Inverse Problem Over Probability Measure Space

Yunan Yang

June 12, 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).

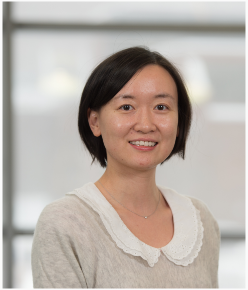
- Qin Li, Li Wang, and Y., 2024. Differential Equation–Constrained Optimization with Stochasticity. *SIAM/ASA Journal on Uncertainty Quantification, 12(2), pp.549-578.*
- Li, Q., Oprea, M., Wang, L. and Y., 2025. Inverse problems over probability measure space. *arXiv preprint arXiv:2504.18999.*

IMSI Workshop “Statistical and Computational Challenges in Probabilistic Scientific Machine Learning”

Collaborators

Qin Li
(UW Madison)

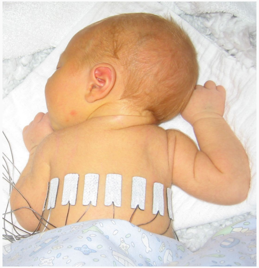
Li Wang
(UMN Twin Cities)

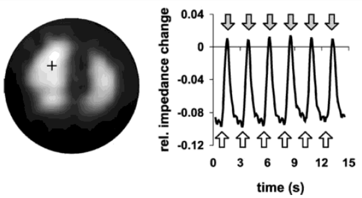


Maria Oprea
(Cornell)

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

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

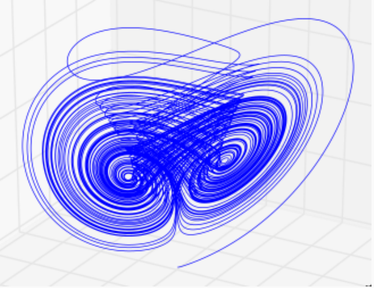
the goal is to find

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

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.

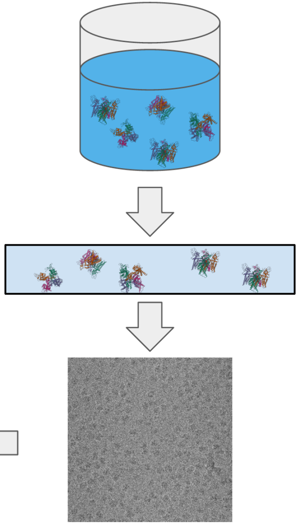
Examples

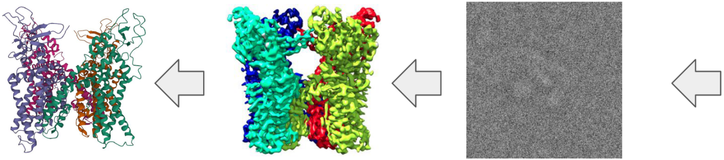
- In image processing, θ is the clean image and g is the noisy/blurred image.
- Calderón's Problem: $\begin{cases} \nabla \cdot (\theta \nabla u) = 0 & \text{on } \Omega \\ u = \phi & \text{on } \partial\Omega \end{cases}$, g is the DtN map.
- In dynamical system modeling, θ parameterizes the drift/diffusion,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

New Types of Inverse Problem: Sand Percentage in River

Cryo-Electron Microscopy

1. Snap-freeze solution of a biomolecule into a thin layer of vitreous ice
2. Image with transmission electron microscope
3. Extract images of individual biomolecules
4. Back out electron density
5. Fit atomistic structure





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.

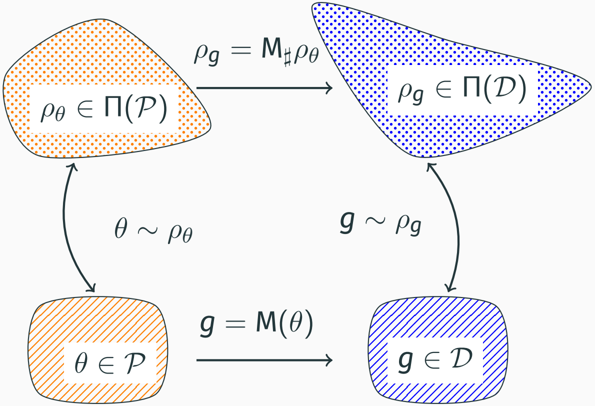
For forward problem is a push-forward map and ρ_θ is the unknown:

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

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 problem (1) and the stochastic problem (2).

Computational Aspects

Stochastic Inverse Problem — Solvers

- **Deterministic** Inverse problem:

$$M(\theta) = g$$

- Optimization problem:

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

- 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 $J(\rho_\theta) := D(M_\sharp \rho_\theta, \rho_g^*)$ under the metric $\mathfrak{G}$ is

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

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 $D = \mathrm{KL}$:

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

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

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

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 $D = \mathrm{KL}$:

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

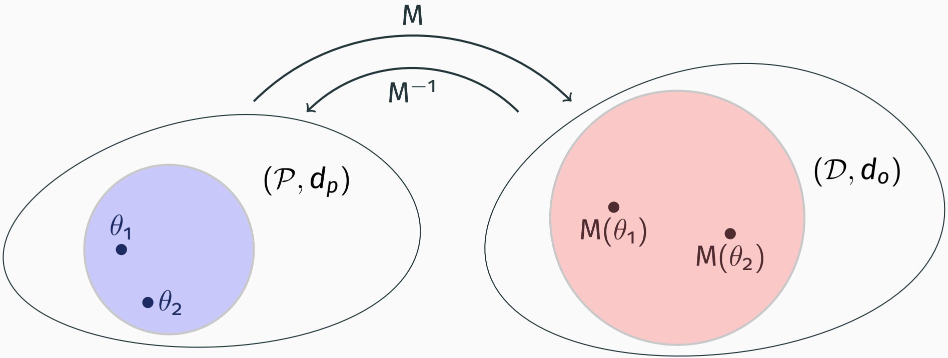
Example 2: Consider $\mathfrak{G} = H^2$ (Hellinger) and $D = \chi^2$:

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

Many **sampling algorithms** and **particle methods** can be derived from these gradient flow PDEs.

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. (Deterministic inverse problem is well-posed.)

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

M **is invertible**

Suppose M^{-1} exists and is Hölder. (Deterministic inverse problem is well-posed.)

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

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. (Deterministic inverse problem is well-posed.)

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

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

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

Outlook: use various probability metrics to balance stability and accuracy.

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 — loss of "**uniqueness**"
2. M is "over-determined", i.e., M is not surjective — loss of "**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 — loss of "**uniqueness**"
2. M is "over-determined", i.e., M is not surjective — loss of "**existence**"

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

1. M is under-determined:

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

2. M is over-determined:

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

[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. Then for any $g \in$ supp(ρ_g), we denote its preimage under M by $\Theta_g := \{\theta : M(\theta) = g\}$. Then

$$\rho_\theta^*(\cdot|\Theta_g) \quad \textit{is a uniform distribution} .$$

That is, 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. For any $g \in supp(\rho_g)$, define $\mathcal{H}$ such that

$$\mathcal{H}(g) := \underset{M(\theta)=g}{\operatorname{argmin}} \ |\theta|^p \quad (\mathcal{H} \text{ is minimum-norm soln. operator}) \tag{5}$$

Then

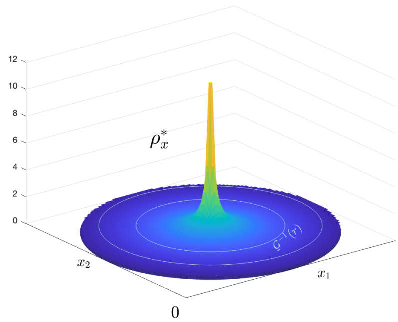
$$\rho_\theta^* = \mathcal{H}_\# \rho_g\,.$$

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

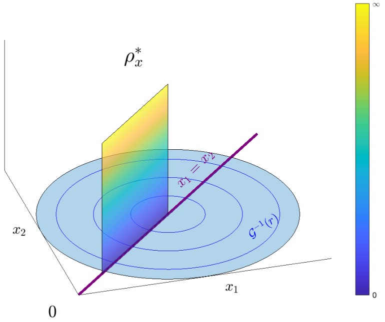
[Li, Oprea, Wang, Y., 2025]

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) $\mathcal{E} = \int \rho_\theta \log \rho_\theta \mathrm{d}\theta$



(b) $\mathcal{E} = \int |x|^2 \rho_\theta \mathrm{d}\theta$

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

Over-determined Case (f-divergence)

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

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

Over-determined Case (f-divergence)

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

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

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

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

Over-determined Case (Wasserstein distance)

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

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

Over-determined Case (Wasserstein distance)

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

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*

$$\mathcal{P}_M(g) = \underset{y \in \mathcal{R}}{\mathrm{argmin}}\, d(y, g)\,.$$

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

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

[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*

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

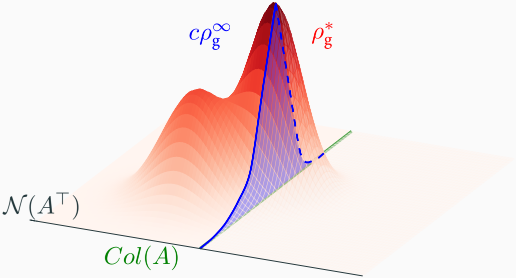
Then we have the reconstructed ***data*** *distribution*

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

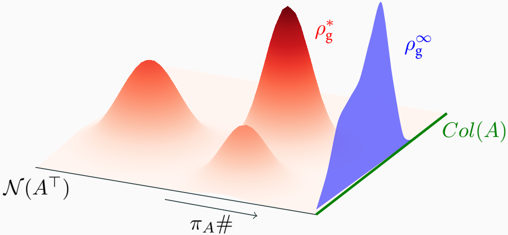
Extract the "marginal" distribution of ρ_g along the projection direction.

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

Over-determined Case: Illustrations (linear)



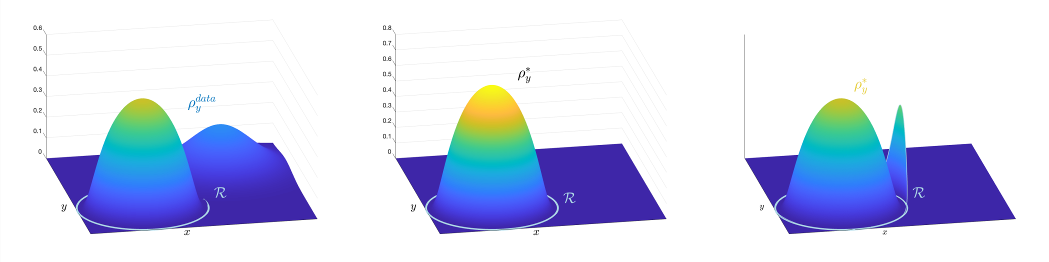
(c) f-divergence optimizer



(d) W_p optimizer

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

Over-determined Case: Illustrations (nonlinear)



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

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

Regularization

Entropy-Entropy Pair

Assume that the data ρ_y^δ contains noise. We would like to add a regularization.

$$\rho_x^* = \operatorname*{argmin}_{\rho_x \in \mathcal{P}_{2,ac}} \mathcal{L}(\rho_x)\,, \qquad \mathcal{L}(\rho_x) := \mathrm{KL}(M_\# \rho_x || \rho_y^\delta) + \alpha\, \mathrm{KL}(\rho_x || \rho_0)\,. \tag{6}$$

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

Entropy-Entropy Pair

Assume that the data ρ_y^δ contains noise. We would like to add a regularization.

$$\rho_x^* = \underset{\rho_x \in \mathcal{P}_{2,\mathrm{ac}}}{\operatorname{argmin}} \; \mathcal{L}(\rho_x)\,, \qquad \mathcal{L}(\rho_x) := \mathrm{KL}(M_\# \rho_x || \rho_y^\delta) + \alpha\, \mathrm{KL}(\rho_x || \rho_0)\,. \tag{6}$$

Theorem

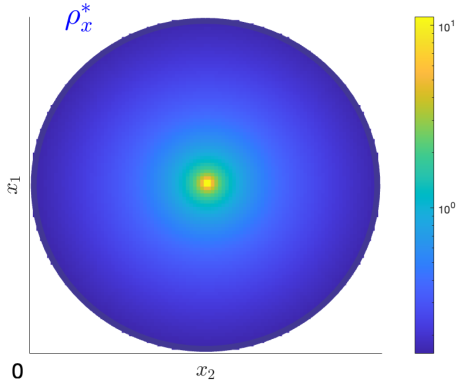
The optimizer ρ_x^* of (6) has the following property:

$$\frac{\rho_x^*(\,\cdot\,|M^{-1}(y))}{\rho_0(\,\cdot\,|M^{-1}(y))} = g(y)\,, \quad \forall y \in \mathcal{R}\,,$$

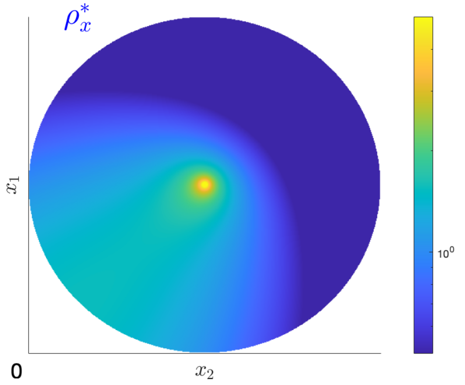
where $g(y)$ is a constant only depending on y. In particular:

$$\rho_x^*(x) \propto \rho_0(x) \left[\frac{\rho_y^\delta(M(x))}{\rho_0^y(M(x))} \right]^{\frac{1}{1+\alpha}}\,, \quad \rho_0^y = M_\# \rho_0\,.$$

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



(h) $\alpha = 0$



(i) $\alpha = 1$

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

W_p-W_p Pair

Consider a different data-matching loss and regularization term pair:

$$\rho_x^* = \operatorname*{argmin}_{\rho_x \in \mathcal{P}_p} \mathcal{L}(\rho_x), \qquad \mathcal{L}(\rho_x) = \mathcal{W}_p^p(M_\# \rho_x, \rho_y^\delta) + \alpha \underbrace{\int |x|^p \mathrm{d}\rho_x(x)}_{=\mathcal{W}_p^p(\rho_x, \delta_0)}. \tag{7}$$

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

Consider a different data-matching loss and regularization term pair:

$$\rho_x^* = \underset{\rho_x \in \mathcal{P}_p}{\operatorname{argmin}} \; \mathcal{L}(\rho_x)\,, \qquad \mathcal{L}(\rho_x) = \mathcal{W}_p^p(M_\# \rho_x, \rho_y^\delta) + \alpha \underbrace{\int |x|^p \mathrm{d}\rho_x(x)}_{=\mathcal{W}_p^p(\rho_x, \delta_0)}\,. \tag{7}$$

Theorem

Assume $\mathcal{R}$ and Θ are compact. The minimizer ρ_x^* to problem (7) satisfies

$$\rho_x^* = \tilde{\mathcal{F}}_\# \rho_y^\delta\,.$$

where $\tilde{\mathcal{F}}(y) = \underset{x \in \Theta}{\operatorname{argmin}} \{|M(x) - y|^p + \alpha |x|^p\}$.

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

W_p-W_p Pair

Consider a different data-matching loss and regularization term pair:

$$\rho_x^* = \underset{\rho_x \in \mathcal{P}_p}{\operatorname{argmin}} \ \mathcal{L}(\rho_x)\,, \qquad \mathcal{L}(\rho_x) = \mathcal{W}_p^p(M_\# \rho_x, \rho_y^\delta) + \alpha \underbrace{\int |x|^p \mathrm{d}\rho_x(x)}_{=\mathcal{W}_p^p(\rho_x, \delta_0)}\,. \tag{7}$$

Theorem

Assume $\mathcal{R}$ and Θ are compact. The minimizer ρ_x^* to problem (7) satisfies

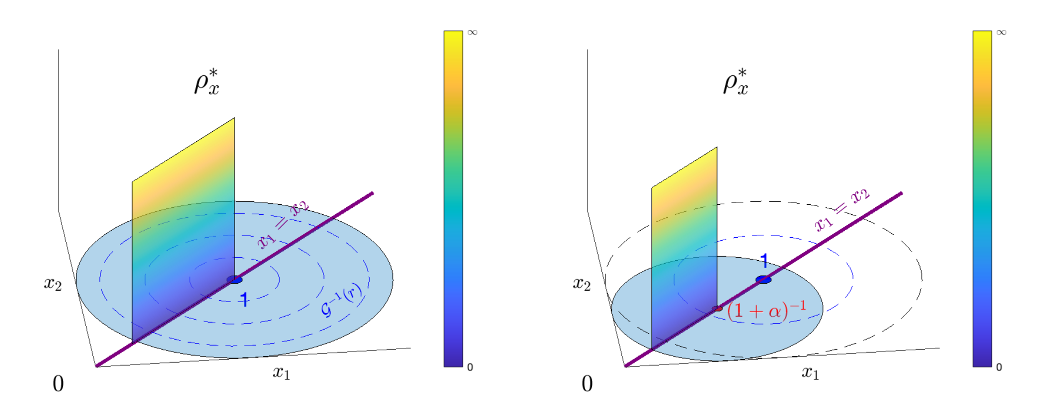
$$\rho_x^* = \tilde{\mathcal{F}}_\# \rho_y^\delta\,.$$

where $\tilde{\mathcal{F}}(y) = \underset{x \in \Theta}{\operatorname{argmin}} \{|M(x) - y|^p + \alpha |x|^p\}$.

Main proof idea: $\mathcal{L}(\rho_x)$ can be written as a *single* $\mathcal{W}_p$ metric matching problem.

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

Numerical Illustration



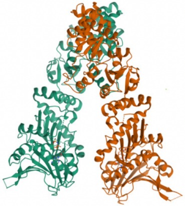
(j) $\alpha = 0$ **(k)** $\alpha \neq 0$

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

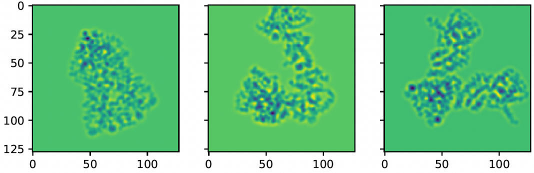
Large-Scale Example — Cryo-EM

Cryo-EM as a Stochastic Inverse Problem

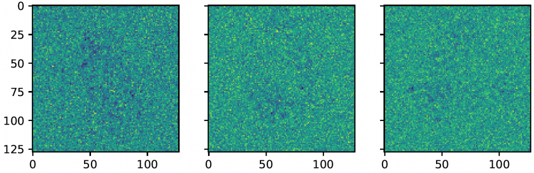
Ongoing work with Erik Thiede (Cornell Chemistry) and Diego Sanchez Espinosa (Cornell CAM)



(a) Protein HSP90 structure.

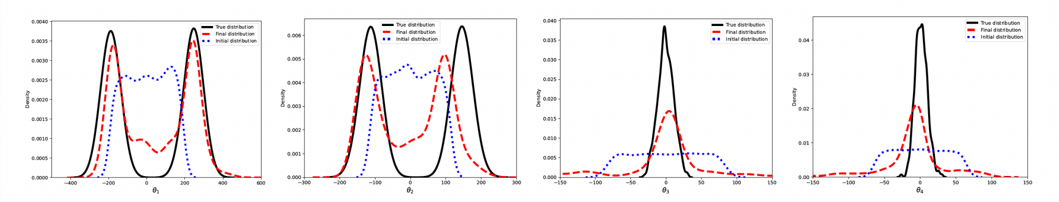


(b) Simulated cryo-EM images without noise.



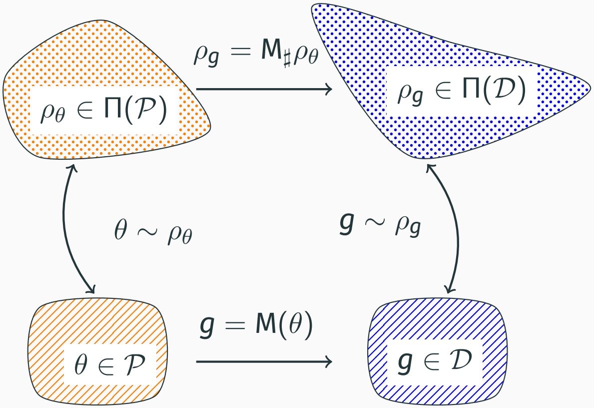
(c) Simulated cryo-EM images with added noise.

Cryo-EM as a Stochastic Inverse Problem



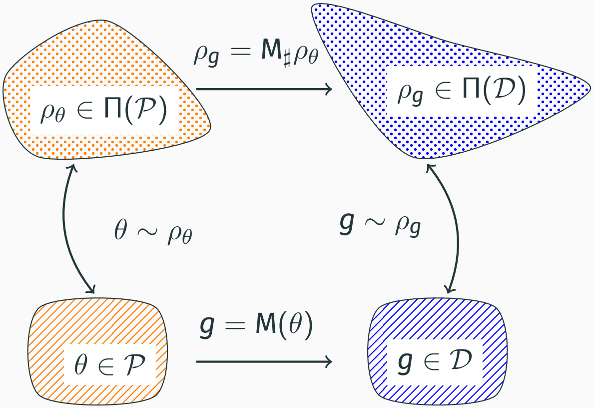
Comparison between the true parameter distribution (black), the estimated distribution (red) and the the initial guess (blue). From left to right are Mode 1, 2, 3 and 4.

Conclusions



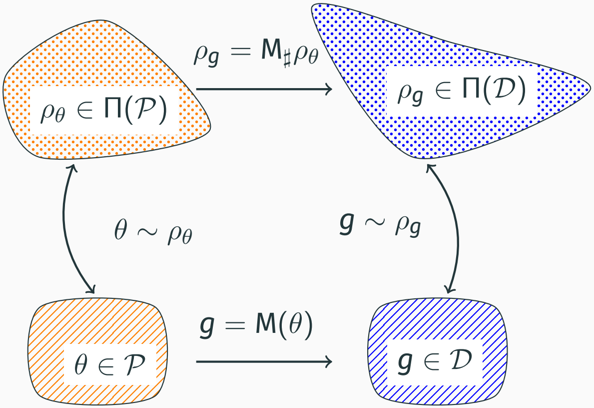
- A different stochastic framework with respect to Bayesian Inversion

Conclusions



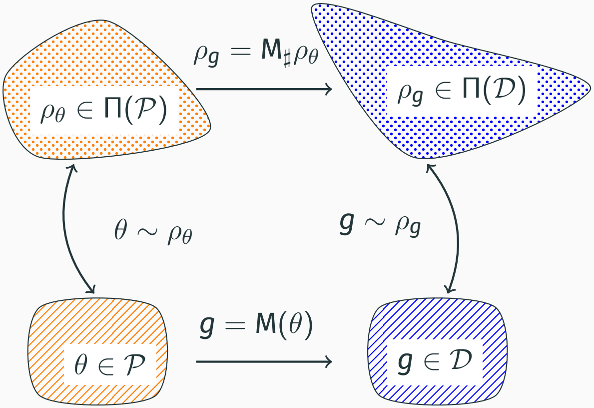
- A different stochastic framework with respect to Bayesian Inversion
- Well-posedness: metric/divergence-dependent stability

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 $\mathcal{P}$.