Bayesian inference using generative adversarial networks

Deep Ray
Department of Aerospace & Mechanical Engineering

with Dhruv V. Patel, Harisankar Ramaswamy & Assad A. Oberai

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

87th Annual Conference of the IMS
December 7, 2021

USC Viterbi
School of Engineering

Outline

Outline

Bayesian approach for solving inverse problems

Consider the forward/direct map

$$\boldsymbol{f} : \Omega_X \subset \mathbb{R}^{N_X} \to \Omega_Y \subset \mathbb{R}^{N_Y}$$

Task: Given a noisy measurement/output $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$ with $\eta \sim P_\eta$, infer $\boldsymbol{x}$.

Example: Consider the PDE for temperature u

$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) &= b(\boldsymbol{\xi}), \qquad && \forall\, \boldsymbol{\xi} \in \Omega \subset \mathbb{R}^2 \\ u(\boldsymbol{\xi}) &= 0, && \forall\, \boldsymbol{\xi} \in \partial\Omega \end{aligned}$$

Problem setup:

- Measurement $\boldsymbol{y}$, noisy temperature field u on a 2D grid.
- Infer $\boldsymbol{x}$, nodal values of conductivity κ.
- Non-linear forward map $\boldsymbol{f}$ solves the PDE (FEM solver).

Bayesian approach for solving inverse problems

Consider the forward/direct map

$$\boldsymbol{f} : \Omega_X \subset \mathbb{R}^{N_X} \to \Omega_Y \subset \mathbb{R}^{N_Y}$$

Task: Given a noisy measurement/output $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$ with $\eta \sim P_\eta$, infer $\boldsymbol{x}$.

Bayesian approach: Model $\boldsymbol{x}$ and $\boldsymbol{y}$ using random variables

$$P_X^{\text{post}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{P_Y^{\text{like}}(\boldsymbol{y}|\boldsymbol{x}) P_X^{\text{prior}}(\boldsymbol{x})}{P_Y(\boldsymbol{y})} \propto P_\eta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x})) P_X^{\text{prior}}(\boldsymbol{x})$$

Main steps

- Construct P_X^{prior} based on prior data or constraints.
- Given $\boldsymbol{y}$, use Markov Chain Monte Carlo (MCMC) to construct a Markov chain whose stationary distribution is P_X^{post}.
- For the chain $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\}$, evaluate empirical statistics

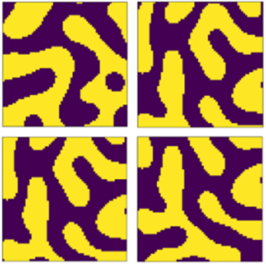
$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim P_X^{\text{post}}} [\phi(\boldsymbol{x})] \approx \frac{1}{N} \sum_{i=1}^{N} \phi(\boldsymbol{x}_i).$$

Bayesian formulation: challenges

- MCMC is prohibitively expensive when N_x is large.
- Characterization of priors for complex data.

Typical Gaussian prior $P_X^{\text{prior}}(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma^2}\right)$

However, prior knowledge may be samples like:



Microstructure profile for κ

Representing this data in the form of a prior is hard!

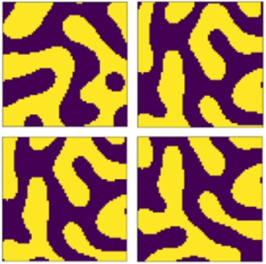
Bayesian formulation: challenges

- MCMC is prohibitively expensive when N_x is large.
- Characterization of priors for complex data.

$\Big\}$ Overcome using Generative Adversarial Networks

Typical Gaussian prior $P_X^{\text{prior}}(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma^2}\right)$

However, prior knowledge may be samples like:

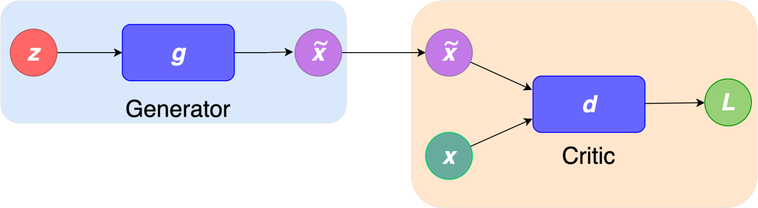


Microstructure profile for κ

Representing this data in the form of a prior is hard!

Generative adversarial network (GAN)

Designed by Goodfellow et al. (2014) to learn and sample from a target P_X.



Generator network:

- Generates fake samples of $\boldsymbol{x}$
- Latent variable $\boldsymbol{z} \in \Omega_Z \subset \mathbb{R}^{N_z}$, $N_z \ll N_x$.
- $\boldsymbol{g} : \Omega_Z \to \Omega_X$.
- $\boldsymbol{z} \sim P_Z$ simple distribution.

Critic network:

- Distinguishes fake samples from real
- $d : \Omega_X \to \mathbb{R}$.
- $\boldsymbol{x} \sim P_X$.
- $d(\boldsymbol{x})$ large for $\boldsymbol{x} \sim P_X$, small otherwise.

Wasserstein GAN

Proposed by Arjovsky et al. (2017)

- Objective function

$$L(\boldsymbol{g}, d) = \underset{\boldsymbol{x}\sim P_X}{\mathbb{E}}[d(\boldsymbol{x})] - \underset{\boldsymbol{z}\sim P_Z}{\mathbb{E}}[d(\boldsymbol{g}(\boldsymbol{z}))]$$

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

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

- For the optimal generator $\boldsymbol{g}^*$, using Kantorovich-Rubinstein dual characterization

$$\boldsymbol{g}^* = \arg\min_{\boldsymbol{g}} W_1(P_X, \boldsymbol{g}_\# P_Z)$$

- Convergence in W_1 implies weak convergence

$$\underset{\boldsymbol{x}\sim P_X}{\mathbb{E}}[\phi(\boldsymbol{x})] = \underset{\boldsymbol{z}\sim P_Z}{\mathbb{E}}[\phi(\boldsymbol{g}^*(\boldsymbol{z}))], \quad \forall\, \phi \in C_b(\Omega_X).$$

Outline

GANs as priors

Given:

- A set $\mathcal{S} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i \sim P_X^{\text{prior}}$.
- The direct map $\boldsymbol{f}(\boldsymbol{x})$ (exactly or approximately).
- The noise distribution P_η
- A noisy measurement $\boldsymbol{y}$

Goal: Determine P_X^{post} and evaluate statistics wrt it.

Solution of Physics-based Bayesian Inverse Problems with Deep Generative Priors, by Patel, Ray & Oberai, arXiv:2107.02926, 2021

GANs as priors

Step 1: Using $\mathcal{S}$, train a WGAN with generator $\boldsymbol{g}^*$.

Assume:

- $\boldsymbol{g}^*$ is the optimal generator satisfying the weak relation

$$\mathbb{E}_{\boldsymbol{x}\sim P_X^{\text{prior}}}[\phi(\boldsymbol{x})] = \mathbb{E}_{\boldsymbol{z}\sim P_Z}[\phi(\boldsymbol{g}^*(\boldsymbol{z}))], \quad \forall\, \phi \in C_b(\Omega_X).$$

- $\boldsymbol{f}$ and P_η are continuous.

Theorem

If the above assumptions hold, then we get a weak expression for P_X^{post}

$$\mathbb{E}_{\boldsymbol{x}\sim P_X^{\text{post}}}[\phi(\boldsymbol{x})] = \mathbb{E}_{\boldsymbol{z}\sim P_Z^{\text{post}}}[\phi(\boldsymbol{g}^*(\boldsymbol{z}))], \quad \forall\, \phi \in C_b(\Omega_X),$$

where

$$P_Z^{\text{post}}(\boldsymbol{z}|\boldsymbol{y}) \propto P_\eta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{g}^*(\boldsymbol{z})))P_Z(\boldsymbol{z}).$$

Sampling $\boldsymbol{x}$ from P_X^{post} $\equiv$ sampling $\boldsymbol{z}$ from P_Z^{post} and evaluating $\boldsymbol{x} = \boldsymbol{g}^*(\boldsymbol{z})$.

GANs as priors

Step 2: Generate a chain $\{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$ using MCMC.

Step 3: Evaluate statistics using Monte Carlo

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim P_X^{\text{post}}} [\phi(\boldsymbol{x})] \approx \frac{1}{n} \sum_{i=1}^{n} \phi(\boldsymbol{g}^*(\boldsymbol{z}_i))).$$

What do we gain?

- Ability to represent complex prior, if $\mathcal{S}$ is available.
- $N_z \ll N_x$ makes MCMC computationally tractable.

Inferring thermal conductivity

Given u, find κ satisfying

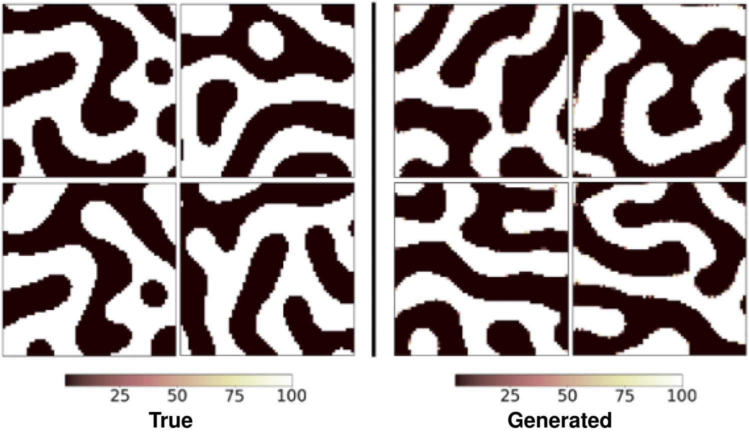
$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) &= b(\boldsymbol{\xi}), \qquad && \forall\, \boldsymbol{\xi} \in \Omega \subset \mathbb{R}^2 \\ u(\boldsymbol{\xi}) &= 0, \qquad && \forall\, \boldsymbol{\xi} \in \partial\Omega \end{aligned}$$

Problem setup:

- Measurement $\boldsymbol{y}$, noisy temperature field u on a 2D grid.
- Infer $\boldsymbol{x}$, nodal values of conductivity κ.
- Non-linear forward map $\boldsymbol{f}$ solves the PDE (FEM solver).
- Noise is assumed to be Gaussian iid.

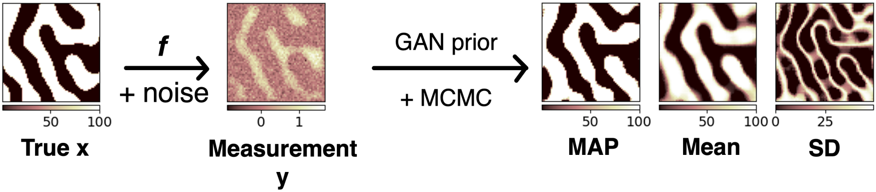
Inferring thermal conductivity ($b(\xi) = 10^3$)

Microstructure profile given by Cahn-Hilliard ($N_x = 4096, N_Z = 100$)

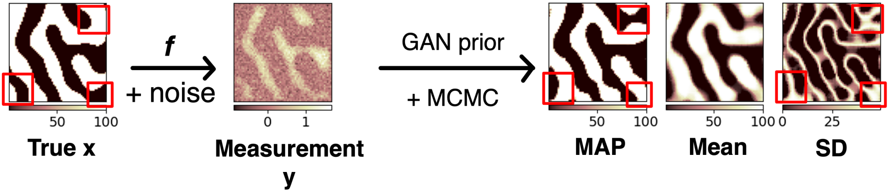


Inferring thermal conductivity ($b(\xi) = 10^3$)

Solving the inference problem on a test sample

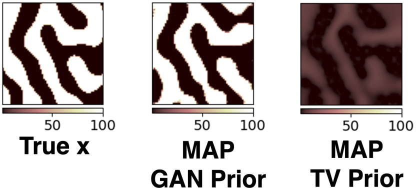


Solving the inference problem on a test sample



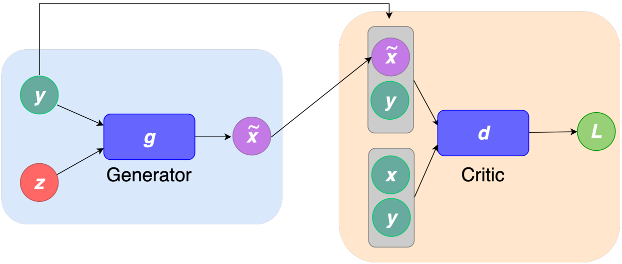
Inferring thermal conductivity ($b(\xi) = 10^3$)

Solving the inference problem on a test sample



Outline

Learning distributions conditioned on another field. Based on work by Adler et al. (2018) & Almahairi et al. (2018).



Generator network:

- $\boldsymbol{g} : \Omega_Z \times \Omega_Y \rightarrow \Omega_X$.
- $\boldsymbol{z} \sim P_Z$, $N_z \ll N_x$.
- $(\boldsymbol{x}, \boldsymbol{y}) \sim P_{XY}$

Critic network:

- $d : \Omega_X \times \Omega_Y \rightarrow \mathbb{R}$.
- $d(\boldsymbol{x}, \boldsymbol{y})$ large for real $\boldsymbol{x}$, small otherwise.

Fix $\boldsymbol{y}$, take many sample of $\boldsymbol{z} \rightarrow$ many samples of $\boldsymbol{x}$

Conditional WGANs

- Objective function

$$L(\boldsymbol{g}, d) = \mathop{\mathbb{E}}_{\substack{(\boldsymbol{x},\boldsymbol{y})\sim P_{XY} \\ \boldsymbol{z}\sim P_Z}} \left[d(\boldsymbol{x}, \boldsymbol{y}) - d\big(\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y}\big) \right]$$

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

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

- For the optimal generator $\boldsymbol{g}^*$

$$\boldsymbol{g}^*(., \boldsymbol{y}) = \arg\min_{\boldsymbol{g}} W_1(P_{X|Y}, \boldsymbol{g}_{\#}(., \boldsymbol{y})P_Z)$$

- Convergence in W_1 implies weak convergence

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P_{X|Y}} [\ell(\boldsymbol{x})] = \mathop{\mathbb{E}}_{\boldsymbol{z}\sim P_Z} [\ell(\boldsymbol{g}^*(\boldsymbol{z}, \boldsymbol{y}))], \quad \forall\, \ell \in C_b(\Omega_X).$$

GANs as posterior

Given:

- A set $\mathcal{S} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_n, \boldsymbol{y}_n)\}$, where $\boldsymbol{x}_i \sim P_X^{\text{prior}}$ and $\boldsymbol{y}_i \sim P_{Y|X}$.
- A noisy measurement $\boldsymbol{y}$

Goal: Determine P_X^{post} and evaluate statistics wrt it.

Step 1: Using $\mathcal{S}$, train a WGAN with generator $\boldsymbol{g}^*(\boldsymbol{z}, \boldsymbol{y})$.

Using Bayes and weak convergence of conditional WGAN for a given $\boldsymbol{y}$

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim P_X^{\text{post}}} [\ell(\boldsymbol{x})] = \mathop{\mathbb{E}}_{\boldsymbol{z} \sim P_Z} [\ell(\boldsymbol{g}^*(\boldsymbol{z}, \boldsymbol{y}))], \quad \forall\, \ell \in C_b(\Omega_X)$$

Sampling $\boldsymbol{x}$ from $P_X^{\text{post}} \equiv$ sampling $\boldsymbol{z}$ from P_Z and evaluating $\boldsymbol{x} = \boldsymbol{g}^*(\boldsymbol{z}, \boldsymbol{y})$.

Step 2: Query $\boldsymbol{g}^*$ and evaluate statistics using Monte Carlo

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim P_X^{\text{post}}} [\ell(\boldsymbol{x})] \approx \frac{1}{n} \sum_{i=1}^{n} \ell(\boldsymbol{g}^*(\boldsymbol{z}_i, \boldsymbol{y})), \quad \boldsymbol{z}_i \sim P_Z.$$

GANs as posterior

What do we gain?

- Ability to represent complex prior, if $\mathcal{S}$ is available.
- $N_z \ll N_x$.
- Sampling from a GAN is very simple.

Architecture of generator:

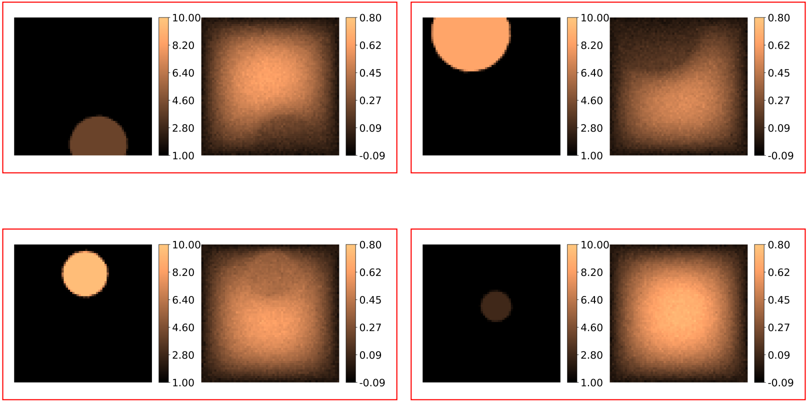
- U-net structure: control the spatial locality of $\boldsymbol{g}(\boldsymbol{z}, .)$ map.
- Conditional instance normalization: inject randomness at multiple scales.

Efficient posterior inference & generalization in physics-based Bayesian inference with conditional GANs, by Ray, Patel, Ramaswamy & Oberai; NeurIPS 2021 Deep Inverse Workshop, 2021.

Inferring thermal conductivity ($b(\xi) = 10$)

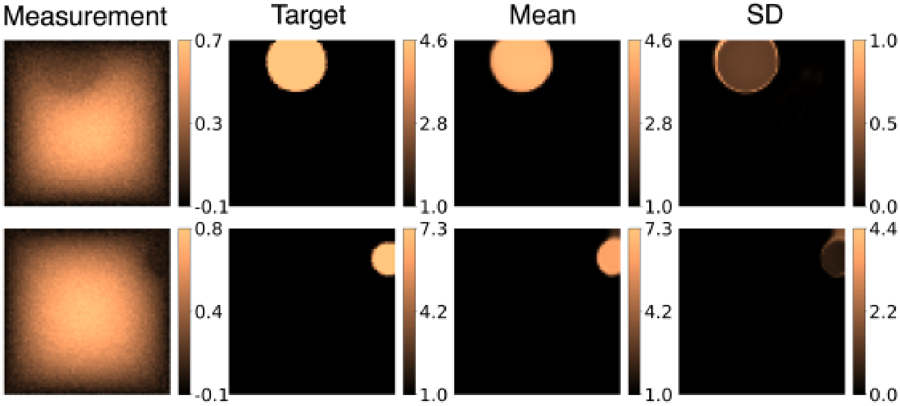
Assume κ is given by circular inclusions ($N_x = N_y = 4096, N_Z = 50$)

Training sample pairs: $(\boldsymbol{x}, \boldsymbol{y})$, $\quad \boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$

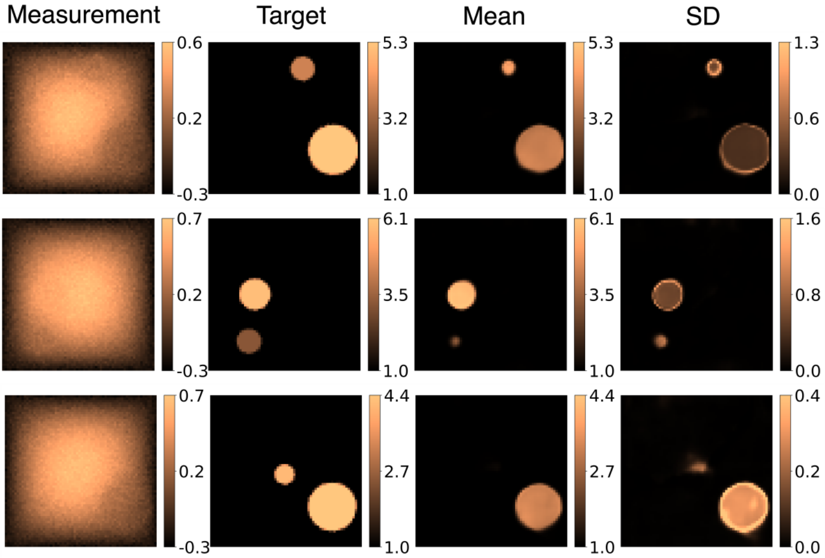


Inferring thermal conductivity ($b(\xi) = 10$)

Solving the inference problem with test $\boldsymbol{y}$



Inferring thermal conductivity ($b(\xi) = 10$) on O.O.D. samples



Conclusions

- GANs as priors and posterior in physics-based Bayesian inference.
- Ability to capture complex prior information.
- Dimensionality reduction using latent space.
- Generate point estimates to quantify uncertainty in inferred field.
- Hints at generalizability using cGAN with special architecture.
- Many more applications ...

Supported by: Army Research Office, Airbus Institute for Engineering Research (AIER, USC), Center for Advanced Research Computing (CARC, USC).

References

J. B. Keller.
Inverse Problems.
The American Mathematical Monthly, 83:107–118, 1976.

I. J. Goodfellow, J. P. -Abadie, M. Mirza, B. Xu, D. W. -Farley, S. Ozair, A. Courville, Y. Bengio
Generative Adversarial Networks.
Advances in Neural Information Processing Systems, 2672–2680, 2014.

M. Arjovsky, S. Chintala, L. Bottou
Wasserstein Generative Adversarial Networks.
PMLR, 70:214-223, 2017

T. Karras, T. Aila, S. Laine, J. Lehtinen
Progressive Growing of GANs for Improved Quality, Stability, and Variation.
arXiv:1710.10196, 2018

J. Adler, O. Öktem
Deep Bayesian Inversion.
arXiv:1811.05910, 2018

References

A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, A. Courville
Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data.
PMLR, 80:195-204, 2018

D. Patel, A. A. Oberai
GAN-based Priors for Uncertainty Quantification
arXiv:2003.12597, 2020

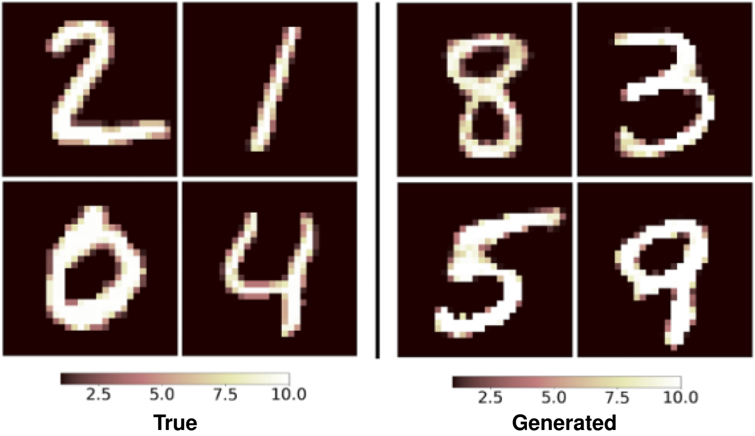
D. Patel, D. Ray, A. A. Oberai
Solution of Physics-based Bayesian Inverse Problems with Deep Generative Priors
arXiv:2107.02926, 2021

Comparing the two approaches

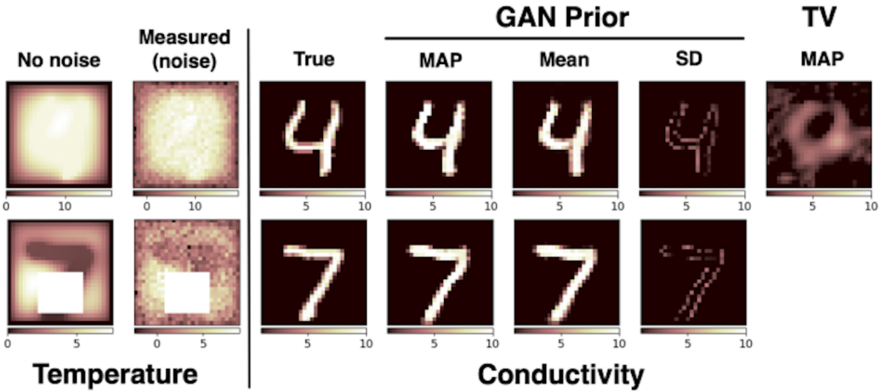
	GAN as prior	**GAN as posterior**
Data generation	$\boldsymbol{x} \sim P_X^{\text{prior}}$	$\boldsymbol{x} \sim P_X^{\text{prior}}$, $\boldsymbol{y} \sim P_{Y\|X}$
Forward model	Need $\boldsymbol{f}$ and $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$	Possibly need $\boldsymbol{f}$ to generate data
Sampling	GAN and MCMC	Only GAN
Generalizability	Hard to control	Better control

Inferring thermal conductivity (MNIST)

Assume that κ is given by MNIST digits ($N_x = 784$, $N_Z = 100$)



Solving the inference problem on test data



Inverse Radon transform (CT)

Find the tissue density $\rho : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ given the line Radon transforms

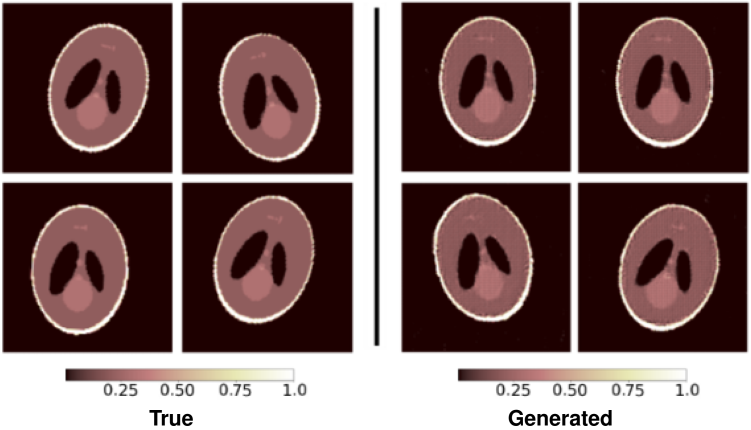
$$\mathcal{R}_{t,\psi} = \int_{\gamma_{t,\psi}} \rho \mathrm{d}\gamma$$

where $\gamma_{t,\psi}$ is the line at an angle ψ and at a signed-distance of t from the center of Ω.

Problem setup:

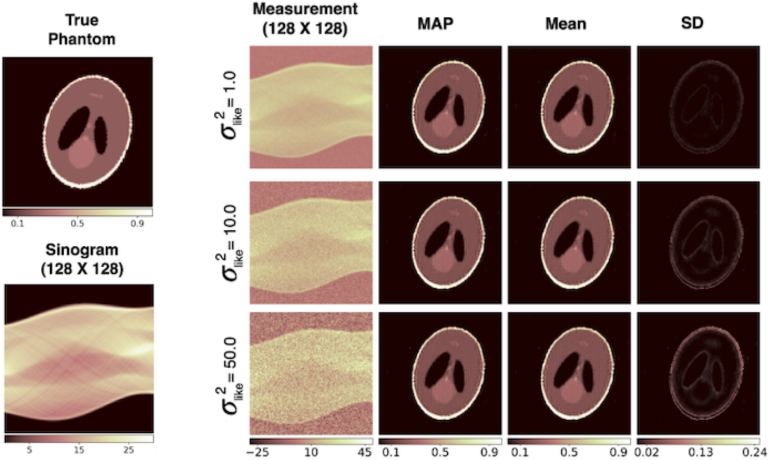
- Infer $\boldsymbol{x}$, nodal values of ρ.
- Linear forward map $\boldsymbol{f}$, Radon transform.
- Measurement $\boldsymbol{y}$, noisy Radon transforms on a set of lines.
- Noise is assumed to be Gaussian iid.

ρ given by perturbed Shepp-Logan phantoms ($N_x = 16384, N_Z = 100$)



Inverse Radon transform (CT)

Solving the inference problem



Inferring the initial condition

Given $u(\xi, T)$, find u_0

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

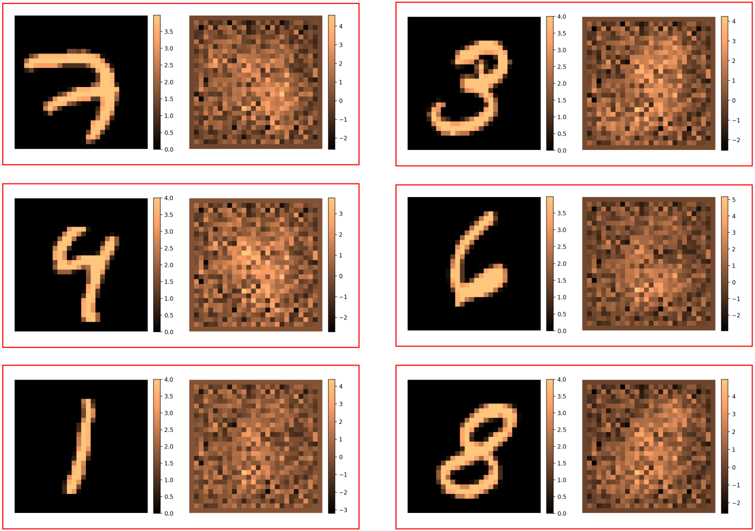
Severely ill-posed problem!

Problem setup:

- Infer $\boldsymbol{x}$, initial temperature field u on a 2D grid.
- Measurement $\boldsymbol{y}$, noisy temperature field u on a 2D grid.
- Generate $\mathcal{S}$ by sampling $\boldsymbol{x} \sim P_X^{\text{prior}}$ and evaluating $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$.
- Train WGAN on $\mathcal{S}$

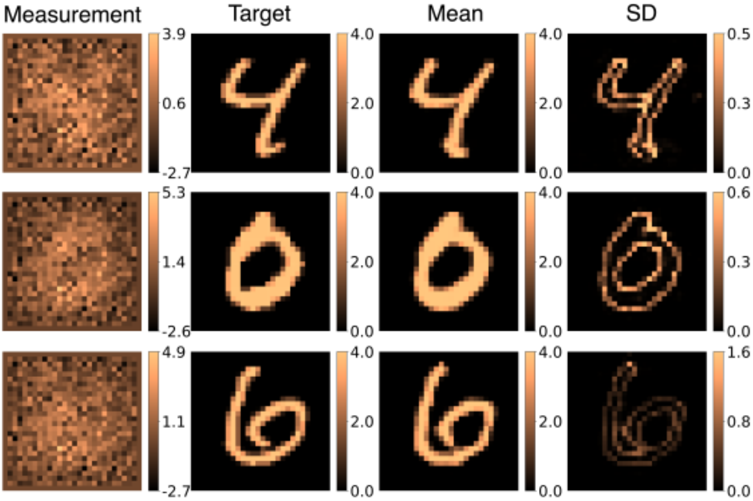
Inferring the initial condition

Assume u_0 is given by MNIST ($N_x = N_y = 784$, $N_Z = 100$)

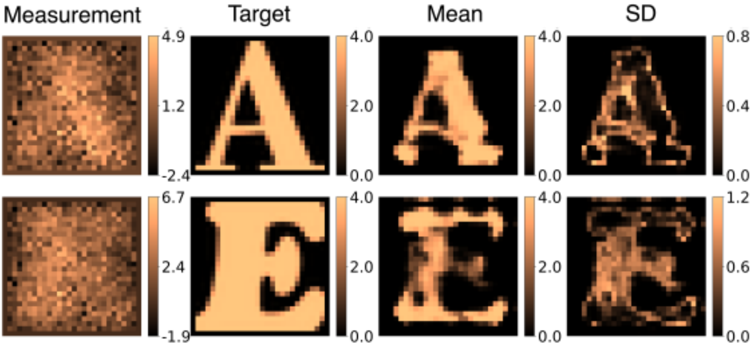


Inferring the initial condition

Solving the inference problem



Solving the inference problem



Local nature of learned inverse

Evaluate average gradient magnitude for the k-th pixel of $\boldsymbol{x}$

$$\overline{\text{grad}}_k = \frac{1}{1000}\sum_{i=1}^{100}\sum_{j=1}^{10}\left|\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}(\boldsymbol{z}^{(j)}, \boldsymbol{y}^{(i)})\right|, \quad \boldsymbol{y}^{(i)} \sim P_Y, \quad \boldsymbol{z}^{(j)} \sim P_Z, \quad 1 \leq k \leq 4096.$$