Measure-Theoretic Approaches for Stochastic Inverse Problems

Yunan Yang

May 9, 2024

Department of Mathematics, Cornell University

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

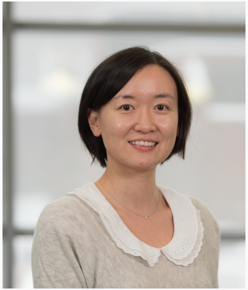
- *Qin Li, Li Wang, and Y., Differential-equation constrained optimization with stochasticity. To appear in SIAM/ASA JUQ* `https://arxiv.org/pdf/2305.04024.pdf`
- An on-going work

ICERM Workshop: Interacting Particle Systems: Analysis, Control, Learning and Computation May, 2024

Collaborators

Qin Li
(UW Madison)

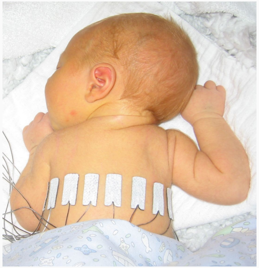
Li Wang
(UMN Twin Cities)

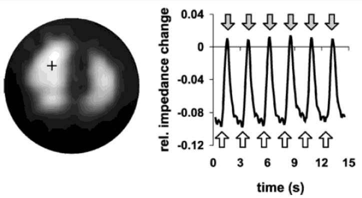


Maria Oprea
(Cornell)

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

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

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.

f_ϵ

u

Denoising, Deblurring, Blind Deconvolution (nonlinear)...

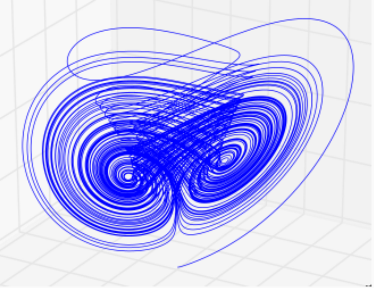
$$f_\epsilon = A(\sigma)u + \epsilon$$

where $A(\sigma)$ could be

- Identity I (denoising)
- Known Kernel K (deblurring)
- Unknown Kernel $A(\sigma)$ (blind deconvolution, nonlinear)

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

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

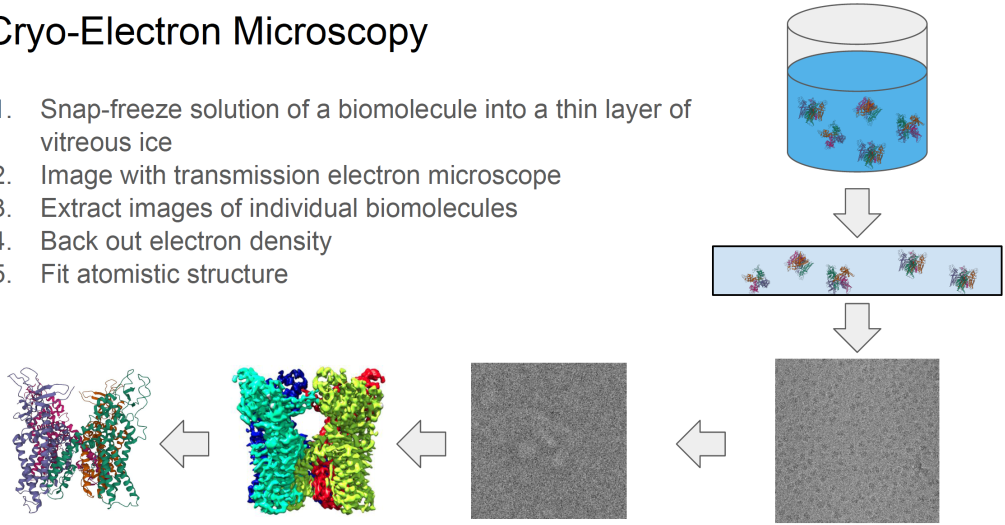
where $\theta \in \mathcal{P}$ is the function space of parameters, M is the forward operator, with $g \in \mathcal{D}$, the function space of data. M can be implicitly defined.

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 cryo-electron microscopy (cryo-EM): θ is the 3D protein structure, g is the noisy 2D projection image with an unknown random rotation.

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



Sand Percentage in River

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.

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.

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.

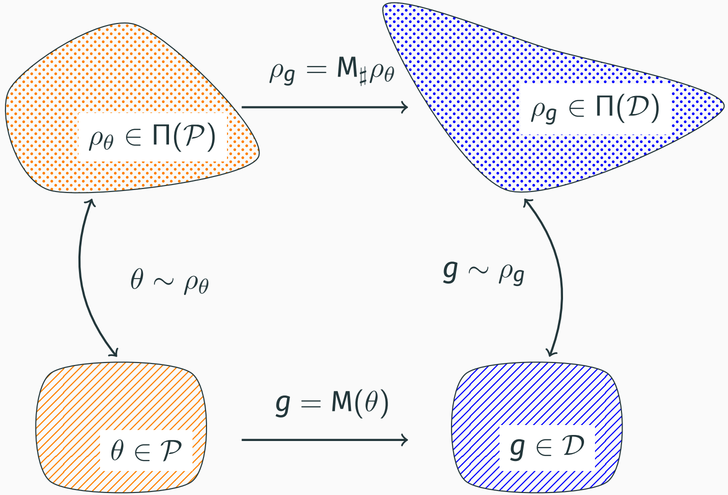
Thus, one must employ a model that incorporates **a parameter distribution**, which gives rise to the so-called Stochastic Inverse Problem.

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

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

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

Deterministic Inverse Problem to Stochastic Inverse Problem



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

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

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 $\mathcal{P}$.

Some Metrics & Divergences

Probability metric and divergence

Definition of the Wasserstein Distance

For $g_1, g_2 \in \Pi(\mathcal{P})$ ($g_1, g_2 \geq 0$ and $\int g_1 = \int g_2 = 1$), the Wasserstein distance is

$$W_p(g_1, g_2) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p \, g_1(x) dx \right)^{\frac{1}{p}} \tag{3}$$

$\mathcal{M}$: the set of all maps that rearrange the distribution g_1 into g_2.

The problem of optimal transportation was first raised by Monge in 1781.

Probability metric and divergence

Definition of the Wasserstein Distance

For $g_1, g_2 \in \Pi(\mathcal{P})$ ($g_1, g_2 \geq 0$ and $\int g_1 = \int g_2 = 1$), the Wasserstein distance is

$$W_p(g_1, g_2) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p \, g_1(x) dx \right)^{\frac{1}{p}} \tag{3}$$

$\mathcal{M}$: the set of all maps that rearrange the distribution g_1 into g_2.

The problem of optimal transportation was first raised by Monge in 1781.

When $p = 2$ (the W_2 metric), we can have a Wasserstein gradient flow of any functional E

$$\partial_t \rho = -\nabla_{W_2} E(\rho) = \nabla \cdot \left(\rho \nabla \frac{\delta E}{\delta \rho} \right).$$

Definition of the Hellinger Distance

Consider two probability measures ν_1 and ν_2 both defined on a measure space $\mathcal{P}$ that are absolutely continuous with respect to an auxiliary measure μ, i.e.,

$$\nu_1(\mathrm{d}x) = g_1(x)\mu(\mathrm{d}x), \quad \nu_2(\mathrm{d}x) = g_2(x)\mu(\mathrm{d}x)\,.$$

The Hellinger distance between ν_1 and ν_2 is

$$H(\nu_1, \nu_2) = \sqrt{\frac{1}{2}\int_M \left(\sqrt{g_1(x)} - \sqrt{g_2(x)}\right)^2 \mu(\mathrm{d}x)}\,.$$

Definition of the f-Divergence

Consider $\nu_1, \nu_2 \in \Pi(\mathcal{P})$ from the previous slide. Consider a convex function $f : \mathbb{R}^+ \mapsto (-\infty, +\infty]$ such that $f(x) < \infty$ for any $x > 0$, $f(1) = 0$ and $f(0)$ could be $+\infty$. The f-divergence of ν_1 from ν_2 is

$$D_f(\nu_1 || \nu_2) = D_f(g_1 || g_2) = \int f\left(\frac{g_1}{g_2}\right) g_2 \mu(\mathrm{d}x)\,. \tag{4}$$

Probability metric and divergence

Definition of the f-Divergence
Consider $\nu_1, \nu_2 \in \Pi(\mathcal{P})$ from the previous slide. Consider a convex function $f : \mathbb{R}^+ \mapsto (-\infty, +\infty]$ such that $f(x) < \infty$ for any $x > 0$, $f(1) = 0$ and $f(0)$ could be $+\infty$. The f-divergence of ν_1 from ν_2 is

$$D_f(\nu_1 || \nu_2) = D_f(g_1 || g_2) = \int f\left(\frac{g_1}{g_2}\right) g_2 \mu(\mathrm{d}x) \,. \tag{4}$$

Examples:
The case $f(x) = x \log x$ is the well-known Kullback–Leibler (KL) divergence.
The case $f(x) = \frac{1}{2}|x - 1|$ is the total variation (TV) distance.
The case $f(x) = (x - 1)^2$ is the χ^2 divergence.

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 $\mathfrak{G}$):

$$\boxed{\rho_\theta^* = \operatorname*{argmin}_{\rho_\theta \in (\Pi(\mathcal{P}), \mathfrak{G})} D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)\,.} \tag{5}$$

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

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{6}$$

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

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{6}$$

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

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

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

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{6}$$

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 $D = W_2$:

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

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

$$\boxed{\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}} J(\rho_\theta) = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)}\,. \tag{6}$$

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 $D = W_2$:

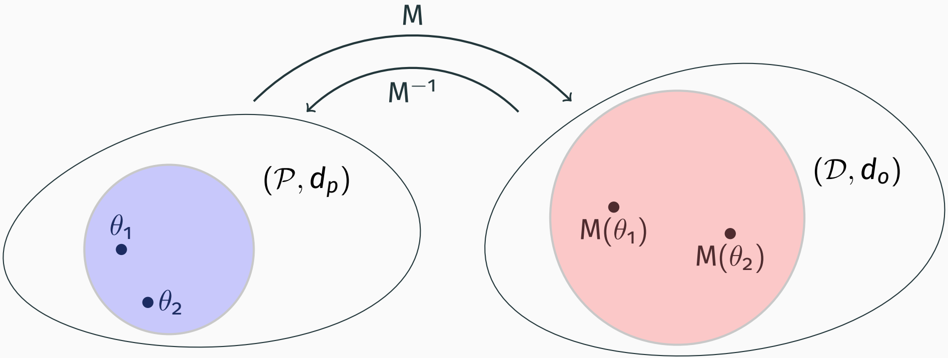
$$\partial_t \rho_\theta = \nabla_\theta \cdot (\rho_\theta \nabla_\theta\, \phi(M(\theta))) \qquad \phi \text{ is the Kantorovich potential}$$

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

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

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.

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

On the other hand, under the total variation distance of measures (TV), we have

$$TV\left(\rho_\theta,\ \widehat{\rho_\theta}\right) = TV\left(\rho_g,\ \widehat{\rho_g}\right) \Longrightarrow \textit{can be generalized to any } D_f\ .$$

M **is non-invertible**

For simplicity, consider M is linear. Then we have two cases

1. M is under-determined
2. M is over-determined

[Li, Wang, Y., 2024]

M **is non-invertible**

For simplicity, consider M is linear. Then we have two cases

1. M is under-determined
2. M is over-determined

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

[Li, Wang, Y., 2024]

M is non-invertible

For simplicity, consider M is linear. Then we have two cases

1. M is under-determined
2. M is over-determined

In the under-determined case, we lose **uniqueness**.
In the over-determined case, we may not have **existence**.

[Li, Wang, Y., 2024]

M is non-invertible

For simplicity, consider <u>M is linear</u>. Then we have two cases

1. M is under-determined
2. M is over-determined

In the under-determined case, we lose **uniqueness**.
In the over-determined case, we may not have **existence**.

Both can be *implicitly* “regularized” by considering an optimization framework!

Optimization framework: $J(\rho_\theta) := D(\mathrm{M}_\sharp \rho_\theta, \rho_g^*)$
Gradient Flow framework: $\partial_t \rho_\theta = -\mathrm{grad}_{\mathfrak{G}}\, D(\mathrm{M}_\sharp \rho_\theta, \rho_g^*)$, with initial guess $\rho_\theta(0)$.

[Li, Wang, Y., 2024]

Under-determined Case (Deterministic Case)

We first augment $\mathsf{A} \in \mathbb{R}^{n\times m}$, $n < m$, $\mathsf{A} = \mathsf{VSU}^\top$. We use $\tilde{\mathsf{A}}$ to form a rank-m matrix, and define the augmented g^{ex}:

$$\mathsf{A}^{\text{ex}} = \begin{bmatrix} \mathsf{A} \\ \tilde{\mathsf{A}} \end{bmatrix} \in \mathbb{R}^{m\times m}, \quad g^{\text{ex}} = \mathsf{A}^{\text{ex}}\theta = \begin{bmatrix} \mathsf{A}\theta \\ \tilde{\mathsf{A}}\theta \end{bmatrix} = \begin{bmatrix} g \\ \tilde{g} \end{bmatrix} \in \mathbb{R}^{m}. \tag{7}$$

Here, $\mathsf{U}^\perp$ is the orthogonal complement of U.

Under-determined Case (Deterministic Case)

We first augment $\mathrm{A} \in \mathbb{R}^{n\times m}$, $n < m$, $\mathrm{A} = \mathrm{VSU}^\top$. We use $\tilde{\mathrm{A}}$ to form a rank-m matrix, and define the augmented g^{ex}:

$$\mathrm{A}^{\mathrm{ex}} = \begin{bmatrix} \mathrm{A} \\ \tilde{\mathrm{A}} \end{bmatrix} \in \mathbb{R}^{m\times m}, \quad g^{\mathrm{ex}} = \mathrm{A}^{\mathrm{ex}}\theta = \begin{bmatrix} \mathrm{A}\theta \\ \tilde{\mathrm{A}}\theta \end{bmatrix} = \begin{bmatrix} g \\ \tilde{g} \end{bmatrix} \in \mathbb{R}^{m}. \tag{7}$$

Here, $\mathrm{U}^\perp$ is the orthogonal complement of U.

Suppose $\theta^* \in \{\theta : \mathrm{A}\theta = g^*\}$. Then the solution set can be written as

$$\mathcal{S} = \{\theta^* + \tilde{\theta} : \quad \mathrm{A}\tilde{\theta} = 0\} = \{\theta^* + \mathrm{span}\mathrm{U}^\perp\}. \tag{8}$$

The GD solution for $\min \|A\theta - g\|^2$ given the initial guess θ_0 is

$$\theta^\infty = \underbrace{\mathrm{U}\mathrm{U}^\top\theta^*}_{\in \mathrm{col}(A^\top),\text{ deteremined by } g^*} + \underbrace{\mathrm{U}^\perp(\mathrm{U}^\perp)^\top\theta_0}_{\in \mathrm{null}(A),,\text{ deteremined by } \theta_0}.$$

Under-determined Case (Stochastic Case)

Theorem (Sketch)

$J(\rho_\theta) := D(M_\sharp\rho_\theta, \rho_g^*)$ with $D =$KL or W_2. Let ρ_θ^∞ be the equilibrium solution to

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) .$$

with initial guess ρ_θ^0, and let $\rho_{g^{ex}}^\infty = A_\sharp^{ex} \rho_\theta^\infty$. Then we can uniquely determine the marginal distribution of $\rho_{g^{ex}}^\infty$:

- *The marginal distribution on g of $\rho_{g^{ex}}^\infty$ entirely recovers that of the data ρ_g^*,*
- *The marginal distribution on $\tilde{g}$ of $\rho_{g^{ex}}^\infty$ is uniquely determined by that of ρ_g^0.*

Over-determined Case (Deterministic Case)

Consider the configuration that provides the minimum misfit under the vector 2-norm. That is,

$$\min_{\theta} \frac{1}{2} \|A\theta - g^*\|_2^2 \,.$$

For a linear system like this, the minimizer is explicit:

$$\theta^* = (A^\top A)^{-1} A^\top g^* =: A^\dagger g^* \,,$$

and hence, with $A = VSU^\top$,

$$g = A\theta^* = AA^\dagger g^* = VV^\top g^* \,, \quad \text{or equivalently} \quad g = g^*_A = \mathrm{Proj}_V \, g^* \,.$$

(Column space of A is also the column space of V).

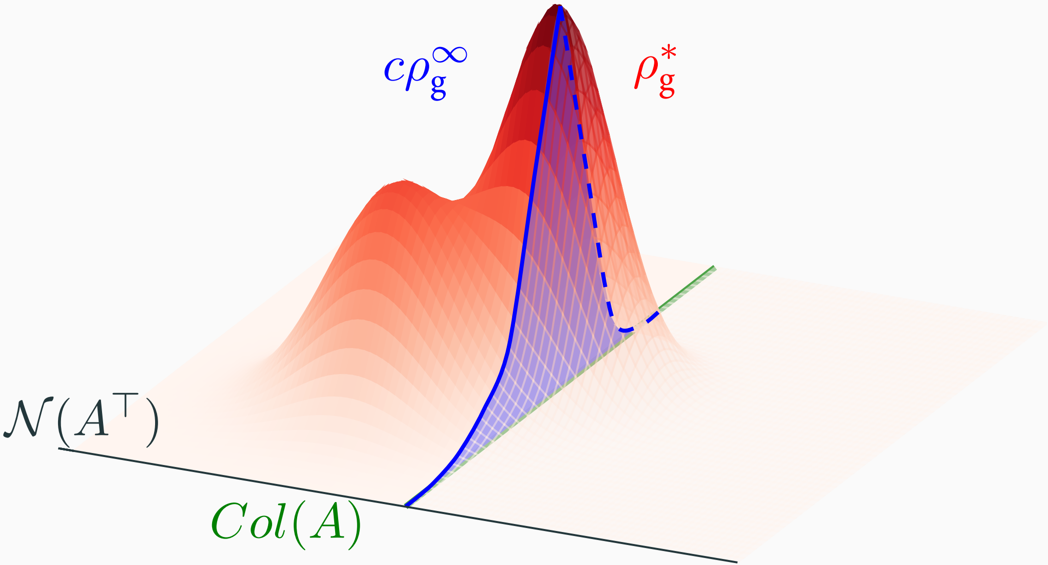
Over-determined Case: KL loss under W_2 gradient flow

Theorem (Sketch)
Let ρ_θ^∞ be the equilibrium solution to the Wasserstein gradient flow of the KL divergence between synthetic data and reference data distributions,

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) .$$

The equilibrium data distribution $\rho_g^\infty = A_\sharp \rho_\theta^\infty$ recovers ρ_g^* conditioned on $\mathrm{col}(A)$.

Over-determined Case: KL loss under W_2 gradient flow



Over-determined Case: W_2 loss under W_2 gradient flow

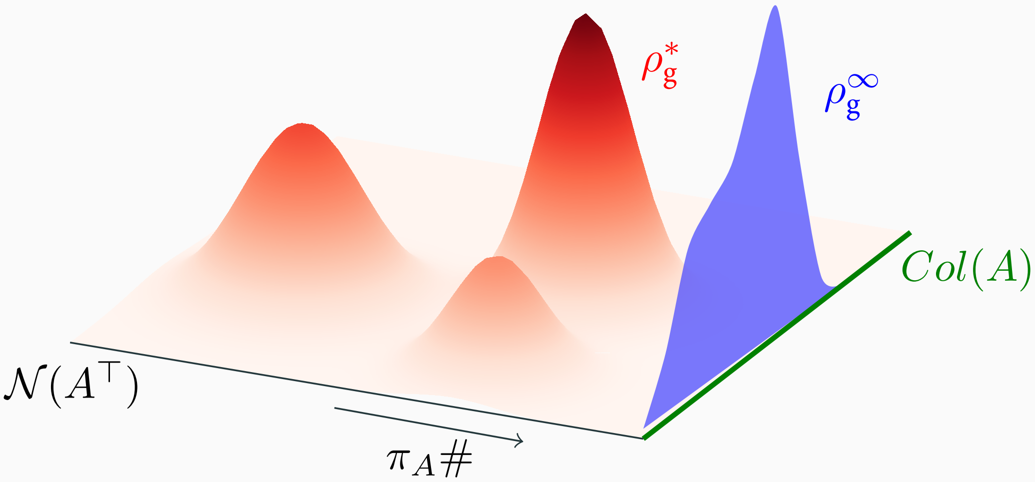
Theorem (Sketch)
Let ρ_θ^∞ be the equilibrium solution to the Wasserstein gradient flow of the squared W_2 metric between synthetic data and reference data distributions,

$$\partial_t \rho_\theta = \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) .$$

The equilibrium data distribution $\rho_g^\infty = A_\sharp \rho_\theta^\infty = A_\sharp^\dagger \rho_g^*$.

That is, ρ_g^∞ recovers the marginal distribution of ρ_g^* on $\mathrm{col}(A)$.

Over-determined Case: W_2 loss under W_2 gradient flow



Particle Method

Numerical Example: Particle Method

To solve the Wasserstein gradient flow equation, $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$,

$$\partial_t \rho_\theta - \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) = 0 \,,$$

we propose a particle method, $j = 1, 2, \ldots, N$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_j = -\nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta}(\mathsf{M}(\theta_j)) \right) = -\, {\color{red}\nabla_\theta \mathsf{M}^\top}\Big|_{\theta_j(t)} \nabla_g \frac{\delta J}{\delta \rho_\theta}(g(t)) \,, \ \text{ where } g(t) = \mathsf{M}(\theta_j(t)) \,,$$

but there are many other deterministic/stochastic variants.

[Li, Wang, Y., 2024]

Numerical Example: Particle Method

To solve the Wasserstein gradient flow equation, $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$,

$$\partial_t \rho_\theta - \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) = 0\,,$$

we propose a particle method, $j = 1, 2, \ldots, N$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_j = -\nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta}(\mathsf{M}(\theta_j)) \right) = - \nabla_\theta \mathsf{M}^\top \Big|_{\theta_j(t)} \nabla_g \frac{\delta J}{\delta \rho_\theta}(g(t))\,, \text{ where } g(t) = \mathsf{M}(\theta_j(t))\,,$$

but there are many other deterministic/stochastic variants.

- (Interactive) The trajectory of particle θ_j is also correlated with all the other particles $\{\theta_i\}_{i \neq j}$ due to the mean-field term "density" — $\rho_g = M_\sharp \rho_\theta$, & ρ_g^*.

[Li, Wang, Y., 2024]

Numerical Example: Particle Method

To solve the Wasserstein gradient flow equation, $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$,

$$\partial_t \rho_\theta - \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) = 0\,,$$

we propose a particle method, $j = 1, 2, \dots, N$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_j = -\nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta}(\mathsf{M}(\theta_j)) \right) = - \nabla_\theta \mathsf{M}^\top \Big|_{\theta_j(t)} \nabla_g \frac{\delta J}{\delta \rho_\theta}(g(t))\,,\ \text{where } g(t) = \mathsf{M}(\theta_j(t))\,,$$

but there are many other deterministic/stochastic variants.

- (Interactive) The trajectory of particle θ_j is also correlated with all the other particles $\{\theta_i\}_{i\neq j}$ due to the mean-field term "density" — $\rho_g = M_\sharp \rho_\theta$, & ρ_g^*.
- We essentially designed an ensemble particle method.

[Li, Wang, Y., 2024]

Numerical Example: Particle Method

To solve the Wasserstein gradient flow equation, $J(\rho_\theta) := D(\mathsf{M}_\sharp \rho_\theta, \rho_g^*)$,

$$\partial_t \rho_\theta - \nabla_\theta \cdot \left(\rho_\theta \nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} \right) \right) = 0\,,$$

we propose a particle method, $j = 1, 2, \ldots, N$,

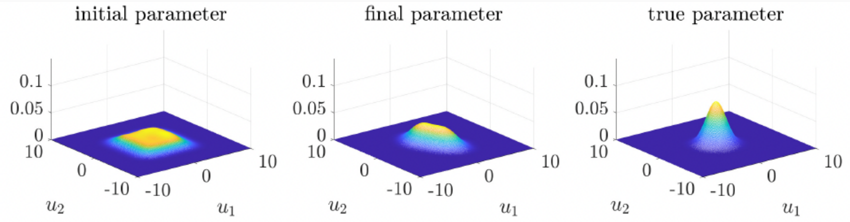
$$\frac{\mathrm{d}}{\mathrm{d}t} \theta_j = -\nabla_\theta \left(\frac{\delta J}{\delta \rho_\theta} (\mathsf{M}(\theta_j)) \right) = - \nabla_\theta \mathsf{M}^\top \Big|_{\theta_j(t)} \nabla_g \frac{\delta J}{\delta \rho_\theta} (g(t))\,, \text{ where } g(t) = \mathsf{M}(\theta_j(t))\,,$$

but there are many other deterministic/stochastic variants.

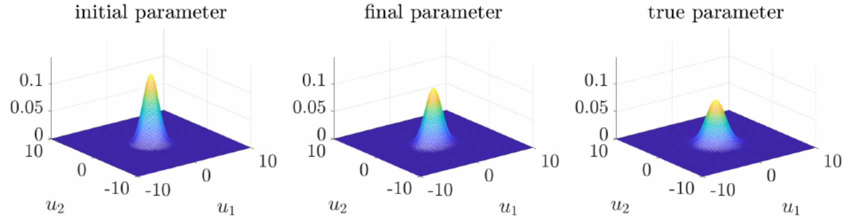
- (Interactive) The trajectory of particle θ_j is also correlated with all the other particles $\{\theta_i\}_{i \neq j}$ due to the mean-field term "density" — $\rho_g = M_\sharp \rho_\theta$, & ρ_g^*.
- We essentially designed an ensemble particle method.
- The red term can be computed using the adjoint-state method.

[Li, Wang, Y., 2024]

Example: under-determined case, W_2 gradient flow of KL

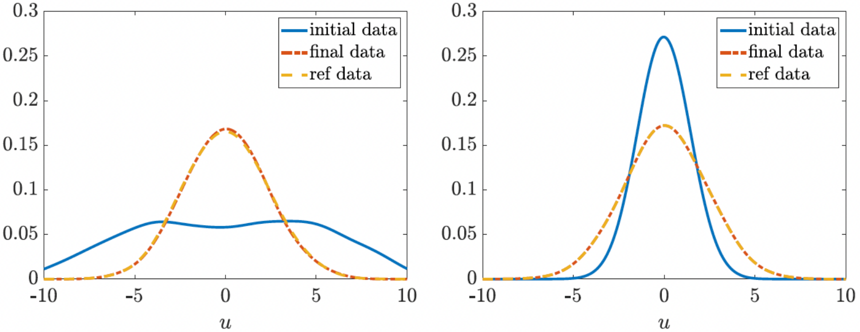


(a) Parameter distribution with initial guess u^1



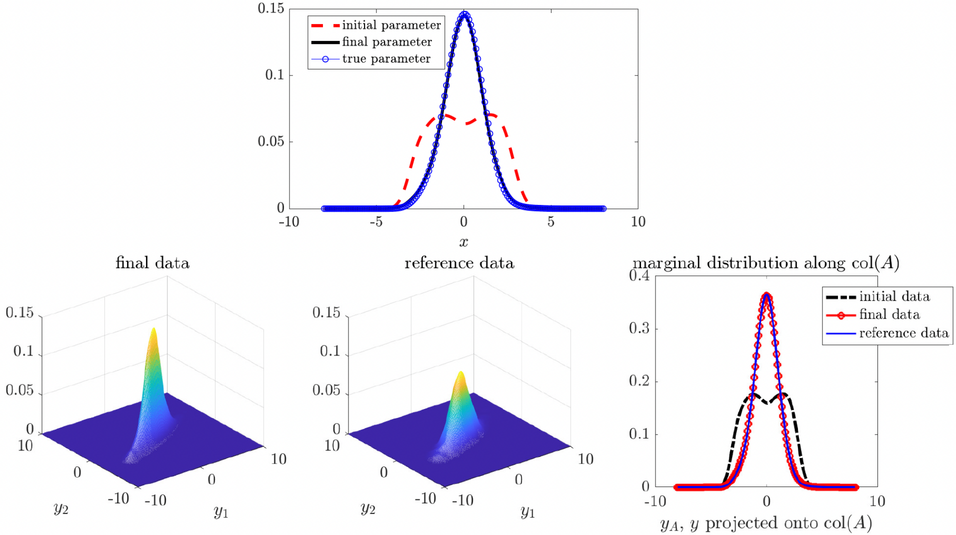
(b) Parameter distribution with initial guess u^2

Example: under-determined case, W_2 gradient flow of KL

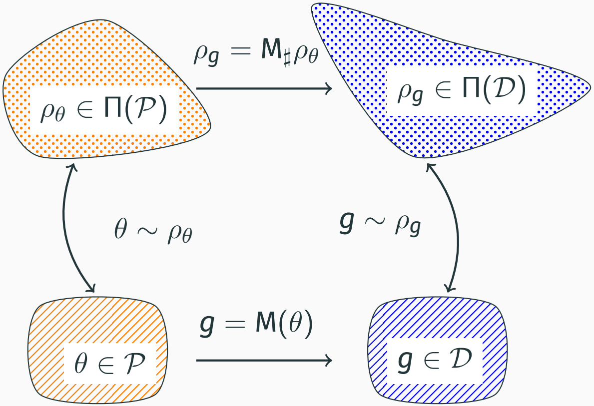


(c) Data with initial guess u^2 (d) Data with initial guess u_2

Example: over-determined case, W_2 gradient flow of KL

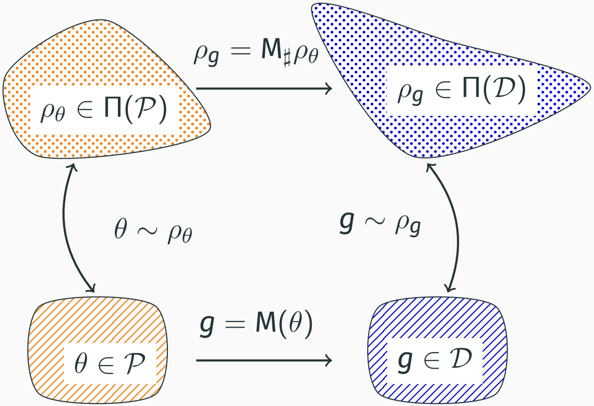


Conclusions



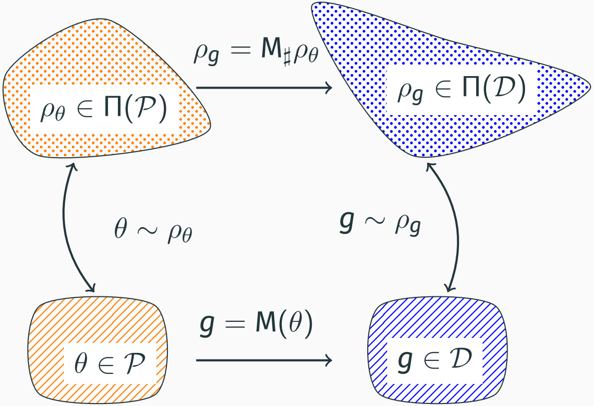
- A different stochastic framework with respect to Bayesian Inversion

Conclusions



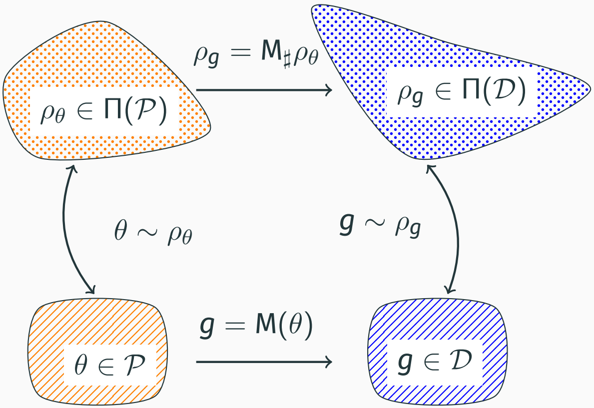
- A different stochastic framework with respect to Bayesian Inversion
- Well-posedness: metric/divergence-dependent stability

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!