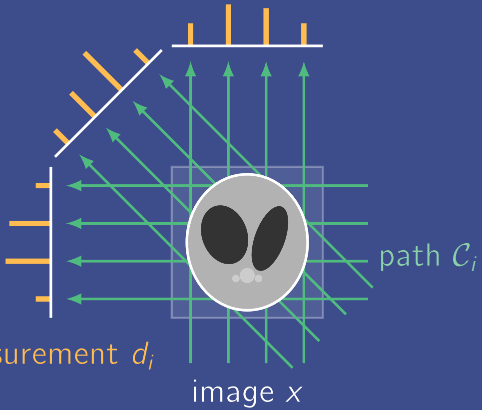
Why Inverse Problems are Hard

Given $d_i = \int_{\mathcal{C}_i} x \, ds$

Find x



Howard Heaton

Typal Academy

What is an Inverse Problem?

Key Terms

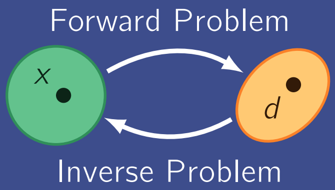
- forward operator G
- parameters x
- measurement data d
- mathematical model $G(x) = d$

Forward Problem

Use G to compute d from x

Inverse Problem

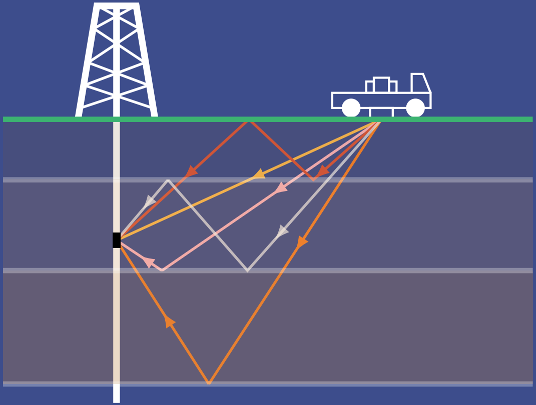
Use G to compute x from d



Examples

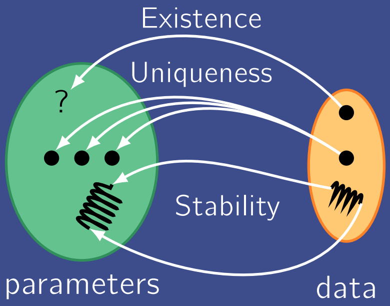
A short list of inverse problems:

- Super Resolution Imaging
- Vertical Seismic Profiling
- X-ray Computed Tomography
- Electrical Resistivity Tomography



(See appendix for more details)

Hadamard's Properties



- Existence – Problem has a solution
- Uniqueness – There is at most one solution
- Stability – Solution changes continuously with data

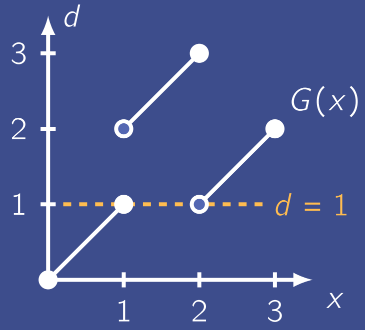
Well-Posed	$\Longleftrightarrow$	Above 3 Properties Hold
Ill-Posed	$\Longleftrightarrow$	Not Well-Posed

Toy Examples – Existence and Uniqueness

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}}_{\text{solution does not exist}} \qquad \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix}}_{\text{solutions are not unique}}$$

Toy Example – Stability

$$G(x) = \begin{cases} x & \text{if } 0 \le x \le 1 \\ x+1 & \text{if } 1 < x \le 2 \\ x-1 & \text{if } 2 < x \le 3 \end{cases}$$



Inverse problem is instable as G^{-1} is discontinuous at $d = 1$, *i.e.*

$$\lim_{\varepsilon \to 0^+} G^{-1}(1+\varepsilon) = \lim_{\varepsilon \to 0^+} 2 + \varepsilon = 2 \neq 1 = G^{-1}(1)$$

Condition Numbers

The relative condition number for a well-posed inverse problem is

$$\kappa(d) = \lim_{\delta \to 0^+} \sup_{||p|| \leq \delta} \frac{||G^{-1}(d+p) - G^{-1}(d)||}{||G^{-1}(d)||} \Bigg/ \frac{||p||}{||d||}$$

This is a limit of the supremum over all infinitesimal perturbations p

If G^{-1} is differentiable, then

$$\kappa(d) = \left\| \frac{\partial G^{-1}(d)}{\partial d} \right\| \cdot \frac{||d||}{||G^{-1}(d)||}$$

If G is linear, then $\kappa(d) = ||G|| \cdot ||G^{-1}|| \cdot ||d|| = \frac{\sigma_{max}(G)}{\sigma_{min}(G)} \cdot ||d||$

Ill-Conditioned Problems

Problem is well-conditioned provided $\kappa(d)$ is small and ill-conditioned otherwise

To illustrate, consider a well-posed linear problem with

$$G = \begin{bmatrix} 1 & 1 \\ 1 & 1+\varepsilon \end{bmatrix} \text{ and } d = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies \kappa(d) \approx \frac{4+\varepsilon}{\varepsilon} \cdot \sqrt{2} \text{ when } \varepsilon \text{ is small}$$

With small ε, small changes in d yield large changes in $x \implies$ ill-conditioned

$$G^{-1}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad G^{-1}\left(\begin{bmatrix} 1+\varepsilon \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2+\varepsilon \\ -1 \end{bmatrix}$$

Why Inverse Problems are Hard

Inverse problems in the wild are often ill-posed and/or ill-conditioned

To address this, practitioners reformulate the problem in various ways:

- No solution? Weaken hard constraint $G(x) = d$
- Not unique? Embed more prior knowledge (*e.g.* sparsity, minimal norm)
- Instable or ill-conditioned? Introduce more regularization or "convexify"

Regularization

Regularizers are used to define a well-posed and well-conditioned optimization problem as a surrogate for the hard inverse problem; namely, one assumes

$$x \approx \operatorname*{argmin}_z \big(\text{data misfit}\big) + \big(\text{regularizer}\big)$$

Data Misfit: Closeness measurement of $G(z)$ and d, *e.g.* $||G(z) - d||_2^2$

Regularizer: A regularizer can take the form of a Tikhonov term $||Lz||_2^2$, ℓ_1 norm $||z||_1$, total variation $||\nabla z||_1$, nonnegativity constraint, and more

Algorithmic Challenges

Distinct problems may require different algorithms to compute solutions,

which have varying characteristics with respect to meeting

- Time constraints (*e.g.* need to compute solutions in milliseconds, hours)
- Memory constraints (*e.g.* avoiding Hessian if x is high dimensional)
- Output tolerances* (*e.g.* $||G(x) - d|| \leq \varepsilon$)

*Algorithm may create a sequence of solution estimates that satisfy hard constraints of surrogate model for each estimate or only asymptotically

Example – Sparse Recovery

Consider recovery of sparse x from noisy linear measurements $d = Ax + \varepsilon$

Multiple optimization problems can used as surrogates:

$$x \approx \underset{z}{\operatorname{argmin}} \frac{1}{2}\|Az - d\|_2^2 + \lambda\|z\|_1 \qquad \text{(LASSO)}$$

$$x \approx \underset{z}{\operatorname{argmin}} \|z\|_1 \quad \text{s.t.} \quad \|Az - d\| \leq \delta \qquad \text{(BPDN)}$$

LASSO can be solved via proximal gradient and BPDN via ADMM

Surrogate model influences algorithm influences performance guarantees

Summary

Several factors can make inverse problems hard to solve

- Inverse problems are often ill-posed and/or ill-conditioned
- Well-posed surrogates must be carefully crafted with G and prior knowledge
- Various constraints may be imposed on algorithms for computing solutions

Appendix

Appendix – Super Resolution Imaging

Inverse Problem: Estimate high-resolution image from low-resolution image

Let x and d be vectorizations of the high and low resolution images, respectively

One simple model makes the estimate

$$x \approx \underset{z}{\text{argmin}} \; ||Dz - d||_2^2 + \lambda ||\nabla z||_1$$

with $\lambda > 0$ a regularization parameter, D a linear downsampling operator and ∇ a discrete differencing operator

Appendix – Vertical Seismic Profiling

Source of seismic waves is at surface and arrival times are measured in borehole

Travel time $u(x, z)$ of seismic wave to a point (x, z) satisfies Eikonal equation

$$\begin{cases} \|\nabla u\| = \frac{1}{v} & \text{in } \Omega \\ u = q & \text{on } \partial\Omega \end{cases}$$

with $\Omega \subset \mathbb{R}^2$ the subsurface region, $v(x, z)$ the seismic velocity and

$q(x, z)$ the travel time on the surface $\partial\Omega$

Inverse Problem: Estimate subsurface velocity v via observed travel times

Appendix – X-ray Computed Tomography

X-rays attentuations by different portions of an object are measured

$$(i\text{-th measurement}) = d_i = \int_{\mathcal{C}_i} x \, \mathrm{ds} \approx \sum_j A_{ij} x_j,$$

with x_j the value for the j-th voxel in a discretization of the continuous object

and A_{ij} the length of the i-th ray path through the j-th voxel

Inverse Problem: Recover image x from X-ray attentuation measurements d

For a tolerance $\varepsilon > 0$, one way to estimate x is via

$$x \approx \operatorname*{argmin}_{z \geq 0} \|\nabla z\|_1 \quad \text{s.t.} \quad \|Az - d\| \leq \varepsilon$$

Appendix – Electrical Resistivity Tomography

For subsurface region $\Omega \subset \mathbb{R}^3$ with conductivity $\sigma : \Omega \to \mathbb{R}$ and injected current I,

electric potential $\phi : \Omega \to \mathbb{R}$ satisfies

$$\begin{cases} \nabla \cdot \sigma \nabla \phi = I\delta & \text{in } \Omega \\ \nabla \phi \cdot n = 0 & \text{on } \partial\Omega \end{cases}$$

with n the normal vector at the surface $\partial\Omega$ and δ a Dirac delta function

Inverse Problem: Recover σ from surface measurements of potential difference