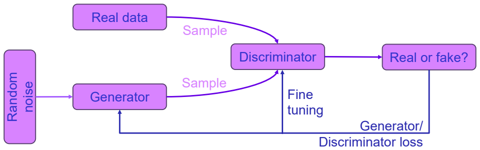
Engagement: Hyperparameter Optimization of Generative Adversarial Network Models for High-Energy Physics Simulations

FASTMath Team Members: Juliane Mueller (NREL); HEP partner: Xiangyang Ju (LBNL); Other contributors: Vincent Dumont (prev. LBNL, now Apple)

GANs can accelerate computational simulations



- By swapping out compute-intensive modules from large simulations, we can significantly reduce the simulation time
- Usefulness of this approach is limited to performance of the GAN
- GAN performance depends on architecture -> Optimization needed

Tuning GAN architectures is computationally challenging: Bilevel problem

$$\min_{p \in \Omega} U(p, w^*; \mathcal{D}_{val}) \quad (1)$$

Minimize a performance metric over hyperparameters p…

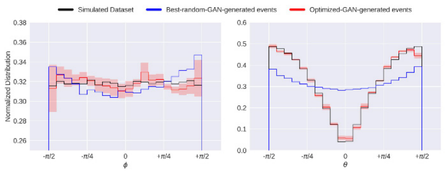
$$s.t. \quad w^* \in \min_{w \in W} l(w; p, \mathcal{D}_{train}) \quad (2)$$

…with a nested optimization problem over weights and biases w as constraint

- U and l measure the Wasserstein distance between generated and target distributions

GAN hyperparameters include:

- noise dimension;
- batch size;
- generator & discriminator learning rates;
- generator and discriminator layer sizes



- Commonly used hyperparameter tuning methods include random sampling, trial and error, intuition
- Stochasticity in model trainings -> how reliable are the results?

Algorithm: from broad search for reducing parameter space to focused search

Evaluating (2) for many different hyperparameter sets p over many training epochs is computationally too expensive

Broad search: Optimize p over wide parameter bounds Ω using only 30 epochs in (2) → Shrink Ω to $\Omega' \subset \Omega$ → **Focused search:** Optimize p over narrow parameter bounds Ω' using 1000 epochs in (2)

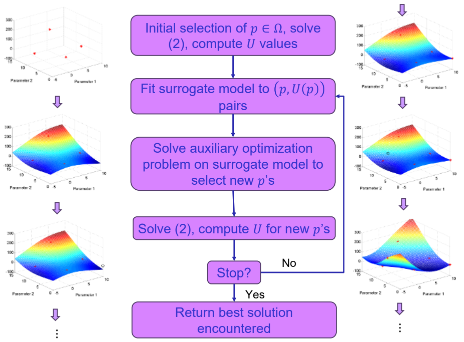
- Broad search is fast, identifies hyperparameters p that are promising (low Wasserstein distance, low performance variability)
- Focused search is slower and more detailed over smaller parameter space

1. Broad search is across all of Ω, yields best solution p_{broad}
2. Define $\Omega' \subset \Omega$ as $\Omega' = [p_{broad} \pm \epsilon] \cap \Omega$
3. Focused search is across Ω', yields final solution p^*

More Information please contact:
Juliane Mueller, Juliane.Mueller@nrel.gov

Surrogate model guided optimization of p

We treat (2) like a black box and use surrogate model optimizer

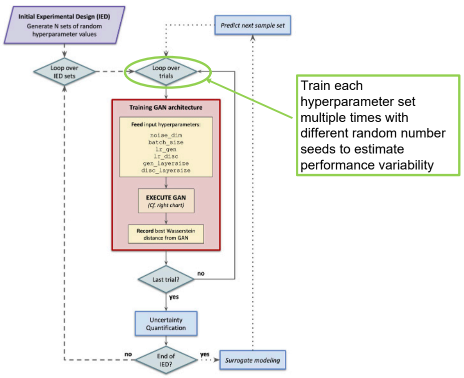


We use the same approach for both broad and focused search:

- Broad search: pick candidate p's from Ω
- Focused search: pick candidate p's from $\Omega' \subset \Omega$

Hyperparameter optimization workflow implementation

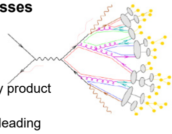
hyppo



Train each hyperparameter set multiple times with different random number seeds to estimate performance variability

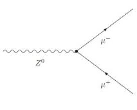
Need for optimized GANs in HEP

Test Case 1: Simulation of Hadronization processes



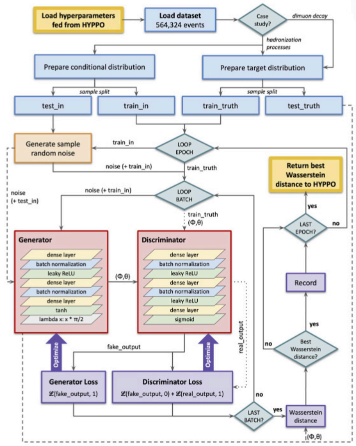
- Current hadronization models are physics inspired phenomenological models with many parameters tuned to experimental measurements
- We use GAN to learn kinematic properties of the decay product of hardonization cluster simulated with Herwig
- 564,000 events, predict polar and azimuthal angles of leading pion's momentum

Test Case 2: Simulation of Drell-Yan Events

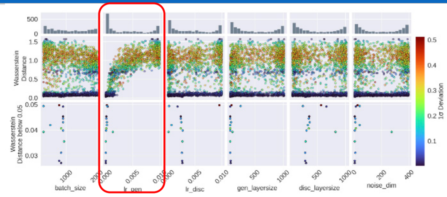


- Generating billions of Drell-Yan events is computationally expensive but necessary for ATLAS to achieve a desired precision for the search for Higgs boson decaying to two muons
- We use GANs to generate the 4-momentum distributions of the outgoing muons
- Train on 100,000 dimuon events

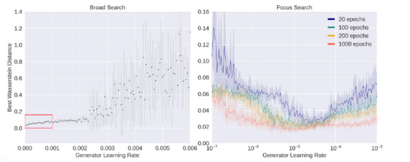
Training the GAN model for HEP simulations



Not all GAN hyperparameters are equally important

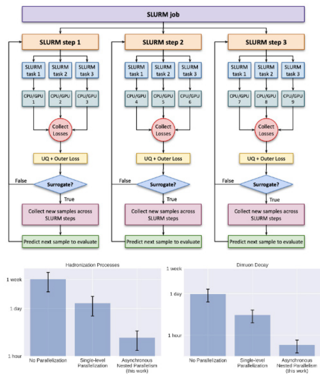


Generator learning rate appears to be the most important hyperparameter



There may be diminishing returns with increasing the number of epochs

Asynchronous parallelism accelerates the optimization



SciDAC-5 Principal Investigator (PI) Meeting
Rockville, Maryland
September 12-14, 2023
NREL/PO-2C00-87403