Numerical Methods for Bayesian Inverse Problems

Lecture 1: Inverse Problems

Robert Scheichl

Institute for Mathematics & IWR, Heidelberg University

Autumn School – "Uncertainty Quantification for High-Dimensional Problems"
CWI Amsterdam, October 7-11, 2024

(Thanks to Björn Sprungk, TU Freiberg)

(Rough) Outline of the Minicourse

- **Lecture 1:** Inverse Problems — Mon 13:30
- **Lecture 2:** Bayesian Approach to Inverse Problems — Mon 15:30
- **Lecture 3:** Markov Chain Monte Carlo — Tue 13:30
- **Lecture 4:** Wrap-up & Tutorial — Tue 15:30

Stochastic Modelling and Uncertainty Quantification

Stochastic Modelling

Many reasons for stochastic modelling:

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)

Stochastic Modelling

Many reasons for stochastic modelling:

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)

Input: best knowledge about system (differential eqn.), statistics of input parameters, measured ouput data with error statistics,. . .

Stochastic Modelling

Many reasons for stochastic modelling:

- **lack of data** (e.g. data assimilation for weather prediction)
- **data uncertainty** (e.g. uncertainty quantification in subsurface flow)
- **parameter identification** (e.g. Bayesian inference in engineering)
- **unresolvable scales** (e.g. atmospheric dispersion modelling)

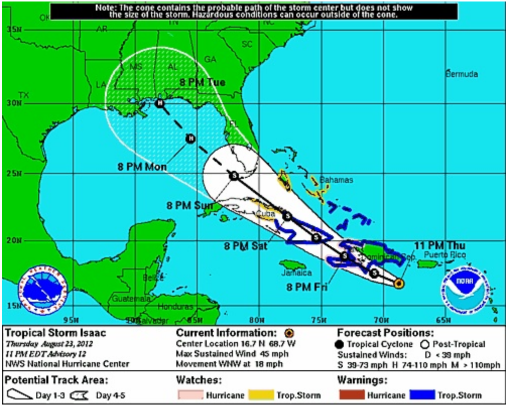
Input: best knowledge about system (differential eqn.), statistics of input parameters, measured ouput data with error statistics,...

Output: statistics of QoIs or of entire state space, e.g.

- Data misfit or rainfall at some location in numerical weather prediction
- Flow or 'breakthrough' time at a radioactive waste repository
- Amount of ash over Heathrow due to atmospheric dispersion
- Certification of carbon fibre composite wing in aeronautical engineering

But: Often data is very sparse/uncertain $\rightarrow$ Need good physical models!

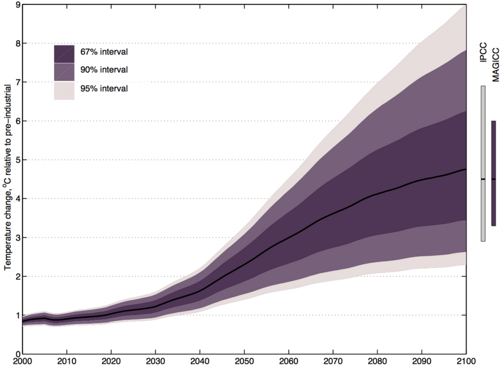
Examples



Source: National Hurricane Center, USA

Predicted storm path with uncertainty cones.

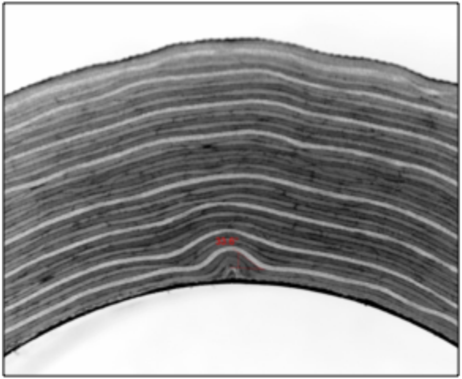
Examples



Source: Brodman & Karoly, 2013

Global-mean temperature change for a business-as-usual emission scenario, relative to pre-industrial. Black line: median, shaded regions 67% (dark), 90% (medium) and 95% (light) confidence intervals.

Examples



Source: GKN Aerospace

Performance "knock-down" factors through wrinkling defects in carbon fibre composite aeroplane wing

A model problem: Predator-prey dynamical system

Consider the popular **Lotka-Volterra** (or **predator-prey**) model of the dynamics of two interacting populations

$$\dot{\boldsymbol{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} \theta_1 u_1 \ - \ \theta_{12} u_1 u_2 \\ \theta_{21} u_1 u_2 \ - \ \theta_2 u_2 \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}), \quad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

where u_1 is the number of *prey*, u_2 is the number of *predator* and $\theta_1, \theta_2, \theta_{12}, \theta_{21} \geq 0$ are *parameters* describing the interaction of the two species.

A model problem: Predator-prey dynamical system

Consider the popular **Lotka-Volterra** (or **predator-prey**) model of the dynamics of two interacting populations

$$\dot{\boldsymbol{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} \theta_1 u_1 \; - \; \theta_{12} u_1 u_2 \\ \theta_{21} u_1 u_2 \; - \; \theta_2 u_2 \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}), \quad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

where u_1 is the number of *prey*, u_2 is the number of *predator* and $\theta_1, \theta_2, \theta_{12}, \theta_{21} \geq 0$ are *parameters* describing the interaction of the two species.

- For simplicity, we assume that

  $$\theta_1 = \theta_2 = \theta_{12} = \theta_{21} = 1$$

  and **only** the vector of initial conditions $\boldsymbol{u}_0$ is **uncertain**.
- Typically not interested in whole solution of system. Our quantity of interest (QoI) might be the number of prey $Q = u_1(T)$ at time T.

A model problem: Predator-prey dynamical system

Consider the popular **Lotka-Volterra** (or **predator-prey**) model of the dynamics of two interacting populations

$$\dot{\boldsymbol{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} \theta_1 u_1 \; - \; \theta_{12} u_1 u_2 \\ \theta_{21} u_1 u_2 \; - \; \theta_2 u_2 \end{bmatrix} = \boldsymbol{f}(\boldsymbol{u}), \quad \boldsymbol{u}(0) = \boldsymbol{u}_0,$$

where u_1 is the number of *prey*, u_2 is the number of *predator* and $\theta_1, \theta_2, \theta_{12}, \theta_{21} \geq 0$ are *parameters* describing the interaction of the two species.

- For simplicity, we assume that

$$\theta_1 = \theta_2 = \theta_{12} = \theta_{21} = 1$$

  and **only** the vector of initial conditions $\boldsymbol{u}_0$ is **uncertain**.

- Typically not interested in whole solution of system. Our quantity of interest (QoI) might be the number of prey $Q = u_1(T)$ at time T.

In what way can we address uncertainty in the initial condition $\boldsymbol{u}_0$?

Stochastic Modelling of Uncertainty

Could introduce δ-ball around given initial vector $\overline{\boldsymbol{u}}_0$ (in a suitable norm):

$$S := \{\boldsymbol{u}_0 \in \mathbb{R}^2 : \|\boldsymbol{u}_0 - \overline{\boldsymbol{u}}_0\| \leq \delta\}.$$

Stochastic Modelling of Uncertainty

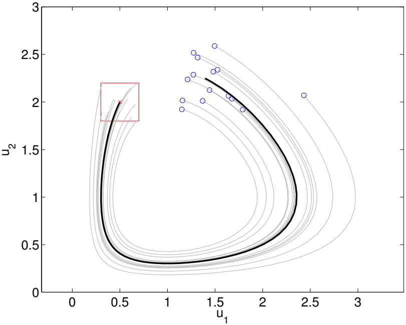
Could introduce δ-ball around given initial vector $\overline{\boldsymbol{u}}_0$ (in a suitable norm):

$$S := \{\boldsymbol{u}_0 \in \mathbb{R}^2 : \|\boldsymbol{u}_0 - \overline{\boldsymbol{u}}_0\| \leq \delta\}.$$

Worst case analysis: determine uncertainty interval

$$I = \left[\min_{\boldsymbol{u}_0 \in S} Q(\boldsymbol{u}_0), \max_{\boldsymbol{u}_0 \in S} Q(\boldsymbol{u}_0)\right].$$

The uncertainty range of $Q(\boldsymbol{u}_0) = u_1(T)$ is then the length of I.



Stochastic Modelling of Uncertainty

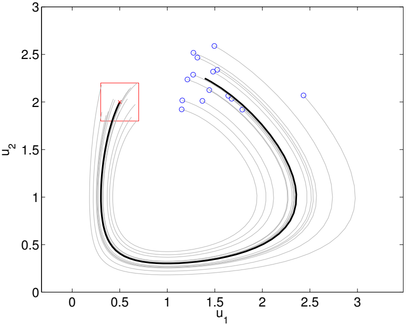
Could introduce δ-ball around given initial vector $\overline{\boldsymbol{u}}_0$ (in a suitable norm):

$$S := \{\boldsymbol{u}_0 \in \mathbb{R}^2 : \|\boldsymbol{u}_0 - \overline{\boldsymbol{u}}_0\| \leq \delta\}.$$

Worst case analysis: determine uncertainty interval

$$I = \left[\min_{\boldsymbol{u}_0 \in S} Q(\boldsymbol{u}_0), \max_{\boldsymbol{u}_0 \in S} Q(\boldsymbol{u}_0)\right].$$

The uncertainty range of $Q(\boldsymbol{u}_0) = u_1(T)$ is then the length of I.



But how to find (or bound) min and max? (especially in high dimensions)

And some initial conditions $\boldsymbol{u}_0 \in S$ are in general more likely than others.

Stochastic Modelling of Uncertainty

Probabilistic approach:

- Introduce a probability measure on S.
- The (measurable) map Q from S to the output set $\{Q(\boldsymbol{u}_0) : \boldsymbol{u}_0 \in S\}$ induces a probability distribution for $u_1(T)$ ("uncertainty propagation")
- Then we can calculate statistics of Q: (mean, variance, empirical CDF, ...)

$$\mathbb{E}\left[f(Q)\right] = \int_S f(Q(\boldsymbol{u}_0))\, \mathrm{d}\pi_{\boldsymbol{u}_0}$$

Stochastic Modelling of Uncertainty

Probabilistic approach:

- Introduce a probability measure on S.
- The (measurable) map Q from S to the output set $\{Q(\boldsymbol{u}_0) : \boldsymbol{u}_0 \in S\}$ induces a probability distribution for $u_1(T)$ ("uncertainty propagation")
- Then we can calculate statistics of Q: (mean, variance, empirical CDF, ...)

$$\mathbb{E}\left[f(Q)\right] = \int_S f(Q(\boldsymbol{u}_0))\,\mathsf{d}\pi_{\boldsymbol{u}_0}$$

- **Big issue:** choice of distribution, too much subjective information?
- Some classical guidelines: Laplace's principle of insufficient reason, maximum entropy, etc.

Stochastic Modelling of Uncertainty

Probabilistic approach:

- Introduce a probability measure on S.
- The (measurable) map Q from S to the output set $\{Q(\boldsymbol{u}_0) : \boldsymbol{u}_0 \in S\}$ induces a probability distribution for $u_1(T)$ ("uncertainty propagation")
- Then we can calculate statistics of Q: (mean, variance, empirical CDF, ...)

$$\mathbb{E}\left[f(Q)\right] = \int_S f(Q(\boldsymbol{u}_0))\, \mathrm{d}\pi_{\boldsymbol{u}_0}$$

- **Big issue:** choice of distribution, too much subjective information?
- Some classical guidelines: Laplace's principle of insufficient reason, maximum entropy, etc.

Instead, choosing distribution **conditioned on data** is the basis of Bayesian inference (genuine "uncertainty quantification")

Computational Challenges

Computing statistics of QoIs:

$$\mathbb{E}[f(Q)] = \int_S f(Q(\boldsymbol{u}_0))\,\mathsf{d}\pi_{\boldsymbol{u}_0}$$

- **High-Dimensional Quadrature**
  - Monte Carlo, Quasi-Monte Carlo, sparse grids, ...
  - Bayesian setting: Markov chain Monte Carlo

Computational Challenges

Computing statistics of QoIs:

$$\mathbb{E}[f(Q)] = \int_S f(Q(\boldsymbol{u}_0))\, \mathsf{d}\pi_{\boldsymbol{u}_0}$$

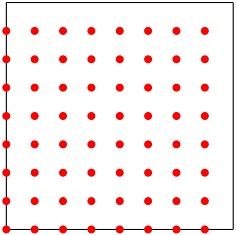
- **High-Dimensional Quadrature**
  - Monte Carlo, Quasi-Monte Carlo, sparse grids, ...
  - Bayesian setting: Markov chain Monte Carlo
- **Solve** large number of **complex** dynamical systems:
  - Numerical solver (e.g., Euler with step size h) $\rightarrow$ high cost per sample
  - Significantly more difficult for spatio-temporal models and PDEs

High-Dimensional Quadrature

Quadrature rules and the **curse of dimensionality**:

$$\int_{[0,1]^d} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{j=0}^{N-1} w_j f(\boldsymbol{t}_j) =: Q_{N,d}(f)$$

Product rules



error $\sim \mathcal{O}(N^{-k/d})$

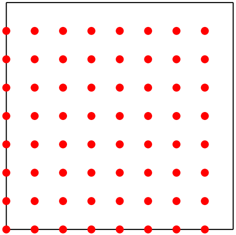
Example: In d=100 dimensions a rule using two points per dimension needs
$2^{100} \approx 1,000,000,000,000,000,000,000,000,000,000$ function evaluations!

High-Dimensional Quadrature

Quadrature rules and the **curse of dimensionality**:

$$\int_{[0,1]^d} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{j=0}^{N-1} w_j f(\boldsymbol{t}_j) =: Q_{N,d}(f)$$

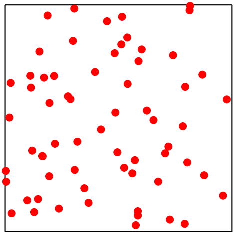
Product rules



error $\sim \mathcal{O}(N^{-k/d})$

Improved w. **"sparse grids"**

Monte Carlo



error $\sim \mathcal{O}(N^{-1/2})$

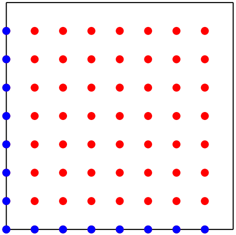
independent of d!

High-Dimensional Quadrature

Quadrature rules and the **curse of dimensionality**:

$$\int_{[0,1]^d} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{j=0}^{N-1} w_j f(\boldsymbol{t}_j) =: Q_{N,d}(f)$$

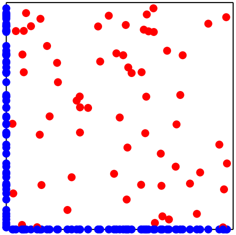
Product rules



error $\sim \mathcal{O}(N^{-k/d})$

Improved w. **"sparse grids"**

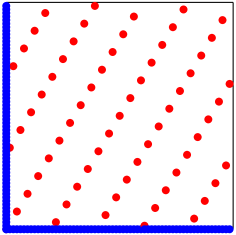
Monte Carlo



error $\sim \mathcal{O}(N^{-1/2})$

independent of d!

Quasi–Monte Carlo



error $\sim \mathcal{O}(N^{-1+\delta})$

independent of d!

What is an Inverse Problem?

Inverse Problem

You can't see the cause, but you can observe the effect.

Inverse Problem

Physical Model

$$y = \mathcal{G}(u)$$

- $u \in \mathcal{X}$ parameter vector / parameter function
- $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ forward response operator
- $y \in \mathcal{Y}$ result / observations
- Evaluation of $\mathcal{G}$ usually expensive (e.g., solving PDE)

Inverse Problem

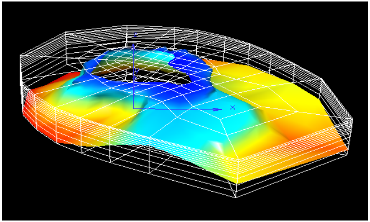
Physical Model

$$y = \mathcal{G}(u)$$

- $u \in \mathcal{X}$ parameter vector / parameter function
- $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ forward response operator
- $y \in \mathcal{Y}$ result / observations
- Evaluation of $\mathcal{G}$ usually expensive (e.g., solving PDE)

Data y

Parameter u



Inverse Problem

Physical Model

$$y = \mathcal{G}(u)$$

- $u \in \mathcal{X}$ parameter vector / parameter function
- $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ forward response operator
- $y \in \mathcal{Y}$ result / observations
- Evaluation of $\mathcal{G}$ usually expensive (e.g., solving PDE)

Forward Problem

Compute/predict output $y \in \mathcal{Y}$ given parameter $u \in \mathcal{X}$

$\to$ **well-posed**

Inverse Problem

Find parameter $u \in \mathcal{X}$ given observations $y \in \mathcal{Y}$

$\to$ **ill-posed**

Inverse Problem

Observational model

$$y = \mathcal{G}(u) + \eta$$

- $u \in \mathcal{X}$ parameter vector / parameter function
- $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$ forward response operator
- $y \in \mathcal{Y}$ result / observations
- Evaluation of $\mathcal{G}$ usually expensive (e.g., solving PDE)
- $\eta \in \mathcal{Y}$ observational noise

Forward Problem

Compute/predict output $y \in \mathcal{Y}$ given parameter $u \in \mathcal{X}$

$\to$ **well-posed**

Inverse Problem

Find parameter $u \in \mathcal{X}$ given (noisy) observations $y \in \mathcal{Y}$

$\to$ **ill-posed**

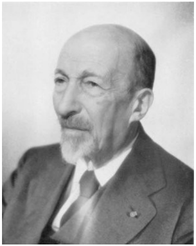
What does 'well-posed' mean?

Well-posedness

A mathematical problem is **well-posed** if

1. it has a solution,
2. this solution is unique
3. and it depends continuosly on the (input) data.

A problem which is not well-posed is called **ill-posed**.



Jacques Hadamard
(1865 – 1963)

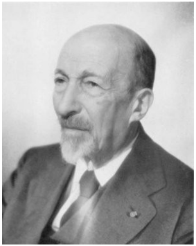
What does 'well-posed' mean?

Well-posedness

A mathematical problem is **well-posed** if

1. it has a solution,
2. this solution is unique
3. and it depends continuosly on the (input) data.

A problem which is not well-posed is called **ill-posed**.



Jacques Hadamard
(1865 – 1963)

Forward Problem

Given u compute $y = \mathcal{G}(u)$: In general

1. $\mathcal{G}(u)$ exists,
2. $\mathcal{G}(u)$ is unique,
3. if $\mathcal{G} \in C(\mathcal{X}, \mathcal{Y})$ it depends continuously on u.

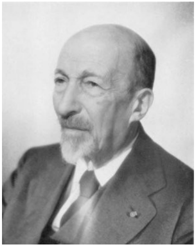
What does 'well-posed' mean?

Well-posedness

A mathematical problem is **well-posed** if

1. it has a solution,
2. this solution is unique
3. and it depends continuosly on the (input) data.

A problem which is not well-posed is called **ill-posed**.



Jacques Hadamard
(1865 – 1963)

Forward Problem

Given u compute $y = \mathcal{G}(u)$: In general

1. $\mathcal{G}(u)$ exists,
2. $\mathcal{G}(u)$ is unique,
3. if $\mathcal{G} \in C(\mathcal{X}, \mathcal{Y})$ it depends continuously on u.

Inverse Problem

Given $y = \mathcal{G}(u) + \eta$ compute u: In general

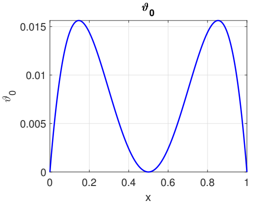
1. $y \notin \mathcal{G}(\mathcal{X})$
2. $\mathcal{G} \colon \mathcal{X} \to \mathcal{Y}$ not injective
3. $\mathcal{G}^{-1} \colon \mathcal{Y} \to \mathcal{X}$ not continuous (even if it exists and is unique!)

Classical Example

Consider rod with thermal conductivity & length one.
Temperature distribution $\theta(x,t)$ satisfies heat equation

$$\frac{\partial \theta}{\partial t} - \frac{\partial^2 \theta}{\partial x^2} = 0, \qquad 0 < x < 1,\ t > 0$$

with boundary condition $\theta(0,t) = \theta(1,t) = 0$ and
initial data $\theta(x,0) = u(x)$.



Classical Example

Consider rod with thermal conductivity & length one. Temperature distribution $\theta(x,t)$ satisfies heat equation

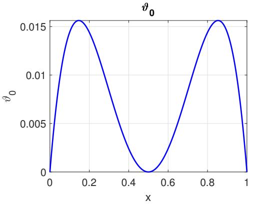
$$\frac{\partial \theta}{\partial t} - \frac{\partial^2 \theta}{\partial x^2} = 0, \qquad 0 < x < 1,\ t > 0$$

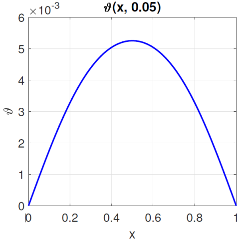
with boundary condition $\theta(0,t) = \theta(1,t) = 0$ and initial data $\theta(x,0) = u(x)$.

Forward map $\mathcal{G}$: Solution can be expressed in terms of its Fourier components

$$\theta(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \ ,$$

with c_n the Fourier coeffs of $u(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.



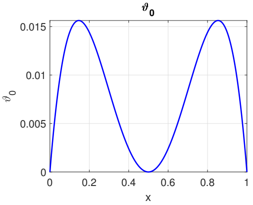


Classical Example

Consider rod with thermal conductivity & length one. Temperature distribution $\theta(x,t)$ satisfies heat equation

$$\frac{\partial \theta}{\partial t} - \frac{\partial^2 \theta}{\partial x^2} = 0, \qquad 0 < x < 1,\ t > 0$$

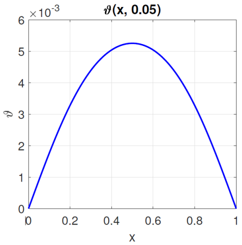
with boundary condition $\theta(0,t) = \theta(1,t) = 0$ and initial data $\theta(x,0) = u(x)$.



Forward map $\mathcal{G}$: Solution can be expressed in terms of its Fourier components

$$\theta(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \ ,$$

with c_n the Fourier coeffs of $u(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.



Inverse problem: Given temperature distribution $\theta(\cdot,T)$ at time $T > 0$ find the initial temperature distribution u?

Classical Example

Assume that we have two initial states, $u^{(1)},\ u^{(2)}$ differing only by a single high frequency component

$$u^{(1)}(x) - u^{(2)}(x) = c_N \sin(N\pi x)$$

for some $N \in \mathbb{N}$.

Classical Example

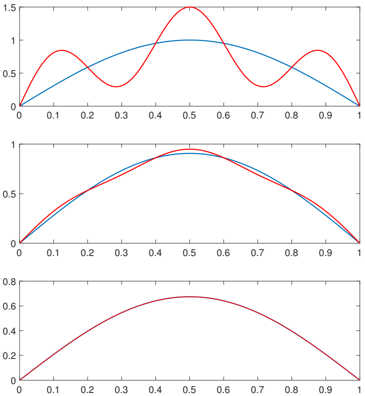
Assume that we have two initial states, $u^{(1)}$, $u^{(2)}$ differing only by a single high frequency component

$$u^{(1)}(x) - u^{(2)}(x) = c_N \sin(N\pi x)$$

for some $N \in \mathbb{N}$.

At time T, solutions will differ by

$$\theta^{(1)}(x,T) - \theta^{(2)}(x,T) = c_N e^{-(N\pi)^2 T} \sin(N\pi x) \, .$$



$\theta^{(1)}$ (red) and $\theta^{(2)}$ (blue) at times $t = 0, 1$ and 4.

Classical Example

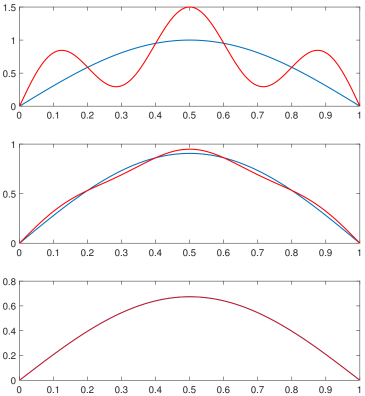
Assume that we have two initial states, $u^{(1)}$, $u^{(2)}$ differing only by a single high frequency component

$$u^{(1)}(x) - u^{(2)}(x) = c_N \sin(N\pi x)$$

for some $N \in \mathbb{N}$.

At time T, solutions will differ by

$$\theta^{(1)}(x,T) - \theta^{(2)}(x,T) = c_N e^{-(N\pi)^2 T} \sin(N\pi x) \; .$$



$\theta^{(1)}$ (red) and $\theta^{(2)}$ (blue) at times $t = 0, 1$ and 4.

Any information about high-frequency components lost in the presence of measurement errors

Classical Example

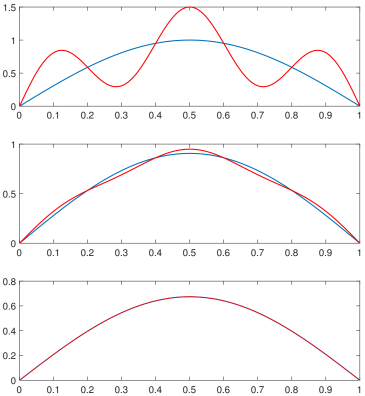
Assume that we have two initial states, $u^{(1)}$, $u^{(2)}$ differing only by a single high frequency component

$$u^{(1)}(x) - u^{(2)}(x) = c_N \sin(N\pi x)$$

for some $N \in \mathbb{N}$.

At time T, solutions will differ by

$$\theta^{(1)}(x,T) - \theta^{(2)}(x,T) = c_N e^{-(N\pi)^2 T} \sin(N\pi x) \, .$$

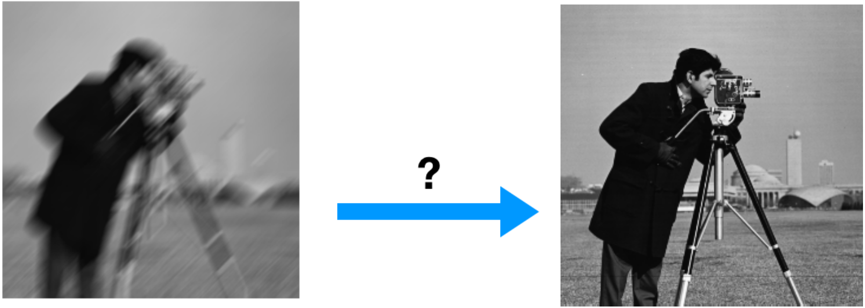


$\theta^{(1)}$ (red) and $\theta^{(2)}$ (blue) at times $t = 0, 1$ and 4.

Any information about high-frequency components lost in the presence of measurement errors and $\mathcal{G}^{-1}$ not continuous due to smoothing effect of $\mathcal{G}$.

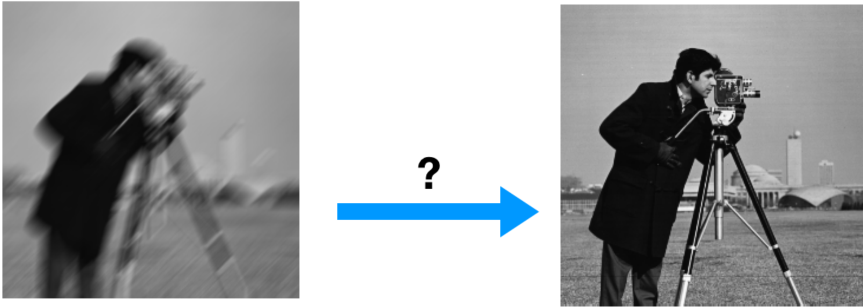
Other Classical Examples: Imaging

Goal: Reconstruct an image u given a noisy, blurred or partial observation y



Other Classical Examples: Imaging

Goal: Reconstruct an image u given a noisy, blurred or partial observation y



- Unknown $u \in \mathbb{R}^{d_u}$: the pixel values of the image
- Forward map $\mathcal{G}$: The map $\mathcal{G} = G \in \mathbb{R}^{d_y \times d_u}$ is linear in many imaging problems, where G incorporates mechanisms such as
  - blurring ($\rightarrow$ averaging over a neighbourhood of pixels)
  - Fourier transform ($\rightarrow$ observations in frequency domain)
  - a mask ($\rightarrow$ partial observations)

Other Classical Examples: Imaging

Particular example: Deblurring (without noise).

Map G defined by its action on each pixel value:

$$y_i = \sum_{j \in N(i)} w_{ij} u_j,$$

where

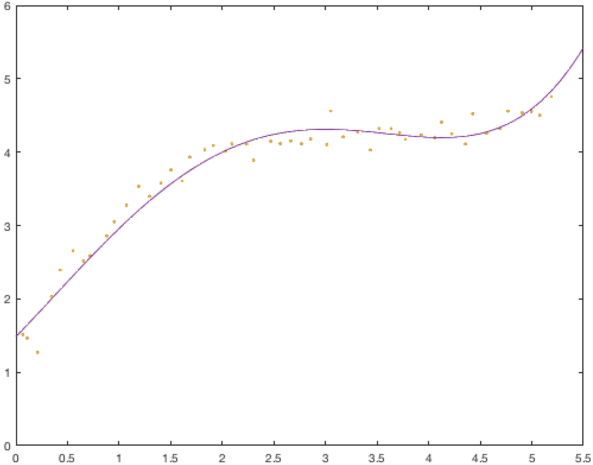
- u_j is jth pixel value of the original image.
- y_i is ith pixel value of the blurred image,
- $N(i)$ is neighbourhood of pixel i,
- w_{ij} are weights indicating the contribution of original pixel $j \in N(i)$ to the 'blur' in pixel i.

Original

Blurred

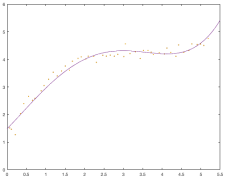
Other Classical Examples: Regression

Goal: Reconstruct a function f given noisy point values $\{f(x_i)\}$.



Other Classical Examples: Regression

Goal: Reconstruct a function f given noisy point values $\{f(x_i)\}$.



- Unknown $u \in \mathbb{R}^{d_u}$: Coefficients in a basis expansion

$$f(x;u) = \sum_{j=1}^{d_u} u_j \phi_j(x),$$

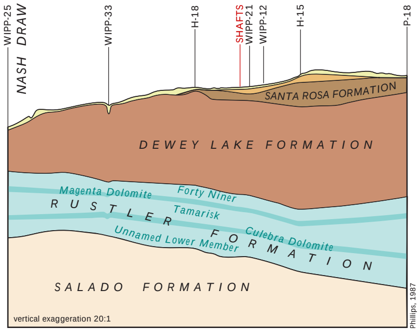
where $\{\phi_j\}_{j=1}^{d_u}$ are linearly independent, e.g. $\phi_j(x) = x^{j-1}$.

- Forward map $\mathcal{G}$: Implicitly defined by $u \mapsto \{f(x_i;u)\}_{i=1}^{d_y}$. In fact, $\mathcal{G} = G$ linear with $a_{ij} = \phi_j(x_i)$ (**Vandermonde-type matrix**).

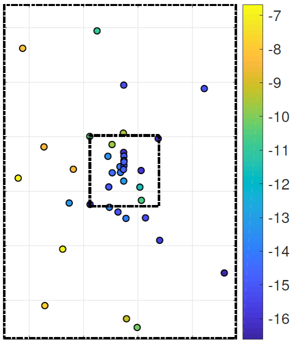
- Observations: $y \in \mathbb{R}^{d_y}$ with $y_i = f(x_i;u) + \eta_i, \; i = 1, \ldots, d_y$.

A PDE Example: Porous Media Flow

Goal: Reconstruct the hydraulic conductivity k of the subsurface given noisy measurements of the water pressure $\{p(x_i)\}$.



Cross-section at WIPP



A PDE Example: Porous Media Flow

Goal: Reconstruct the hydraulic conductivity k of the subsurface given noisy measurements of the water pressure $\{p(x_i)\}$.

- Unknown $u \in \mathbb{R}^{d_u}$: Coefficients in a basis expansion

$$k(x;u) = \phi_0(x) + \sum_{j=1}^{d_u} u_j \phi_j(x),$$

  where $\{\phi_j\}_{j=1}^{d_u}$ are linearly independent and ϕ_0 is s.t. k is (strictly) positive.

- Forward map $\mathcal{G}$: Implicitly defined by $u \mapsto \{p(x_i;u)\}_{i=1}^{d_y}$, where p is the solution of

$$-\nabla \cdot (k(x;u)\nabla p(x;u)) = h(x).$$

  (This equation comes from Darcy's law plus conservation of mass: conductivity k, pressure head p, sources/sinks h.)

- Observations: $y \in \mathbb{R}^{d_y}$ with $y_i = p(x_i;u) + \eta_i,\ i = 1, \ldots, d_y$.

Deterministic Approach to Inverse Problems

Deterministic Approach to Inverse Problems

Find the unknown data $u \in \mathcal{X}$ from noisy observations

$$y = Au + \eta \in \mathcal{Y}, \qquad A \in \mathcal{L}(\mathcal{X}, \mathcal{Y}),$$

where $\mathcal{X}, \mathcal{Y}$ are (separable) Hilbert spaces and $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ bounded, linear operators.

Deterministic Approach to Inverse Problems

Find the unknown data $u \in \mathcal{X}$ from noisy observations

$$y = Au + \eta \in \mathcal{Y}, \qquad A \in \mathcal{L}(\mathcal{X}, \mathcal{Y}),$$

where $\mathcal{X}, \mathcal{Y}$ are (separable) Hilbert spaces and $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ bounded, linear operators.

Least-Squares Solution (Existence)

In order to cope with the problem that the data $y \in \mathcal{Y}$ does, in general, not belong to the range $\mathcal{R}(A)$, we do not seek u such that $Au = y$, but rather ask for

$$u \in \operatorname*{argmin}_u \tfrac{1}{2}\|y - Au\|^2.$$

Deterministic Approach to Inverse Problems

Find the unknown data $u \in \mathcal{X}$ from noisy observations

$$y = Au + \eta \in \mathcal{Y}, \qquad A \in \mathcal{L}(\mathcal{X}, \mathcal{Y}),$$

where $\mathcal{X}, \mathcal{Y}$ are (separable) Hilbert spaces and $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ bounded, linear operators.

Least-Squares Solution (Existence)

In order to cope with the problem that the data $y \in \mathcal{Y}$ does, in general, not belong to the range $\mathcal{R}(A)$, we do not seek u such that $Au = y$, but rather ask for

$$u \in \operatorname*{argmin}_u \tfrac{1}{2}\|y - Au\|^2.$$

Minimal-Norm Solution (Uniqueness)

In order to avoide multiple solutions we further define, if existing,

$$u^\dagger = \operatorname{argmin}\left\{\|u\| : u \in \operatorname{argmin}_u \tfrac{1}{2}\|y - Au\|^2\right\}.$$

Remarks

- The set of least-squares solutions $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ is convex.
- Therefore, if the set is non-empty, the minimal-norm solution $u^\dagger$ is unique.
- However, the set $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ can be empty if $\mathcal{R}(A)$ is not closed.

Remarks

- The set of least-squares solutions $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ is convex.
- Therefore, if the set is non-empty, the minimal-norm solution $u^\dagger$ is unique.
- However, the set $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ can be empty if $\mathcal{R}(A)$ is not closed.
- If the data y belongs to $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, then $u^\dagger$ exists and is given by $u^\dagger = A^\dagger y$ or, equivalently

$$u^\dagger = (A^* A)^\dagger A^* y$$

  where $A^\dagger$ denotes the **Moore-Penrose inverse** of A (and A^* the adjoint).
- Thus, the concept of minimal-norm solutions allows to satisfy, at least, the **first two conditions for well-posedness** (under suitable assumptions).

Remarks

- The set of least-squares solutions $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ is convex.
- Therefore, if the set is non-empty, the minimal-norm solution $u^\dagger$ is unique.
- However, the set $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ can be empty if $\mathcal{R}(A)$ is not closed.
- If the data y belongs to $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, then $u^\dagger$ exists and is given by $u^\dagger = A^\dagger y$ or, equivalently

  $$u^\dagger = (A^*A)^\dagger A^*y$$

  where $A^\dagger$ denotes the **Moore-Penrose inverse** of A (and A^* the adjoint).
- Thus, the concept of minimal-norm solutions allows to satisfy, at least, the **first two conditions for well-posedness** (under suitable assumptions).
- What about the **third one (continuos dependence)?**

Remarks

- The set of least-squares solutions $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ is convex.
- Therefore, if the set is non-empty, the minimal-norm solution $u^\dagger$ is unique.
- However, the set $\operatorname{argmin}_u \frac{1}{2}\|y - Au\|^2$ can be empty if $\mathcal{R}(A)$ is not closed.
- If the data y belongs to $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, then $u^\dagger$ exists and is given by $u^\dagger = A^\dagger y$ or, equivalently

$$u^\dagger = (A^* A)^\dagger A^* y$$

  where $A^\dagger$ denotes the **Moore-Penrose inverse** of A (and A^* the adjoint).
- Thus, the concept of minimal-norm solutions allows to satisfy, at least, the **first two conditions for well-posedness** (under suitable assumptions).
- What about the **third one (continuos dependence)?**
- Focus in the following only on **compact** $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ (smoothing effect)

Compact operators

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is called **compact** if for any bounded sequence $(u_k)_{k\in\mathbb{N}} \subset \mathcal{X}$ the sequence of images $(Au_k)_{k\in\mathbb{N}}$ has a converging subsequence.

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is compact if and only if there exists

(i) a decaying sequence $\lambda_k \to 0$, $k \in \mathbb{N}$ of singular values

(ii) a complete ONS $(v_k)_{k\in\mathbb{N}}$ of $\overline{\mathcal{R}(A)} \subseteq \mathcal{Y}$,

(iii) a complete ONS $(u_k)_{k\in\mathbb{N}}$ of $\mathcal{N}(A)^\perp \subseteq \mathcal{X}$ with $\mathcal{N}(A)$ the null space of A,

Compact operators

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is called **compact** if for any bounded sequence $(u_k)_{k\in\mathbb{N}} \subset \mathcal{X}$ the sequence of images $(Au_k)_{k\in\mathbb{N}}$ has a converging subsequence.

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is compact if and only if there exists

(i) a decaying sequence $\lambda_k \to 0$, $k \in \mathbb{N}$ of singular values

(ii) a complete ONS $(v_k)_{k\in\mathbb{N}}$ of $\overline{\mathcal{R}(A)} \subseteq \mathcal{Y}$,

(iii) a complete ONS $(u_k)_{k\in\mathbb{N}}$ of $\mathcal{N}(A)^\perp \subseteq \mathcal{X}$ with $\mathcal{N}(A)$ the null space of A,

sucht that $Au_k = \lambda_k v_k$ for each $k \in \mathbb{N}$ and

$$Au = \sum_{k\in\mathbb{N}} \lambda_k \langle u, u_k\rangle v_k \qquad \forall u \in \mathcal{X}.$$

Compact operators

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is called **compact** if for any bounded sequence $(u_k)_{k\in\mathbb{N}} \subset \mathcal{X}$ the sequence of images $(Au_k)_{k\in\mathbb{N}}$ has a converging subsequence.

A linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ is compact if and only if there exists

(i) a decaying sequence $\lambda_k \to 0$, $k \in \mathbb{N}$ of singular values

(ii) a complete ONS $(v_k)_{k\in\mathbb{N}}$ of $\overline{\mathcal{R}(A)} \subseteq \mathcal{Y}$,

(iii) a complete ONS $(u_k)_{k\in\mathbb{N}}$ of $\mathcal{N}(A)^\perp \subseteq \mathcal{X}$ with $\mathcal{N}(A)$ the null space of A,

sucht that $Au_k = \lambda_k v_k$ for each $k \in \mathbb{N}$ and

$$Au = \sum_{k\in\mathbb{N}} \lambda_k \langle u, u_k \rangle v_k \qquad \forall u \in \mathcal{X}.$$

We then call $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$ the **singular system** of A.

Examples

Heat equation

Let $A_t \colon \mathcal{X} \to \mathcal{Y}$, $\mathcal{X}, \mathcal{Y} \subseteq L^2(D)$, $D = [0,1]$, be the solution operator for the heat equation

$$\frac{\partial \theta}{\partial t} - \frac{\partial^2 \theta}{\partial x^2} = 0, \qquad \theta(0,t) = \theta(1,t) = 0,$$

which maps the inital data $\theta(x,0) = u(x)$ to the solution $\theta(\cdot, t)$ at time $t > 0$, i.e.,

$$A_t u(x) = \sum_{k=1}^{\infty} c_k \mathrm{e}^{-(k\pi)^2 t} \sin(k\pi x), \qquad u(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

with

$$u_k = v_k = \sin(k\pi x), \qquad \lambda_k = \mathrm{e}^{-(k\pi)^2 t}.$$

The operator A_t is compact.

Examples

Integral operators

Another classical example for compact operators are integral operators $A\colon L^2(D) \to L^2(D)$ given by

$$Au(x) = \int_D c(x,y)\; u(x')\,\mathrm{d}x'$$

where $D \subseteq \mathbb{R}^d$ and $c \in L^2(D \times D)$ is a given kernel function. A common kernel is, e.g., the Gaussian kernel

$$c(x,x') := \sigma^2 \exp\left(-\tfrac{1}{2}|x-x'|^2\right)$$

which is used for image blurring (continuous setting).

Inverse problems and compact operators

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be compact with singular system $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$. Then, the inverse problem for $y = Au + \eta$ admits a minimum-norm solution, if and only if the data y satisfies the Picard condition

$$\sum_{k\in\mathbb{N}} \frac{|\langle y, v_k\rangle|^2}{\lambda_k^2} < \infty$$

Inverse problems and compact operators

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be compact with singular system $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$. Then, the inverse problem for $y = Au + \eta$ admits a minimum-norm solution, if and only if the data y satisfies the Picard condition

$$\sum_{k\in\mathbb{N}} \frac{|\langle y, v_k\rangle|^2}{\lambda_k^2} < \infty$$

in which case the minimum-norm solution is given by

$$u^\dagger = \sum_{k\in\mathbb{N}} \frac{1}{\lambda_k} \langle y, v_k\rangle \, u_k \, .$$

Inverse problems and compact operators

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be compact with singular system $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$. Then, the inverse problem for $y = Au + \eta$ admits a minimum-norm solution, if and only if the data y satisfies the Picard condition

$$\sum_{k\in\mathbb{N}} \frac{|\langle y, v_k\rangle|^2}{\lambda_k^2} < \infty$$

in which case the minimum-norm solution is given by

$$u^\dagger = \sum_{k\in\mathbb{N}} \frac{1}{\lambda_k} \langle y, v_k\rangle \, u_k \, .$$

Example: For the heat equation inverse problem the Picard condition implies that we have to observe very smooth data $y \in L^2(D)$ (**exponentially ill-posed**),

$$\sum_{k\in\mathbb{N}} \mathrm{e}^{2(k\pi)^2 t} \, y_k^2 < \infty, \qquad y_k := \int_0^1 y(x) \; \sin(k\pi x) \; \mathrm{d}x.$$

Instability of minimum-norm solutions

- The previous theorem also implies the lack of continuous dependence of $u^\dagger$ on the data y.
- Suppose, we are given data y and perturbed data $y^{(n)}$, $n \in \mathbb{N}$, with y satisfying the Picard condition and $y^{(n)}$ given by

$$y^{(n)} = y + \sqrt{\lambda_n} v_n$$

where $(\lambda_k, u_k, v_k)_{k \in \mathbb{N}}$ denotes singular system of $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$.

Instability of minimum-norm solutions

- The previous theorem also implies the lack of continuous dependence of $u^\dagger$ on the data y.
- Suppose, we are given data y and perturbed data $y^{(n)}$, $n \in \mathbb{N}$, with y satisfying the Picard condition and $y^{(n)}$ given by

$$y^{(n)} = y + \sqrt{\lambda_n} v_n$$

  where $(\lambda_k, u_k, v_k)_{k \in \mathbb{N}}$ denotes singular system of $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$.
- Then, $y^{(n)}$ also satisfies the Picard condition and

$$\left\| y - y^{(n)} \right\| = \sqrt{\lambda_n} \to 0 \qquad \text{as} \;\; n \to \infty$$

Instability of minimum-norm solutions

- The previous theorem also implies the lack of continuous dependence of $u^\dagger$ on the data y.
- Suppose, we are given data y and perturbed data $y^{(n)}$, $n \in \mathbb{N}$, with y satisfying the Picard condition and $y^{(n)}$ given by

$$y^{(n)} = y + \sqrt{\lambda_n} v_n$$

where $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$ denotes singular system of $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$.

- Then, $y^{(n)}$ also satisfies the Picard condition and

$$\left\| y - y^{(n)} \right\| = \sqrt{\lambda_n} \to 0 \qquad \text{as} \quad n \to \infty$$

but for the respective minimum-norm solutions $u^\dagger$ and $u^{(n)}$ we have

$$\left\| u^\dagger - u^{(n)} \right\| = \frac{\sqrt{\lambda_n}}{\lambda_n} = \frac{1}{\sqrt{\lambda_n}} \to \infty \qquad \text{as} \quad n \to \infty.$$

Instability of minimum-norm solutions

- The previous theorem also implies the lack of continuous dependence of $u^\dagger$ on the data y.
- Suppose, we are given data y and perturbed data $y^{(n)}$, $n \in \mathbb{N}$, with y satisfying the Picard condition and $y^{(n)}$ given by

$$y^{(n)} = y + \sqrt{\lambda_n} v_n$$

  where $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$ denotes singular system of $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$.
- Then, $y^{(n)}$ also satisfies the Picard condition and

$$\left\| y - y^{(n)} \right\| = \sqrt{\lambda_n} \to 0 \qquad \text{as} \quad n \to \infty$$

  but for the respective minimum-norm solutions $u^\dagger$ and $u^{(n)}$ we have

$$\left\| u^\dagger - u^{(n)} \right\| = \frac{\sqrt{\lambda_n}}{\lambda_n} = \frac{1}{\sqrt{\lambda_n}} \to \infty \qquad \text{as} \quad n \to \infty.$$

- **Inverse problems with minimum-norm solution still not well-posed!**

Regularization of Inverse Problems

Tikhonov Regularization

In order to "stabilize" the least-squares approach to inverse problems we apply **regularization** via a regularizing functional $R\colon \mathcal{X} \to [0,\infty]$

$$u \in \operatorname*{argmin}_u \frac{1}{2}\|y - Au\|^2 + R(u).$$

Tikhonov Regularization

In order to "stabilize" the least-squares approach to inverse problems we apply **regularization** via a regularizing functional $R\colon \mathcal{X} \to [0, \infty]$

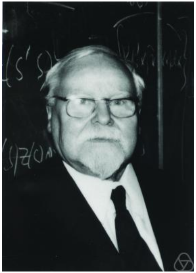
$$u \in \operatorname*{argmin}_u \frac{1}{2}\|y - Au\|^2 + R(u).$$

Tikhonov regularization

The **Tikhonov-regularized solution** to the inverse problem for $y = Au + \eta$ is defined by

$$u^\alpha = \operatorname*{argmin}_{u \in \mathcal{X}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2$$

where $\alpha > 0$ denotes a **regularization parameter**.



Andrey N. Tikhonov
(1906 – 1993)

Properties of Tikhonov Regularization

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then for any $y \in \mathcal{Y}$ and any $\alpha > 0$ there exists a unique Tikhonov-regularized solution

$$u^\alpha = \underset{u \in \mathcal{X}}{\operatorname{argmin}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2 = (A^*A + 2\alpha I)^{-1} A^* y.$$

Properties of Tikhonov Regularization

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then for any $y \in \mathcal{Y}$ and any $\alpha > 0$ there exists a unique Tikhonov-regularized solution

$$u^\alpha = \operatorname*{argmin}_{u \in \mathcal{X}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2 = (A^*A + 2\alpha I)^{-1}A^*y.$$

In particular, if A is compact with singular system $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$, then

$$u^\alpha = \sum_{k\in\mathbb{N}} \frac{\lambda_k}{\lambda_k^2 + \alpha} \langle y, v_n \rangle \, u_n$$

Properties of Tikhonov Regularization

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then for any $y \in \mathcal{Y}$ and any $\alpha > 0$ there exists a unique Tikhonov-regularized solution

$$u^\alpha = \operatorname*{argmin}_{u \in \mathcal{X}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2 = (A^*A + 2\alpha I)^{-1} A^* y.$$

In particular, if A is compact with singular system $(\lambda_k, u_k, v_k)_{k\in\mathbb{N}}$, then

$$u^\alpha = \sum_{k\in\mathbb{N}} \frac{\lambda_k}{\lambda_k^2 + \alpha} \langle y, v_n \rangle \, u_n$$

and, if y satisfies the Picard condition, then also $\lim_{\alpha\to 0} u^\alpha = u^\dagger$.

Properties of Tikhonov Regularization

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then for any $y \in \mathcal{Y}$ and any $\alpha > 0$ there exists a unique Tikhonov-regularized solution

$$u^\alpha = \operatorname*{argmin}_{u \in \mathcal{X}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2 = (A^*A + 2\alpha I)^{-1} A^* y.$$

In particular, if A is compact with singular system $(\lambda_k, u_k, v_k)_{k \in \mathbb{N}}$, then

$$u^\alpha = \sum_{k \in \mathbb{N}} \frac{\lambda_k}{\lambda_k^2 + \alpha} \langle y, v_n \rangle \, u_n$$

and, if y satisfies the Picard condition, then also $\lim_{\alpha \to 0} u^\alpha = u^\dagger$.

Moreover, for the regularization (operator) $\mathcal{R}_\alpha := (A^*A + 2\alpha I)^{-1} A^*$ mapping $y \in \mathcal{Y}$ to $u^\alpha \in \mathcal{X}$ we have $\mathcal{R}_\alpha \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ with $\|\mathcal{R}_\alpha\| \le \sqrt{2}\alpha^{-1/2}$.

Properties of Tikhonov Regularization

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then for any $y \in \mathcal{Y}$ and any $\alpha > 0$ there exists a unique Tikhonov-regularized solution

$$u^\alpha = \operatorname*{argmin}_{u \in \mathcal{X}} \frac{1}{2}\|y - Au\|^2 + \alpha\|u\|^2 = (A^*A + 2\alpha I)^{-1} A^* y.$$

In particular, if A is compact with singular system $(\lambda_k, u_k, v_k)_{k \in \mathbb{N}}$, then

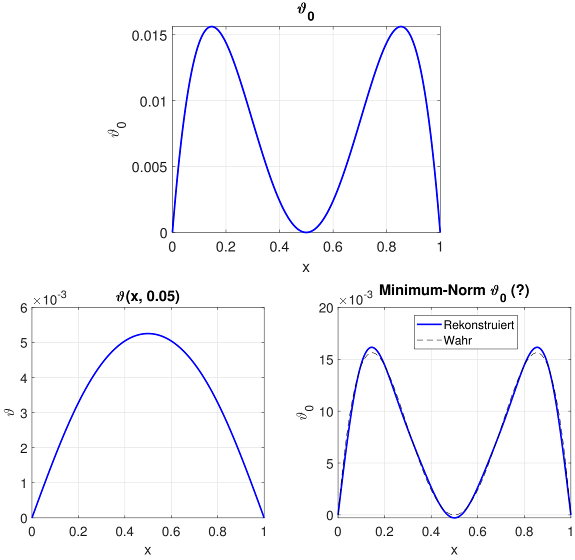
$$u^\alpha = \sum_{k \in \mathbb{N}} \frac{\lambda_k}{\lambda_k^2 + \alpha} \langle y, v_n \rangle \, u_n$$

and, if y satisfies the Picard condition, then also $\lim_{\alpha \to 0} u^\alpha = u^\dagger$.

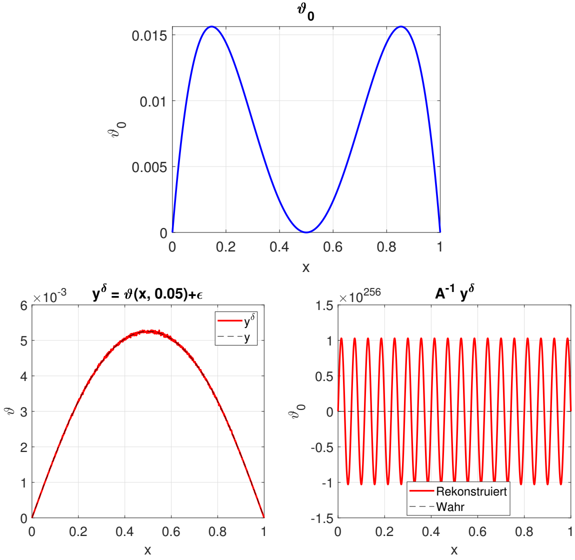
Moreover, for the regularization (operator) $\mathcal{R}_\alpha := (A^*A + 2\alpha I)^{-1} A^*$ mapping $y \in \mathcal{Y}$ to $u^\alpha \in \mathcal{X}$ we have $\mathcal{R}_\alpha \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ with $\|\mathcal{R}_\alpha\| \leq \sqrt{2}\alpha^{-1/2}$.

The Tikhonov-regularized inverse problem is well-posed!

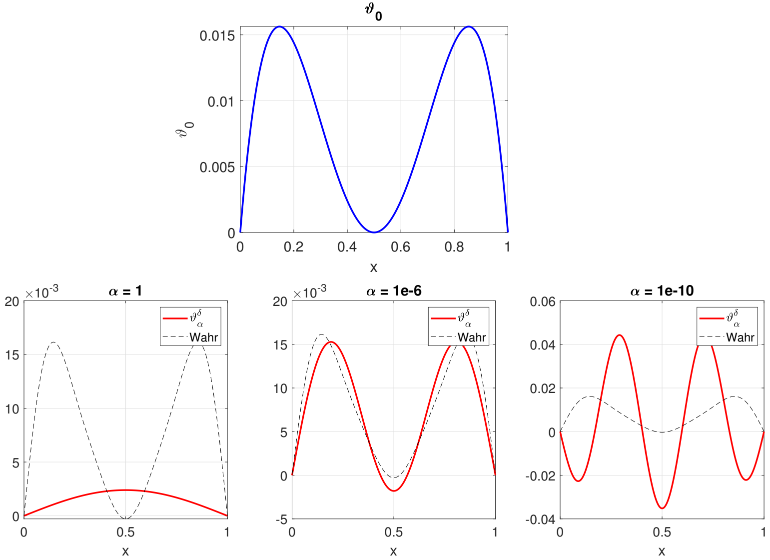
Example: Heat equation



Example: Heat equation



Example: Heat equation



Choice of α

- The regularization parameter steers the influence of the regularization and balances two main error terms.

Choice of α

- The regularization parameter steers the influence of the regularization and balances two main error terms.
- To that end, let $u \in \mathcal{X}$ and $y := Au$ as well as $y^\delta := y + \eta$ with $\delta := \|\eta\|$.
- Then,

$$\begin{aligned}\|u^\alpha - u\| = \|\mathcal{R}_\alpha y^\delta - u\| &\leq \|\mathcal{R}_\alpha y^\delta - \mathcal{R}_\alpha y\| + \|\mathcal{R}_\alpha y - A^\dagger y\| \\ &\leq \underbrace{\|\mathcal{R}_\alpha\|\,\delta}_{\text{data error}} + \underbrace{\|\mathcal{R}_\alpha y - A^\dagger y\|}_{\text{approximation error}}\end{aligned}$$

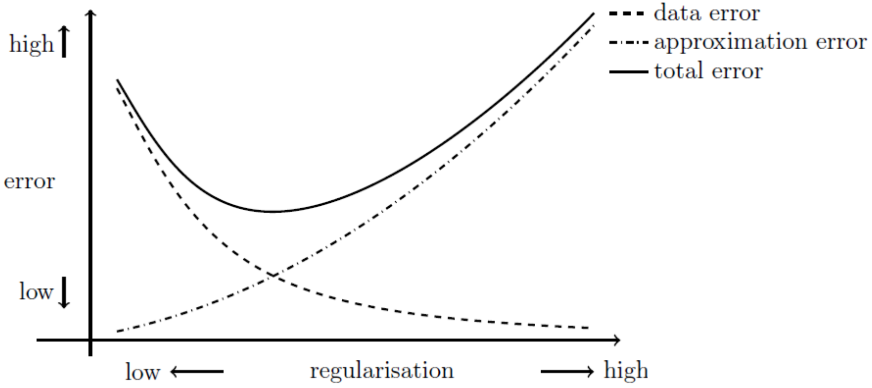
Choice of α

- The regularization parameter steers the influence of the regularization and balances two main error terms.
- To that end, let $u \in \mathcal{X}$ and $y := Au$ as well as $y^\delta := y + \eta$ with $\delta := \|\eta\|$.
- Then,

$$\begin{aligned} \|u^\alpha - u\| = \|\mathcal{R}_\alpha y^\delta - u\| &\leq \|\mathcal{R}_\alpha y^\delta - \mathcal{R}_\alpha y\| + \|\mathcal{R}_\alpha y - A^\dagger y\| \\ &\leq \underbrace{\|\mathcal{R}_\alpha\|\,\delta}_{\text{data error}} + \underbrace{\|\mathcal{R}_\alpha y - A^\dagger y\|}_{\text{approximation error}} \end{aligned}$$

- For α **too small**, the data error dominates (**under-regularized**).
- For α **too large**, the approximation error dominates (**over-regularized**).
- In practice, have to find "sweet spot" depending on the noise level $\alpha = \alpha(\delta)$.

Illustration



Mozorov's discrepancy principle

- For given noisy data $y^\delta = y + \eta$, $y = Au$, the discrepancy from the ground truth satisfies
$$d(u) := \|y^\delta - Au\| = \|\eta\| = \delta.$$

Mozorov's discrepancy principle

- For given noisy data $y^\delta = y + \eta$, $y = Au$, the discrepancy from the ground truth satisfies
$$d(u) := \|y^\delta - Au\| = \|\eta\| = \delta.$$

- Thus, if δ is known, it is sound to search for u^α or α, respectively, with similar resulting discrepancy
$$\delta \leq \|y^\delta - Au^\alpha\| \leq \tau\delta, \qquad \tau > 1. \tag{MDP}$$

Mozorov's discrepancy principle

- For given noisy data $y^\delta = y + \eta$, $y = Au$, the discrepancy from the ground truth satisfies

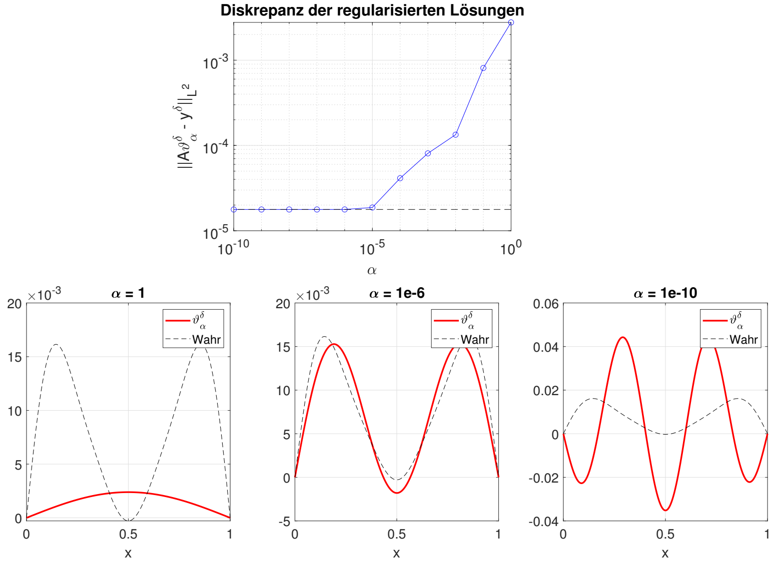
$$d(u) := \|y^\delta - Au\| = \|\eta\| = \delta.$$

- Thus, if δ is known, it is sound to search for u^α or α, respectively, with similar resulting discrepancy

$$\delta \leq \|y^\delta - Au^\alpha\| \leq \tau\delta, \qquad \tau > 1. \tag{MDP}$$

- In practice:
  1. consider null sequence $\alpha_n \to 0$, $n \in \mathbb{N}$,
  2. Compute for each α_n the Tikhonov-regularized solution u^{α_n}
  3. Stop , we u^{α_n} satisfies (MDP).

Example: Heat equation



Convergence of Tikhonov regularization

- It is natural to ask, if in the limit of vanishing noise, i.e., $\delta \to 0$, we recover the ground truth, i.e.,

$$\lim_{\delta \to 0} \|\mathcal{R}_\alpha y^\delta - u\| = 0 \qquad \forall u \in \mathcal{X},$$

where $y^\delta = Au + \eta$ with $\|\eta\| = \delta$

Convergence of Tikhonov regularization

- It is natural to ask, if in the limit of vanishing noise, i.e., $\delta \to 0$, we recover the ground truth, i.e.,

$$\lim_{\delta \to 0} \|\mathcal{R}_\alpha y^\delta - u\| = 0 \qquad \forall u \in \mathcal{X},$$

  where $y^\delta = Au + \eta$ with $\|\eta\| = \delta$

- This requires a suitable noise-adaptive choice for $\alpha = \alpha(\delta)$.

Convergence of Tikhonov regularization

- It is natural to ask, if in the limit of vanishing noise, i.e., $\delta \to 0$, we recover the ground truth, i.e.,

$$\lim_{\delta \to 0} \|\mathcal{R}_\alpha y^\delta - u\| = 0 \qquad \forall u \in \mathcal{X},$$

  where $y^\delta = Au + \eta$ with $\|\eta\| = \delta$
- This requires a suitable noise-adaptive choice for $\alpha = \alpha(\delta)$.

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be compact. Then, we obtain for $\alpha(\delta) = \alpha_0 \delta^{2-\epsilon}$ with $\alpha_0 > 0$ and $\epsilon \in (0, 2)$ that

$$\lim_{\delta \to 0} \sup \left\{ \|R_{\alpha(\delta)} y^\delta - u\| \ : \ y^\delta \in \mathcal{Y}, \|y^\delta - Au\| \le \delta \right\} = 0.$$

Convergence of Tikhonov regularization

- It is natural to ask, if in the limit of vanishing noise, i.e., $\delta \to 0$, we recover the ground truth, i.e.,

$$\lim_{\delta \to 0} \|\mathcal{R}_\alpha y^\delta - u\| = 0 \qquad \forall u \in \mathcal{X},$$

  where $y^\delta = Au + \eta$ with $\|\eta\| = \delta$

- This requires a suitable noise-adaptive choice for $\alpha = \alpha(\delta)$.

Theorem

Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be compact. Then, we obtain for $\alpha(\delta) = \alpha_0 \delta^{2-\epsilon}$ with $\alpha_0 > 0$ and $\epsilon \in (0, 2)$ that

$$\lim_{\delta \to 0} \sup \left\{ \|R_{\alpha(\delta)} y^\delta - u\| \; : \; y^\delta \in \mathcal{Y}, \|y^\delta - Au\| \le \delta \right\} = 0.$$

This holds, if $\|y^\delta - Au\| \le \delta \le \|y^\delta\|$, also for the a-posteriori choice for α given by Mozorov's discrepancy principle.

Extensions

- For nonlinear forward mappings $\mathcal{G}\colon \mathcal{X} \to \mathcal{Y}$ we have

$$u^\alpha \in \underset{u \in \mathcal{X}}{\operatorname{argmin}} \, \tfrac{1}{2} \|y - \mathcal{G}(u)\|^2 + \alpha \|u\|^2$$

- Also for the nonlinear case theoretical results are available, e.g., on convergence, but they require suitable assumptions on $\mathcal{G}$

Extensions

- For nonlinear forward mappings $\mathcal{G}\colon \mathcal{X} \to \mathcal{Y}$ we have

$$u^\alpha \in \underset{u\in\mathcal{X}}{\operatorname{argmin}} \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2 + \alpha\|u\|^2$$

- Also for the nonlinear case theoretical results are available, e.g., on convergence, but they require suitable assumptions on $\mathcal{G}$

- Given a suitable linear operator $Q\colon \mathcal{X} \to Z$ we can also again use **Tikhonov–Philipps regularization**

$$u^\alpha = \mathcal{R}_\alpha y \in \underset{u\in\mathcal{X}}{\operatorname{argmin}} \tfrac{1}{2}\|y - \mathcal{G}(x)\|^2 + \|Qu\|^2,$$

  e.g., $Q = \nabla$ for $\mathcal{X} = H^1(D)$ and $Z = L^2(D)$

- Moreover, also Banach space norms, such as $\|\cdot\|_{L^1(D)}$ or $\|\cdot\|_{\ell^1}$ are common. They promote sparsity in the regularized solution!

Summary – Part I

Inverse problems are typically ill-posed!

Observational model

$$y = \mathcal{G}(u) + \eta,$$

Summary – Part I

Inverse problems are typically ill-posed!

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- Deterministic approach yields regularized least-squares problem

$$\underset{u}{\operatorname{argmin}} \frac{1}{2}\|y - \mathcal{G}(u)\|^2 + \alpha R(u)$$

which is a large-scale, deterministic (nonlinear) optimization problem

Summary – Part I

Inverse problems are typically ill-posed!

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- Deterministic approach yields regularized least-squares problem

$$\operatorname*{argmin}_u \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2 + \alpha R(u)$$

  which is a large-scale, deterministic (nonlinear) optimization problem

- **Crucial** to properly choose the regularization term R and the regularization parameter α!

Summary – Part I

Inverse problems are typically ill-posed!

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- Deterministic approach yields regularized least-squares problem

$$\underset{u}{\operatorname{argmin}} \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2 + \alpha R(u)$$

  which is a large-scale, deterministic (nonlinear) optimization problem

- **Crucial** to properly choose the regularization term R and the regularization parameter α!

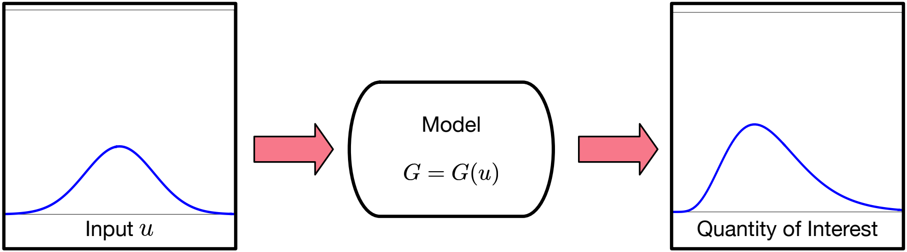
- **No quantification of the uncertainty** in the unknown u!

Statistical Approach to Inverse Problems and Uncertainty Quantification

Uncertainty Quantification

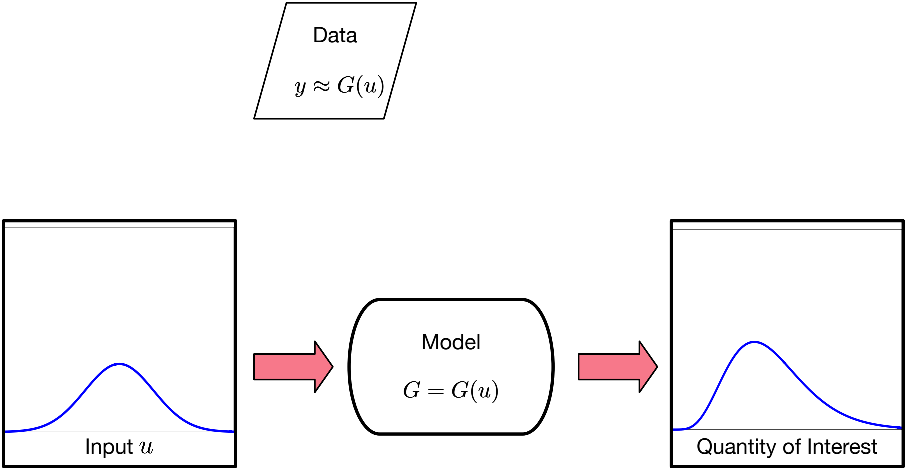
UQ forward problem: Propagation of uncertainty

Wouter Edeling's lectures



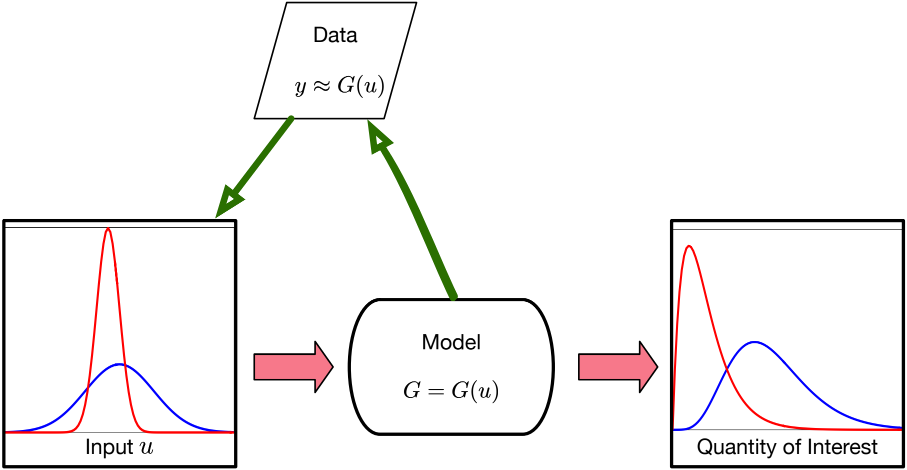
Uncertainty Quantification

In practice: Noisy observations available



Uncertainty Quantification

Bayesian inference: Updating the uncertainty



Statistical Approach to Inverse Problems

We now take a statistical perspective on inverse problems.

Observational model

$$y = \mathcal{G}(u) + \eta,$$

Statistical Approach to Inverse Problems

We now take a statistical perspective on inverse problems.

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- The variables u, y, η above are not considered deterministic anymore...

  ...but as random variables (or realizations of those).

- This will allow us to quantify the remaining uncertainty about u.

Statistical Approach to Inverse Problems

We now take a statistical perspective on inverse problems.

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- The variables u, y, η above are not considered deterministic anymore...

  ... but as random variables (or realizations of those).

- This will allow us to quantify the remaining uncertainty about u.

- For simplicity, we first discuss the finite-dimensional setting

  $$\mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \qquad \mathcal{X} \subseteq \mathbb{R}^n$$

  and start with the **"frequentist"** statistical approach before entering the Bayesian framework

Statistical Approach to Inverse Problems

We now take a statistical perspective on inverse problems.

Observational model

$$y = \mathcal{G}(u) + \eta,$$

- The variables u, y, η above are not considered deterministic anymore...

  ...but as random variables (or realizations of those).

- This will allow us to quantify the remaining uncertainty about u.

- For simplicity, we first discuss the finite-dimensional setting

  $$\mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \qquad \mathcal{X} \subseteq \mathbb{R}^n$$

  and start with the **"frequentist"** statistical approach before entering the Bayesian framework – that is, u still a deterministic parameter and only y, η random!

Recalling the basics

- We assume an underlying abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$:
  1. Ω: set of abstract elementary events $\omega \in \Omega$
  2. $\mathcal{A}$: σ-algebra of (measurable) events, $\mathcal{A} \subset 2^\Omega$
  3. $\mathbb{P}$: abstract probability measure $\mathbb{P} \colon \mathcal{A} \to [0, 1]$

Recalling the basics

- We assume an underlying abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$:
  1. Ω: set of abstract elementary events $\omega \in \Omega$
  2. $\mathcal{A}$: σ-algebra of (measurable) events, $\mathcal{A} \subset 2^{\Omega}$
  3. $\mathbb{P}$: abstract probability measure $\mathbb{P} \colon \mathcal{A} \to [0, 1]$
- A $\mathbb{R}^n$-valued random variable is a (measurable) mapping $U \colon \Omega \to \mathbb{R}^n$
- Its distribution $U \sim \pi$ is the probability measure on $\mathbb{R}^n$ given by

$$\pi(A) := \mathbb{P}(U \in A), \qquad A \subseteq \mathbb{R}^n.$$

Recalling the basics

- We assume an underlying abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$:
  1. Ω: set of abstract elementary events $\omega \in \Omega$
  2. $\mathcal{A}$: σ-algebra of (measurable) events, $\mathcal{A} \subset 2^\Omega$
  3. $\mathbb{P}$: abstract probability measure $\mathbb{P}\colon \mathcal{A} \to [0,1]$

- A $\mathbb{R}^n$-valued random variable is a (measurable) mapping $U\colon \Omega \to \mathbb{R}^n$

- Its distribution $U \sim \pi$ is the probability measure on $\mathbb{R}^n$ given by

$$\pi(A) := \mathbb{P}(U \in A), \qquad A \subseteq \mathbb{R}^n.$$

- Will mainly work with absolutely continuous distributions in this lecture, i.e., π allows for a probability density function also denoted by $\pi\colon \mathbb{R}^n \to [0,\infty)$

$$\pi(A) = \int_A \pi(u)\,\mathrm{d}u, \qquad A \subseteq \mathbb{R}^n.$$

Example: Multivariate Normal Distribution

$$\pi = \mathrm{N}(m, \Sigma)$$

with mean $m \in \mathbb{R}^d$ and symmetric positive (semi-)definite covariance $\Sigma \in \mathbb{R}^{d\times d}$

Probability density function:

$$\pi(y) = \pi_{m,\Sigma}(y) := \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y-m)^\top \Sigma^{-1}(y-m)\right)$$

Example: Multivariate Normal Distribution

$$\pi = \mathrm{N}(m, \Sigma)$$

with mean $m \in \mathbb{R}^d$ and symmetric positive (semi-)definite covariance $\Sigma \in \mathbb{R}^{d\times d}$

Probability density function:

$$\pi(y) = \pi_{m,\Sigma}(y) := \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y-m)^\top \Sigma^{-1}(y-m)\right)$$

Expectation: $Y \sim \mathrm{N}(m, \Sigma)$

$$m = \mathbb{E}\left[Y\right] = \int_{\mathbb{R}^d} y\, \pi_{m,\Sigma}(y)\, \mathrm{d}y$$

Covariance: $Y \sim \mathrm{N}(m, \Sigma)$

$$\Sigma = \mathbb{C}\mathrm{ov}(Y) = \mathbb{E}\left[(Y - \mathbb{E}\left[Y\right])\,(Y - \mathbb{E}\left[Y\right])^\top\right] = \int_{\mathbb{R}^d} (y-m)\,(y-m)^\top\, \pi_{m,\Sigma}(y)\, \mathrm{d}y$$

Example: Multivariate Normal Distribution

$$\pi = \mathrm{N}(m, \Sigma)$$

with mean $m \in \mathbb{R}^d$ and symmetric positive (semi-)definite covariance $\Sigma \in \mathbb{R}^{d\times d}$

Probability density function:

$$\pi(y) = \pi_{m,\Sigma}(y) := \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y-m)^\top \Sigma^{-1}(y-m)\right)$$

Expectation: $Y \sim \mathrm{N}(m, \Sigma)$

$$m = \mathbb{E}\left[Y\right] = \int_{\mathbb{R}^d} y\, \pi_{m,\Sigma}(y)\ \mathrm{d}y$$

Covariance: $Y \sim \mathrm{N}(m, \Sigma)$

$$\Sigma = \mathbb{C}\mathrm{ov}(Y) = \mathbb{E}\left[(Y - \mathbb{E}\left[Y\right])\,(Y - \mathbb{E}\left[Y\right])^\top\right] = \int_{\mathbb{R}^d} (y-m)\,(y-m)^\top\, \pi_{m,\Sigma}(y)\ \mathrm{d}y$$

Affine invariance: $Y \sim \mathrm{N}(m, \Sigma)$ and $A \colon \mathbb{R}^d \to \mathbb{R}^k$, $b \in \mathbb{R}^k$,

$$Y' := AY + b \ \sim\ \mathrm{N}(Am + b, A\Sigma A^\top)$$

Statistical Models and Inverse Problems

Observational model

$$y = \mathcal{G}(u) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad \mathcal{X} \subseteq \mathbb{R}^n$$

Observational model

$$y = \mathcal{G}(u) + \eta, \qquad \mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \quad \mathcal{X} \subseteq \mathbb{R}^n$$

If noise η is assumed to follow a certain probability distribution, e.g., $\eta \sim \mathrm{N}(0, \Sigma)$, **but** u is still a determenistic parameter, then we obtain a

Statistical Models and Inverse Problems

Observational model

$$y = \mathcal{G}(u) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad \mathcal{X} \subseteq \mathbb{R}^n$$

If noise η is assumed to follow a certain probability distribution, e.g., $\eta \sim \mathrm{N}(0, \Sigma)$, **but** u is still a determenistic parameter, then we obtain a

(Frequentist) Statistical Model

A (parametric) statistical model is a pair $(\mathcal{Y}, \mathcal{P})$ consisting of a data space $\mathcal{Y}$ and a family $\mathcal{P}$ of probability distributions or measures $\mathcal{P} = \{\pi_{y|u}\colon u \in \mathcal{X}\}$, $\mathcal{X} \subseteq \mathbb{R}^n$.

Statistical Models and Inverse Problems

Observational model

$$y = \mathcal{G}(u) + \eta, \qquad \mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \quad \mathcal{X} \subseteq \mathbb{R}^n$$

If noise η is assumed to follow a certain probability distribution, e.g., $\eta \sim \mathrm{N}(0, \Sigma)$, **but** u is still a determenistic parameter, then we obtain a

(Frequentist) Statistical Model

A (parametric) statistical model is a pair $(\mathcal{Y}, \mathcal{P})$ consisting of a data space $\mathcal{Y}$ and a family $\mathcal{P}$ of probability distributions or measures $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$, $\mathcal{X} \subseteq \mathbb{R}^n$.

(Additive Gaussian) **inverse problem as statistical model:** Assuming $\eta \sim \mathrm{N}(0, \Sigma)$ we have

$$\mathcal{Y} = \mathbb{R}^d, \qquad \mathcal{P} = \{\mathrm{N}(\mathcal{G}(u), \Sigma) \colon u \in \mathcal{X}\}$$

Statistical Models and Inverse Problems

Observational model

$$y = \mathcal{G}(u) + \eta, \qquad \mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \quad \mathcal{X} \subseteq \mathbb{R}^n$$

If noise η is assumed to follow a certain probability distribution, e.g., $\eta \sim \mathrm{N}(0, \Sigma)$, **but** u is still a determenistic parameter, then we obtain a

(Frequentist) Statistical Model

A (parametric) statistical model is a pair $(\mathcal{Y}, \mathcal{P})$ consisting of a data space $\mathcal{Y}$ and a family $\mathcal{P}$ of probability distributions or measures $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$, $\mathcal{X} \subseteq \mathbb{R}^n$.

(Additive Gaussian) **inverse problem as statistical model:** Assuming $\eta \sim \mathrm{N}(0, \Sigma)$ we have

$$\mathcal{Y} = \mathbb{R}^d, \qquad \mathcal{P} = \{\mathrm{N}(\mathcal{G}(u), \Sigma) \colon u \in \mathcal{X}\}$$

Task: Given samples/realizations $y_1, \ldots, y_m$ of an observable $Y \sim \pi_{y|u} \in \mathcal{P}$, determine $\pi_{y|u}$ as well as u.

Illustration

Consider $\mathcal{Y} = \mathbb{R}$ and $\mathcal{G}(u) = u$ as well as $\eta \sim \mathrm{N}(0,1)$, i.e.,

$$y = u + \eta \qquad \text{or} \qquad \mathcal{P} = \{\mathrm{N}(u,1) \colon u \in \mathcal{X}\}.$$

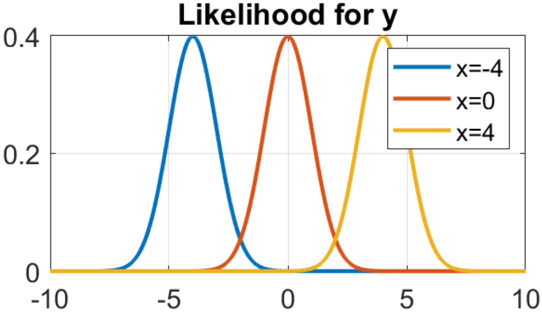
Let $\mathcal{X} = \{-4, 0, 4\}$.

Illustration

Consider $\mathcal{Y} = \mathbb{R}$ and $\mathcal{G}(u) = u$ as well as $\eta \sim \mathrm{N}(0,1)$, i.e.,

$$y = u + \eta \qquad \text{or} \qquad \mathcal{P} = \{\mathrm{N}(u,1) \colon u \in \mathcal{X}\}.$$

Let $\mathcal{X} = \{-4, 0, 4\}$.

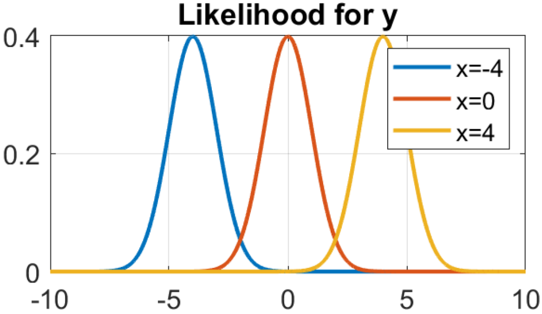


Illustration

Consider $\mathcal{Y} = \mathbb{R}$ and $\mathcal{G}(u) = u$ as well as $\eta \sim \mathrm{N}(0,1)$, i.e.,

$$y = u + \eta \qquad \text{or} \qquad \mathcal{P} = \{\mathrm{N}(u,1) \colon u \in \mathcal{X}\}.$$

Let $\mathcal{X} = \{-4, 0, 4\}$.



Suppose now, you are given the data $y = 2.73$. Which value for u would you infer?

'Frequentist' Answer: Maximum Likelihood Estimation

Maximum Likelihood Principle

Given samples $y_1, \ldots, y_m$ of $Y \sim \pi_{y|u} \in \mathcal{P}$, choose that $u \in \mathcal{X}$ for which the observed data $y_1, \ldots, y_m$ is most probable or most likely.

'Frequentist' Answer: Maximum Likelihood Estimation

Maximum Likelihood Principle

Given samples $y_1, \ldots, y_m$ of $Y \sim \pi_{y|u} \in \mathcal{P}$, choose that $u \in \mathcal{X}$ for which the observed data $y_1, \ldots, y_m$ is most probable or most likely.

If $\pi_{y|u}$ admits a probability density function $\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty)$, e.g., for $\pi_{y|u} = \mathrm{N}(\mathcal{G}(u), \Sigma)$,

$$\pi_{y|u}(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y - \mathcal{G}(u))^\top \Sigma^{-1}(y - \mathcal{G}(u))\right),$$

we define $\pi_{y|u}(y) =: L(y; u)$ as the likelihood of observing $y \in \mathbb{R}^d$ given $u \in \mathcal{X}$.

'Frequentist' Answer: Maximum Likelihood Estimation

Maximum Likelihood Principle

Given samples $y_1, \ldots, y_m$ of $Y \sim \pi_{y|u} \in \mathcal{P}$, choose that $u \in \mathcal{X}$ for which the observed data $y_1, \ldots, y_m$ is most probable or most likely.

If $\pi_{y|u}$ admits a probability density function $\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty)$, e.g., for $\pi_{y|u} = \mathrm{N}(\mathcal{G}(u), \Sigma)$,

$$\pi_{y|u}(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y - \mathcal{G}(u))^\top \Sigma^{-1}(y - \mathcal{G}(u))\right),$$

we define $\pi_{y|u}(y) =: L(y; u)$ as the likelihood of observing $y \in \mathbb{R}^d$ given $u \in \mathcal{X}$.

Maximum Likelihood Estimate

Given data $y \in \mathbb{R}^d$ of $Y \sim \mathrm{N}(\mathcal{G}(u), \Sigma)$ the maximum likelihood estimate for u is

'Frequentist' Answer: Maximum Likelihood Estimation

Maximum Likelihood Principle

Given samples $y_1, \ldots, y_m$ of $Y \sim \pi_{y|u} \in \mathcal{P}$, choose that $u \in \mathcal{X}$ for which the observed data $y_1, \ldots, y_m$ is most probable or most likely.

If $\pi_{y|u}$ admits a probability density function $\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty)$, e.g., for $\pi_{y|u} = \mathrm{N}(\mathcal{G}(u), \Sigma)$,

$$\pi_{y|u}(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y - \mathcal{G}(u))^\top \Sigma^{-1}(y - \mathcal{G}(u))\right),$$

we define $\pi_{y|u}(y) =: L(y; u)$ as the likelihood of observing $y \in \mathbb{R}^d$ given $u \in \mathcal{X}$.

Maximum Likelihood Estimate

Given data $y \in \mathbb{R}^d$ of $Y \sim \mathrm{N}(\mathcal{G}(u), \Sigma)$ the maximum likelihood estimate for u is

$$u_{\mathsf{MLE}} \in \operatorname*{argmax}_u L(y; u) = \operatorname*{argmin}_u \tfrac{1}{2}(y - \mathcal{G}(u))^\top \Sigma^{-1}(y - \mathcal{G}(u))$$

'Frequentist' Answer: Maximum Likelihood Estimation

Maximum Likelihood Principle

Given samples $y_1, \ldots, y_m$ of $Y \sim \pi_{y|u} \in \mathcal{P}$, choose that $u \in \mathcal{X}$ for which the observed data $y_1, \ldots, y_m$ is most probable or most likely.

If $\pi_{y|u}$ admits a probability density function $\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty)$, e.g., for $\pi_{y|u} = \mathrm{N}(\mathcal{G}(u), \Sigma)$,

$$\pi_{y|u}(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\tfrac{1}{2}(y - \mathcal{G}(u))^\top \Sigma^{-1}(y - \mathcal{G}(u))\right),$$

we define $\pi_{y|u}(y) =: L(y; u)$ as the likelihood of observing $y \in \mathbb{R}^d$ given $u \in \mathcal{X}$.

Maximum Likelihood Estimate

Given data $y \in \mathbb{R}^d$ of $Y \sim \mathrm{N}(\mathcal{G}(u), \Sigma)$ the maximum likelihood estimate for u is

$$u_{\mathsf{MLE}} \in \operatorname*{argmax}_u L(y; u) = \operatorname*{argmin}_u \tfrac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

Summary – Part II

- In the (frequentist) statistical approach (following maximum likelihood) we end up again with a (generalized) least squares solution

$$\frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

Summary – Part II

- In the (frequentist) statistical approach (following maximum likelihood) we end up again with a (generalized) least squares solution

$$\frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

- We inherit similar problems such as non-uniqueness and (potentially) ill-conditoned dependence on y:
  - For underdetermined cases, i.e., single data $y \in \mathbb{R}^d$ where $d < n$ ($u \in \mathbb{R}^n$) we require regularization again and
  - the (frequentist) statistical approach yields no uncertainty quantification for u!

Summary – Part II

- In the (frequentist) statistical approach (following maximum likelihood) we end up again with a (generalized) least squares solution

$$\frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

- We inherit similar problems such as non-uniqueness and (potentially) ill-conditoned dependence on y:
  - For underdetermined cases, i.e., single data $y \in \mathbb{R}^d$ where $d < n$ ($u \in \mathbb{R}^n$) we require regularization again and
  - the (frequentist) statistical approach yields no uncertainty quantification for u!

- A more elegant way to regularization in the statistical setting that also yields a way to quantify uncertainty is the Bayesian approach to inverse problems.

$\longrightarrow$ **Lecture 2**