Cyclically Consistent Adversarial Networks for Reliable Surrogates in Intertial Confinement Fusion

Rushil Anirudh, Jayaraman J. Thiagarajan, Peer-Timo Bremer, Brian Spears

Center for Applied Scientific Computing (CASC), Lawrence Livermore National Laboratory
Email: {anirudh1, jjayaram, bremer5, spears9}@llnl.gov

I. Introduction

Computational modeling of complex dynamical systems is prevalent in a wide variety of scientific applications. Such simulations are routinely used to interpret experimental data, explore the impacts of different scientific models, or to inform design decisions. Since these systems are generally too complex to fully understand, they are treated as black boxes that require a set of input parameters and produce various quantities of interest as outputs. Given the need to obtain insights into the the variability and sensitivity of the outputs to changing inputs, machine learning techniques are being adopted to build mathematical surrogates to the actual simulator. In many scientific workflows surrogate models are essential for a wide range of critical tasks including sequential optimization in the parameter space, uncertainty quantification, and more importantly to fuse information from experimental data into the analysis.

Modern deep neural networks are capable of learning highly non-linear maps between high dimensional spaces, and as a result their utility is extending beyond conventional AI applications into scientific machine learning. Despite their success in AI, an important, yet often overlooked, consideration while adopting DNNs in scientific applications is that the networks must be conditioned to produce "physically meaningful" predictions. For example, in incompressible fluid dynamics problems, any non-realistic flow solution that violate the conservation of mass principle should be discarded. While in some cases, this conditioning can be achieved by placing geometric constraints on the space of outputs, in other scenarios, it can be very challenging to ensure that the scientific assumptions are not violated. In practice, it is often non-trivial to translate prior knowledge about the characteristics of the output space into mathematical constraints that can be integrated into the optimization process of machine learning systems. Consequently, there is a strong disconnect between highly parameterized surrogate models such as deep neural networks and the actual blackbox simulator. In this work, we propose to adopt a principled, data-driven approach for enabling DNNs to be consistent with the physical processes governing the data. More specifically, by utilizing a combination of three different loss functions, we obtain highly reliable surrogate models – (a) *Surrogate fidelity*: For the given training data from a simulator, this ensures that the prediction from the surrogate model is *close* to the expected output from the simulator, where closeness can typically measured using conventional fidelity measures; (b) *Physical consistency*: This utilizes an *adversarial* loss to ensure that the output from the surrogate is *realistic*, i.e., it lies on the (possibly curved) manifold induced by all observed simulator outputs; (c) *Self-consistency*: By jointly constructing an inverse model that maps back to the input parameter space, this ensures that the surrogate is self-consistent and hence adheres to the underlying data generation process. Note that, existing surrogate modeling techniques rely solely on surrogate fidelity minimization. While this enables the models to make accurate predictions in expectation, often measured using traditional metrics such as the mean squared error, the resulting models are not guaranteed to respect the physical constraints and can produce faulty mappings, thus making subsequent analysis and inferencing unreliable.

II. Approach

A class of methods called generative adversarial networks (GANs) [1] have emerged in the recent few years that are able to build effective generative models that capture crucial latent characteristics of high-dimensional signals. These characteristics are implicitly learned, i.e. they are inferred from data rather than having knowledge of them in closed form. Generally, GANs are a system of two neural networks – a generator, which acts as a traditional decoder learning to map from a low-dimensional latent space to the data space, and a discriminator, which acts as a classifier that tries to distinguish between a realization from the true data distribution and "fake" data produced by the generator. After training, the resulting generator learns to produce highly *realistic* outputs, i.e., respecting constraints that are deemed to be important by the domain experts. In a scientific machine learning scenario, this can translate to having outputs that are consistent with underlying physical processes.

Similar to [2], we propose to utilize 3 neural networks, surrogate model $\mathcal{F}: \mathbb{R}^d \mapsto \mathbb{R}^D$, inverse model $\mathcal{G}: \mathbb{R}^D \mapsto \mathbb{R}^d$ and a discriminator $\mathcal{D}: \mathbb{R}^D \mapsto \{real, fake\}$ that validates if a sample was actually from the true data distribution. Here, d and D refer to the dimensions of the input parameter space and the space of simulator outputs respectively. Denoting the inputs by X and the outputs by Y, we describe the proposed approach as follows:

$$X \xrightarrow{\mathcal{F}} \hat{Y} \xrightarrow{\mathcal{G}} \hat{X}; \quad \ell = \ell^{sf} + \ell^{pc} + \ell^{sc}, \tag{1}$$
$$\ell^{sf} = \|Y - \mathcal{F}(X)\|_2^2; \quad \ell^{pc} = -E_{Y \sim P_{true}(Y)}[log(\mathcal{D}(Y))] - E_X[log(1 - \mathcal{D}(\mathcal{F}(X)))]; \quad \ell^{sc} = \|X - \mathcal{G}(\mathcal{F}(X))\|_2^2.$$

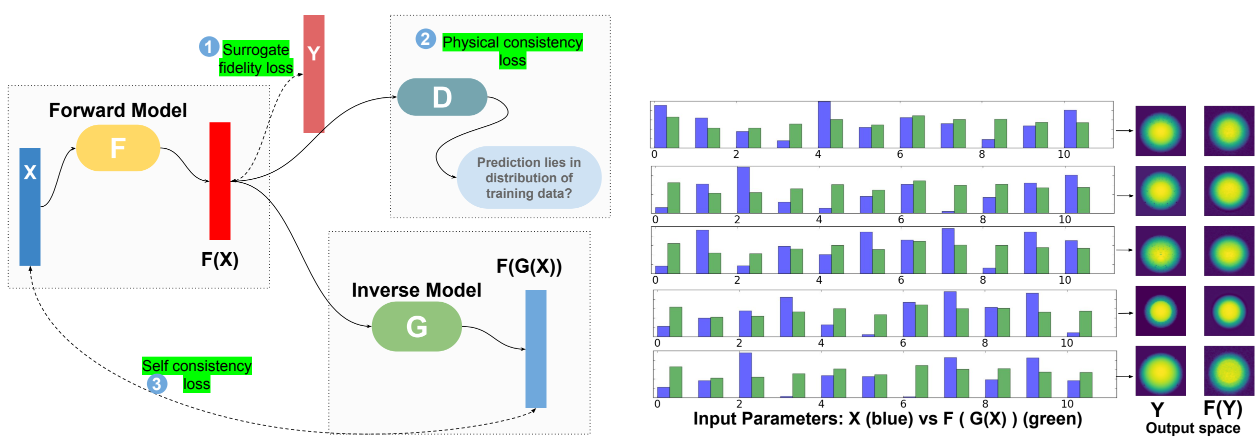
All the three networks are trained jointly using back-propagation and the actual loss functions for each of the cases can be chosen to be different than the ones employed. For example, the simple cross-entropy loss for ℓ^{pc} can be replaced with the Wasserstein metric while ℓ^{sf} and ℓ^{sc} can utilize more meaningful fidelity metrics than the L_2 error. An overview of the system is provided in figure 1(a).

III. Dataset Description & Results

We consider the problem of designing surrogate models for an inertial confinement fusion (ICF) simulator developed at the National Ignition Facility (NIF). The NIF is aimed at demonstrating ICF, that is, thermonuclear ignition and energy gain in a laboratory setting. The goal is to focus 192 beams of the most energetic laser built so far onto a tiny capsule containing frozen deuterium. Under the right conditions, the resulting pressure will collapse the target to the point of ignition where hydrogen starts to fuse and produce massive amounts of energy, effectively creating a small star which can be harnessed for energy production. Though significant progress has been made, the ultimate goal of "ignition" has not yet been reached.

NIF employs an adaptive pipeline: perform experiments, use post-shot simulations to understand the experimental results, and design new experiments with parameter settings that are expected to improve performance. From an analysis viewpoint, the goal is to search the parameter space to find the region that leads to near-optimal performance. The dataset considered here is a so called engineering or macro-physics simulation ensemble in which an implosion is simulated using different input parameters, such as, laser power, pulse shape etc. From these simulations, scientists extract a set of drivers, physical quantities believed to determine the behavior of the resulting implosion. These drivers are then analyzed with respect to the energy yield to better understand how to optimize future experiments.

In our experiments, we used data collected from an analytical simulator for ICF that accepts configurations for 11 input parameters and produces intensity images of resolution 50×50. For training the proposed architecture, we collected the simulator outputs $100K$ random samples in the $11-$D parameter space obtained using latin hypercube sampling. Note that, we treated the images as a 2500 dimensional vectors in these experiments, and used only fully connected network architectures for $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{D}$. We show samples generated by the forward and inverse models on a hidden test dataset, in figure 1(b). It shows that our model is able to perform well in both domains, while capturing interesting characteristics in images such as size, position, and skewness.



(a) Cycle Consistency for Joint Training

(b) Sample results from the JAG 100K dataset

Fig. 1. An overview of the system (left) and performance of forward and inverse models (right), after training with 90K intput-output pairs from an analytical ICF simulator.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems (NIPS)*, pages 2672–2680, 2014.

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344