Solving physics-based inverse problems using GANs

Deep Ray
Department of Aerospace & Mechanical Engineering

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

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

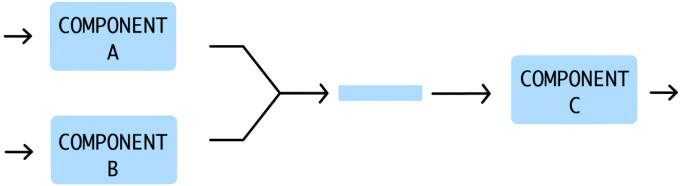
Charlotte
October 8, 2021

USC Viterbi
School of Engineering

Motivation

Theoretically sound techniques available for computational physics

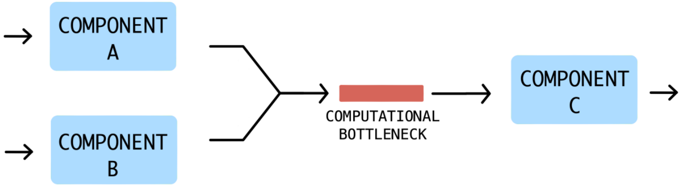
- Solving differential equations
- Optimization and control
- Uncertainty quantification
- Inverse problems
- ...



Motivation

Theoretically sound techniques available for computational physics

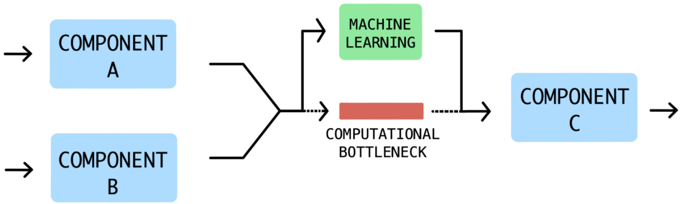
- Solving differential equations
- Optimization and control
- Uncertainty quantification
- Inverse problems
- ...



Motivation

Theoretically sound techniques available for computational physics

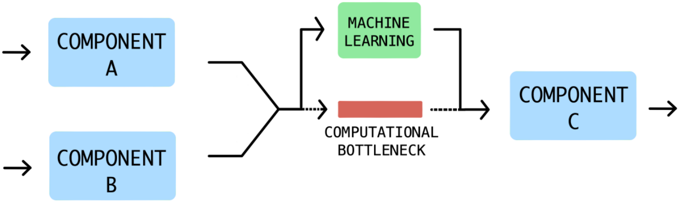
- Solving differential equations
- Optimization and control
- Uncertainty quantification
- Inverse problems
- ...



Motivation

Theoretically sound techniques available for computational physics

- Solving differential equations
- Optimization and control
- Uncertainty quantification
- Inverse problems
- ...



Key ingredients: Inverse problems, Bayesian inference and Generative Adversarial Networks (GANs).

Outline

1. Ingredients

2. GANs as priors

3. GANs as posterior

4. Conclusions

Outline

1. Ingredients
2. GANs as priors
3. GANs as posterior
4. Conclusions

Inverse problems

"We call two problems inverses of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for sometime, while the other is newer and not so well understood. In such cases, the former is called the direct problem, while the latter is called the inverse problem."

– Joseph Keller, 1976

Inverse problems

Consider the heat conduction equation for the temperature field u

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

Direct: Given {PDE, b, u_0, κ} $\longrightarrow u(T)$.

Inverse: Given {PDE, b, u_0, $u(T)$} $\longrightarrow \kappa$.

Challenges with inverse problems:

- Inverse map is not well posed.
- Noisy measurements from direct problem.
- Need to encode prior knowledge about inferred field.

Two approaches: regularization and Bayesian inference.

Notations:

- Parameter we wish to infer $\boldsymbol{x} \in \Omega_X \subset \mathbb{R}^{N_x}$.
- Measured response from direct problem $\boldsymbol{y} \in \Omega_Y \subset \mathbb{R}^{N_y}$.
- Direct map $\boldsymbol{f} : \Omega_X \to \Omega_Y$. Sometimes,

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\eta}$$

where $\boldsymbol{\eta}$ is noise with distribution P_η.

Bayesian inference

Bayes theorem gives us:

$$P_{X|Y}(\boldsymbol{x}|\boldsymbol{y}) = \frac{P_{Y|X}(\boldsymbol{y}|\boldsymbol{x})P_X(\boldsymbol{x})}{P_Y(\boldsymbol{y})}$$

Applying this to the inverse problem: given $\boldsymbol{y}$ and prior information, infer $\boldsymbol{x}$

$$P_X^{\text{post}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Q}P_\eta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}))P_X^{\text{prior}}(\boldsymbol{x})$$

where

- P_X^{prior}: prior distribution, obtained from samples or other constraints.
- P_η: distribution of measurement noise.
- Q: evidence.

Bayesian inference

Posterior distribution

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

Steps:

- Construct P_X^{prior}.
- For a given $\boldsymbol{y}$, use MCMC to generate Markov chains for $\boldsymbol{x}$ containing samples from P_X^{post}.

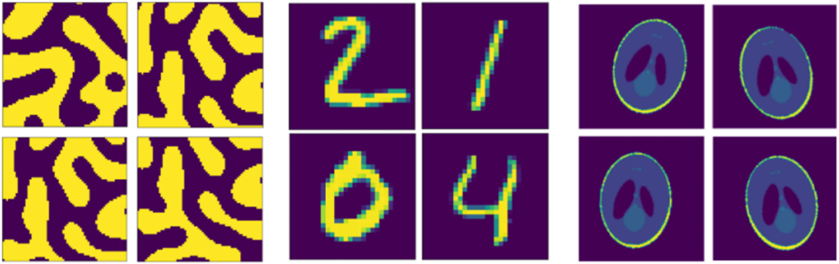
Challenges:

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

Bayesian inference

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

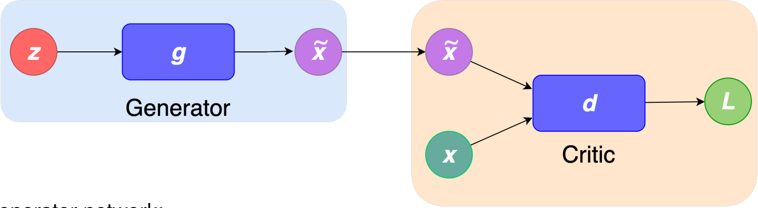
However, prior knowledge may be samples like:



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 $\boldsymbol{z} \in \Omega_Z \subset \mathbb{R}^{N_z}$.
- $\boldsymbol{g} : \Omega_Z \to \Omega_X$.
- $\boldsymbol{z} \sim P_Z$ simple distribution.
- $N_z \ll N_x$.

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

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

- $\boldsymbol{g}$ 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

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

What a GAN can do

Results by Karras et al. (2018) from NVIDIA.
CELEBA-HQ dataset, $N_z = 512$, $N_x = 3.14 \times 10^6$.

Outline

1. Ingredients

2. GANs as priors

3. GANs as posterior

4. Conclusions

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

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

Assume:

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

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

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

Then, we can show using the expression for P_X^{post}

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

where

$$P_Z^{\text{post}}(\boldsymbol{z}|\boldsymbol{y}) = \frac{1}{Q} P_\eta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{g}^*(\boldsymbol{z})))P_Z(\boldsymbol{z})$$

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

Step 2: Generate an MCMC approximation $P_Z^{\text{mcmc}}(\boldsymbol{z}|\boldsymbol{y}) \approx P_Z^{\text{post}}(\boldsymbol{z}|\boldsymbol{y})$.

Step 3: Evaluate statistics using Monte Carlo

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

What do we gain?

- Ability to represent complex prior, if $\mathcal{S}$ is available.
- $N_z \ll N_x$ makes MCMC computational 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. Implemented in Fenics.
- Noise is assumed to be Gaussian iid.

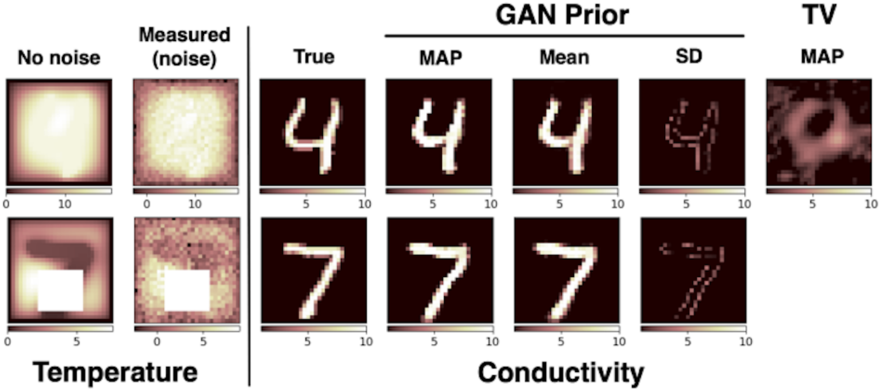
Inferring thermal conductivity (MNIST)

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

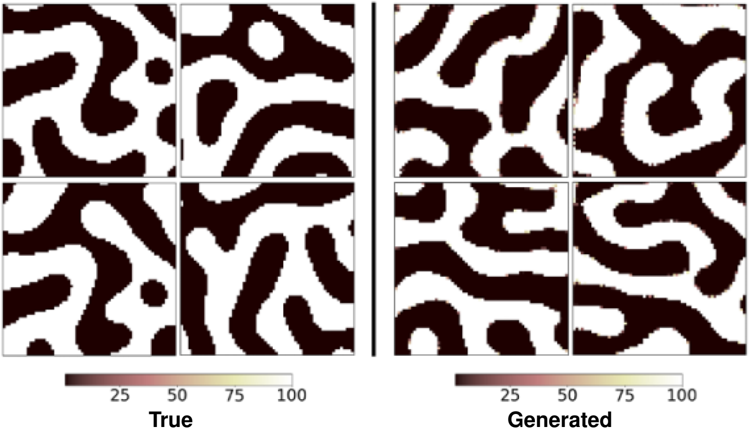
True

Generated

Solving the inference problem on test data

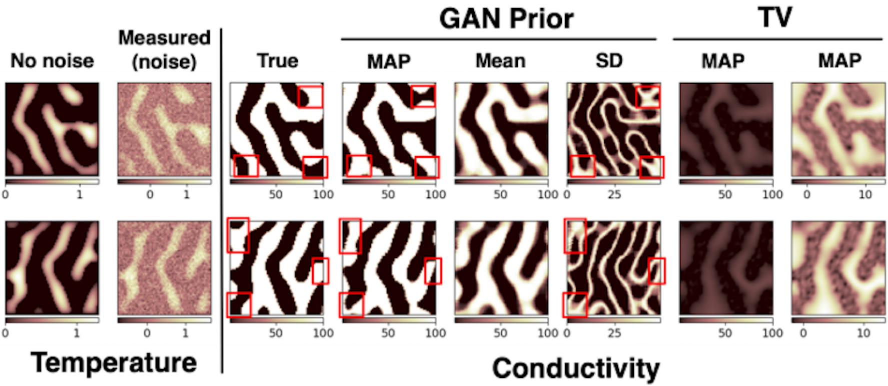


Microstructure profile given by Cahn-Hilliard ($N_x = 4096, N_Z = 100$)



Inferring thermal conductivity (microsctructure)

Solving the inference problem



Inverse Radon transform (CT)

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

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

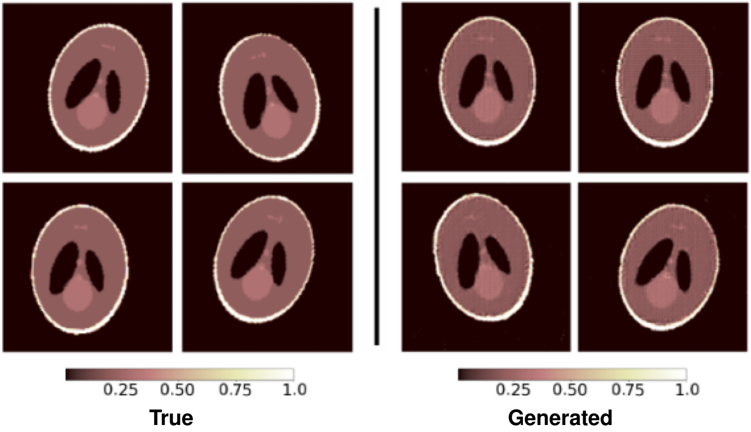
where $\gamma_{t,\psi}$ is the line at an angle ψ and at a signed-distance of t from the center of Ω.

Problem setup:

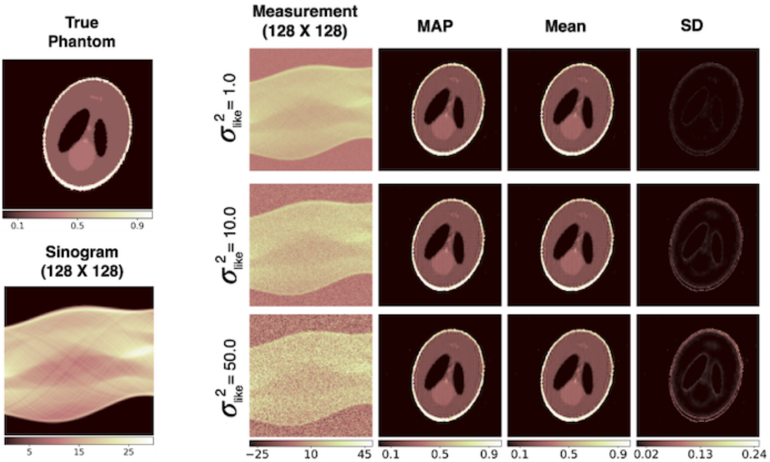
- Infer $\boldsymbol{x}$, nodal values of ρ.
- Linear forward map $\boldsymbol{f}$, Radon transform.
- Measurement $\boldsymbol{y}$, noisy Radon transforms on a set of lines.
- Noise is assumed to be Gaussian iid.

Inverse Radon transform (CT)

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



Solving the inference problem



Outline

1. Ingredients

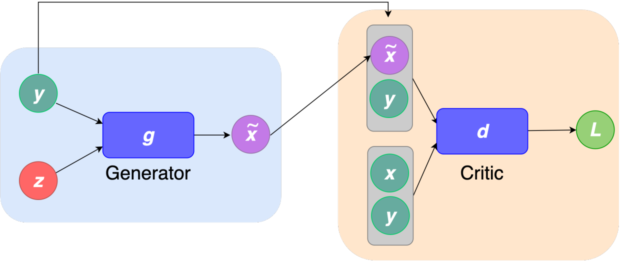
2. GANs as priors

3. GANs as posterior

4. Conclusions

Conditional WGANs

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



Generator network:

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

Critic network:

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

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

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

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 $\boldsymbol{y}$

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

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

GANs as posterior

Step 2: Evaluate statistics using Monte Carlo

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

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.

Inferring thermal conductivity

Given u, find κ satisfying

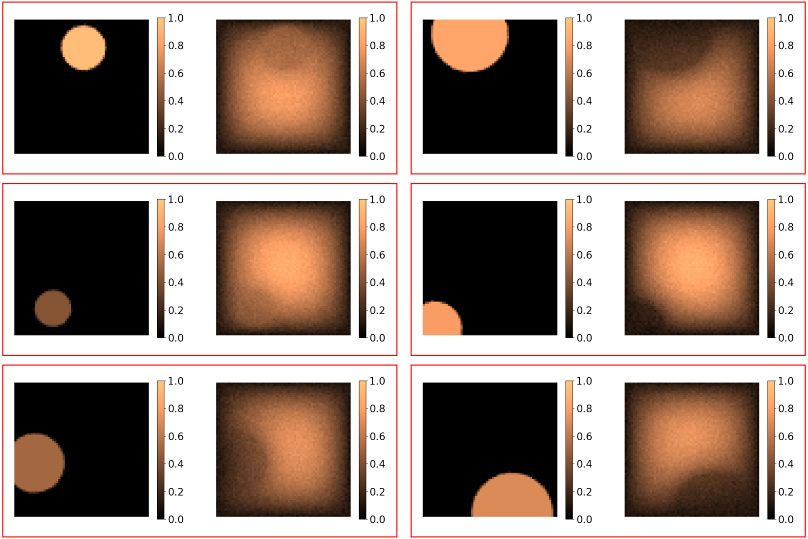
$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) &= 10, \qquad && \forall\, \boldsymbol{\xi} \in \Omega \subset \mathbb{R}^2 \\ u(\boldsymbol{\xi}) &= 0, && \forall\, \boldsymbol{\xi} \in \partial\Omega \end{aligned}$$

Problem setup:

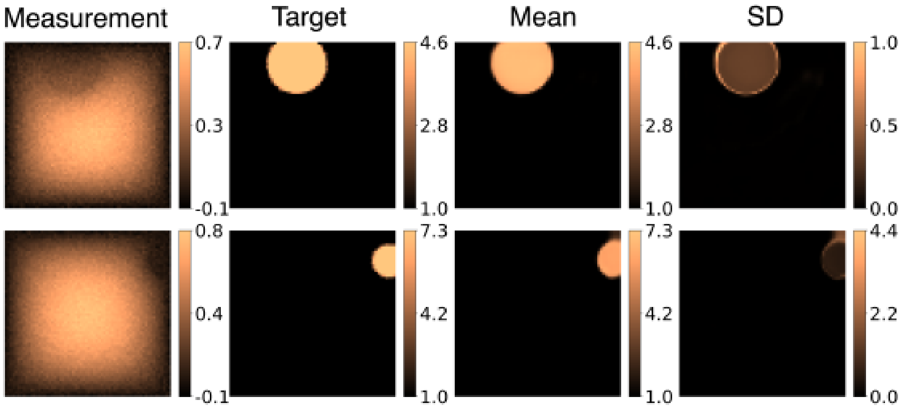
- Infer $\boldsymbol{x}$, nodal values of conductivity κ.
- Measurement $\boldsymbol{y}$, noisy temperature field u on a 2D grid.
- Generate $\mathcal{S}$ by sampling $\boldsymbol{x} \sim P_X^{\text{prior}}$ and evaluating $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$.
- Train WGAN on $\mathcal{S}$

Inferring thermal conductivity

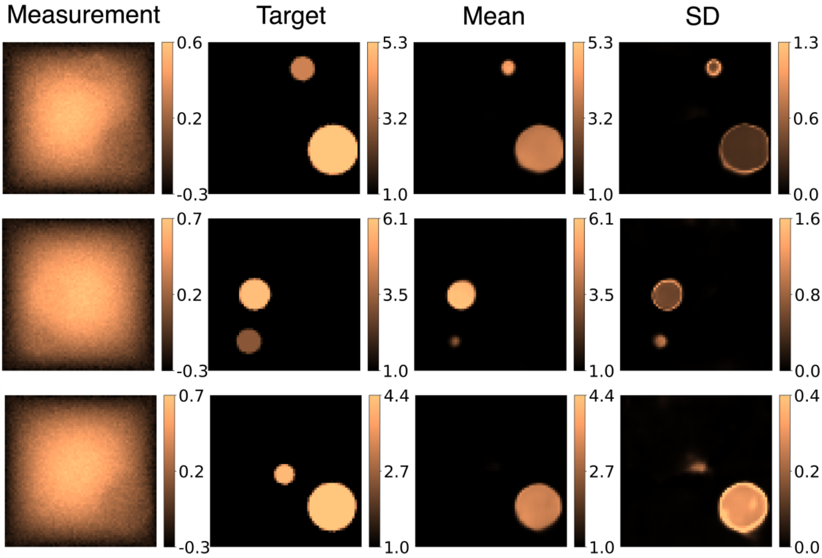
Assume κ is given by circular inclusions ($N_x = N_y = 4096$, $N_Z = 50$)



Solving the inference problem



Inferring thermal conductivity



Inferring the initial condition

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

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

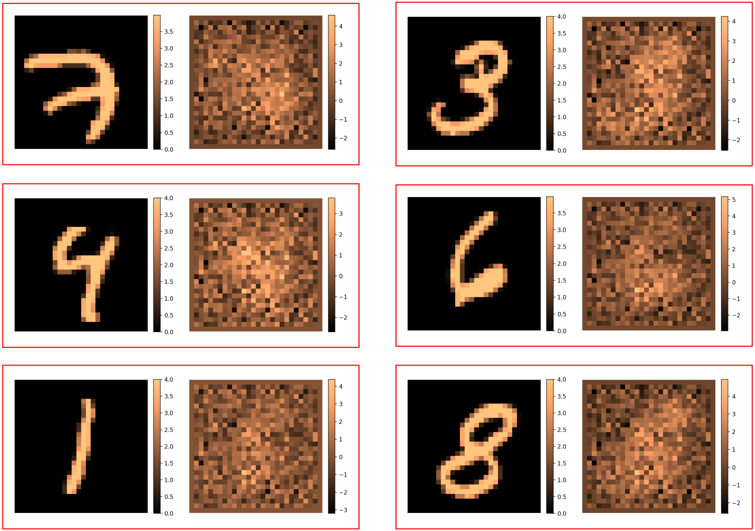
Severely ill-posed problem!

Problem setup:

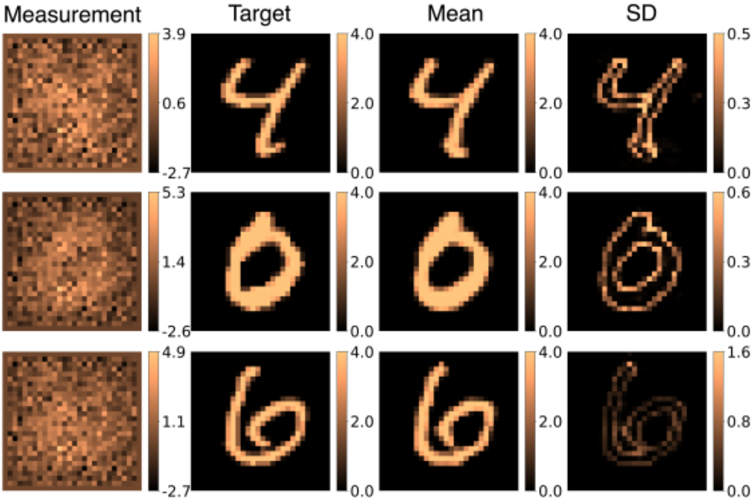
- Infer $\boldsymbol{x}$, initial temperature field u on a 2D grid.
- Measurement $\boldsymbol{y}$, noisy temperature field u on a 2D grid.
- Generate $\mathcal{S}$ by sampling $\boldsymbol{x} \sim P_X^{\text{prior}}$ and evaluating $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \eta$.
- Train WGAN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

Inferring the initial condition

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

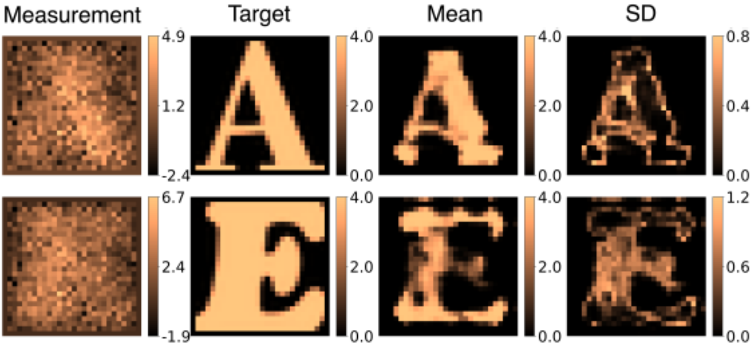


Solving the inference problem



Inferring the initial condition

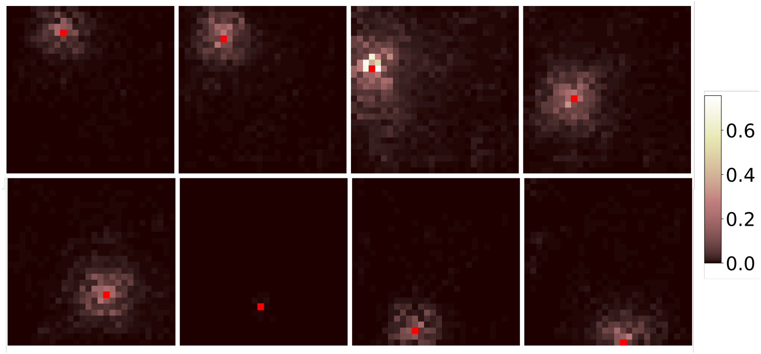
Solving the inference problem



Local-nature of learned inverse

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

$$\overline{\text{grad}}_k = \frac{1}{100} \sum_{i=1}^{10} \sum_{j=1}^{10} \left| \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}(\boldsymbol{z}^{(j)}, \boldsymbol{y}^{(i)}) \right|, \quad \boldsymbol{y}^{(i)} \sim P_Y, \quad \boldsymbol{z}^{(j)} \sim P_Z, \quad 1 \leq k \leq 784.$$



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

Conclusions

- GANs as priors and posterior in physics-based Bayesian inference.
- Ability to capture complex prior information.
- Dimensional 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