Inverse Problems
Introduction and examples

Clarice Poon
University of Bath

January 30, 2020

There are 3 lectures a week for 10 weeks.

Two guest lectures:

- On Tuesday 10th March, Jingwei Liang from University of Cambridge will talk about variational techniques for imaging applications.
- On Tuesday 31st March, Chris Budd will talk about data assimilation techniques in weather forecasting.

Assessment

1. 40% for 2 sets of exercises. First deadline will be 28th February.
2. 30% for group presentations (approx. 15 minutes) in week 11. To be announced.
3. 30% for oral examination after the end of lectures (in May).

Inverse problems

This course is concerned with determining causes for a desired or observed effect, given indirect measurements.

Inverse problems

This course is concerned with determining causes for a desired or observed effect, given indirect measurements.

We will often see two problems, a direct problem and an inverse problem.

This course is concerned with determining causes for a desired or observed effect, given indirect measurements.

We will often see two problems, a direct problem and an inverse problem.

Find u given measurements f:

$$Au = f \tag{0.1}$$

where $A : \mathcal{U} \to \mathcal{V}$ is the forward operator acting between some spaces $\mathcal{U}$ and $\mathcal{V}$.

Typically, A models the physics of data acquisition and is called the **forward model**.

Ill posed inverse problems according to Hadamard (1902)

Definition 1

The problem (0.1) is well-posed if

- it has a solution for all $f \in \mathcal{V}$.
- the solution is unique
- the solution depends continuously on the data, i.e. small errors in the data f result in small errors in the reconstruction.

If any one of these properties is violated, then the problem is called ill-posed.

Ill posed inverse problems according to Hadamard (1902)

Definition 1

The problem (0.1) is well-posed if

- it has a solution for all $f \in \mathcal{V}$.
- the solution is unique
- the solution depends continuously on the data, i.e. small errors in the data f result in small errors in the reconstruction.

If any one of these properties is violated, then the problem is called ill-posed.

Need to quantify some of these notions with appropriate norms or topologies, e.g. we require that $\|u_f - u_g\|_{\mathcal{U}} \leqslant C \|f - g\|_{\mathcal{V}}$.

Ill posed inverse problems according to Hadamard (1902)

Definition 1

The problem (0.1) is well-posed if

- it has a solution for all $f \in \mathcal{V}$.
- the solution is unique
- the solution depends continuously on the data, i.e. small errors in the data f result in small errors in the reconstruction.

If any one of these properties is violated, then the problem is called ill-posed.

Need to quantify some of these notions with appropriate norms or topologies, e.g. we require that $\|u_f - u_g\|_{\mathcal{U}} \leqslant C \|f - g\|_{\mathcal{V}}$.

Of these, it is perhaps the last point which is the most pertinent. Without continuity, any numerical method will be unstable.

Ill posed inverse problems according to Hadamard (1902)

Definition 1

The problem (0.1) is well-posed if

- it has a solution for all $f \in \mathcal{V}$.
- the solution is unique
- the solution depends continuously on the data, i.e. small errors in the data f result in small errors in the reconstruction.

If any one of these properties is violated, then the problem is called ill-posed.

Need to quantify some of these notions with appropriate norms or topologies, e.g. we require that $\|u_f - u_g\|_{\mathcal{U}} \leqslant C \|f - g\|_{\mathcal{V}}$.

Of these, it is perhaps the last point which is the most pertinent. Without continuity, any numerical method will be unstable.

Under realistic choices of norms, many problems are ill-posed. We will discuss how to deal with this ill-posedness using regularization.

Let's consider the problem of finding the derivative of $f \in C^1[0,1]$.

Ill-posed

Let $f \in C^1[0,1]$, $\delta \in (0,1)$ and $n \in \mathbb{N}$. Define

$$f_n^\delta(x) \overset{\text{def.}}{=} f(x) + \delta \sin\left(\frac{nx}{\delta}\right)$$
$$(f_n^\delta)'(x) = f'(x) + n \cos\left(\frac{nx}{\delta}\right).$$

Now, $\left\|f - f_n^\delta\right\|_\infty = \delta$ but $\left\|(f_n^\delta)' - f'\right\|_\infty = n$.

We can of course make the problem stable by considering continuity with respect to $\|g\|_{C^1} \overset{\text{def.}}{=} \max\left(\|g\|_\infty, \|g'\|_\infty\right)$ but this is kind of cheating...

Differentiation

Given f, we seek to find f' by solving

$$(Au)(s) = \int_0^s u(t)\mathrm{d}t = f(s) - f(0).$$

This is solvable in $C[0,1]$ if $f \in C^1[0,1]$. Note that A is a continuous linear operator on $C[0,1]$ but we just saw that its inverse defined on $C^1[0,1]$ is unbounded.

How can we make the problem stable? Need to exclude the presence of high frequency error (e.g. a bound on f'').

Let's restrict A to

$$\mathcal{X} \overset{\text{def.}}{=} \left\{ u \in C^1[0,1] \; ; \; \|u\|_\infty + \left\|u'\right\|_\infty \leqslant \gamma \right\}$$

which is a compact set in $C[0,1]$ by Arzela-Ascoli, so its inverse is continuous on its range $A(\mathcal{X})$.

Differentiation via forward differences

Let $f \in C^1[0,1]$ and f^δ be such that $\left\|f - f^\delta\right\|_\infty \leqslant \delta$.

If $f \in C^1[0,1]$, then by Taylor expansion:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h).$$

Differentiation via forward differences

Let $f \in C^1[0,1]$ and f^δ be such that $\left\|f - f^\delta\right\|_\infty \leqslant \delta$.

If $f \in C^1[0,1]$, then by Taylor expansion:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h).$$

If $f \in C^2[0,1]$, then

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2).$$

Differentiation via forward differences

Let $f \in C^1[0,1]$ and f^δ be such that $\left\| f - f^\delta \right\|_\infty \leqslant \delta$.

If $f \in C^1[0,1]$, then by Taylor expansion:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h).$$

If $f \in C^2[0,1]$, then

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2).$$

Note that

$$\frac{f^\delta(x+h) - f^\delta(x-h)}{2h} = \frac{f(x+h) - f(x-h)}{2h} + \mathcal{O}\left(\frac{\delta}{h}\right).$$

Let $f \in C^1[0,1]$ and f^δ be such that $\left\| f - f^\delta \right\|_\infty \leqslant \delta$.

If $f \in C^1[0,1]$, then by Taylor expansion:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h).$$

If $f \in C^2[0,1]$, then

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2).$$

Note that

$$\frac{f^\delta(x+h) - f^\delta(x-h)}{2h} = \frac{f(x+h) - f(x-h)}{2h} + \mathcal{O}\left(\frac{\delta}{h}\right).$$

Rem 1: The approximation error is bad when h is large, while the data error blows up when h is small.

Differentiation via forward differences

Let $f \in C^1[0,1]$ and f^δ be such that $\left\| f - f^\delta \right\|_\infty \leqslant \delta$.

If $f \in C^1[0,1]$, then by Taylor expansion:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h).$$

If $f \in C^2[0,1]$, then

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2).$$

Note that

$$\frac{f^\delta(x+h) - f^\delta(x-h)}{2h} = \frac{f(x+h) - f(x-h)}{2h} + \mathcal{O}\left(\frac{\delta}{h}\right).$$

Rem 1: The approximation error is bad when h is large, while the data error blows up when h is small.
Rem 2: The error depends on the smoothness of f and is $h^\nu + \delta/h$ for $\nu \in \{1,2\}$. The error is $\mathcal{O}(\delta^{\nu/(\nu+1)})$ under the choice $h = (\delta/\nu)^{1/(\nu+1)}$.

Let's summarise:

- There is an amplification of high frequency errors. (The forward operator, integration, is a "smoothing" process)
- We can restore stability by a-priori information.

For the simple finite differences:

- There are 2 error terms: approximation and data errors and there is ultimately a tradeoff.
- The optimal choice of stepsize h depends on a-priori information.
- The optimal error is $\mathcal{O}(\delta^{\nu/(\nu+1)})$, so $\mathcal{O}(\delta^{2/3})$ at best, so there is a loss of information.

When a camera records an image:

$$f(x) = (Au)(x) \stackrel{\text{def.}}{=} \int K(x,\xi)u(\xi)\mathrm{d}\xi$$

where u is the true image, $K(x,\xi)$ is the point-spread function, which models the optics of the camera.

Theorem 2

Let $A : L^2(\Omega) \to L^2(\Omega)$ with $K(\cdot,\cdot) \in L^2(\Omega \times \Omega)$. Then, A is compact.

We shall see later that the inversion of a compact operator is always ill-posed.

Deconvolution

Special case is the spatially invariant kernel $K(x,\xi) = \kappa(x-\xi)$. Then, Au is a convolution and this problem is called **deconvolution**.


Ill-posed: $f \overset{\text{def.}}{=} Au = u \star \varphi$, where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \implies \hat{\varphi}(\omega) = e^{-\frac{\omega^2}{2}}$.

Deconvolution

Special case is the spatially invariant kernel $K(x, \xi) = \kappa(x - \xi)$. Then, Au is a convolution and this problem is called **deconvolution**.

Ill-posed: $f \stackrel{\text{def.}}{=} Au = u \star \varphi$, where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \implies \hat{\varphi}(\omega) = e^{-\frac{\omega^2}{2}}$.

- So, $\hat{f}(\omega) = \hat{u}(\omega)\hat{\varphi}(\omega)$ and reconstruction is $\hat{u}(\omega) = \hat{f}(\omega) e^{\frac{\omega^2}{2}}$.

Deconvolution

Special case is the spatially invariant kernel $K(x, \xi) = \kappa(x - \xi)$. Then, Au is a convolution and this problem is called **deconvolution**.

Ill-posed: $f \stackrel{\text{def.}}{=} Au = u \star \varphi$, where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \implies \hat{\varphi}(\omega) = e^{-\frac{\omega^2}{2}}$.

- So, $\hat{f}(\omega) = \hat{u}(\omega)\hat{\varphi}(\omega)$ and reconstruction is $\hat{u}(\omega) = \hat{f}(\omega)e^{\frac{\omega^2}{2}}$.
- Given noisy observations, $f_\delta = u \star \varphi + z$, so $\hat{f}_\delta(\omega)e^{\frac{\omega^2}{2}} = \hat{u}(\omega) + \hat{z}(\omega)e^{\frac{\omega^2}{2}}$, error is amplified exponentially in frequency!

Matrix inversion

Suppose $u, f \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix.

There exists eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0$ and (orthonormal) eigenvectors a_j such that

$$A = \sum_{j=1}^{n} \lambda_j a_j a_j^\top \quad \text{and} \quad \|A\| = \lambda_1.$$

Matrix inversion

Suppose $u, f \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix.

There exists eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0$ and (orthonormal) eigenvectors a_j such that

$$A = \sum_{j=1}^{n} \lambda_j a_j a_j^\top \quad \text{and} \quad \|A\| = \lambda_1.$$

Given perturbed data $f_\delta = f + \delta a_n$, so $\|f_\delta - f\| = \delta$, $u_\delta \overset{\text{def.}}{=} A^{-1} f_\delta$ satisfies

$$u - u_\delta = \sum_{j=1}^{n} \lambda_j^{-1} a_j a_j^\top (f - f_\delta) = \lambda_n^{-1} \delta a_n.$$

$$\implies \|u - u_\delta\|_2 = \delta/\lambda_n = \kappa\delta.$$

Matrix inversion

Suppose $u, f \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix.

There exists eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0$ and (orthonormal) eigenvectors a_j such that

$$A = \sum_{j=1}^{n} \lambda_j a_j a_j^\top \quad \text{and} \quad \|A\| = \lambda_1.$$

Given perturbed data $f_\delta = f + \delta a_n$, so $\|f_\delta - f\| = \delta$, $u_\delta \stackrel{\text{def.}}{=} A^{-1} f_\delta$ satisfies

$$u - u_\delta = \sum_{j=1}^{n} \lambda_j^{-1} a_j a_j^\top (f - f_\delta) = \lambda_n^{-1} \delta a_n.$$

$$\implies \|u - u_\delta\|_2 = \delta/\lambda_n = \kappa\delta.$$

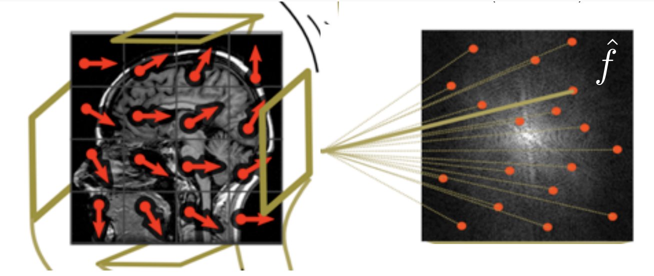
In the worst case, an error of δ is amplified by the condition number κ of A.

A matrix with large κ is called ill-conditioned.

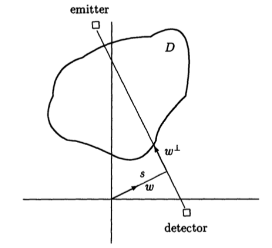
Matrix inversion

Another problem is where $u \in \mathbb{R}^n$, $f \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ where $m \ll n$.

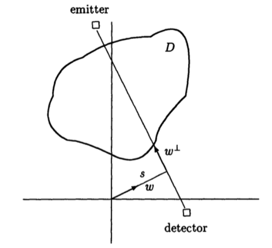
- This is clearly ill-posed as there may be multiple solutions.
- This arises in many situation where we have limited data (such as certain medical imaging and astronomy applications where the acquisition of data might be very expensive).



Example: Magnetic resonance imaging.



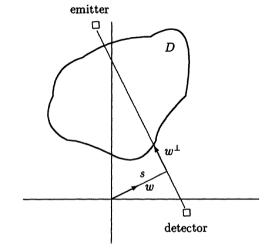
Let $\mathcal{D} \subset \mathbb{R}^2$ be a compact domain with a spatially varying density u inside.
We have X-rays travelling along lines parameterized by $w \in \mathbb{R}^2$ with $\|w\| = 1$ and their distance $s > 0$ to the origin.



Let $\mathcal{D} \subset \mathbb{R}^2$ be a compact domain with a spatially varying density u inside.
We have X-rays travelling along lines parameterized by $w \in \mathbb{R}^2$ with $\|w\| = 1$ and their distance $s > 0$ to the origin.

Assume that the decay $-\Delta I$ of an X-ray beam along distance Δt is proportional to density u, intensity I and Δt, so

$$\Delta I(sw + tw^\perp) = -I(sw + tw^\perp)u(sw + tw^\perp)\Delta t$$

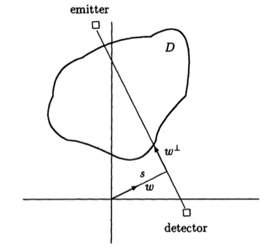


Let $\mathcal{D} \subset \mathbb{R}^2$ be a compact domain with a spatially varying density u inside.
We have X-rays travelling along lines parameterized by $w \in \mathbb{R}^2$ with $\|w\| = 1$ and their distance $s > 0$ to the origin.

Assume that the decay $-\Delta I$ of an X-ray beam along distance Δt is proportional to density u, intensity I and Δt, so

$$\frac{\mathrm{d}}{\mathrm{d}t} I(sw + tw^\perp) = -I(sw + tw^\perp)u(sw + tw^\perp).$$

The X-ray transform



Let $\mathcal{D} \subset \mathbb{R}^2$ be a compact domain with a spatially varying density u inside.
We have X-rays travelling along lines parameterized by $w \in \mathbb{R}^2$ with $\|w\| = 1$ and their distance $s > 0$ to the origin.

Assume that the decay $-\Delta I$ of an X-ray beam along distance Δt is proportional to density u, intensity I and Δt, so

$$\frac{\mathrm{d}}{\mathrm{d}t} I(sw + tw^\perp) = -I(sw + tw^\perp)u(sw + tw^\perp).$$

Let $I_L(s, w)$ and $I_0(s, w)$ be the intensities at the detector and emitter (at infinity). Then

$$\log I_L(s, w) - \log I_0(s, w) = -\int u(sw + tw^\perp)\mathrm{d}t$$

Our observations are

$$f(s,w) = (\mathcal{X}u)(s,w) \stackrel{\text{def.}}{=} \int u(sw + tw^{\perp})\mathrm{d}t = -\log\left(\frac{I_L(s,w)}{I_0(s,w)}\right)$$

Simple case: $\mathcal{D} = B(0,\rho)$ and $u(s,w) = U(s)$ is circularly symmetric. Let $w = (0,\pm 1)$. We have to solve for $s \in (0,\rho]$,

$$(\mathcal{X}u)(s,w) = 2\int_s^{\rho} \frac{rU(r)}{\sqrt{r^2 - s^2}}\mathrm{d}r = f(s) \stackrel{\text{def.}}{=} -\log\left(\frac{I_L(s,w)}{I_0(s,w)}\right)$$

This is the Abel integral of first kind and if $f(\rho) = 0$, then

$$U(r) = \frac{-1}{\pi}\int_r^{\rho} \frac{g'(s)}{\sqrt{s^2 - r^2}}\mathrm{d}s$$

and the solution involves f' which we know is ill-posed (subsequent integration only partially annihilates the effects of differentiation).

The Radon transform

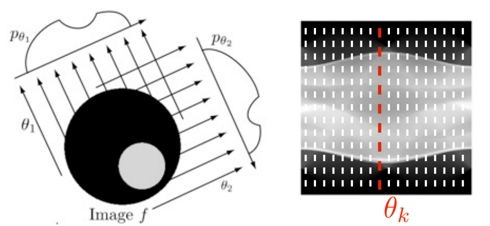
For $s \in \mathbb{R}$, $\theta \in \mathbb{S}^{n-1}$, the Radon transform $\mathcal{R} : C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{S}^{n-1} \times \mathbb{R})$ integrates over hyperplanes and is defined as

$$f(\theta, s) = (\mathcal{R}u)(\theta, s) = \int_{x \cdot \theta = s} u(x) \mathrm{d}x.$$

The Radon transform

For $s \in \mathbb{R}$, $\theta \in \mathbb{S}^{n-1}$, the Radon transform $\mathcal{R} : C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{S}^{n-1} \times \mathbb{R})$ integrates over hyperplanes and is defined as

$$f(\theta, s) = (\mathcal{R}u)(\theta, s) = \int_{x\cdot\theta=s} u(x)\mathrm{d}x.$$

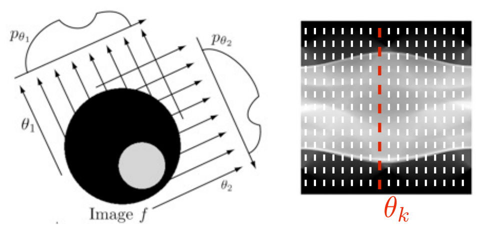


Parallel beam computed tomography

The Radon transform

For $s \in \mathbb{R}$, $\theta \in \mathbb{S}^{n-1}$, the Radon transform $\mathcal{R} : C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{S}^{n-1} \times \mathbb{R})$ integrates over hyperplanes and is defined as

$$f(\theta, s) = (\mathcal{R}u)(\theta, s) = \int_{x\cdot\theta=s} u(x)\mathrm{d}x.$$



Parallel beam computed tomography

- The Radon transform is linear, continuous and compact from L^2 to L^2.
- For $n = 2$, this is $(\mathcal{R}u)(\theta, s) = \int_t u(s\theta_y + t\theta^\perp)\mathrm{d}y$ where $\theta \in \mathbb{S}^{n-1}$ and $\theta^\perp$ is the vector orthogonal to θ. So, it integrates along lines and coincides with the X-ray transform.

Course outline

Core theory:

1. Least squares solutions.
2. Spectral regularisation techniques.
3. Structured regularisation techniques, e.g. Total variation.

Special cases of inverse problems:

1. Data assimilation – how do we optimally combine observations with some physical models for more accurate predictions?
2. Compressed sensing – recovering from very few measurements under the assumption of sparsity.