Data-Driven Modeling with Generative Adversarial Networks

Yumeng Wang (2nd year Ph.D. candidate) [1] Yanzhi Zhang (Advisor) [2]

[1]Missouri University of Science and Technology, Rolla, USA [2]Missouri University of Science and Technology, Rolla, USA

Problem and Motivation

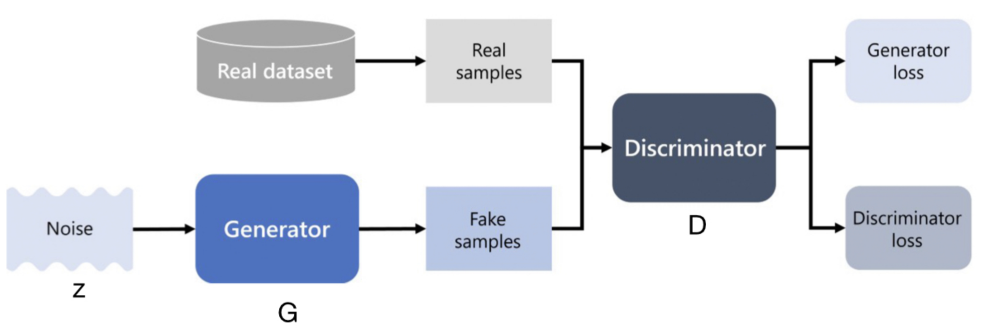
Problem setting: Predict solution behavior when limited spatio-temporal observed data are available, but the physical model is partially or completely unknown.

- **Physics-informed approach:** It depends on accurate physical model (such as PDEs) and may be implemented together with observed data.
- **Data-driven approach:** It predicts the solution behavior purely based on observed data.

Goal: Develop a new data-driven modeling approach with generative adversarial networks to generate solutions without physical models.

Generative Adversarial Network

Generative Adversarial Networks (GAN) : GAN generates model data with an adversarial strategy, such that its probability distribution P_{model} approximates the implicit probability distribution of observed data P_{data} [1]. It usually includes two sub-networks: generator network (G) and discriminator network (D); see the illustration as below.



Kullback–Leibler (KL) divergence: It measures the distance between two probability distributions, and the KL divergence from P_{model} to P_{data} is defined.

$$KL(P_{\text{data}} \| P_{\text{model}}) = \mathbb{E}_{x\sim P_{\text{data}}}\left[\log \frac{P_{\text{data}}(x)}{P_{\text{model}}(x)}\right]$$

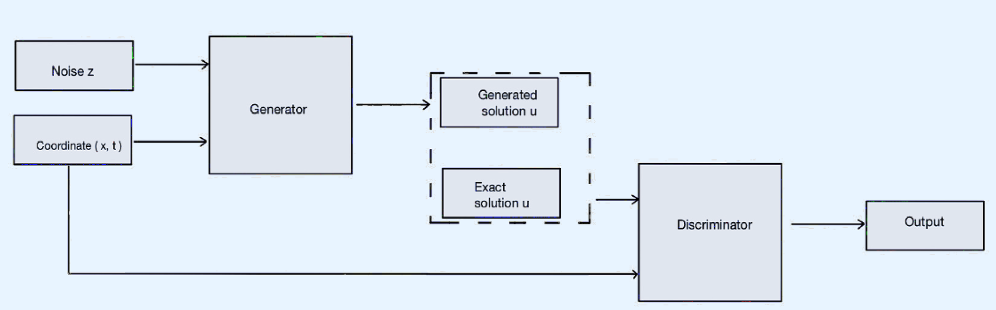
Minimax problem of generator and discriminator:

$$\min_G \max_D V(D,G) = \min_G \max_D \left[\mathbb{E}_{x\sim P_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim P_z}[\log(1-D(G(z)))]\right]$$

where z is random noise, and P_z denotes its distribution.

Conditional Wasserstein Generative Adversarial Network

Conditional Wasserstein Generative Adversarial Network (cWGAN): Use the Wasserstein distance to measure the probability distribution similarity under the conditions of spatial and temporal coordinates (x,t).



Wasserstein distance: Wasserstein distance has a smoother gradient, which is defined by

$$D_w(P_{\text{data}} \| P_{\text{model}}) = \sup_{\|f\|_L \le 1} \left[\mathbb{E}_{x\sim P_{\text{data}}}[f(x)] - \mathbb{E}_{x\sim P_{\text{model}}}[f(x)]\right]$$

Corresponding optimization problem:

$$\max_{G,D} V(G,D) = \max_{G,D} \left[\mathbb{E}_{x\sim P_{\text{data}}}[D(u,x,t)] - \mathbb{E}_{z\sim P_z}[D(G(z,x,t),x,t)]\right]. \quad (1)$$

Proposed method: Improve the performance of cWGAN by adding the mean square error (MSE) regularization to the loss function.

Note that the mean square error is added only into the generator. In practice, the optimization problem is realized by two sub-problems:

$$\min_G V(G,D) = \min_G \mathbb{E}_{z\sim P_z}[D(G(z,x,t),x,t) + (u - G(z,x,t))^2], \quad (2)$$

$$\min_D V(G,D) = \min_D \mathbb{E}_{z\sim P_z}[D(G(z,x,t),x,t)] - \mathbb{E}_{x\sim P_{\text{data}}}[D(u,x,t)]. \quad (3)$$

Optimization: Gradient descent algorithm update parameters of both G and D iteratively.

Numerical Experiments and Comparison

Data preparation: We use synthetic spatio-temporal data that are prepared by numerically solving PDE problems.

Experiment 1: Data are sampled from numerical solution of the fractional Allen–Cahn equation

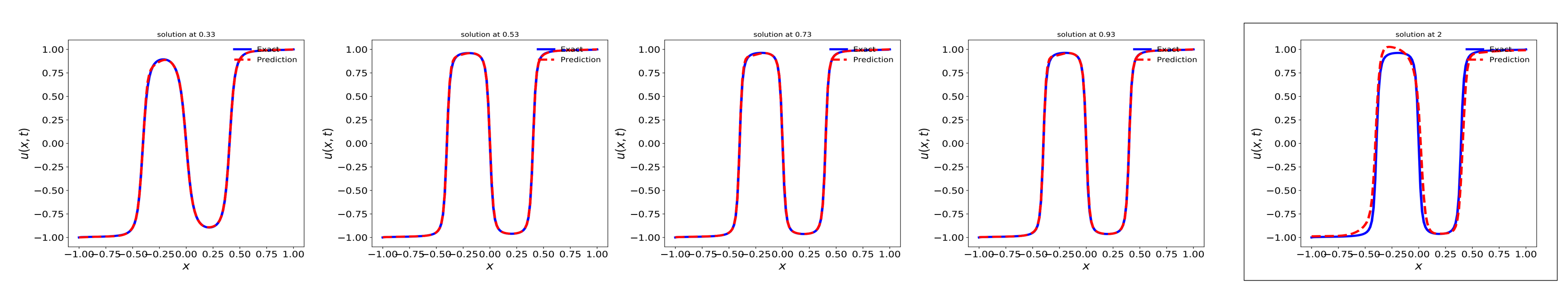
$$u_t = -\gamma(-\Delta)^{\frac{\alpha}{2}}u + \frac{1}{\epsilon^2}(1-u^2)u, \qquad \text{for } x\in(-1,1), \quad t>0$$

$$u(x,0) = 0.68x + 0.32\sin(-\frac{3}{2}\pi x), \qquad \text{for } x\in[-1,1],$$

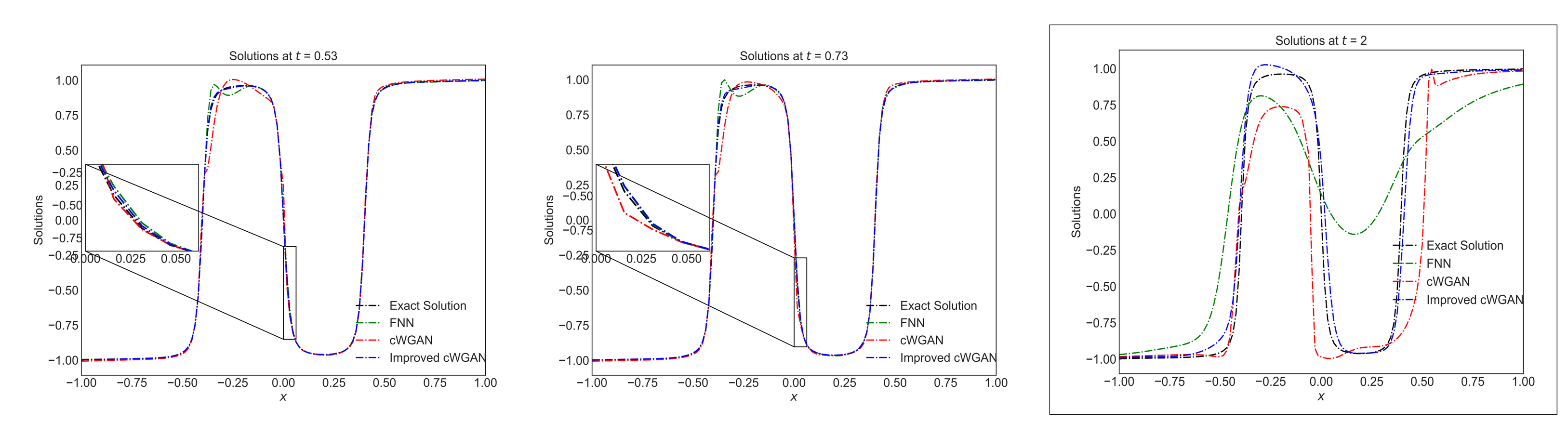
with extended nonhomogeneous boundary conditions on $(-1,1)^c$, where we choose $\gamma = 0.01$, $\epsilon = 0.1$, and $\alpha = 1.2$.

Sampling points: 1050 observed data points are sampled at randomly fixed spatial coordinates and uniformly temporal coordinates $t = 0, 0.05, 0.1, ..., 1$.

Our results in comparison to the true solution at time $t = 0.33, 0.53, 0.73, 0.93$ and prediction at $t = 2$.



Our results in comparison to the predictions from FNN and cWGAN:



Experiment 2: Data are sampled from numerical solution of the viscous Burgers' equation.

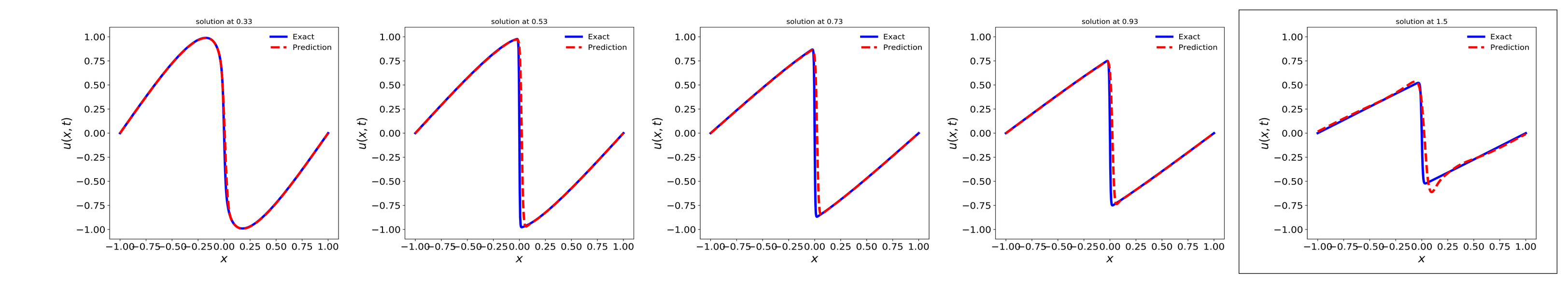
$$-u_t + \frac{1}{2x}(u^2) = \nu u_{xx}, \qquad \text{for } x\in(-1,1), \quad t>0,$$

$$u(x,0) = \sin(\pi x), \qquad \text{for } x\in[-1,1],$$

$$u(-1,t) = u(1,t) = 0, \qquad \text{for } t\ge 0,$$

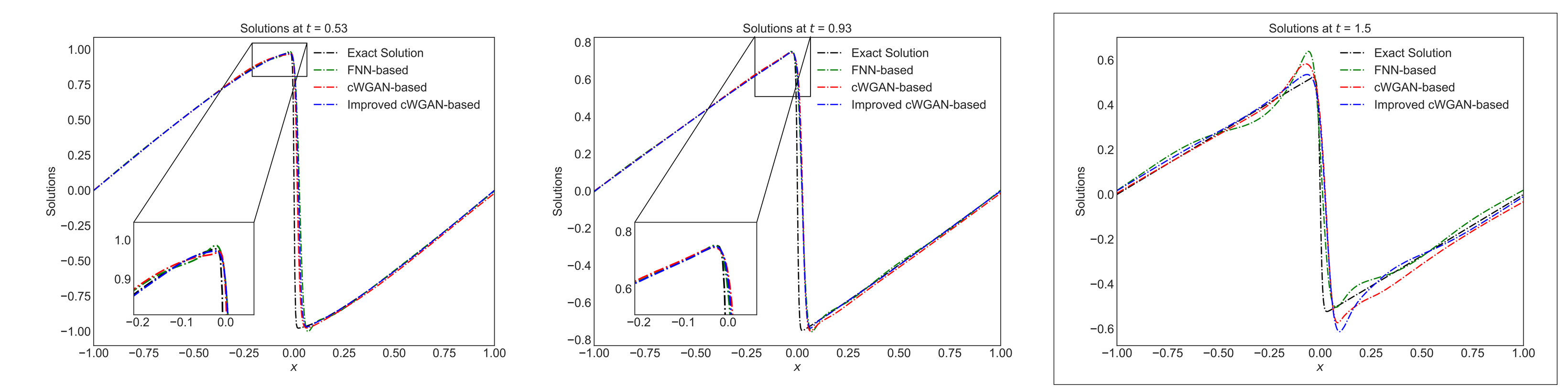
where we choose the viscosity parameter $\nu = \frac{0.01}{\pi}$.

Our results in comparison to the true solution at time $t = 0.33, 0.53, 0.73, 0.93$ and prediction at $t = 1.5$

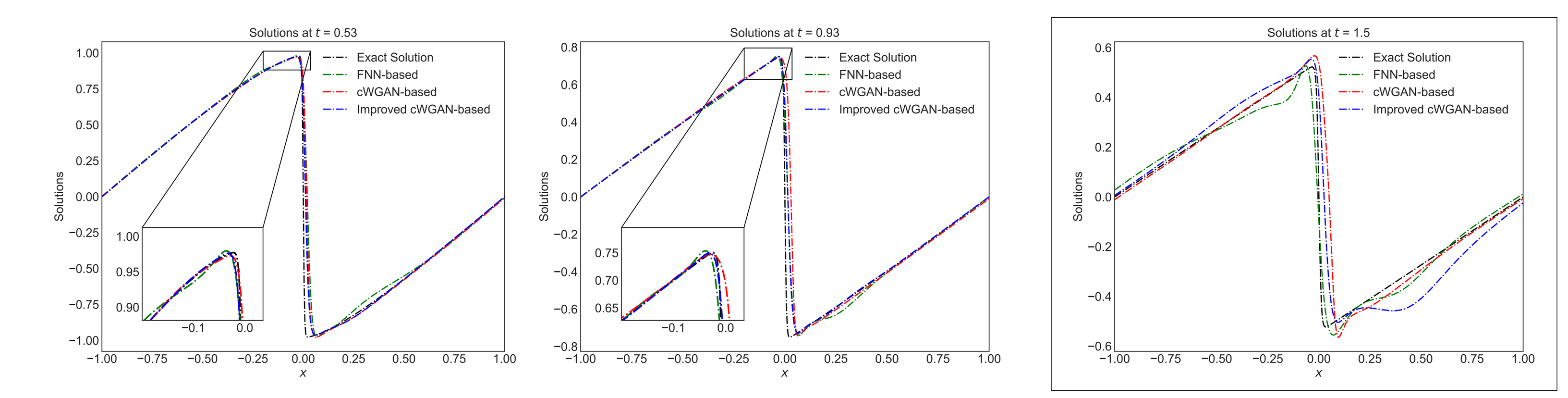


Our results in comparison to the predictions from FNN and cWGAN:

- **Case 1:** More (i.e., $N = 1050$) data points



- **Case 2:** Less (i.e. $N = 462$) data points



Results Analysis and Summary

- cWGAN can avoid model collapses and direct the generation under specific spatial and temporal coordinate compared with GAN.
- **Our method can achieve higher accuracy than cWGAN and FNN,** especially when fewer data are available.
- Loss function for cWGAN is a distribution based method and for FNN is point-to-point algorithm.
- **Probability distribution based algorithms give better solution than point-to-point based algorithm.**
- Using mean square error as a regularizer can improve the performance of cWGAN.
- These three algorithms cannot make long-time prediction out of domain.

Acknowledgement

This work was supported by the National Science Foundation DMS-1953177.

Reference

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. pages 2672–2680. Curran Associates, Inc.

[2] Y. Wang, Y. Zhang. Data Driven Modeling with Generative Adversarial Networks.In preparation (2023).

yw2bc@umsystem.edu