Stochastic Inverse Problems in Digital Twins

Yunan Yang

April 17, 2025

Department of Mathematics, Cornell University

This is a joint work with Qin Li (UW Madison), Li Wang (UMN Twin Cities) and Maria Oprea (Cornell).

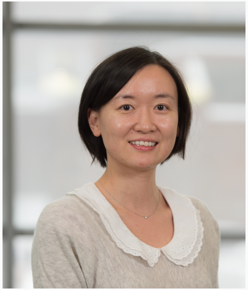
- Qin Li, Li Wang, and Y., 2024. Differential Equation–Constrained Optimization with Stochasticity. *SIAM/ASA Journal on Uncertainty Quantification, 12(2), pp.549-578.*
- Li, Q., Oprea, M., Wang, L. and Y., 2024. Stochastic Inverse Problem: stability, regularization and Wasserstein gradient flow. *arXiv preprint arXiv:2410.00229.*
- Li, Q., Wang, L. and Y., 2025. Least-Squares Problem Over Probability Measure Space. *arXiv preprint arXiv:2501.09097.*

East Coast Optimization Meeting (ECOM) April 17 to 18, 2025

Collaborators

Qin Li
(UW Madison)

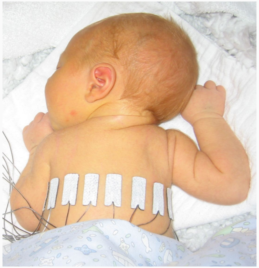
Li Wang
(UMN Twin Cities)

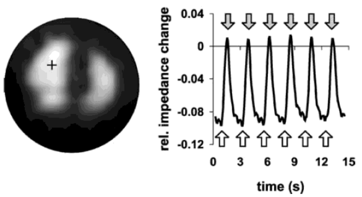


Maria Oprea
(Cornell)

Motivation

Calderón's Problem (Electrical Impedance Tomography, EIT)





$$\begin{cases} \nabla \cdot (\gamma(x) \nabla u) = 0, & x \in \Omega \\ u(x) = \psi, & x \in \partial\Omega \end{cases}$$

Given "Dirichlet-to-Neumann" map

$$\Lambda_\gamma : \quad \mathcal{H}^{1/2}(\partial\Omega) \longrightarrow \mathcal{H}^{-1/2}(\partial\Omega)$$
$$\Lambda_\gamma : \quad \psi \longrightarrow \gamma \nabla u_\psi \cdot \mathbf{n}\big|_{\partial\Omega}$$

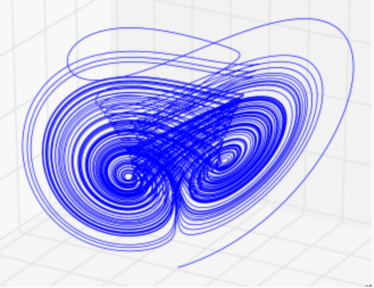
the goal is to find

$$\gamma(x), \quad x \in \Omega.$$

Kohn, R. V., & Vogelius, M. (1987). Relaxation of a variational method for impedance computed tomography. CPAM.

Learning the Dynamics

"Chen" System [Chen-Ueta, 1999]



Y.-Nurbekyan-Negrini-Martin-Pasha, 2023. SIADS.

Botvinick-Greenhouse, J., Martin, R. & Y., 2023. Chaos.

Parameterized dynamical system in the Lagrangian form

$$\dot{\mathbf{x}} = v(\mathbf{x};\theta) \quad \text{or} \quad dX_t = v(\mathbf{x};\theta)dt + \sigma dW_t$$

or the Eulerian form (Fokker–Planck Eqn.)

$$\partial_t \rho(\mathbf{x},t) + \nabla \cdot (v(\mathbf{x};\theta)\rho(\mathbf{x},t)) = \frac{\sigma^2}{2}\Delta\rho(\mathbf{x},t)$$

where θ can correspond to

- basis coefficients
  e.g., SINDy [Brunton-Proctor-Kutz, 2016],
- neural network weights
  e.g., Neural-ODE [Chen et al., 2018],
- other parameterizations [Lu-Maggioni-Tang,2021]
- or nonparametric using Frobenius–Perron or Koopman operators [Kloeckner, 2018]

Deterministic Inverse Problem

$$\mathsf{M}(\theta) = g\,, \quad \mathsf{M} : \mathcal{P} \mapsto \mathcal{D}\,, \tag{1}$$

where $\theta \in \mathcal{P}$ is the function space of parameters, M is the forward operator, with $g \in \mathcal{D}$, the function space of data. M can be implicitly defined.

Deterministic Inverse Problem

$$\mathsf{M}(\theta) = g\,, \quad \mathsf{M} : \mathcal{P} \mapsto \mathcal{D}\,, \tag{1}$$

where $\theta \in \mathcal{P}$ is the function space of parameters, M is the forward operator, with $g \in \mathcal{D}$, the function space of data. M can be implicitly defined.

Examples

- In image processing, θ is the clean image and g is the noisy/blurred image.

Deterministic Inverse Problem

$$\mathsf{M}(\theta) = g\,, \quad \mathsf{M} : \mathcal{P} \mapsto \mathcal{D}\,, \tag{1}$$

where $\theta \in \mathcal{P}$ is the function space of parameters, M is the forward operator, with $g \in \mathcal{D}$, the function space of data. M can be implicitly defined.

Examples

- In image processing, θ is the clean image and g is the noisy/blurred image.
- Calderón's Problem: $\begin{cases} \nabla \cdot (\theta \nabla u) = 0 & \text{on } \Omega \\ u = \phi & \text{on } \partial\Omega \end{cases}$, g is the DtN map.

Deterministic Inverse Problem

$$\mathsf{M}(\theta) = g\,, \quad \mathsf{M} : \mathcal{P} \mapsto \mathcal{D}\,, \tag{1}$$

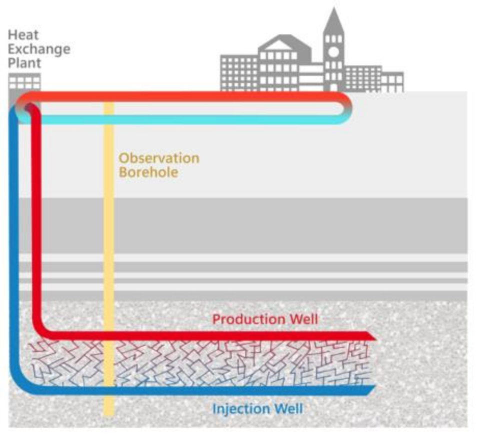
where $\theta \in \mathcal{P}$ is the function space of parameters, M is the forward operator, with $g \in \mathcal{D}$, the function space of data. M can be implicitly defined.

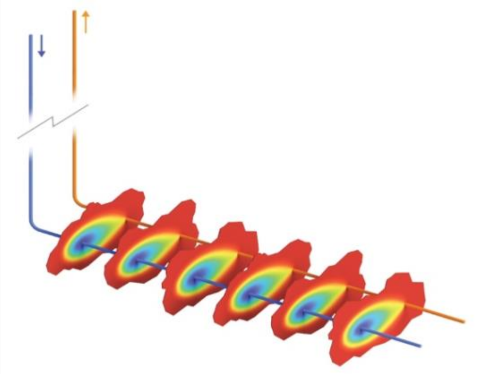
Examples

- In image processing, θ is the clean image and g is the noisy/blurred image.
- Calderón's Problem: $\begin{cases} \nabla \cdot (\theta \nabla u) = 0 & \text{on } \Omega \\ u = \phi & \text{on } \partial\Omega \end{cases}$, g is the DtN map.
- In dynamical system modeling, θ parameterizes the drift/diffusion, and g is the observed trajectory.

New Types of Inverse Problem: Sand Percentage in River

Enhanced Geothermal Systems





[Arson et al., 2024;arXiv:2412.11421]

Stochastic Inverse Problem [Breidt-Butler-Estep, 2011]

In certain applications, the deterministic framework is challenging.

- The math modeling is based on data gathered from a variety of subjects.
- It is impractical to conduct *repeated* measurements on a single subject.

Stochastic Inverse Problem [Breidt-Butler-Estep, 2011]

In certain applications, the deterministic framework is challenging.

- The math modeling is based on data gathered from a variety of subjects.
- It is impractical to conduct *repeated* measurements on a single subject.

Thus, one must employ a model that incorporates **a parameter distribution**, which gives rise to the so-called Stochastic Inverse Problem.

Stochastic Inverse Problem [Breidt-Butler-Estep, 2011]

In certain applications, the deterministic framework is challenging.

- The math modeling is based on data gathered from a variety of subjects.
- It is impractical to conduct *repeated* measurements on a single subject.

Thus, one must employ a model that incorporates **a parameter distribution**, which gives rise to the so-called Stochastic Inverse Problem.

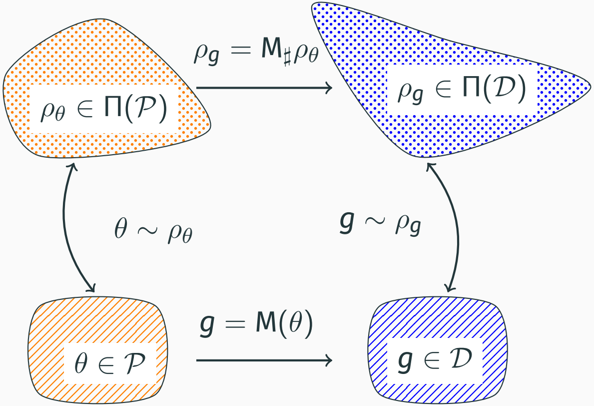
For forward problem is a push-forward map and ρ_θ is the unknown:

$$\boxed{\rho_g = \mathsf{M}_\sharp \rho_\theta =: F_\mathsf{M}(\rho_\theta)}, \quad F_\mathsf{M} : \Pi(\mathcal{P}) \mapsto \Pi(\mathcal{D}). \tag{2}$$

We say $\nu = \mathsf{M}_\sharp \mu$ if for any Borel measurable set B, $\nu(B) = \mu(\mathsf{M}^{-1}(B))$.

Intuitively, a change of variable through the map M

Deterministic Inverse Problem to Stochastic Inverse Problem



A diagram showing the relations between the deterministic (1) and the stochastic problem (2).

Computational Aspects

Stochastic Inverse Problem — Solvers

- **Deterministic** Inverse problem:

$$\mathsf{M}(\theta) = g$$

- Optimization problem:

$$\min_{\theta} d_o(\mathsf{M}(\theta), g^*)$$

- Optimization algorithms: gradient descent, nonlinear CG, etc.

Stochastic Inverse Problem — Solvers

- **Deterministic** Inverse problem:

$$\mathsf{M}(\theta) = g$$

- Optimization problem:

$$\min_{\theta} d_o(\mathsf{M}(\theta), g^*)$$

- Optimization algorithms: gradient descent, nonlinear CG, etc.

- **Stochastic** Inverse problem:

$$\rho_g = \mathsf{M}_\sharp \rho_\theta$$

- Optimization problem:

$$\min_{\rho_\theta} \mathcal{D}(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$$

- Optimization algorithms: ??? over the probability space

Stochastic Inverse Problem — Solvers

- **Deterministic** Inverse problem:

$$\mathsf{M}(\theta) = g$$

- Optimization problem:

$$\min_{\theta} d_o(\mathsf{M}(\theta), g^*)$$

- Optimization algorithms: gradient descent, nonlinear CG, etc.

- **Stochastic** Inverse problem:

$$\rho_g = \mathsf{M}_\sharp \rho_\theta$$

- Optimization problem:

$$\min_{\rho_\theta} \mathcal{D}(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$$

- Optimization algorithms: ??? over the probability space

There are two important metric/divergence that matter here (D and $\mathfrak{G}$):

$$\boxed{\rho_\theta^* = \operatorname*{argmin}_{\rho_\theta \in (\Pi(\mathcal{P}), \mathfrak{G})} D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)\,.} \tag{3}$$

Gradient Flow (Analogous to Gradient Descent)

The gradient flow for the energy $J(\rho_\theta) := D(\mathsf{M}_\sharp\rho_\theta, \rho_g^*)$ under the metric $\mathfrak{G}$ is

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp\rho_\theta, \rho_g^*)}\,. \tag{4}$$

Gradient Flow (Analogous to Gradient Descent)

The gradient flow for the energy $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$ under the metric $\mathfrak{G}$ is

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{4}$$

Example 1: Consider $\mathfrak{G} = W_2$ and $D = \mathrm{KL}$:

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\log \frac{\rho_g}{\rho_g^*}(M(\theta)) \right) \right)\,.$$

Gradient Flow (Analogous to Gradient Descent)

The gradient flow for the energy $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$ under the metric $\mathfrak{G}$ is

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{4}$$

Example 1: Consider $\mathfrak{G} = W_2$ and $D = $ KL:

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\log \frac{\rho_g}{\rho_g^*}(M(\theta)) \right) \right) .$$

Example 2: Consider $\mathfrak{G} = W_2$ and $D = W_2$:

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta\, \phi(M(\theta)) \right) \qquad \phi \text{ is the Kantorovich potential}$$

Gradient Flow (Analogous to Gradient Descent)

The gradient flow for the energy $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$ under the metric $\mathfrak{G}$ is

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{4}$$

Example 1: Consider $\mathfrak{G} = W_2$ and $D = \mathbf{KL}$:

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\log \frac{\rho_g}{\rho_g^*}(M(\theta)) \right) \right)\,.$$

Example 2: Consider $\mathfrak{G} = W_2$ and $D = W_2$:

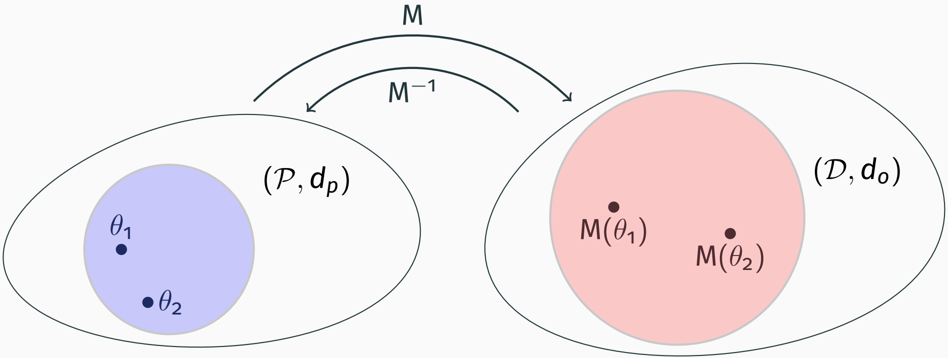
$$\partial_t \rho_\theta = \nabla_\theta \cdot (\rho_\theta \nabla_\theta\, \phi(M(\theta))) \qquad \phi \text{ is the Kantorovich potential}$$

Example 3: Consider $\mathfrak{G} = H^2$ (Hellinger) and $D = \chi^2$:

$$\partial_t \rho_\theta = 8 \rho_\theta \left[\int \frac{\rho_g}{\rho_g^*}(M(\theta)) \rho_\theta \mathrm{d}\theta - \frac{\rho_g}{\rho_g^*}(M(\theta)) \right]\,.$$

Well-Posedness: Stability

Stability



We need probability metrics to quantify the size of the blue and red balls.

M is invertible

Suppose M^{-1} exists and is Hölder continuous:

$$\|\mathsf{M}^{-1}(g_1) - \mathsf{M}^{-1}(g_2)\| \leq C_{\mathsf{M}^{-1}} \|g_1 - g_2\|^{\beta}\,, \quad \beta \in (0, 1]\,.$$

(Deterministic inverse problem is well-posed.)

M **is invertible**

Suppose M^{-1} exists and is Hölder continuous:

$$\|\mathsf{M}^{-1}(g_1) - \mathsf{M}^{-1}(g_2)\| \leq C_{\mathsf{M}^{-1}} \|g_1 - g_2\|^{\beta}\,, \quad \beta \in (0, 1]\,.$$

(Deterministic inverse problem is well-posed.)

Let $\rho_g, \widehat{\rho_g} \in \Pi(\mathbb{R}^n)$ be two data distributions. Their parameter distributions are

$$\rho_\theta = \mathsf{M}_\sharp^{-1} \rho_g, \quad \text{and} \quad \widehat{\rho_\theta} = \mathsf{M}_\sharp^{-1} \widehat{\rho_g}$$

M is invertible

Suppose M^{-1} exists and is Hölder continuous:

$$\|\mathsf{M}^{-1}(g_1) - \mathsf{M}^{-1}(g_2)\| \leq C_{\mathsf{M}^{-1}} \|g_1 - g_2\|^{\beta}, \quad \beta \in (0, 1].$$

(Deterministic inverse problem is well-posed.)

Let $\rho_g, \widehat{\rho_g} \in \Pi(\mathbb{R}^n)$ be two data distributions. Their parameter distributions are

$$\rho_\theta = \mathsf{M}_\sharp^{-1} \rho_g, \quad \text{and} \quad \widehat{\rho_\theta} = \mathsf{M}_\sharp^{-1} \widehat{\rho_g}$$

Theorem (Ernst et al.,2022)
Consider the p-Wasserstein metric.

$$\boxed{W_p\left(\rho_\theta, \ \widehat{\rho_\theta}\right) \leq C_{\mathsf{M}^{-1}} W_p\left(\rho_g, \ \widehat{\rho_g}\right)^{\beta}.}$$

Theorem (Qin-Oprea-Wang-Y.,2024)
Under any f-divergence ($\mathcal{D}_f$), we have

$$\boxed{\mathcal{D}_f\left(\rho_\theta || \widehat{\rho_\theta}\right) = \mathcal{D}_f\left(\rho_g || \widehat{\rho_g}\right)}$$

Solution Characterization

M is non-invertible

We have two cases for a general nonlinear M:

1. M is “under-determined”, i.e., M is not injective
2. M is “over-determined”, i.e., M is not surjective

[Li, Oprea, Wang, Y., 2025]

M is non-invertible

We have two cases for a general nonlinear M:

1. M is "under-determined", i.e., M is not injective
2. M is "over-determined", i.e., M is not surjective

In the under-determined case, we lose **uniqueness**.

[Li, Oprea, Wang, Y., 2025]

M is non-invertible

We have two cases for a general nonlinear M:

1. M is "under-determined", i.e., M is not injective
2. M is "over-determined", i.e., M is not surjective

In the under-determined case, we lose **uniqueness**.
In the over-determined case, we lose **existence**.

[Li, Oprea, Wang, Y., 2025]

M is non-invertible

We have two cases for a general nonlinear M:

1. M is "under-determined", i.e., M is not injective
2. M is "over-determined", i.e., M is not surjective

In the under-determined case, we lose **uniqueness**.
In the over-determined case, we lose **existence**.

Both can be "regularized" by considering an optimization framework:

1. M is under-determined:

$$\rho_\theta^* = \operatorname*{argmin}_{S=\{\rho_\theta : M_\# \rho_\theta = \rho_g\}} \mathcal{E}[\rho_\theta]$$

2. M is over-determined:

$$\rho_\theta^* = \operatorname*{argmin}_{\rho_\theta} \mathcal{D}(M_\# \rho_\theta, \rho_g)$$

[Li, Oprea, Wang, Y., 2025]

Under-determined Case (Entropy)

$$\rho_\theta^* = \underset{M_\# \rho_\theta = \rho_g}{\operatorname{argmin}} \; \mathcal{E}[\rho_\theta], \qquad \mathcal{E}[\rho_\theta] = \int \rho_\theta \log \rho_\theta \mathrm{d}\theta$$

Theorem (Sketch)

Denote the optimizer ρ_θ^* to the problem above with $\mathcal{E}[\rho_\theta] = \int \rho_\theta \log \rho_\theta \mathrm{d}\theta$. Then for any $g \in \mathrm{supp}(\rho_g)$, we have its preimage under M given by

$$\Theta_g := \{\theta : M(\theta) = g\} .$$

ρ_θ^* is constant on the set Θ_g.

The recovered ρ_θ^* is a **uniform distribution** conditioned on each level set!

[Li, Oprea, Wang, Y., 2025]

Under-determined Case (p-th Moment)

$$\rho_\theta^* = \underset{M_\# \rho_\theta = \rho_g}{\operatorname{argmin}} \ \mathcal{E}[\rho_\theta] \,, \qquad \mathcal{E}[\rho_\theta] = \int |\theta|^p \rho_\theta \mathrm{d}\theta \,.$$

Theorem (Sketch)

Denote the optimizer ρ_θ^* to the problem above with $\mathcal{E}[\rho_\theta] = \int |\theta|^p \rho_\theta \mathrm{d}\theta$. For any $g \in \mathrm{supp}(\rho_g)$, define $\mathcal{H}$ such that

$$\mathcal{H}(g) := \underset{M(\theta)=g}{\operatorname{argmin}} |\theta|^p \,. \tag{5}$$

Then

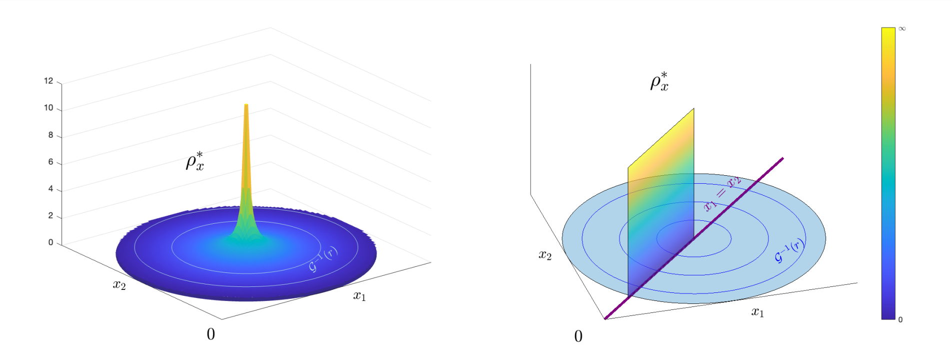
$$\rho_\theta^* = \mathcal{H}_\# \rho_g \,.$$

The recovered ρ_θ^* is supported only on a (least p-norm) point at each level set!

[Li, Oprea, Wang, Y., 2025]

Under-determined Case: Illustrations

$$r = M(x_1, x_2) = \sqrt{(x_1 - 1)^2 + (x_2 - 1)^2} \quad \mu_r = \mathcal{U}([0, 1])$$



(a) $\mathcal{E} = \int \rho_x \log \rho_x \mathrm{d}x$ **(b)** $\mathcal{E} = \int |x|^2 \rho_x \mathrm{d}x$

[Li, Oprea, Wang, Y., 2025]

Over-determined Case (f-divergence)

$$\rho_\theta^* = \operatorname*{argmin}_{\rho_\theta} \mathcal{D}(M_\# \rho_\theta, \rho_g) \,, \quad \mathcal{D}(\mu, \nu) = \int f\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\nu$$

Theorem (Sketch)

Denote the optimizer ρ_θ^* to the problem with $\mathcal{D}$ being the f-divergence. Let $\mathcal{R}$ be the range of M. Then we have

$$M_\# \rho_\theta^* = \textit{conditional distribution of } \rho_g \textit{ on } \mathcal{R}.$$

[Li, Oprea, Wang, Y., 2025]

Over-determined Case (Wasserstein distance)

$$\rho_\theta^* = \operatorname*{argmin}_{\rho_\theta} \mathcal{D}(M_\# \rho_\theta, \rho_g)\,, \quad \mathcal{D}(M_\# \rho_\theta, \rho_g) = W_d(M_\# \rho_\theta, \rho_g)\,.$$

Theorem (Sketch)
*Denote the optimizer ρ_θ^* to the problem with $\mathcal{D}$ being the Wasserstein metric of cost function d. Define the **projection operator** $\mathcal{P}_M : \mathbb{R}^n \to \mathcal{R}$ as*

$$\mathcal{P}_M(g) = \operatorname*{argmin}_{y \in \mathcal{R}} d(y, g)\,.$$

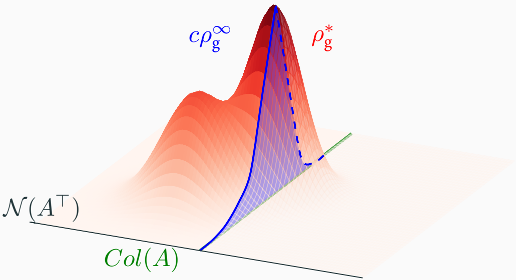
*Then we have the reconstructed **data** distribution*

$$M_\# \rho_\theta^* = \mathcal{P}_{M\,\#} \rho_g$$

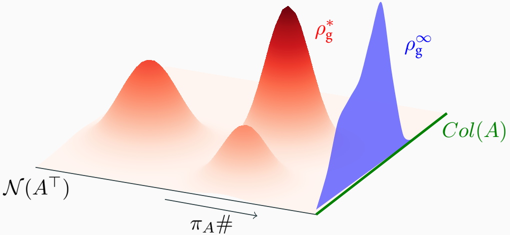
Extract the "marginal" distribution of ρ_g along the projection direction.

[Li, Oprea, Wang, Y., 2025]

Over-determined Case: Illustrations (linear)



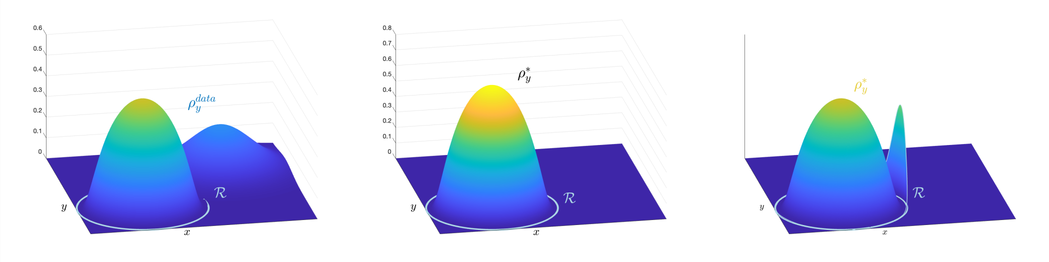
(c) f-divergence optimizer



(d) W_p optimizer

[Li, Oprea, Wang, Y., 2025]

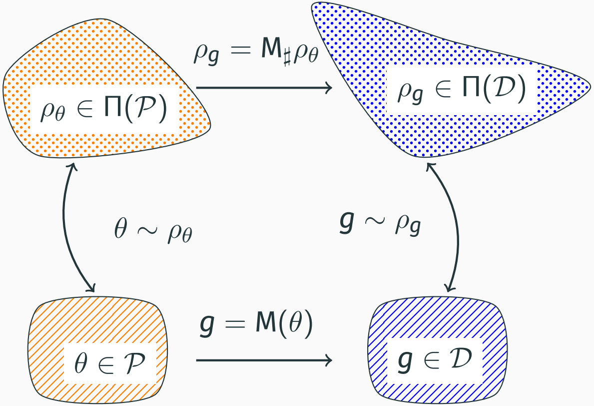
Over-determined Case: Illustrations (nonlinear)



(e) Data distribution ρ_y **(f)** f-divergence optimizer **(g)** W_p optimizer

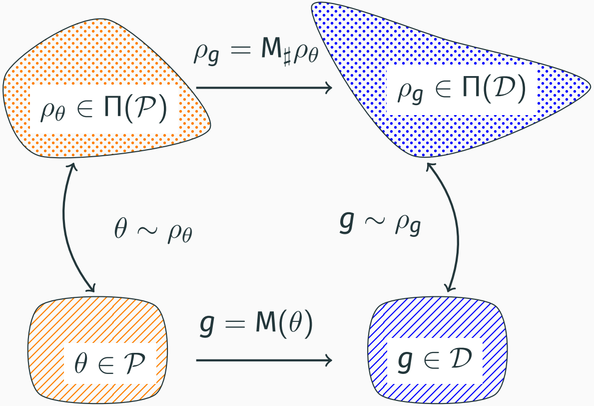
[Li, Oprea, Wang, Y., 2025]

Conclusions



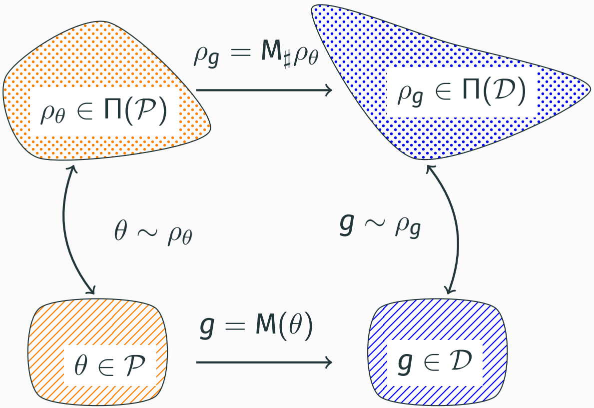
- A different stochastic framework with respect to Bayesian Inversion

Conclusions



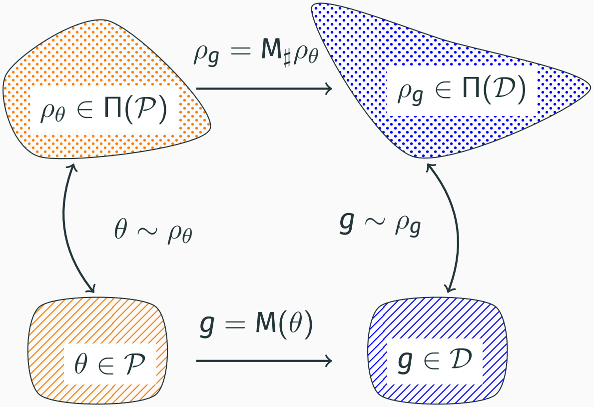
- A different stochastic framework with respect to Bayesian Inversion
- Well-posedness: metric/divergence-dependent stability

Conclusions



- A different stochastic framework with respect to Bayesian Inversion
- Well-posedness: metric/divergence-dependent stability
- Implicit Regularization: depending on both D (energy) and $\mathfrak{G}$ (dissipation)

Conclusions



- A different stochastic framework with respect to Bayesian Inversion
- Well-posedness: metric/divergence-dependent stability
- Implicit Regularization: depending on both D (energy) and $\mathfrak{G}$ (dissipation)
- Rich geometry in probability space yields various (ensemble) particle methods

Future Work

Inverse Problem Analysis

- Existence
- Uniqueness
- Stability

Inverse Problem Computation

- Objective/loss functional (convexity and stability)
- Regularization (implicit & explicit)
- Optimization algorithm (gradient flow & others)
- UQ for probability-valued soln.

Thanks for your attention!

Comparisons with Bayesian Framework

	Bayesian Framework	Stochastic Inverse Problem
source of noise	prior & measurement	parameter
consistency	Dirac delta	parameter distribution
prior information	Yes	No
measure-theoretic	Yes	Yes
require sampling	Yes	Yes
solution is a distribution	Yes	Yes

One can regard the new setup as a "deterministic inverse problem" over the $\Pi(\mathcal{P})$ (all prob. measures over $\mathcal{P}$) rather than the classic setup over $\mathcal{P}$.