ML-BASED SURROGATES AND EMULATORS

Tareq Alghamdi, Yaohang Li, and Nobuo Sato
Old Dominion University, Jefferson Lab

INTRODUCTION

- An inverse problem is a problem in which the goal is to determine the cause of an observed effect by solving for an unknown input given an output, rather than predicting the effect given the cause.
- Nucleons consist of quarks and gluons, which are basic particles that follow the rules of quantum chromodynamics (QCD) in terms of their interactions and behavior.
- Understanding the mysteries of QCD requires analyzing its multi-dimensional quantum correlation functions (QCFs).
- While the mapping of parameters-to-observables in QCFs is a well-posed problem with unique solutions, the inverse problem at the event-level is a backward mapping which is likely non-unique.

PROBLEM STATEMENT

In the toy problem, the proton and neutron cross-sections become

$$\sigma_p(x) = 4\,u(x) + d(x), \quad \sigma_n(x) = 4\,d(x) + u(x). \tag{1}$$

Consider the following probability density function model for the up and down quarks

$$\begin{aligned} u(x) =& N_u x^{a_u} (1-x)^{b_u} \\ d(x) =& N_d x^{a_d} (1-x)^{b_d} \end{aligned} \tag{2}$$

where $x \in (0,1)$ and parameter vector $= [N_u, a_u, b_u, N_u, a_u, b_u]^\top$ is uncertain. We observe $= [\sigma_p^o, \sigma_n^o]^\top$ events generated by the model(2) and filtered through cross-sections defined in (1). We assume we observe M_{σ_p} proton events and M_{σ_n} neutron events.
The observed events are generated by sampling from the two cross-sections with the "true" parameter values $_T$ used in(2) $_T = [2.1875, -0.5, 3, 1.09375, -0.5, 4]^\top$.

MISSION

To solve the inverse problem of femtoscale imaging at the event level:

- We developed an ML-based surrogate event generator framework, as generative adversarial networks (GANs) based architecture that effectively allows building a posterior distribution sampler for the QCF parameters.
- A dedicated ML-based discriminator that is free of specific choices for the likelihood function, and accurately predict the parameters and the cross sections for better generalization on unseen parameters.
- The problem is to identify the parameters along with their uncertainty that generated the finite number of observed events. This is formulated as an inverse problem: invert the parameters based on limited observations.

DATA DESCRIPTIONS

- In order to train inner GAN, we generate a dataset with 20,000,000 data samples for both parameters and events (σ_p^o, σ_n^o). The same dataset, which has been generated for the inner GAN, is used to train the norm neural network.
- Additionally, we generate three categories of datasets with different sizes for outer GAN to be trained and tested by using θ_T: small, medium, and large.
- These sets include cases A, B, and C described in Table (1). For case A, we use 1000 samples for σ_p and 500 samples for σ_n. In case B, we use 10,000 samples for σ_p and 5,000 samples for σ_n. For case C, we use 100,000 samples for σ_p and 50,000 samples for σ_n.

Case	M_{σ_p}	M_{σ_n}
Case A	1000	500
Case B	10000	5000
Case C	100000	50000

Table 1: Dataset

METHODOLOGY & RESULTS

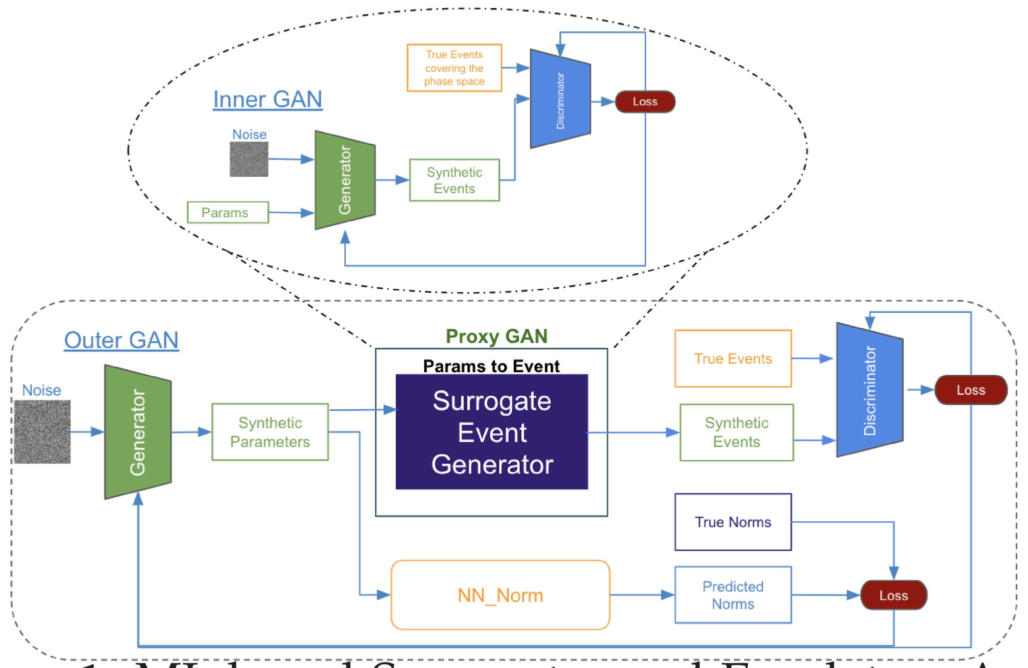

Figure 1: ML-based Surrogates and Emulators Archeticture

Our approach adopts the architecture of a generative adversarial network[1], consisting of:

- An *inner GAN* that maps the parameter space to the observable space (*proxy GAN*).
- An *outer GAN* that maps from the observable space back to the parameter space
- We incorporate a *pre-trained neural network* into our framework in order to help the generator produce the correct parameters.
- The crucial element of this framework is the use of a discriminator that is capable of updating the parameter generator using event-level information.
- The design of the ML-based surrogates and emulators model focuses on learning to mimic cross sections in DIS from protons and neutrons.

- We demonstrate the effectiveness of our framework by applying it to the proxy application for a toy 1D QCF analysis at the event-level.
- As shown in Figure 2, our method is able to correctly learn the mapping between the observable space and the QCF space and shows the cross sections σ_p and σ_n mimic cross sections in DIS from protons and neutrons, respectively.

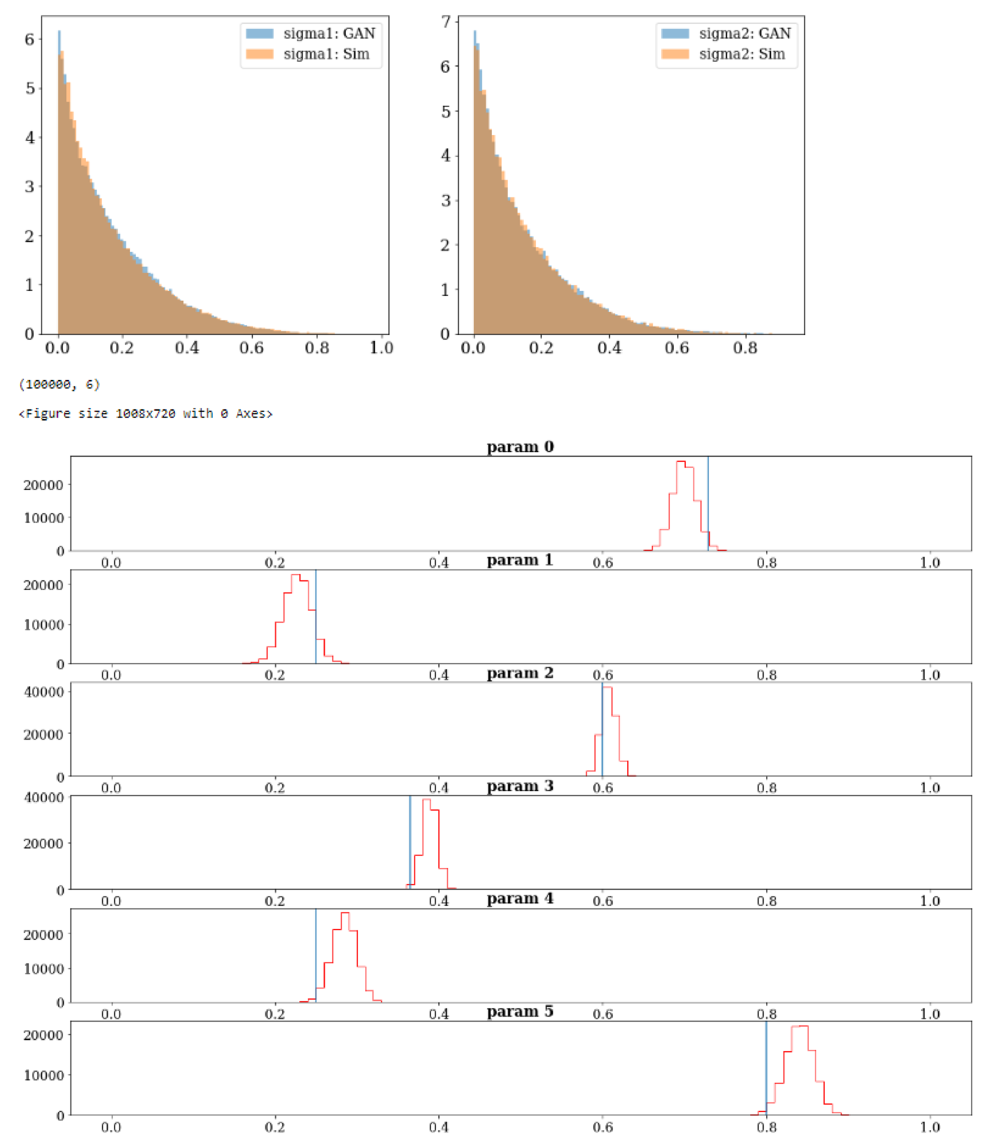

Figure 2: Preliminary result of parameter distribution for dataset C obtained by solving the GAN-based inverse problem. Parameters correspond to the ones in (2).

CONCLUSION& ACKNOWLEDGMENTS

- Our method shows the capability of inverting the parameters based on limited observations and getting back to the correct parameters.
- It effectively allows building a posterior distribution sampler for the QCF parameters free of underlying physics theory.
- **Acknowledgments**

This material is based upon work supported by the SciDAC project No DE-FOA-0002589

REFRENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[2] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.