Linear Inverse Problems

Ajinkya Kadu

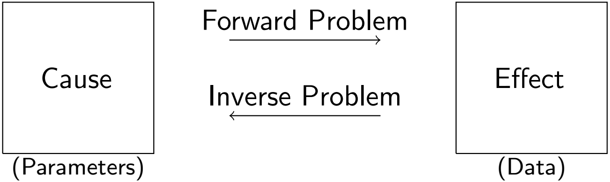
Utrecht University, The Netherlands

February 26, 2018

Outline

What are inverse problems?

Inverse problems are determining cause for an observed effect.



- Forward Problem: $\mathbf{x} \to F(\mathbf{x})$
- Inverse Problem: $F(\mathbf{x}) \to \mathbf{x}$

Properties of Inverse Problems

Forward problems are always *well-posed*, while inverse problems are not!

Well-posedness in terms of Hadamard conditions:

- There exists a solution for all input data.
- If solution exists, it must be unique.
- solution of the problem depends continuously on input datum.

If any of these conditions is violated, problem is called *ill-posed*.

Linear Inverse Problem

$F(x)$ is a linear functional.

Problems of form

$$\mathbf{y} \approx \mathbf{A}\mathbf{x}$$

- we are given $\mathbf{A} \in \mathbb{R}^{M \times N}$, we observe $\mathbf{y} \in \mathbb{R}^{M}$ and want to find (or estimate) $\mathbf{x} \in \mathbb{R}^{N}$.
- most fundamental concept in all of engineering, science, and applied maths!
- two areas of Interests:
  - Supervised Learning
  - Computational Imaging

Supervised Learning

estimate a function $f(\mathbf{t})$ on $\mathbf{R}^D$ from observations of its samples.

$$f(\mathbf{t}_m) \approx y_m, \quad m = 1, \ldots, M$$

- $f : \mathbb{R}^D \to \mathbb{R}$.
- Problem is not well-posed (many f possible).
- Need to define set of functions $\mathcal{F}$ from which to choose f.

Supervised Learning

Example: Linear Regression

$\mathcal{F}$ contain set of all linear functionals on $\mathbb{R}^D$.

linear function $f : \mathbb{R}^D \to \mathbb{R}$ obeys

$$f(\alpha \mathbf{t}_1 + \beta \mathbf{t}_2) = \alpha f(\mathbf{t}_1) + \beta f(\mathbf{t}_2)$$

for all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{t}_1, \mathbf{t}_2 \in \mathbb{R}^D$.

every linear functional on $\mathbb{R}^D$ is uniquely represented by vector $\mathbf{x}_f \in \mathbb{R}^D$ (Riesz Representation Theorem, holds in any Hilbert space)

$$f(\mathbf{t}) = \langle \mathbf{t}, \mathbf{x}_f \rangle$$

Supervised Learning

Example: Linear Regression

given $(\mathbf{t}_m, y_m)$, find $\mathbf{x}$ such that $y_m = \langle \mathbf{t}_m, \mathbf{x} \rangle$.

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad \text{where} \quad \mathbf{A} = \begin{bmatrix} — & \mathbf{t}_1^T & — \\ — & \mathbf{t}_2^T & — \\ & \vdots & \\ — & \mathbf{t}_M^T & — \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}$$

Supervised Learning

Example: Non-Linear Regression using a basis

$\mathcal{F}$ is spanned by basis functions $\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_N$.

$$f(\mathbf{t}) = \sum_{n=1}^{N} x_n \psi_n(\mathbf{t})$$

Again, fitting can be rewritten as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \psi_1(\mathbf{t}_1) & \psi_2(\mathbf{t}_1) & \dots & \psi_N(\mathbf{t}_1) \\ \psi_1(\mathbf{t}_2) & \psi_2(\mathbf{t}_2) & \dots & \psi_N(\mathbf{t}_2) \\ \vdots & & \ddots & \\ \psi_1(\mathbf{t}_M) & \psi_2(\mathbf{t}_M) & \dots & \psi_N(\mathbf{t}_M) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

Computational Imaging

recover a function f that represents some physical structure indexed by location

- similar to regression problem: discretize the problem by representing f using a basis.
- Unlike regression problem: not observe f, but more general linear functions.

Computational Imaging

Example: Range profiling using deconvolution

sending a pulse out (of electromagnetic or acoustic energy) and listening to the echo.

Applications:

- radar imaging
- underwater acoustic imaging
- seismic imaging
- medical imaging
- channel equalization in wireless communications
- image deblurring

Computational Imaging

Example: Range profiling using deconvolution

send a pulse $p(t)$ out, and receive back a signal $y(t)$

$$y(t) = \int_{-\infty}^{\infty} f(s)p(t-s)\,\mathrm{d}s$$

assuming $f(t)$ is *time-limited*, $\{\boldsymbol{\psi}_n\}$ basis for $\mathrm{L}_2([0,T])$

$$f(t) = \sum_{n=1}^{N} x_n \psi_n(t)$$

This leads to

$$y(t) = \sum_{n=1}^{N} x_n \left(\int_{-\infty}^{\infty} \psi_n(s)p(t-s)\,\mathrm{d}s \right)$$

we only observe finite set of samples of $y(t)$:

$$y_m := y(t_m) = \sum_{n=1}^{N} x_n \left(\int_{-\infty}^{\infty} \psi_n(s) p(t_m - s)\, \mathrm{d}s \right)$$

$$= \sum_n A[m,n] x_n$$

$$\text{where} \quad A[m,n] = \int_{-\infty}^{\infty} \psi_n(s) p(t_m - s)\, \mathrm{d}s = \langle \mathbf{p}_m, \boldsymbol{\psi}_n \rangle$$

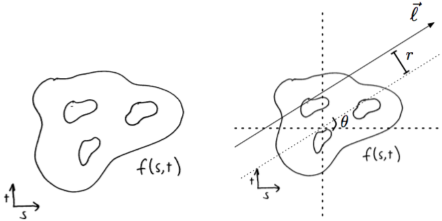
can write deconvolution problem as:

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

a solution can be synthesized using

$$f\hat{(t)} = \sum_{i=1}^{N} \hat{x}_n \psi_n(t)$$

Tomography: learn about the interior of an object while only taking measurements on the exterior

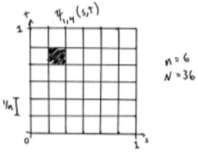


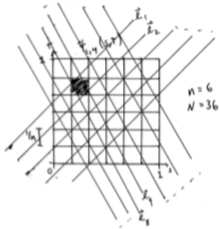
$$\mathcal{R}[\mathbf{f}] = \int f(s,t)\, dl$$

Computational Imaging

Example: Tomographic reconstruction

$$f(s,t) = \sum_{\gamma \in \Gamma} x_\gamma \psi_\gamma(s,t) \implies y_m = \mathcal{R}_{r_m,\theta_m}[f(s,t)]$$





Resulting problem is a linear IP:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad A[m,n] = \mathcal{R}_{r_m,\theta_m}[\Psi_n]$$

Outline

Least-Squares formulation

LS framework: find an $\mathbf{x}$ that minimizes length of residual

$$\mathbf{r} = \mathbf{y} - \mathbf{A}\mathbf{x}$$

solve an optimization problem

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

If $\mathbf{A}$ written using Singular value decomposition:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \quad \mathbf{U} \in \mathbb{R}^{M \times R}, \quad \boldsymbol{\Sigma} \in \mathbb{R}^{R \times R}, \quad \mathbf{V} \in \mathbb{R}^{N \times R}$$

Then the solution to least-squares problem is:

$$\mathbf{x}_{\text{ls}} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T\mathbf{y}$$

Least-squares solution

$$\mathbf{x}_{\mathrm{ls}} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{y}$$

When $\mathbf{y} = \mathbf{A}\mathbf{x}$ has

- exact solution: it must be $\mathbf{x}_{\mathrm{ls}}$.
- no exact solution: $\mathbf{x}_{\mathrm{ls}}$ is a solution to least-squares problem
- infinite solutions: $\mathbf{x}_{\mathrm{ls}}$ is the one with smallest norm.

Solution can be written in compact form:

$$\mathbf{x}_{\mathrm{ls}} = \mathbf{A}^{\dagger}\mathbf{y}$$

$\mathbf{A}^{\dagger}(= \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T)$ is called pseudo-inverse!

Pseudo-inverse

- $\mathbf{A}$ is a square matrix

$$\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$$

- $\mathbf{A}$ has full column rank

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$$

- $\mathbf{A}$ has full row rank

$$\mathbf{A}^{\dagger} = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}$$

- Otherwise (for low-rank matrix)

$$\mathbf{A}^{\dagger} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$$

Outline

Introduction

Least-squares

Reconstruction Methods

Examples

Summary

Stable Reconstructions

Two important methods:

- Truncated SVD
- Tikhonov regularization

Truncated SVD

The solution of $\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ is given by:

$$\mathbf{x}^\star = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T\mathbf{y} = \sum_{r=1}^{R} \frac{\mathbf{u_r}^T\mathbf{y}}{\sigma_r}\mathbf{v}_r = \sum_{r=1}^{R} \frac{\mathbf{u_r}^T(\mathbf{y}^{\text{tr}} + \boldsymbol{\delta})}{\sigma_r}\mathbf{v}_r$$

If $\sigma_r \to 0$, noise $\boldsymbol{\delta}$ affects the solution.

Truncate SVD:

- throw away contributions of $\sigma_r < \epsilon$.
- assuming $\sigma_1, \ldots, \sigma_K > \epsilon$ then $x_{\text{tsvd}} = \sum_{r=1}^{K} \frac{\mathbf{u_r}^T\mathbf{y}}{\sigma_r}\mathbf{v}_r$

Tikhonov regularization

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$$

solution is obtained by setting gradient to zero:

$$\mathbf{x}_{\text{tik}} = \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{y}$$

Tikhonov reconstruction in SVD form:

$$\mathbf{x}_{\text{tik}} = \sum_{r=1}^{R} \frac{\sigma_r}{\sigma_r^2 + \lambda}(\mathbf{u}_r^T\mathbf{y})\mathbf{v}_r$$

Generalized Tikhonov regularization:

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_2^2 \quad \Longrightarrow \quad \mathbf{x}_{\text{gen-tik}} = \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{A}^T\mathbf{y}$$

Iterative methods

Two types of methods:

- specifically for linear problems - Krylov subspace methods { Arnoldi, Lanczos, conjugate gradient (CG, BiCG, NLCG,etc), GMRES, IDR, ... }
- Convex optimization

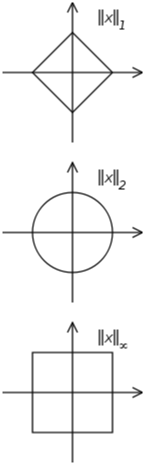
$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{s}^k$$

- $\mathbf{s}_k$ - descent direction
  - Gradient descent method: $\mathbf{s}_k = \nabla f$
  - Newton method: $\mathbf{s}_k = H(f)^{-1} \nabla f$
- α_k - step size, chosen from line search method

Sparse reconstruction

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- ℓ_2 regularization induces smoothness.
- ℓ_1 regularization induces sparsity, ℓ_1 norm in higher dimension is very pointy.
- The above problem is also known as LASSO in statistics.
- widely used in statistics, machine learning, signal processing.



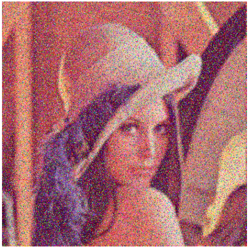
Outline

Image denoising

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{I}\mathbf{x}\|_2^2$$

True Image

Noisy Image



Tikhonov Reg

Image deblurring

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2$$

True Image

Blurred Image

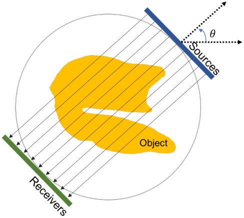
Sparse Reg

X-Ray Tomography

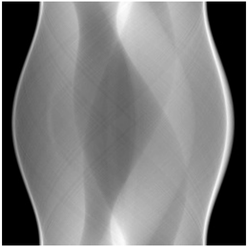
$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2$$

True Image

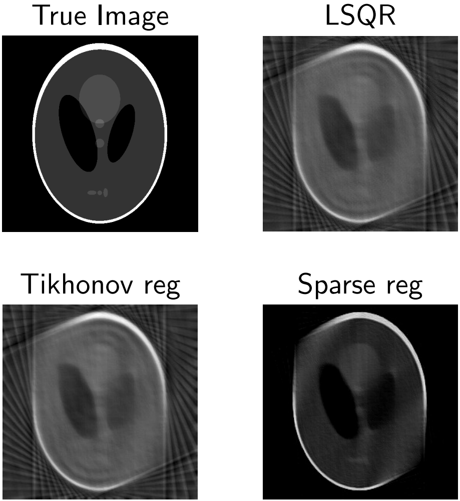
Projection



Data



X-Ray Tomography



Outline

Summary

- Inverse problems : an active field of research
- arises in many applications including computational imaging, machine learning, remote sensing, etc
- Least-squares is a popular choice for inversion.
- Stable reconstructions are important, and hence the regularization.
- Sparse reconstruction methods have gained popularity in last two decades.

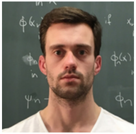
Thank you!

If interested in the topic, Join us in the journey!

Current Team Members:

Tristan van Leeuwen

Sarah Gaaf



Nick Luiken

Ajinkya Kadu

Opportunities:

- Undergraduate/Graduate Thesis
- Summer Research Project