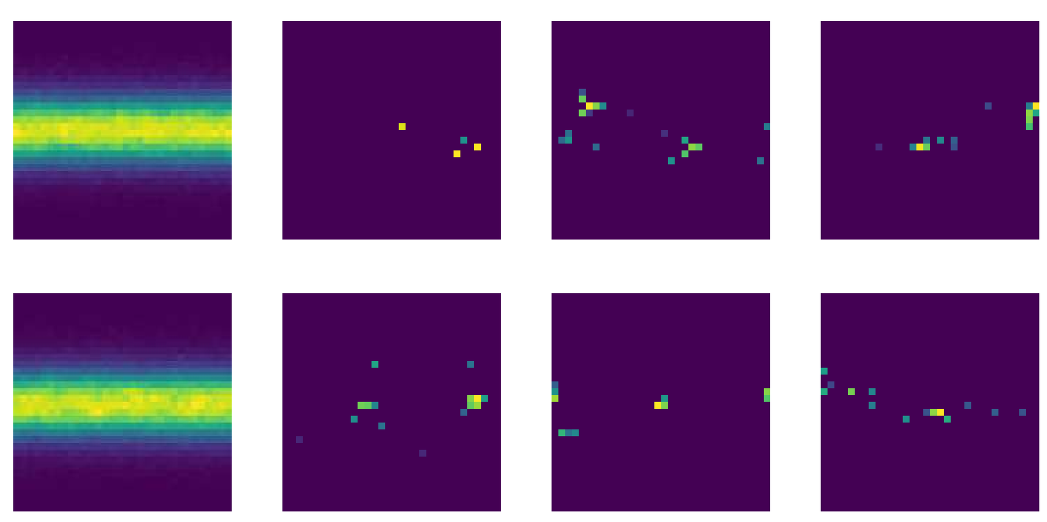
Examples. Particle physics, population genetics, epidemiology, climate science, cosmology.

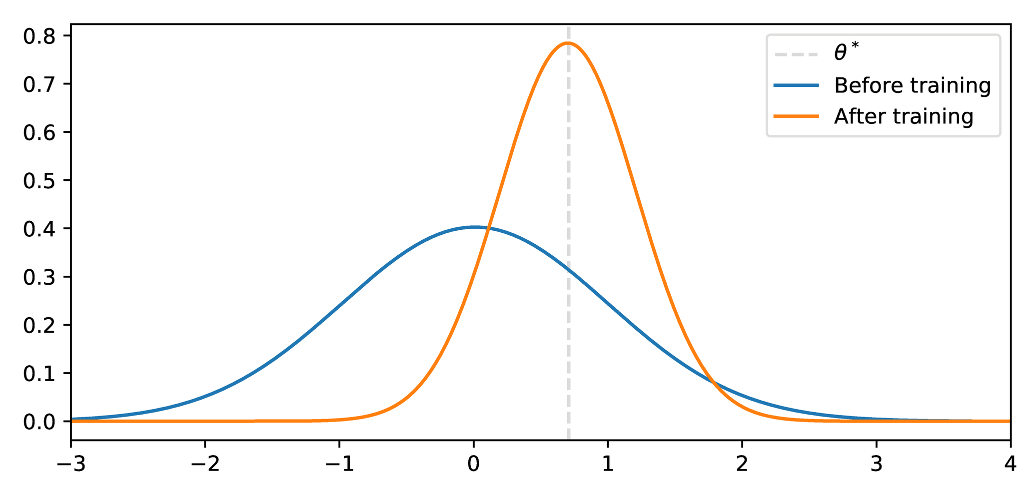
$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &+ i\bar{\psi}D\psi + h.c. \\ &+ \bar{\psi}_i y_{ij}\psi_j\phi + h.c. \\ &+ |D_\mu\phi|^2 - V(\phi)\end{aligned}$$

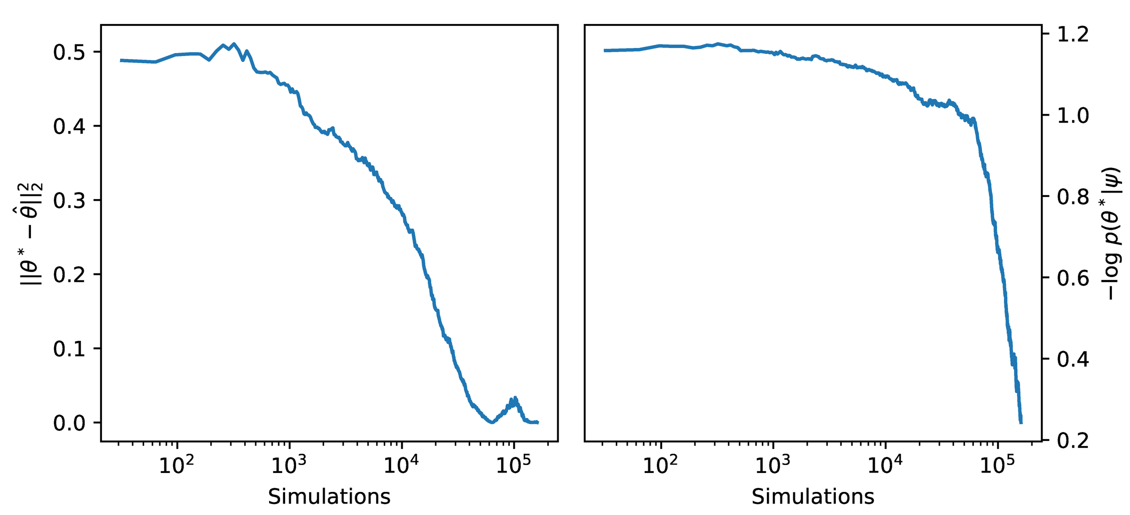


The case of particle physics. The Standard Model defines an implicit distribution $p(\mathbf{x}|\boldsymbol{\theta})$ from which high-dimensional observables can be simulated. Given data collected from Nature, we want to fit the model parameters $\boldsymbol{\theta}$.

Illustration (particle physics)







Particle detector alignment. (*Top and second rows*) : Detector response for the detector offset $\boldsymbol{\theta}=0$ vs $\boldsymbol{\theta}=1$ These plots highlight the difficulty in observing a difference between samples from one or the other parameter setting. (*Bottom*) : Training.

In AVO, the discriminator adapts to the inference problem, regardless of its difficulty. It is **not limited by the sub-optimality of an ad hoc summary statistic**.

Tricks of the trade

Variational optimization/REINFORCE.
Minimize variational upper bounds

$$U_d(\phi) = \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\psi)}[\mathcal{L}_d(\phi)]$$
$$U_g(\psi) = \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\psi)}[\mathcal{L}_g(\boldsymbol{\theta})]$$

defined by a proposal distribution $q(\boldsymbol{\theta}|\psi)$.

- Gradients $\nabla_\psi U_g$ are obtained with REINFORCE estimates, which only requires forward evaluations of the simulator g.
- This effectively results in minimizing
$$\text{JSD}(p_r(\mathbf{x}), q(\mathbf{x}|\psi)),$$
where $q(\mathbf{x}|\psi)) = \int p(\mathbf{x}|\boldsymbol{\theta})q(\boldsymbol{\theta}|\psi)d\boldsymbol{\theta}$.

R_1 regularization (Mescheder et al, 2018).
Penalty added to U_d to improve convergence.

$$R_1(\phi) = \mathbb{E}_{\mathbf{x}\sim p_r(\mathbf{x})}\left[||\nabla_\phi d(\mathbf{x};\phi)||^2\right]$$

Benchmarks