

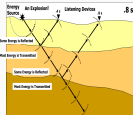
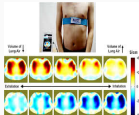
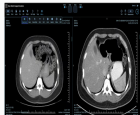
Inverse Problems: A Mathematician's Perspective

Jenn-Nan Wang

Seminar@NCU, Mar 17, 2023

National Taiwan University

Introduction



Inverse problems everyday!

- Medical imaging
- Nondestructive testing
- Radar
- Seismology exploration
- Remote sensing
- Forensics
- and much more....

General setup

Consider

$$y = G(p),$$

where

$$\begin{cases} p \in X \text{ (parameter space),} \\ y \in Y \text{ (data or observation space),} \end{cases}$$

and

$$G : X \rightarrow Y \text{ (parameter to data map).}$$

X and Y are usually infinite dimensions and G may depend on p linearly or nonlinearly. In many cases, G are given by solutions of PDEs.

General setup

Consider

$$y = G(p),$$

where

$$\begin{cases} p \in X \text{ (parameter space),} \\ y \in Y \text{ (data or observation space),} \end{cases}$$

and

$$G : X \rightarrow Y \text{ (parameter to data map).}$$

X and Y are usually infinite dimensions and G may depend on p linearly or nonlinearly. In many cases, G are given by solutions of PDEs.

Forward (or direct) problems: given p , determine y .

Inverse problems: given y , determine p .

General setup

Consider

$$y = G(p),$$

where

$$\begin{cases} p \in X \text{ (parameter space),} \\ y \in Y \text{ (data or observation space),} \end{cases}$$

and

$$G : X \rightarrow Y \text{ (parameter to data map).}$$

X and Y are usually infinite dimensions and G may depend on p linearly or nonlinearly. In many cases, G are given by solutions of PDEs.

Forward (or direct) problems: given p , determine y .

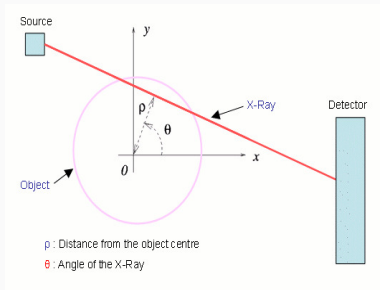
Inverse problems: given y , determine p .

It is quite often that finding G^{-1} is impossible.

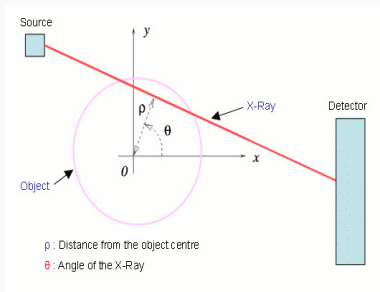


The Nobel Prize in Physiology or Medicine 1979 was awarded jointly to Allan M. Cormack and Godfrey N. Hounsfield "for the development of computer assisted tomography"

X-ray transform



X-ray transform



ℓ is the red straight line.

$$I_{\text{out}} = I_{\text{in}} \exp\left(-\int_{\ell} f(x) ds\right)$$

$I_{\text{out}}, I_{\text{in}}$ are the numbers of photons at the source and the receiver

$f(x)$ is the optical density of the medium

X-ray transform

So we can compute the line integral of $f(x)$ along any X-ray by measuring the numbers of photons at the source and the receiver. The question now is to determine $f(x)$ by all its line integrals. In other words, one would like to invert the map:

$$\mathcal{R} : f(x) \rightarrow Rf(\ell) = \int_{\ell} f(x) ds,$$

where Rf is known to be the **Radon transform** of f in \mathbb{R}^2 .

Reconstruction formula

We can write

$$Rf(\rho, \theta) = \int_{-\infty}^{\infty} f(\rho \cos \theta - s \sin \theta, \rho \sin \theta + s \cos \theta) ds.$$

Let point $P = (x, y)$, define

$$\bar{F}_P(q) = \frac{1}{2\pi} \int_0^{2\pi} Rf(x \cos \theta + y \sin \theta + q, \theta) d\theta,$$

then

$$f(P) = -\frac{1}{\pi} \int_0^{\infty} \frac{d\bar{F}_P(q)}{q}.$$

Reconstruction formula

We can write

$$Rf(\rho, \theta) = \int_{-\infty}^{\infty} f(\rho \cos \theta - s \sin \theta, \rho \sin \theta + s \cos \theta) ds.$$

Let point $P = (x, y)$, define

$$\bar{F}_P(q) = \frac{1}{2\pi} \int_0^{2\pi} Rf(x \cos \theta + y \sin \theta + q, \theta) d\theta,$$

then

$$f(P) = -\frac{1}{\pi} \int_0^{\infty} \frac{d\bar{F}_P(q)}{q}.$$

This is the famous inversion formula derived by Radon in 1917.



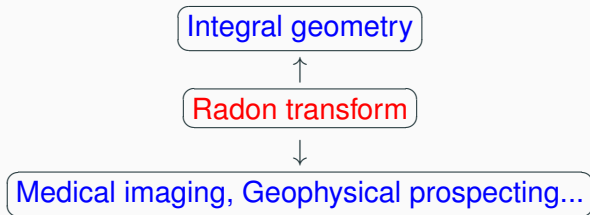
Johann Radon (1887-1956)

Ignored result

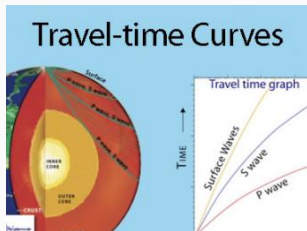


Johann Radon (1887-1956)

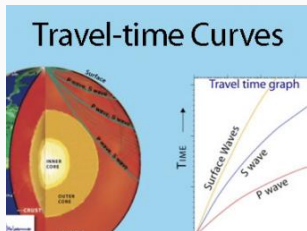
Unaware of Radon's result, Cormack reconsidered this reconstruction problem and published two papers in 1963, 1964. It was not until 1970 that Cormack learned that the problem has been solved by Radon.



Travel time tomography

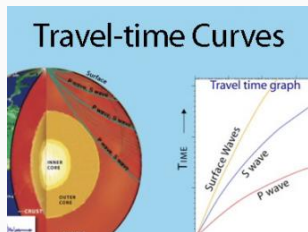


Travel time tomography



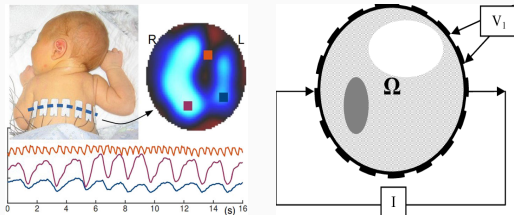
To determine the structure of the core by measuring the travel time. This problem can be recasted into a geometry problem. One of them is the problem of determining the Riemannian metric from the lengths of geodesics joining points in the boundary.

Travel time tomography

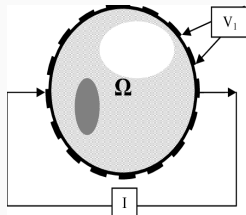
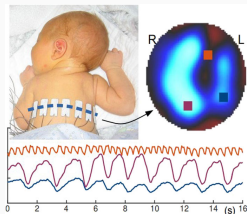


To determine the structure of the core by measuring the travel time. This problem can be recasted into a geometry problem. One of them is the problem of determining the Riemannian metric from the lengths of geodesics joining points in the boundary. However, in general, this problem does not have a unique solution.

EIT or ERT

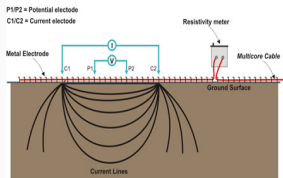


Electrical impedance tomography



Electrical impedance tomography

General resistivity principle



Typical field set-up



Electrical resistivity tomography

Mathematical setup

The boundary value problem:

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

or

$$\begin{cases} \nabla \cdot (\gamma \nabla u) = 0 & \text{in } \Omega, \\ \gamma \frac{\partial u}{\partial \nu} = g & \text{on } \partial\Omega, \quad \int_{\partial\Omega} g d\sigma = 0, \end{cases}$$

where $\gamma > 0$ is the conductivity function (the inverse of the resistivity), f is the voltage potential on the boundary, and g is the current on the boundary. Here ν is the unit outer normal of $\partial\Omega$.

Inverse problem

The problem is to determine γ from the collection of $\{f, \gamma \frac{\partial u}{\partial \nu}|_{\partial\Omega}\}$ (voltage-current pairs) or $\{g, u|_{\partial\Omega}\}$ (current-voltage pairs). In the mathematical term: determine γ from

$$\Lambda_\gamma : \text{voltage} \rightarrow \text{current}, \Lambda_\gamma f = \gamma \frac{\partial u}{\partial \nu}|_{\partial\Omega} \text{ (Dirichlet-to-Neumann map)}$$

or

$$\mathcal{N}_\gamma : \text{current} \rightarrow \text{voltage}, \mathcal{N}_\gamma g = u|_{\partial\Omega} \text{ (Neumann-to-Dirichlet map)}.$$

Inverse problem

The problem is to determine γ from the collection of $\{f, \gamma \frac{\partial u}{\partial \nu}|_{\partial\Omega}\}$ (voltage-current pairs) or $\{g, u|_{\partial\Omega}\}$ (current-voltage pairs). In the mathematical term: determine γ from

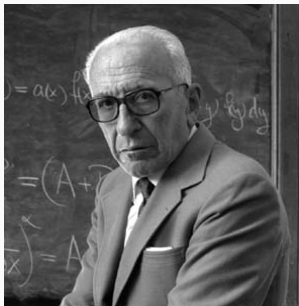
Λ_γ : voltage \rightarrow current, $\Lambda_\gamma f = \gamma \frac{\partial u}{\partial \nu}|_{\partial\Omega}$ (Dirichlet-to-Neumann map)

or

\mathcal{N}_γ : current \rightarrow voltage, $\mathcal{N}_\gamma g = u|_{\partial\Omega}$ (Neumann-to-Dirichlet map).

Basic mathematical questions:

- Uniqueness: $\Lambda_{\gamma_1} = \Lambda_{\gamma_2} \Rightarrow \gamma_1 = \gamma_2$
- Reconstruction: reconstruction formula of determining γ from Λ_γ
- Stability: how error of Λ_γ affects γ



Alberto Calderón (1920-1988)

He studied civil engineering at the University of Buenos Aires and graduated in 1947. He worked at a geophysics research lab at YPF (the state-owned petroleum company) for a few years.

A.P. Calderón, ON AN INVERSE BOUNDARY VALUE PROBLEM, in Seminar on Numerical Analysis and its Applications to Continuum Physics, Rio de Janeiro, Editors W.H. Meyer and M.A. Raupp, Sociedade Brasileira de Matematica, (1980). 65-73.

A.P. Calderón, ON AN INVERSE BOUNDARY VALUE PROBLEM, in Seminar on Numerical Analysis and its Applications to Continuum Physics, Rio de Janeiro, Editors W.H. Meyer and M.A. Raupp, Sociedade Brasileira de Matematica, (1980). 65-73.



Alberto Grünbaum

Calderón's approach

Although the boundary value problem is linear in γ , the problem of determining γ from Λ_γ is *nonlinear*.

Calderón's approach

Although the boundary value problem is linear in γ , the problem of determining γ from Λ_γ is *nonlinear*.

Idea: linearization

Consider the nonlinear function $y = F(x)$. Assume that $y_0 = F(x_0)$. Then Taylor's formula:

$$y - y_0 \approx \nabla F(x_0)(x - x_0),$$

i.e.,

$$x \approx x_0 + (\nabla F(x_0))^{-1}(y - y_0)$$

provided $\nabla F(x_0)$ is invertible. Also, if x_1, x_2 are near x_0 and $\nabla F(x_0)$ is invertible, then $F(x_1) = F(x_2) \Rightarrow x_1 = x_2$.

Green's identity:

$$\int_{\partial\Omega} \Lambda_\gamma f \cdot h d\sigma = \int_{\Omega} \gamma u \cdot v dx := Q_\gamma(f, h),$$

where $v|_{\partial\Omega} = h$. You can think of Q_γ as the energy and

$$\Lambda_\gamma \Leftrightarrow Q_\gamma.$$

Then for $\gamma = 1 + \gamma'$

$$dQ(1)(\gamma') = \int_{\Omega} \gamma' \nabla u' \cdot \nabla v' dx,$$

where $\Delta u' = \Delta v' = 0$ in Ω , i.e., harmonic functions. Plugging suitable harmonic functions into this formula will give us some information of γ' , a perturbation of the identity conductivity.

Global uniqueness

The linearized map $dQ(1)$ will not provide us any uniqueness result even when $dQ(1)$ is "invertible".

Global uniqueness

The linearized map $dQ(1)$ will not provide us any uniqueness result even when $dQ(1)$ is "invertible".



John Sylvester



Gunther Uhlmann

Proved the uniqueness in 1987: $\Lambda_{\gamma_1} = \Lambda_{\gamma_2} \Rightarrow \gamma_1 = \gamma_2$. They introduced a special class of solutions called **complex geometric optics solutions** in the proof. These solutions are also useful in designing a reconstruction algorithm (in theory).

Ill-posedness

EIT problem is ill-posed in the sense of Hadamard.



Jacques Hadamard (1865-1963)

Ill-posedness

EIT problem is ill-posed in the sense of Hadamard.



Jacques Hadamard (1865-1963)

Hadamard defined that a problem is well-posed if it has existence, uniqueness, and continuous dependence (solution depends on data or on coefficients continuously).

Continuous dependence

Who cares about the existence and uniqueness?

Continuous dependence

Who cares about the existence and uniqueness?

Mathematicians!

Continuous dependence

Who cares about the existence and uniqueness?

Mathematicians!

The continuous dependence (or stability) is more serious in the applications.

Continuous dependence

Who cares about the existence and uniqueness?

Mathematicians!

The continuous dependence (or stability) is more serious in the applications.

In fact, the usual sense of continuous dependence **does not** hold. That is, let $(X, |\cdot|_X)$ be the space of γ and $(Y, |\cdot|_Y)$ be the space of Λ_γ , where $|\cdot|_X$ and $|\cdot|_Y$ are distance functions in X and Y , respectively. One **cannot** expect that for any $\gamma_1, \gamma_2 \in X$

$$\text{error}(\Lambda_{\gamma_1}, \Lambda_{\gamma_2}) \rightarrow 0 \Rightarrow \text{error}(\gamma_1, \gamma_2) \rightarrow 0,$$

where

$$\begin{cases} \text{error}(\Lambda_{\gamma_1}, \Lambda_{\gamma_2}) = |\Lambda_{\gamma_1} - \Lambda_{\gamma_2}|_Y, \\ \text{error}(\gamma_1, \gamma_2) = |\gamma_1 - \gamma_2|_X. \end{cases}$$

Stable estimate

But we can find a subspace $E \subset X$ such that if $\gamma_1, \gamma_2 \in E$ such that it holds

$$\text{error}(\Lambda_{\gamma_1}, \Lambda_{\gamma_2}) \rightarrow 0 \Rightarrow \text{error}(\gamma_1, \gamma_2) \rightarrow 0$$

(closely related to the [regularization](#)).

Stable estimate

But we can find a subspace $E \subset X$ such that if $\gamma_1, \gamma_2 \in E$ such that it holds

$$\text{error}(\Lambda_{\gamma_1}, \Lambda_{\gamma_2}) \rightarrow 0 \Rightarrow \text{error}(\gamma_1, \gamma_2) \rightarrow 0$$

(closely related to the [regularization](#)).

In fact, it was proved by Alessandrini that if $\gamma_1, \gamma_2 \in E$ for some suitable $E \subset X$:

$$\text{error}(\gamma_1, \gamma_2) \leq C\omega(\varepsilon), \quad (1)$$

where $\varepsilon = \text{error}(\Lambda_{\gamma_1}, \Lambda_{\gamma_2})$, C is some positive constant, and

$$\omega(\varepsilon) \leq |\log \varepsilon|^{-\eta}, \quad 0 < \eta < 1. \quad (\text{logarithmically stable})$$

This estimate cannot be improved in general.

Difficulty

Estimate (1) implies that even a small error in the measurement will result in high inaccuracy in the conductivity.

Difficulty

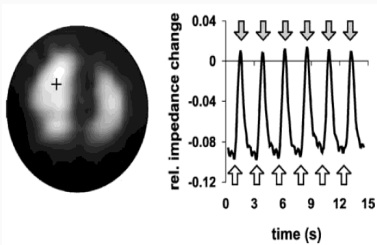
Estimate (1) implies that even a small error in the measurement will result in high inaccuracy in the conductivity.

For example, we take $\eta = 1$, and $\varepsilon = 10^{-10}$, the error in the determination of the conductivity $\text{error}(\gamma_1, \gamma_2)$ can be large as 10^{-1} .

Difficulty

Estimate (1) implies that even a small error in the measurement will result in high inaccuracy in the conductivity.

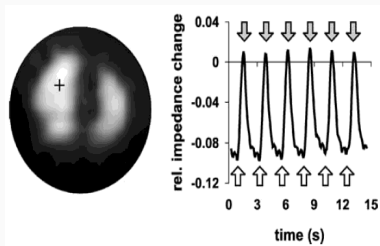
For example, we take $\eta = 1$, and $\varepsilon = 10^{-10}$, the error in the determination of the conductivity $\text{error}(\gamma_1, \gamma_2)$ can be large as 10^{-1} .



Difficulty

Estimate (1) implies that even a small error in the measurement will result in high inaccuracy in the conductivity.

For example, we take $\eta = 1$, and $\varepsilon = 10^{-10}$, the error in the determination of the conductivity $\text{error}(\gamma_1, \gamma_2)$ can be large as 10^{-1} .



How to improve the stability is probably the most challenging problem in EIT or ERT.

Piecewise constant conductivity: $\gamma = \sum_{j=1}^J c_j \mathbb{1}_{D_j}$, where $\mathbb{1}_{D_j}$ is the characteristic function of the region D_j and $\Omega = \cup_{j=1}^J D_j$. Assume that $\{D_j\}_{j=1}^J$ are known, but $\{c_j\}_{j=1}^J$ are unknown. The determination of $\{c_j\}_{j=1}^J$ by the measurement Λ_γ is almost *linearly stable*.

Piecewise constant conductivity: $\gamma = \sum_{j=1}^J c_j \mathbb{1}_{D_j}$, where $\mathbb{1}_{D_j}$ is the characteristic function of the region D_j and $\Omega = \cup_{j=1}^J D_j$. Assume that $\{D_j\}_{j=1}^J$ are known, but $\{c_j\}_{j=1}^J$ are unknown. The determination of $\{c_j\}_{j=1}^J$ by the measurement Λ_γ is almost *linearly stable*.

However, if $\{D_j\}_{j=1}^J$ are unknown, then the determination is still *logarithmically stable*!

Inclusion problem

Related to the case of piecewise conductivity, we consider the following problem: let $\gamma = \gamma_0 \mathbb{1}_{\Omega \setminus D} + \gamma_1 \mathbb{1}_D$. Assume that γ_0 is known, but γ_1 is unknown. You can regard D as the region of an abnormality, e.g., tumor.

Inclusion problem

Related to the case of piecewise conductivity, we consider the following problem: let $\gamma = \gamma_0 \mathbb{1}_{\Omega \setminus D} + \gamma_1 \mathbb{1}_D$. Assume that γ_0 is known, but γ_1 is unknown. You can regard D as the region of an abnormality, e.g., tumor.

The problem is to determine D from the measurement Λ_γ .

Inclusion problem

Related to the case of piecewise conductivity, we consider the following problem: let $\gamma = \gamma_0 \mathbb{1}_{\Omega \setminus D} + \gamma_1 \mathbb{1}_D$. Assume that γ_0 is known, but γ_1 is unknown. You can regard D as the region of an abnormality, e.g., tumor.

The problem is to determine D from the measurement Λ_γ .

This problem is again logarithmically stable. However, we can observe and also prove mathematically that the reconstruction of D is more stable when D is near the boundary.

Stationary equations

We can consider the Helmholtz type equation

$$\Delta u + k^2 u + qu = 0 \quad \text{in } \Omega,$$

where k is the wave number and q is a potential function. We can study the problem of determining q by the measurements

$$\left\{ u|_{\partial\Omega}, \frac{\partial u}{\partial \nu} |_{\partial\Omega} \right\}.$$

Stationary equations

We can consider the Helmholtz type equation

$$\Delta u + k^2 u + qu = 0 \quad \text{in } \Omega,$$

where k is the wave number and q is a potential function. We can study the problem of determining q by the measurements

$$\left\{ u|_{\partial\Omega}, \frac{\partial u}{\partial \nu} |_{\partial\Omega} \right\}.$$

For this case, one can observe that the stability estimate tends to a Hölder type as k increases. That is,

$$\text{error}(q_1, q_2) \leq C(k) (\text{error}(\Lambda_{q_1}, \Lambda_{q_2}))^\alpha + \underbrace{\text{error term}}_{\downarrow 0 \text{ as } k \rightarrow \infty},$$

where $\alpha \in (0, 1)$. Stability is **increasing** if we are prospecting the unknown by the high frequency waves.

Wave equation

Consider the initial-boundary value problem for the wave equation:

$$\begin{cases} \partial_{tt}^2 v - \Delta v + qv = 0 & (x, t) \in \Omega \times (0, T), \\ v(0, x) = 0, \quad \partial_t v(0, x) = 0 & x \in \Omega, \\ v(t, x) = f, & (x, t) \in \partial\Omega \times (0, T). \end{cases}$$

We can define the measurement Λ_q^h :

$$\Lambda_q^h(f) = \frac{\partial v}{\partial \nu} \Big|_{\partial\Omega \times (0, T)}$$

Wave equation

Consider the initial-boundary value problem for the wave equation:

$$\begin{cases} \partial_{tt}^2 v - \Delta v + qv = 0 & (x, t) \in \Omega \times (0, T), \\ v(0, x) = 0, \quad \partial_t v(0, x) = 0 & x \in \Omega, \\ v(t, x) = f, & (x, t) \in \partial\Omega \times (0, T). \end{cases}$$

We can define the measurement Λ_q^h :

$$\Lambda_q^h(f) = \frac{\partial v}{\partial \nu} \Big|_{\partial\Omega \times (0, T)}$$

In this problem, the following stability estimate holds

$$\text{error}(q_1, q_2) \leq C(\text{error}(\Lambda_{q_1}^h, \Lambda_{q_2}^h))^\beta, \quad \beta \in (0, 1).$$

Finite speed of propagation

When k is large, the Helmholtz equation behaves like the wave equation.

Finite speed of propagation

When k is large, the Helmholtz equation behaves like the wave equation.

The better stability for the wave equation is probably due to the property of **finite propagation speed**.

Finite speed of propagation

When k is large, the Helmholtz equation behaves like the wave equation.

The better stability for the wave equation is probably due to the property of **finite propagation speed**.

For the elliptic (e.g., $\Delta u = 0$) or parabolic equations (e.g. $\partial_t u - \Delta u = 0$), the speed of propagation is **infinite**.

Numerical simulations

Determine p from the observation y in $y = G(p)$. Try to solve

$$p = \operatorname{argmin}_{p \in X} \|y - G(p)\|_Y^2.$$

It looks quite straightforward and easy!

Numerical simulations

Determine p from the observation y in $y = G(p)$. Try to solve

$$p = \operatorname{argmin}_{p \in X} \|y - G(p)\|_Y^2.$$

It looks quite straightforward and easy!

Problems:

- Inverse problem is ill-posed. Small perturbations of data will lead to large errors in the determination of p . That is, the numerically determined solution is not stable.
- It is difficult to apply a Newton-type scheme since the derivative of G is quite often hard to compute.

Solve (Tikhonov's regularization)

$$p = \operatorname{argmin}_{p \in X} \left\{ \|y - G(p)\|_Y^2 + \alpha \|p\|_E^2 \right\}$$

$E \subset X$: compact subspace

α : parameter

Introducing a regularizer can stabilize the problem, but G is most likely not convex in p . There are many minimizers. Also, the choice of α is tricky.

Statistical viewpoint

In practice, there exist measurement errors. Thus, it is more reasonable to consider

$$y = G(p) + \xi,$$

where ξ is the noise.

Statistical viewpoint

In practice, there exist measurement errors. Thus, it is more reasonable to consider

$$y = G(p) + \xi,$$

where ξ is the noise.

Precisely, let X, Y be separable Hilbert spaces and $G : X \rightarrow Y$ be a Borel measurable map. Normally, ξ is a centered Gaussian on Y , i.e. $\xi \sim \mathcal{N}(0, \Gamma)$, where Γ the covariance operator which is a positive operator of trace class.

Statistical viewpoint

In practice, there exist measurement errors. Thus, it is more reasonable to consider

$$y = G(p) + \xi,$$

where ξ is the noise.

Precisely, let X, Y be separable Hilbert spaces and $G : X \rightarrow Y$ be a Borel measurable map. Normally, ξ is a centered Gaussian on Y , i.e. $\xi \sim \mathcal{N}(0, \Gamma)$, where Γ the covariance operator which is a positive operator of trace class.

We put a prior on p , i.e. p is a random element on X . It is common to choose $p \sim \mathcal{N}(\bar{p}, C)$, the Gaussian with mean \bar{p} and covariance C .

$$\underbrace{\mathbb{P}(p|y)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(y|p)}_{\text{likelihood}} \underbrace{\mathbb{P}(p)}_{\text{prior}}$$

In most cases, we can express

$$\mathbb{P}(p|y) \propto \exp(-\Phi(p; y)).$$

Given observation y , $-\Phi(\cdot; y)$ is called the *log likelihood*.

$$\underbrace{\mathbb{P}(p|y)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(y|p)}_{\text{likelihood}} \underbrace{\mathbb{P}(p)}_{\text{prior}}$$

In most cases, we can express

$$\mathbb{P}(p|y) \propto \exp(-\Phi(p; y)).$$

Given observation y , $-\Phi(\cdot; y)$ is called the *log likelihood*.

For the case of $\xi \sim \mathcal{N}(0, \Gamma)$ and $u \sim \mathcal{N}(\bar{p}, C)$,

$$\Phi(p; y) = \frac{1}{2} \|\Gamma^{-1/2}(y - G(p))\|_Y^2 + \frac{1}{2} \|C^{-1/2}(p - \bar{p})\|_X^2$$

Maximum likelihood estimation (MLE):

$$\hat{p}_{\text{MLE}} = \underset{p \in X}{\operatorname{argmin}} \|\Gamma^{-1/2}(y - G(p))\|_Y^2.$$

Minimization without regularization

Maximum likelihood estimation (MLE):

$$\hat{p}_{\text{MLE}} = \operatorname{argmin}_{p \in X} \|\Gamma^{-1/2}(y - G(p))\|_Y^2.$$

Minimization without regularization

Maximum a posterior estimation (MAP):

$$\hat{p}_{\text{MAP}} = \operatorname{argmin}_{p \in X} \left\{ \frac{1}{2} \|\Gamma^{-1/2}(y - G(p))\|_Y^2 + \frac{1}{2} \|C^{-1/2}(p - \bar{p})\|_X^2 \right\}.$$

Minimization with regularization

Maximum likelihood estimation (MLE):

$$\hat{p}_{\text{MLE}} = \operatorname{argmin}_{p \in X} \|\Gamma^{-1/2}(y - G(p))\|_Y^2.$$

Minimization without regularization

Maximum a posterior estimation (MAP):

$$\hat{p}_{\text{MAP}} = \operatorname{argmin}_{p \in X} \left\{ \frac{1}{2} \|\Gamma^{-1/2}(y - G(p))\|_Y^2 + \frac{1}{2} \|C^{-1/2}(p - \bar{p})\|_X^2 \right\}.$$

Minimization with regularization

Both estimations provide point estimators.

Posterior distribution

Instead of the minimization technique, we can use the posterior distribution to make inference of the parameter.

Let $\Pi(\cdot|Y)$ be the posterior distribution. The expectation $\mathbb{E}[\Pi(\cdot|Y)]$ is a natural point estimator of p . Formally

$$\mathbb{E}[\Pi(\cdot|Y)] = \int_{\mathcal{X}} p \, d\Pi(p|Y).$$

This integral can be rigorously defined in Bochner's sense. To be useful, we need to find an effective way to approximate it.

Posterior distribution

Instead of the minimization technique, we can use the posterior distribution to make inference of the parameter.

Let $\Pi(\cdot|Y)$ be the posterior distribution. The expectation $\mathbb{E}[\Pi(\cdot|Y)]$ is a natural point estimator of p . Formally

$$\mathbb{E}[\Pi(\cdot|Y)] = \int_{\mathcal{X}} p \, d\Pi(p|Y).$$

This integral can be rigorously defined in Bochner's sense. To be useful, we need to find an effective way to approximate it.

Monte Carlo Integration

Posterior distribution

Instead of the minimization technique, we can use the posterior distribution to make inference of the parameter.

Let $\Pi(\cdot|Y)$ be the posterior distribution. The expectation $\mathbb{E}[\Pi(\cdot|Y)]$ is a natural point estimator of p . Formally

$$\mathbb{E}[\Pi(\cdot|Y)] = \int_{\mathcal{X}} p \, d\Pi(p|Y).$$

This integral can be rigorously defined in Bochner's sense. To be useful, we need to find an effective way to approximate it.

Monte Carlo Integration

Let $p^{(1)}, \dots, p^{(n)}$ be sampled independently from $\Pi(\cdot|Y)$. Then

$$\frac{1}{n} \sum_{k=1}^n p^{(k)} \rightarrow \mathbb{E}[\Pi(\cdot|Y)]$$

(Law of Large Number)

MCMC (sampling)

MCMC algorithms are the most popular and powerful methods to sample probability distributions. However, in dealing with PDE inverse problems, the posteriors in general lie in the infinite dimensional function spaces, it becomes a challenging question to sample such posteriors effectively.

MCMC (sampling)

MCMC algorithms are the most popular and powerful methods to sample probability distributions. However, in dealing with PDE inverse problems, the posteriors in general lie in the infinite dimensional function spaces, it becomes a challenging question to sample such posteriors effectively.

The *preconditioned Crank-Nicolson algorithm* (pCN) is a method developed to sample distributions at high dimensions. The pCN method is a Metropolis-type algorithm. Most importantly, the pCN algorithm has dimension-independent sampling efficiency.

Ensemble Kalman filter

Kalman filter method can also be used to study inverse problems. As before, let $G : X \rightarrow Y$ be the forward map

$$y = G(q).$$

Let $Z = X \times Y$ and the map $\Theta : Z \rightarrow Z$ defined by

$$\Theta(z) = \begin{pmatrix} q \\ G(q) \end{pmatrix} \text{ for } z = \begin{pmatrix} q \\ p \end{pmatrix}, \text{ where } p \in Y.$$

We consider the artificial dynamics given by

$$z_{\ell+1} = \Theta(z_\ell), \quad \ell = 0, 1, 2, \dots \quad (2)$$

Associated to the dynamical equation is the observation equation

$$y_{\ell+1} = Hz_{\ell+1} + \eta_{\ell+1}, \quad (3)$$

where $H : Z \rightarrow Y$ is the projection operator given by $H = \begin{pmatrix} 0 \\ I \end{pmatrix}$ and η_1, η_2, \dots are independent Gaussian random elements on Y .

Associated to the dynamical equation is the observation equation

$$y_{\ell+1} = Hz_{\ell+1} + \eta_{\ell+1}, \quad (3)$$

where $H : Z \rightarrow Y$ is the projection operator given by $H = \begin{pmatrix} 0 \\ I \end{pmatrix}$ and η_1, η_2, \dots are independent Gaussian random elements on Y .

The dynamical system (2) gives rises to a prediction step and the observation equation (3) provides a correction step.

EnKF : initial ensemble

Initial ensemble: $\{z_0^{(j)}\}_{j=1}^J$ are determined by the prior knowledge of q . Let μ_0 be the prior distribution of q , the initial ensemble consists of $\{q_0^{(j)}\}_{j=1}^J$, which are i.i.d. samples of μ_0 . We now define the space of initial ensemble

$$\mathcal{A} = \{q_0^{(j)}\}_{j=1}^J.$$

Alternatively, if $\mu_0 \sim \mathcal{N}(\bar{q}, C)$, we can consider $q_0^{(j)} = \bar{q} + \sqrt{\alpha_j} \zeta_j$, $j \leq J$, where (α_j, ζ_j) are eigenpairs of C . This is known as Karhunen-Loève basis.

EnKF : initial ensemble

Initial ensemble: $\{z_0^{(j)}\}_{j=1}^J$ are determined by the prior knowledge of q . Let μ_0 be the prior distribution of q , the initial ensemble consists of $\{q_0^{(j)}\}_{j=1}^J$, which are i.i.d. samples of μ_0 . We now define the space of initial ensemble

$$\mathcal{A} = \{q_0^{(j)}\}_{j=1}^J.$$

Alternatively, if $\mu_0 \sim \mathcal{N}(\bar{q}, C)$, we can consider $q_0^{(j)} = \bar{q} + \sqrt{\alpha_j} \zeta_j$, $j \leq J$, where (α_j, ζ_j) are eigenpairs of C . This is known as Karhunen-Loève basis.

Assume that the observational data y is given by the "ground truth" $q^\dagger \in X$ with a measurement noise, i.e.,

$$y = G(q^\dagger) + \eta^\dagger$$

where η^\dagger denotes the measurement noise.

Algorithm

Algorithm. Iterative EnKF.

Let the initial ensemble $\{q_0^{(j)}\}_{j=1}^J$ have been chosen. Denote

$$p_0^{(j)} = G(q_0^{(j)}) \text{ and } z_0^{(j)} = \begin{pmatrix} q_0^{(j)} \\ p_0^{(j)} \end{pmatrix}. \text{ For } \ell = 0, 1, 2, \dots.$$

- (1) **Prediction step.** Propagate the ensemble particles by the dynamical system (2):

$$\hat{z}_{\ell+1}^{(j)} = \Theta(z_\ell^{(j)}), \quad j = 1, \dots, J. \quad (4)$$

Sample mean: $\bar{z}_{\ell+1} = \frac{1}{J} \sum_{j=1}^J \hat{z}_{\ell+1}^{(j)}$

Sample covariance:

$$C_{\ell+1} = \frac{1}{J-1} \sum_{j=1}^J \hat{z}_{\ell+1}^{(j)} (\hat{z}_{\ell+1}^{(j)})^T - \bar{z}_{\ell+1} (\bar{z}_{\ell+1})^T.$$

(2) **Analysis step.** Here we would like to update the ensemble particles that were obtained in (1) by the observational data. We first define the Kalman gain K_ℓ by

$$K_{\ell+1} = C_{\ell+1} H^* (H C_{\ell+1} H^* + \Gamma)^{-1},$$

where $H^* = (0, I_N)$. Then each ensemble particle is updated as follows:

$$\begin{aligned} z_{\ell+1}^{(j)} &= \widehat{z}_{\ell+1}^{(j)} + K_{\ell+1} (y_{\ell+1}^{(j)} - H \widehat{z}_{\ell+1}^{(j)}) \\ &= (I - K_{\ell+1} H) \widehat{z}_{\ell+1}^{(j)} + K_{\ell+1} y_{\ell+1}^{(j)}, \end{aligned}$$

where

$$y_{\ell+1}^{(j)} = y + \eta_{\ell+1}^{(j)},$$

where $\eta_{\ell+1}^{(j)}$ are i.i.d. with $\eta_{\ell+1}^{(j)} \sim \eta_{\ell+1}$ for $j = 1, \dots, J$.

(3) Compute the updated sample mean

$$q_{\ell+1} = \frac{1}{J} \sum_{j=1}^J H^\perp z_{\ell+1}^{(j)},$$

where $H^\perp = (I, 0)$.

(3) Compute the updated sample mean

$$q_{\ell+1} = \frac{1}{J} \sum_{j=1}^J H^\perp z_{\ell+1}^{(j)},$$

where $H^\perp = (I, 0)$.

Subspace invariance property: $\{q_\ell^{(j)}\}_{j=1}^J \in \mathcal{A}$ for all $\ell \in \mathbb{N}$.

(3) Compute the updated sample mean

$$q_{\ell+1} = \frac{1}{J} \sum_{j=1}^J H^\perp z_{\ell+1}^{(j)},$$

where $H^\perp = (I, 0)$.

Subspace invariance property: $\{q_\ell^{(j)}\}_{j=1}^J \in \mathcal{A}$ for all $\ell \in \mathbb{N}$.

Interesting questions:

- Choice of initial ensemble
- Stopping criteria
- Convergence theory

Thank you for your attention