Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe[1], Joeri Hermans[1], and Kyle Cranmer[2]

[1] Department of Electrical Engineering and Computer Science, University of Liège
[2] Department of Physics, New York University

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 incorporating ideas from generative adversarial networks, variational optimization and empirical Bayes. 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 of observed data is minimized. We evaluate and compare the method with simulators producing both discrete and continuous data.

Keywords: Likelihood-free Inference · Simulator-based inference · Physics

Adversarial Variational Optimization (AVO). In this work, we introduce Adversarial Variational Optimization [1], a likelihood-free inference algorithm for non-differentiable, implicit generative models. We adapt the adversarial training procedure of generative adversarial networks (GANs) by replacing the implicit generative network with a domain-based scientific simulator, and solve the resulting non-differentiable minimax problem by minimizing variational upper bounds of the adversarial objectives. The objective of the algorithm is to match the marginal distribution of the synthetic data to the empirical distribution of observations.

Method. The alternating stochastic gradient descent on the discriminator and generator losses $\mathcal{L}_d$ and $\mathcal{L}_g$ in GANs implicitly assumes that the generator g is a differentiable function. In the setting where we are interested in estimating the parameters of a fixed non-differentiable simulator, gradients either do not exist or are not accessible. As a result, gradients $\nabla_{\boldsymbol{\theta}} \mathcal{L}_g$ cannot be constructed and the optimization procedure cannot be carried out.

In this work, we rely on variational optimization to minimize $\mathcal{L}_d$ and $\mathcal{L}_g$, thereby bypassing the non-differentiability of g. We consider a proposal distribution $q(\boldsymbol{\theta}|\boldsymbol{\psi})$ over the parameters of the simulator g and alternately minimize the variational upper bounds

$$U_d(\boldsymbol{\phi}) = \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\psi})}[\mathcal{L}_d(\boldsymbol{\phi})] \tag{1}$$

$$U_g(\boldsymbol{\psi}) = \mathbb{E}_{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\psi})}[\mathcal{L}_g(\boldsymbol{\theta})] \tag{2}$$

respectively over $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$. The discriminator d is therefore no longer pit against a single generator g, but instead against a hierarchical family of generators induced by the proposal distribution.

When updating the discriminator parameters $\boldsymbol{\phi}$, unbiased estimates of $\nabla_{\boldsymbol{\phi}} U_d$ can be obtained by directly evaluating the gradient of U_d over mini-batches of real and synthetic data. When updating the proposal parameters $\boldsymbol{\psi}$, $\nabla_{\boldsymbol{\psi}} U_g$ can be estimated as described in the previous section with $f(\boldsymbol{\theta}) = \mathcal{L}_g(\boldsymbol{\theta})$. That is,

$$\nabla_{\boldsymbol{\psi}} U_g = \mathbb{E}_{\substack{\boldsymbol{\theta}\sim q(\boldsymbol{\theta}|\boldsymbol{\psi}), \\ \tilde{\mathbf{x}}\sim p(\mathbf{x}|\boldsymbol{\theta})}}[\nabla_{\boldsymbol{\psi}} \log q(\boldsymbol{\theta}|\boldsymbol{\psi}) \log(1 - d(\tilde{\mathbf{x}}; \boldsymbol{\phi}))], \tag{3}$$

which we can approximate with mini-batches of synthetic data.

Results. Figure 1 summarizes our results for AVO, ABC-SMC and BOLFI on four inference problems. The figure clearly indicates AVO works better on average compared to ABC-SMC and BOLFI. We attribute this superior performance primarily to the fact that AVO is not limited by the deficiencies of a hand-crafted summary statistic. Instead, AVO benefits from a high-capacity discriminator that dynamically adapts to the inference problem and to the current proposal.

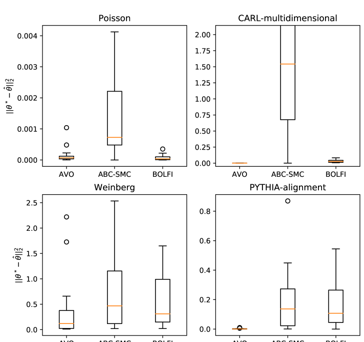

Fig. 1: Benchmark results comparing AVO against ABC-SMC and BOLFI.

References

1. Louppe, G., Hermans, J., Cranmer, K.: Adversarial variational optimization of non-differentiable simulators. In: The 22nd International Conference on Artificial Intelligence and Statistics. pp. 1438–1447 (2019)