Conditional GANs and their generalizability in physics-based inverse problems

Deep Ray
Department of Aerospace & Mechanical Engineering
University of Southern California

Email: deepray@usc.edu
Website: deepray.github.io

Jointly with Assad Oberai, Harisankar Ramaswamy, Dhruv Patel

Supported by ARO grant W911NF2010050, NRL grant N00173-19-P-1119, AIER-USC, CARC-USC

USC Viterbi
School of Engineering

USACM UQ-MLIP
Arlington, Virginia
August 18-19, 2022

Outline

- Challenges with Bayesian inference
- Conditional generative adversarial networks (cGANs)
- Deep posteriors with cGANs
- Generalizability
- Conclusion

Consider a forward problem

$$\mathcal{F} : \boldsymbol{x} \in \Omega_x \mapsto \boldsymbol{y} \in \Omega_y, \quad \Omega_x \in \mathbb{R}^{N_x},\ \Omega_y \in \mathbb{R}^{N_y}$$

For example the **heat conduction** PDE:

$$\frac{\partial u(\boldsymbol{s},t)}{\partial t} - \kappa\Delta u(\boldsymbol{s},t)) = 0$$

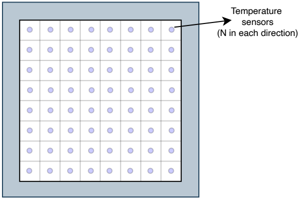
$$u(\boldsymbol{s},0) = u_0(\boldsymbol{s})$$

where

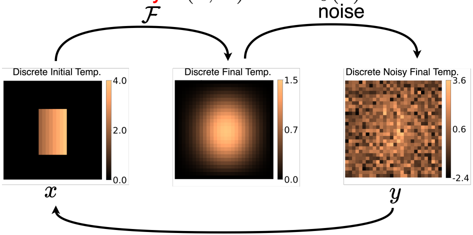
$u(\boldsymbol{s},t) \rightarrow$ temperature at location $\boldsymbol{s}$ at time t

$u_0(\boldsymbol{s}) \rightarrow$ initial temperature at location $\boldsymbol{s}$

$\kappa \rightarrow$ thermal conductivity of material



Inverse problem $\mathcal{F}^{-1}$: Given noisy $u(\boldsymbol{s},T)$ infer $u_0(\boldsymbol{s})$



Bayesian inference

Challenges with inverse problems:

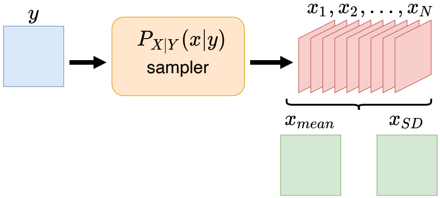
- Inverse map is not well posed.
- Noisy measurements.
- Need to encode prior knowledge about $\boldsymbol{x}$.

Bayesian framework: $\boldsymbol{x}$ and $\boldsymbol{y}$ modelled by random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$.

AIM: Given a measurement $\boldsymbol{Y} = \boldsymbol{y}$ approximate the conditional (posterior) distribution

$$P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$$

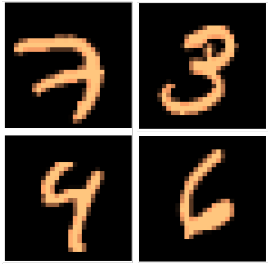
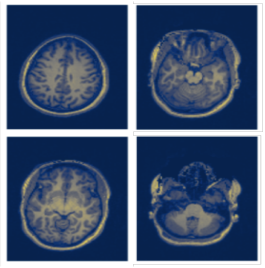
and sample from it.



Bayesian inference: challenges

- Posterior sampling techniques, such as Markov Chain Monte Carlo, are prohibitively expensive when dimension of $\boldsymbol{X}$ is large.
- Characterization of priors for complex data

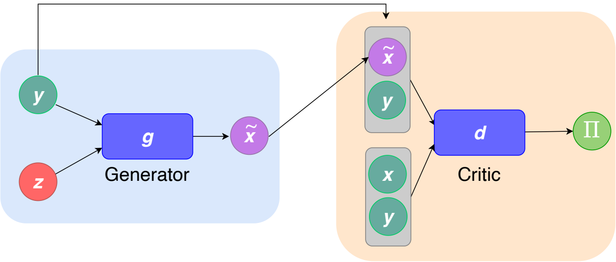
For example, x data might look like:

Representing this data using simple distributions is hard!

Resolve both issues using deep learning

- Learning distributions conditioned on another field.
- Comprises two neural networks, $\boldsymbol{g}$ and d.



Generator network:

- $\boldsymbol{g} : \Omega_z \times \Omega_y \to \Omega_x$.
- Latent variable $\boldsymbol{Z} \sim P_{\boldsymbol{Z}}$, e.g. $N(0, \boldsymbol{I})$. Also $N_z \ll N_x$.
- $(\boldsymbol{x}, \boldsymbol{y})$ sampled from true P_{XY}

Critic network:

- $d : \Omega_x \times \Omega_y \to \mathbb{R}$.
- d tries to detect fake samples.
- $d(\boldsymbol{x}, \boldsymbol{y})$ large for real $\boldsymbol{x}$, small otherwise.

Conditional GANs

- Given $\boldsymbol{Y} = \boldsymbol{y}$ and $\boldsymbol{Z} \sim P_{\boldsymbol{Z}}$ we get a random variable

$$\boldsymbol{X}^{\boldsymbol{g}} = \boldsymbol{g}(\boldsymbol{Z}, \boldsymbol{y}), \quad \boldsymbol{X}^{\boldsymbol{g}} \sim P^{\boldsymbol{g}}_{\boldsymbol{X}|\boldsymbol{Y}}.$$

- Objective function

$$\Pi(\boldsymbol{g}, d) = \mathop{\mathbb{E}}_{\substack{(\boldsymbol{X},\boldsymbol{Y}) \sim P_{\boldsymbol{X}\boldsymbol{Y}} \\ \boldsymbol{Z} \sim P_{\boldsymbol{Z}}}} \left[d(\boldsymbol{X}, \boldsymbol{Y}) - d\big(\boldsymbol{g}(\boldsymbol{Z}, \boldsymbol{Y}), \boldsymbol{Y}\big) \right]$$

- $\boldsymbol{g}$ and d determined (with constraint $\|d\|_{\text{Lip}} \leq 1$) through

$$(\boldsymbol{g}^*, d^*) = \arg\min_{\boldsymbol{g}} \ \arg\max_{d} \Pi(\boldsymbol{g}, d)$$

- Adler et al. (2018) proved that the minmax problem is equivalent to

$$\boldsymbol{g}^* = \arg\min_{\boldsymbol{g}} \mathop{\mathbb{E}}_{\boldsymbol{Y} \sim P_{\boldsymbol{Y}}} \left[W_1(P_{\boldsymbol{X}|\boldsymbol{Y}}, P^{\boldsymbol{g}}_{\boldsymbol{X}|\boldsymbol{Y}}) \right]$$

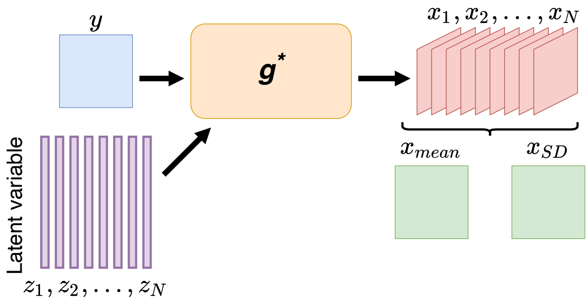
  where W_1 is the Wasserstein-1 distance.

Steps:

- Acquire samples $\mathcal{S}_x = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$, where $\boldsymbol{x}_i \sim P_X^{\text{prior}}$.
- Use forward map $\mathcal{F}$ to generate paired dataset

$$\mathcal{S} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_N, \boldsymbol{y}_N)\} \quad \text{where} \quad \boldsymbol{y}_n = \mathcal{F}(\boldsymbol{x}_n) + \text{ noise.}$$

- Train a cGAN on $\mathcal{S}$.
- For a new test measurement $\boldsymbol{y}$, generate samples using $\boldsymbol{g}^*$.
- Evaluate statistics using Monte Carlo.



Solving the inverse heat conduction equation

Consider the PDE

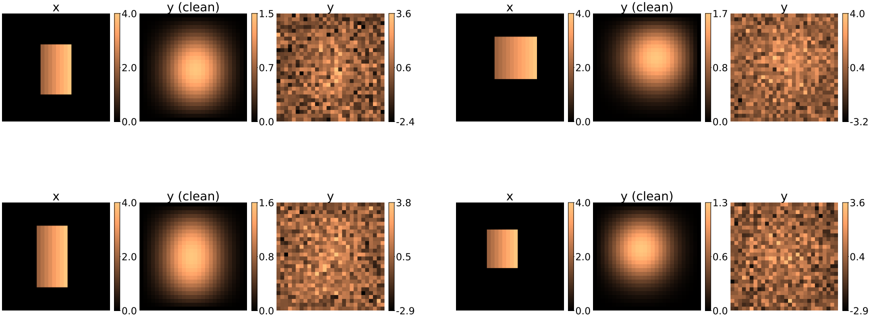
$$\begin{aligned} \frac{\partial u(\boldsymbol{s},t)}{\partial t} - \nabla \cdot (\kappa(\boldsymbol{s}) \nabla u(\boldsymbol{s},t)) = 0, \qquad & \forall\ (\boldsymbol{s},t) \in (0,2\pi)^2 \times (0,1] \\ u(\boldsymbol{\xi},0) = u_0(\boldsymbol{s}), \qquad & \forall\ \boldsymbol{s} \in (0,2\pi)^2 \\ u(\boldsymbol{\xi},t) = 0, \qquad & \forall\ \boldsymbol{s} \in \partial(0,2\pi)^2 \times (0,1] \end{aligned}$$

- $\boldsymbol{x}$: discrete initial temperature field.
- $\boldsymbol{y}$: noisy discrete final temperature field.
- $\boldsymbol{\mathcal{F}}$: Finite difference solver for the PDE.
- We assume a constant κ.

Inferring initial condition: parametric prior

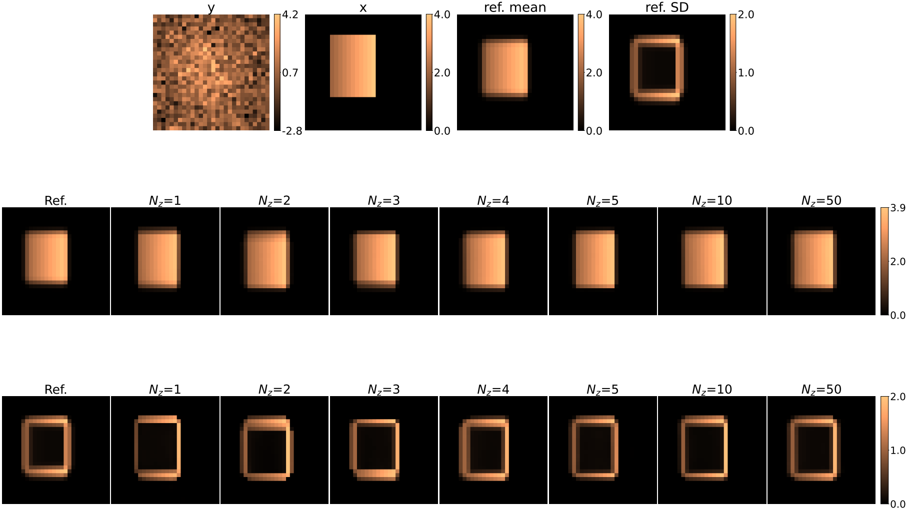
Assuming $\boldsymbol{x}$ to given by a rectangular inclusion and $N_x = N_y = 28 \times 28 = 784$

Training samples:



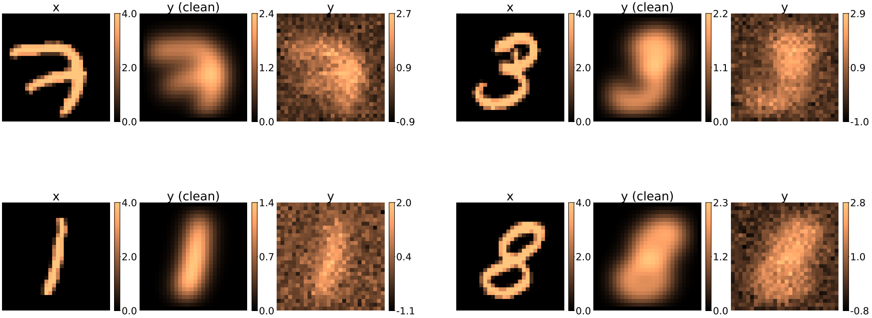
We don't have clean y in real problems!

Testing trained cGAN (statistics with 800 z samples)



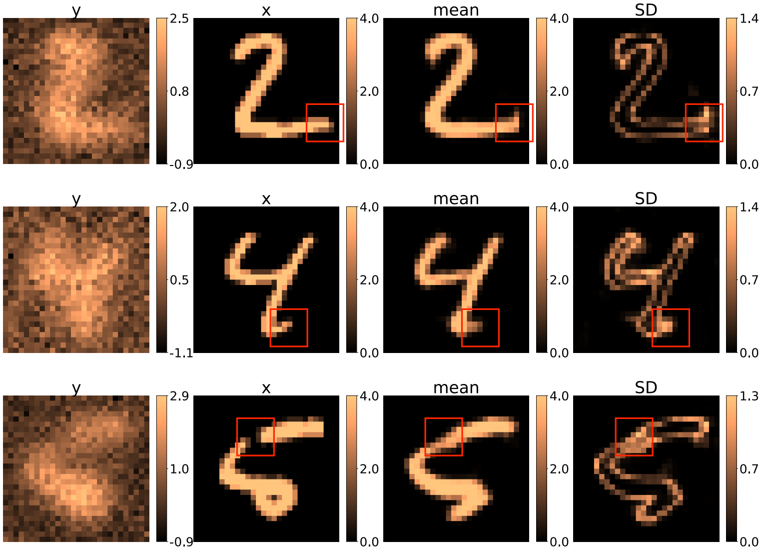
Assuming $\boldsymbol{x}$ to given by (heat stamped) MNIST handwritten digits and $N_x = N_y = 28 \times 28 = 784$

Training samples:



cGAN trained using latent dimension $N_z = 100$.

Testing trained cGAN (statistics with 800 z samples)

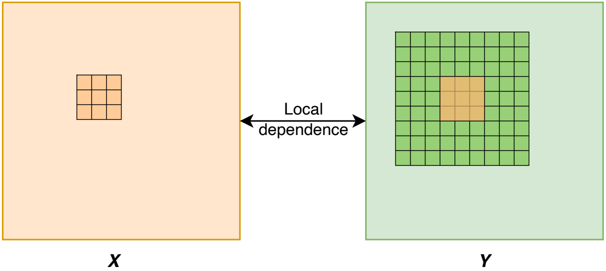


Generalization: cGAN trained on Set A gives good predictions on another distinct Set B, possibility sampled from a different distribution.

We can prove[1] the following: Assume

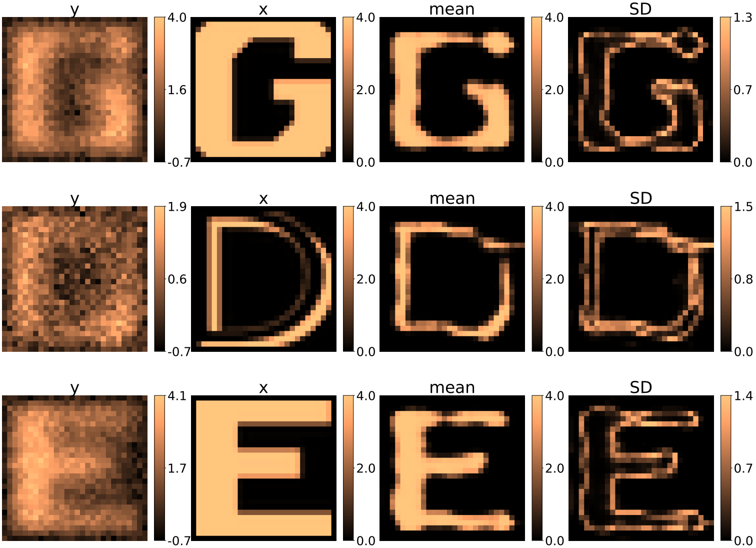
- The true (regularized) inverse map $\mathcal{F}^{-1}$ is spatially local.
- Set A and Set B contain samples with similar local spatial features.
- A cGAN train on Set A, and is also spatially local.

Then the cGAN can generalize well to Set B.



1: *The efficacy and generalizability of conditional GANs for posterior inference in physics-based inverse problems* (D. Ray, D. Patel, H. Ramaswamy, A. A. Oberai); preprint 2022.

cGAN trained on MNIST, tested on notMNIST



Conclusion

- What do we gain?
  - Ability to represent and encode complex prior data.
  - Dimension reduction since $N_z \ll N_x$.
  - Sampling from cGAN is quick and easy.
- Need $(\boldsymbol{x}, \boldsymbol{y})$ pairs to train – supervised algorithm.
- Generalizability – training on smaller dataset.
- Currently testing algorithm on several other physics-based and medical applications.

Questions?

Special architecture of generator

- U-Net architecture when $\boldsymbol{x}$, $\boldsymbol{y}$ have tensored (image-like) structure.
- $\boldsymbol{z}$ injected through conditional instance normalization:
  - Can choose N_z independent of N_y.
  - Introduce stochasticity at multiple scales of U-Net.