Optimal design of experiments for inverse problems: Bayesian approach and linear Bayesian problems

Georg Stadler

Courant Institute, New York University
stadler@cims.nyu.edu

June 28, 2023

Organization

- Part I (today): Bayesian inverse problems, a concrete exaple (illustrating OED), linear Bayesian inverse problems in high/infinite dimensions
- Part II (Thursday): Optimal experimental design for linear Bayesian inverse problems
- Part II (Friday): Towards OED for nonlinear inverse problems and problems with model uncertainty

Please make the most out of these lectures—ask, interrupt, talk to me, etc.

My background

- Applied Math PHD (from Graz, Austria) in *variational inequalities in mechanics, deterministic optimization and optimal control.*
- Researcher at UT Austin, where I got into *Bayesian inverse problems and optimal design, scalable solvers/algorithms, uncertainty quantification.*
- Now at Courant Institute of Mathematical Sciences (part of New York University), working on Math problems in *UQ, optimization under uncertainty*, recently also *rare events and scientific ML.*

Driving applications: Climate, fusion, geophysics, viscous flow

Research focus: (Rigorous) Math and algorithms that enable "scalable" methods to solve real-world problems.

Outline

Inverse problems examples

Inverse problems approaches

A concrete example in 2D

Posterior approximations

Linear/linearized problems

Inverse problems/parameter estimation

real-world process

$$\mathcal{F}(\boldsymbol{m}) \to \boldsymbol{d}$$

- $\mathcal{F}$... physical process
- $\boldsymbol{m}$... non-directly observable parameters/model; the "cause"
- $\boldsymbol{d}$... results/observation; the "effect"

Inverse problems/parameter estimation

real-world process

$$\mathcal{F}(\boldsymbol{m}) \to \boldsymbol{d}$$

- $\mathcal{F}\ldots$ physical process
- $\boldsymbol{m}\ldots$ non-directly observable parameters/model; the "cause"
- $\boldsymbol{d}\ldots$ results/observation; the "effect"

mathematical model

$$\boldsymbol{f}(\boldsymbol{m}) + \boldsymbol{e} = \boldsymbol{d}$$

- $\boldsymbol{f}\ldots$ forward mapping/parameter-to-observable map (mathematical description of physical process)
- $\boldsymbol{m}\ldots$ parameter vector or parameter function
- $\boldsymbol{e}\ \ldots$ measurement and model errors

Inverse problems

mathematical model

$$\boldsymbol{f}(\boldsymbol{m})(+\boldsymbol{e}) = \boldsymbol{d}$$

$$\longrightarrow$$

forward problem

Given the forward map $\boldsymbol{f}$ and the parameters $\boldsymbol{m}$, find output $\boldsymbol{d}$.

- well-posed (unique solution, stable with respect to perturbation)
- causal (in time)
- local (or strongly decaying dependency in space and time)

Inverse problems

mathematical model

$$\boldsymbol{f}(\boldsymbol{m})(+\boldsymbol{e}) = \boldsymbol{d}$$

$$\longleftarrow$$

forward problem

Given the forward map $\boldsymbol{f}$ and the parameters $\boldsymbol{m}$, find output $\boldsymbol{d}$.

- well-posed (unique solution, stable with respect to perturbation)
- causal (in time)
- local (or strongly decaying dependency in space and time)

inverse problem

Given the forward map $\boldsymbol{f}$ and the output $\boldsymbol{d}$, infer parameters $\boldsymbol{m}$.

- ill-posed (few observations/data, many parameters consistent with data)
- non-causal (coupled over entire time horizon)
- global ($\boldsymbol{m}$ depends on model over space and time)

Example I: Image processing

Original image

$\xrightarrow{\mathcal{F}}$

Blurred image

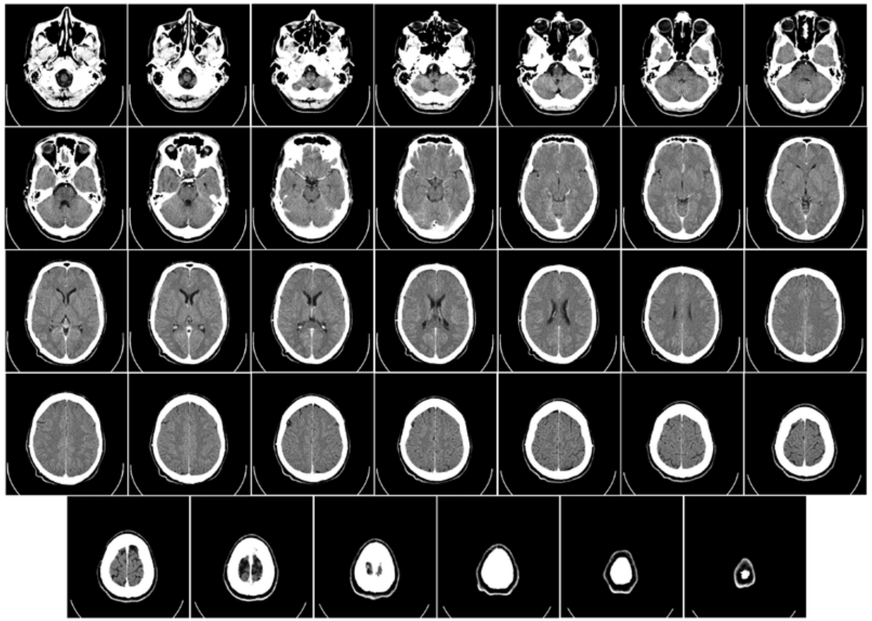
Example I: Image processing

$\xrightarrow{\mathcal{F}}$

- $\mathcal{F}$... blurring due to motion while taking a photo (not a PDE solution operator)
- $\boldsymbol{f} = \boldsymbol{F}$... linear convolution operator
- $\boldsymbol{m}$... left image
- $\boldsymbol{d}$... right image

$\longleftarrow$ inverse problem $\longleftarrow$

Example II: computer tomography (CT)

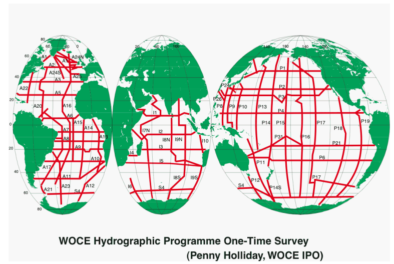


- $\boldsymbol{d}$. . . intensity of X-rays going through human head
- $\boldsymbol{m}$. . . tissue density
- $\boldsymbol{f}$. . . decay of X-rays depending on density

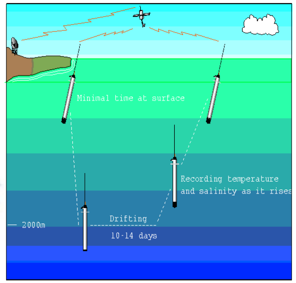
Image shows slices through 3D reconstruction of human brain. Many in vivo imaging methods are inverse problems (MRI, PET scan. . .).

Example III: Ocean dynamics example

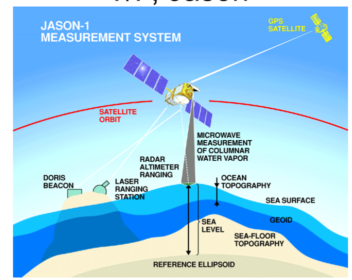
WOCE



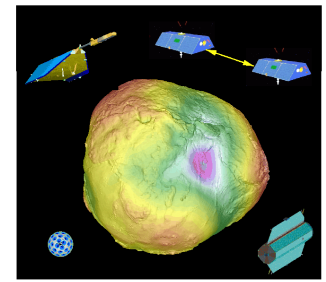
Argo



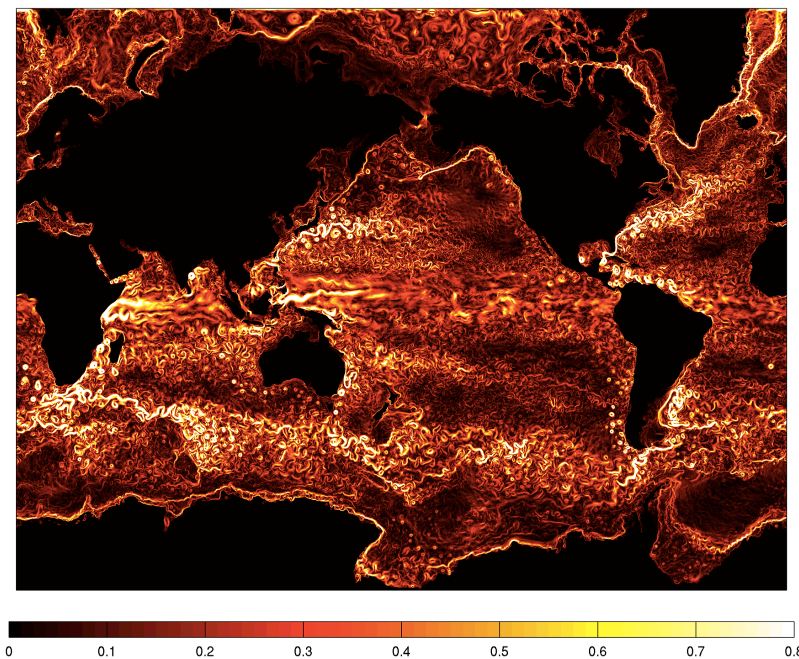
T/P, Jason



GRACE



How to integrate observational data into predictive model?



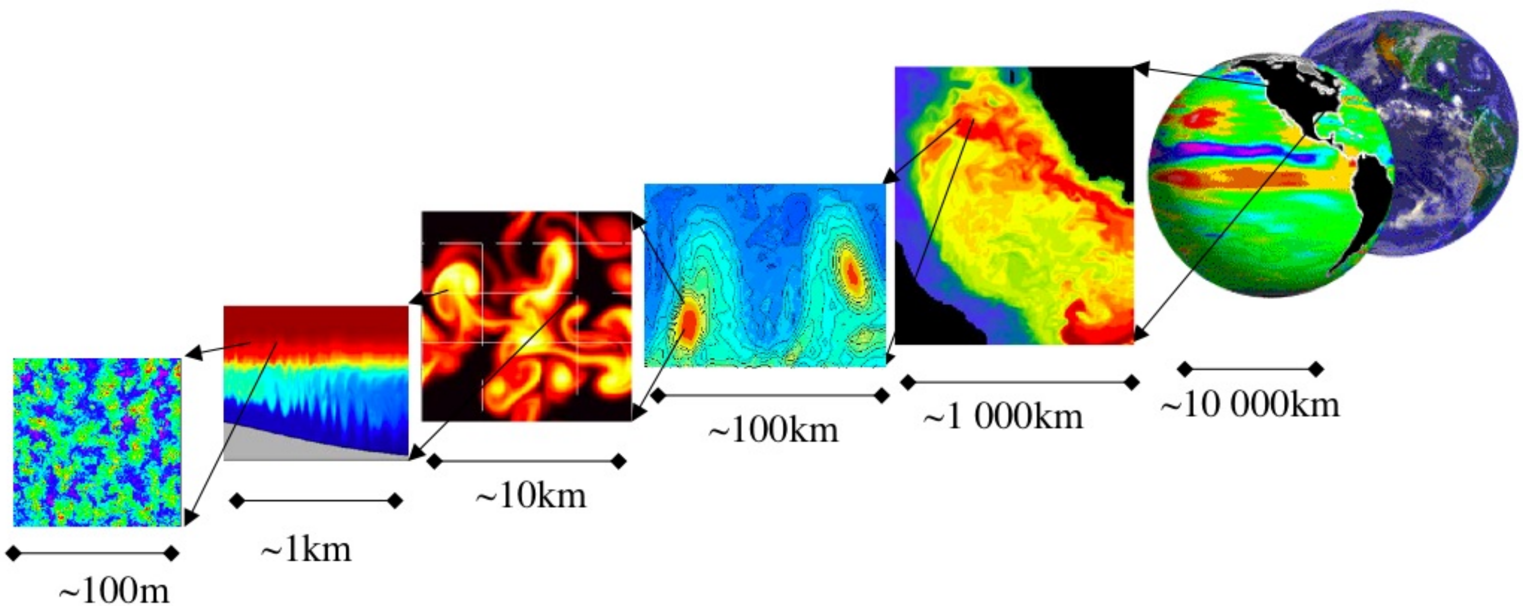
Courtesy Patrick Heimbach, MIT

Example III: Ocean dynamics example

Navier-Stokes equations

- Conservation of mass
- Conservation of momentum
- Conservation of energy
- Conservation of salinity
- Equation of state
- Subgrid parameterizations

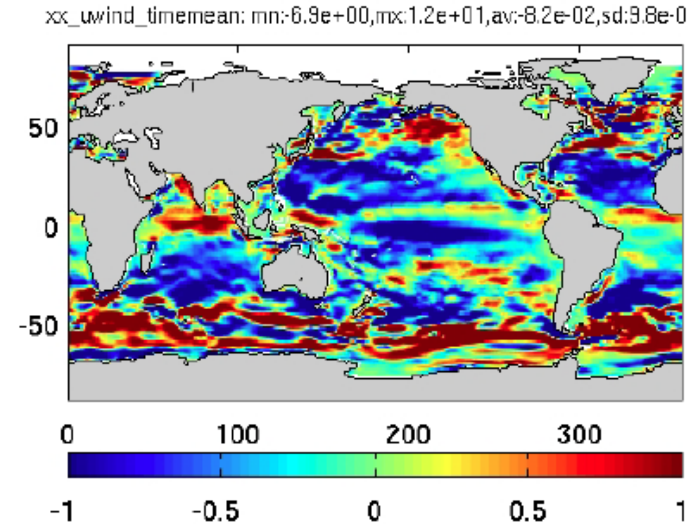
TACC *Stampede* supercomputer

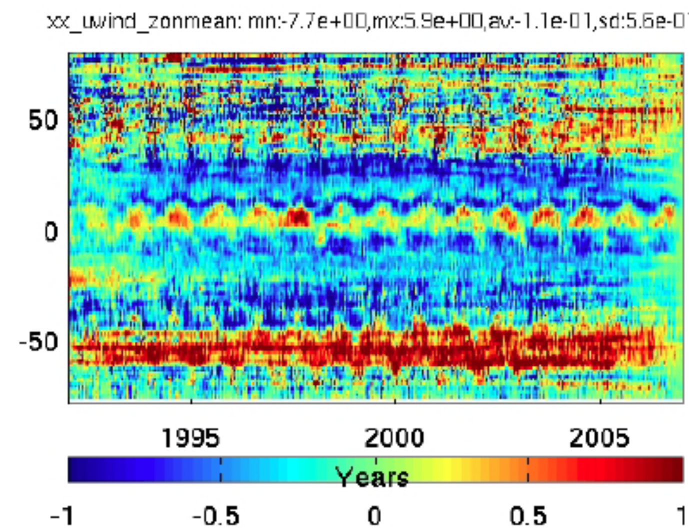


Example III: Ocean dynamics example

2D and 3D parameter fields

- 3D initial temperature and salinity fields
- 2D time-varying atmospheric state at ocean–atmosphere interface
  - surface air temperature
  - specific humidity
  - downwelling shortwave radiation
  - zonal and meridional wind speed
- 3D subgrid model parameter fields
  - vertical mixing coefficient
  - GM coefficient (geostrophic eddy mixing)
  - Redi coefficient (along-isypycnal mixing)





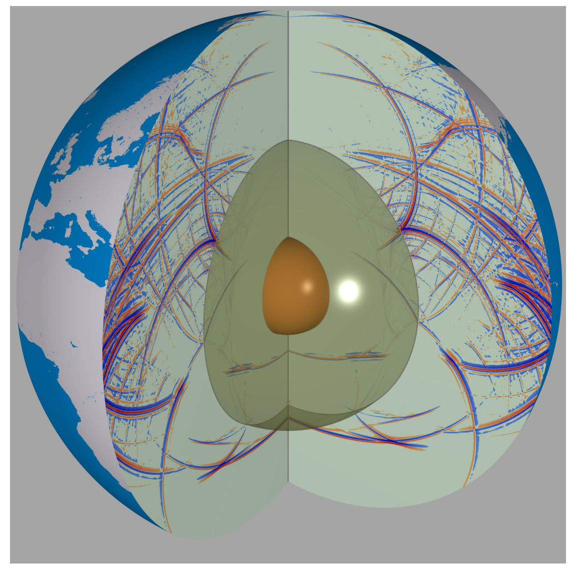
Courtesy Patrick Heimbach, ICES/GEO

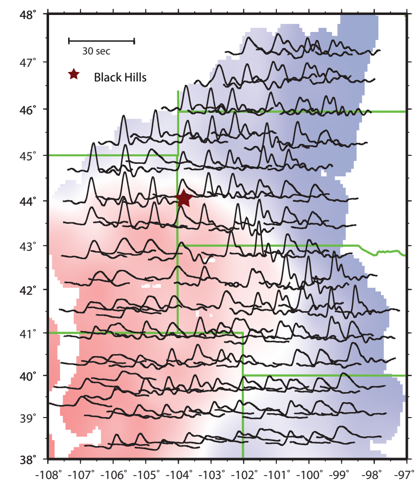
Example IV: Wave-based material inversion (scattering)

Propagate acoustic/elastic/electromagnetic waves through unknown medium

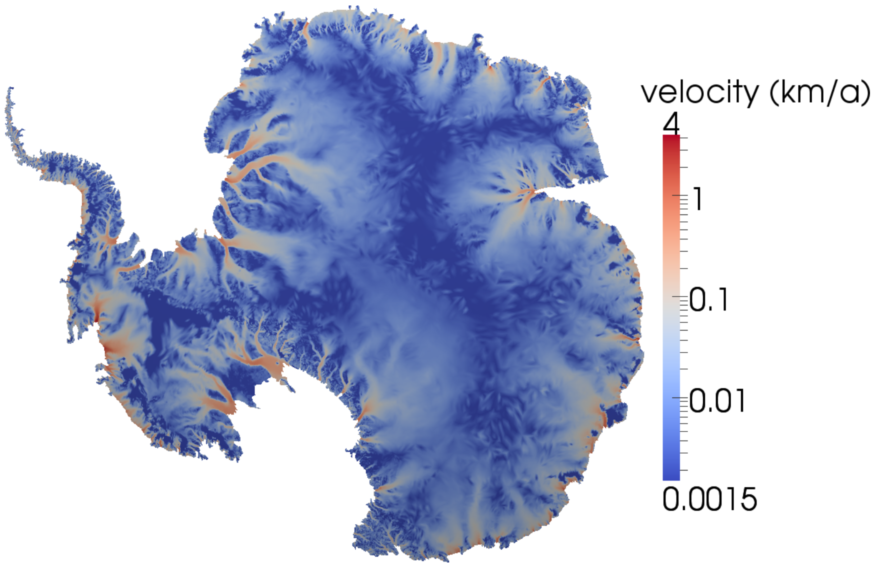
$$u_{tt} - \frac{1}{\boldsymbol{m}(x)^2}\Delta u = 0 \quad \text{in } \Omega \subset \mathbb{R}^d \quad \text{with boundary and initial conditions}$$

- $\boldsymbol{m}(x)$... unknown wave speed
- $\boldsymbol{f} : \boldsymbol{m} \to \boldsymbol{d}$... solution of wave equation with wave speed $\boldsymbol{m}(x)$
- $\boldsymbol{d}$... seismograms, i.e., point measurements of wave field





Example V: Basal condition in ice dynamics

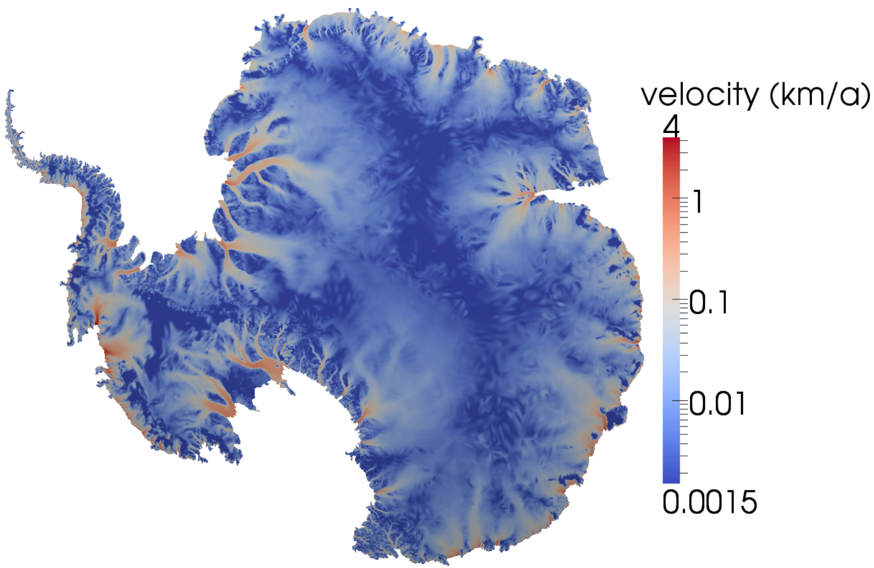


Surface velocity observations

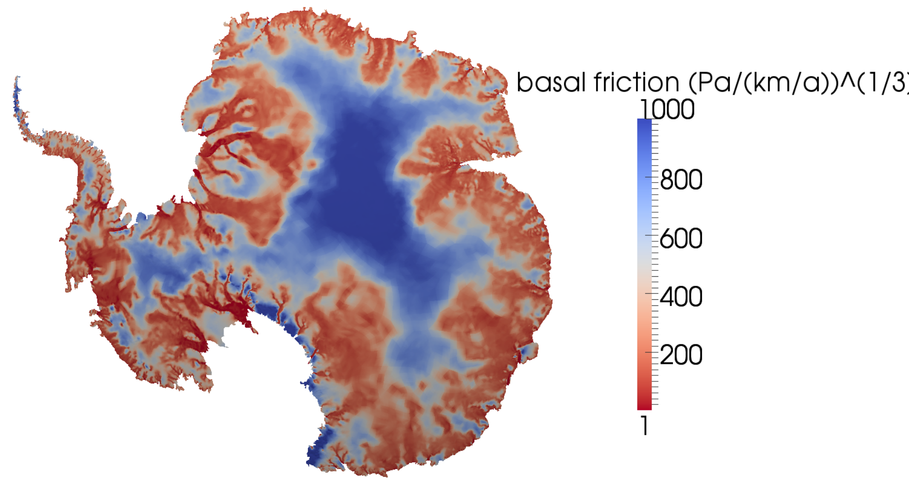
- $\boldsymbol{d}$. . . surface velocity
- $\boldsymbol{m}$. . . basal boundary friction condition
- $\boldsymbol{f}$. . . velocity usually modelled as solution of incompressible nonlinear Stokes equation

- Modeling the dynamics of polar ice sheets is critical for projections of future sea level rise.
- There remain large uncertainties in the basal boundary conditions, which we want to estimate from surface velocity observations.

Example V: Basal condition in ice dynamics



Surface velocity observations



- Modeling the dynamics of polar ice sheets is critical for projections of future sea level rise.
- There remain large uncertainties in the basal boundary conditions, which we want to estimate from surface velocity observations.

Outline

Inversion approach I: Regularization–Occam's razor

Inverse problem ($\boldsymbol{m}\ldots$ parameters, $\boldsymbol{d}\ldots$ data, $\boldsymbol{e}\ldots$ error):

$$\boldsymbol{f}(\boldsymbol{m})(+\boldsymbol{e}) = \boldsymbol{d}$$

Formulate as optimization problem:

$$\min_{\boldsymbol{m}} \frac{1}{2}\|\boldsymbol{f}(\boldsymbol{m}) - \boldsymbol{d}\|^2 + R(\boldsymbol{m})$$

- data misfit term, $\|\cdot\|$ is a measure of the distance between $\boldsymbol{f}(\boldsymbol{m})$ and $\boldsymbol{d}$
- regularization term, stabilizes and picks a specific $\boldsymbol{m}$.

Remarks:

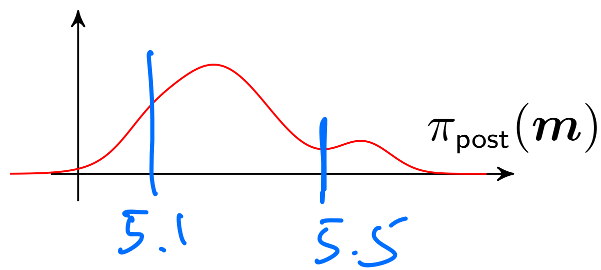
- deterministic problem
- large-scale (PDE-constrained) optimization
- no quantification of uncertainty in $\boldsymbol{m}$
- influence of $R(\cdot)$ on $\boldsymbol{m}$ unclear

Inversion approach II: Statistical–Bayesian inference

Inverse problem

$$\boldsymbol{f}(\boldsymbol{m})(+\boldsymbol{e}) = \boldsymbol{d}$$

Interpret $\boldsymbol{m}$, $\boldsymbol{d}$ as random variables; Solution of inverse problem is a probability density function $\pi_{\text{post}}(\boldsymbol{m})$ for $\boldsymbol{m}$:



Remarks:

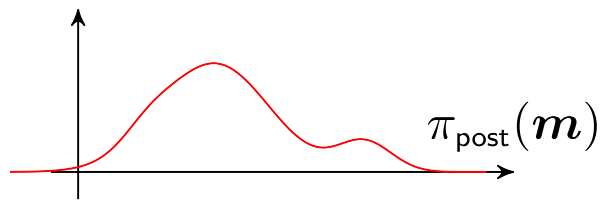
- systematic method to quantify measurement errors and prior knowledge
- allows quantification of uncertainty
- related to regularization approach
- high-dimensional probability density

Inversion approach II: Statistical–Bayesian inference

Inverse problem

$$\boldsymbol{f}(\boldsymbol{m})(+\boldsymbol{e}) = \boldsymbol{d}$$

Interpret $\boldsymbol{m}$, $\boldsymbol{d}$ as random variables; Solution of inverse problem is a probability density function $\pi_{\text{post}}(\boldsymbol{m})$ for $\boldsymbol{m}$:



Target:

- characterize $\pi_{\text{post}}(\boldsymbol{m})$
- for functions $\boldsymbol{m}$ (large vectors after discretization)
- for expensive $\boldsymbol{f}(\cdot)$
- use connection to optimization

Bayes formula (finite dimensions)

Given:

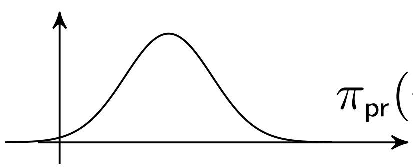
$$\pi_{\text{pr}}(\boldsymbol{m}) \quad : \quad \text{prior p.d.f. of model parameters } \boldsymbol{m}$$

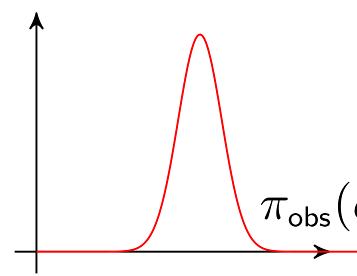
$$\pi_{\text{obs}}(\boldsymbol{d}) \quad : \quad \text{prior p.d.f. of measurement error } \boldsymbol{d}$$

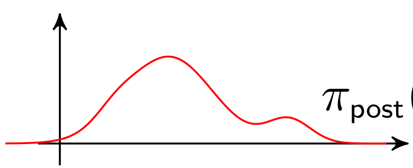
$$\pi_{\text{model}}(\boldsymbol{d}|\boldsymbol{m}) \quad : \quad \text{conditional p.d.f. combining } \boldsymbol{d} \text{ and } \boldsymbol{m} \text{ (model)}$$

Then, the *posterior p.d.f. of the model parameters* is given by:

$$\pi_{\text{post}}(\boldsymbol{m}|\boldsymbol{d}) \propto \pi_{\text{pr}}(\boldsymbol{m})\, \pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{m})$$







model connecting $\boldsymbol{m}$ and $\boldsymbol{d}$

Bayes formula (infinite dimensions)

In infinite dimensional spaces, a Lebesgue measure cannot be defined $\Rightarrow$ alternative formulation of Bayes formula (A. Stuart, Acta Numerica, (2010)).

Given Gaussian random fields:

$$\begin{aligned} \mu_0 := \mathcal{N}(m_0, \mathcal{C}_0) \text{ on } L^2(\Omega) &: \text{ prior measure of parameters } \boldsymbol{m} \\ \mu := \mathcal{N}(0, \Gamma_{\text{noise}}) \text{ on } \mathbb{R}^n &: \text{ prior measure of finite-dimensional data } \boldsymbol{d} \\ \pi_{\text{model}}(\boldsymbol{d}|\boldsymbol{m}) &: \text{ conditional p.d.f. relating } \boldsymbol{d} \text{ and } \boldsymbol{m} \text{ (model)} \end{aligned}$$

Then the *posterior measure* μ^d (i.e., the solution of the statistical inverse problem) is given by:

$$\frac{d\mu^d}{d\mu_0} \propto \pi(\boldsymbol{d}|\boldsymbol{m}),$$

$$R(m) = \int_\Omega |\nabla m|^2 dx$$

$$R(m) = \int_\Omega \left((-\Delta + \alpha I) m\right)^2 dx$$

Left side is *Radon-Nikodym derivative of posterior with respect to the prior.* $\mathcal{C}_0$ *has be of trace class to make problem well-defined (Tikhonov regularization with gradient not sufficient in 2D or 3D)*

Outline

A concrete example

From: Allmaras et al, SIREV, 2013

$v(t)$ vertical vel.

$z(t)$ location $\quad v = \frac{dz}{dt}$

$$F = m\frac{dv}{dt} = mg - mCv^2$$

(m: mass, g: gravity, C: coef. of air resistance)

ODE: $\frac{dv}{dt} = g - Cv^2, \quad \frac{dz}{dt} = v$

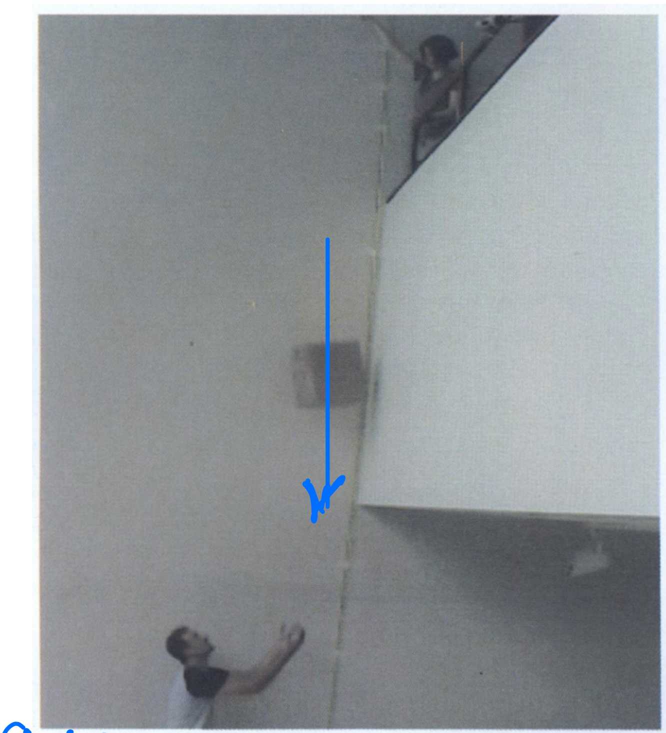
$t \in [t_0, T]$

analytical sol:

$$v(t) = v_\infty \tanh\left(\frac{g(t-t_0)}{v_\infty}\right), \quad z(t) = \frac{1}{C}\log\cosh\left(\frac{g(t-t_0)}{v_\infty}\right)$$

$v_\infty = \sqrt{g/C}$ terminal velocity,

$m = (g, C)$



A concrete example

From: Allmaras et al, SIREV, 2013

Forward: know $(g, C) \in \mathbb{R}^2$, $f(m) = \begin{bmatrix} z(t_0) \\ z(t_1) \\ \vdots \\ z(t_n) \end{bmatrix} \in \mathbb{R}^n$

$t_0, \dots, t_n$ measurement times

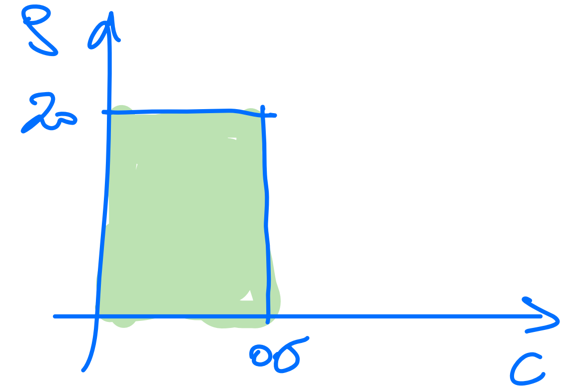
$d = f(m) + \varepsilon$

Prior for g, C:

$0 \le g \le 20 \ m/s$

$0 \le C \le 0.5 \ \frac{1}{m}$

uniform

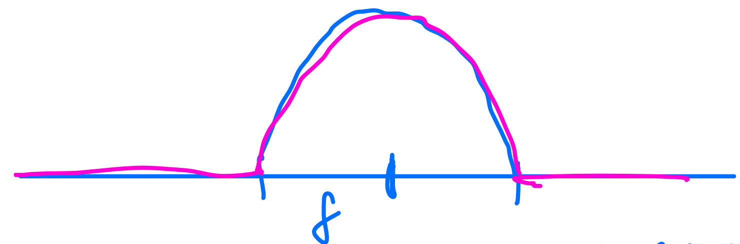


A concrete example

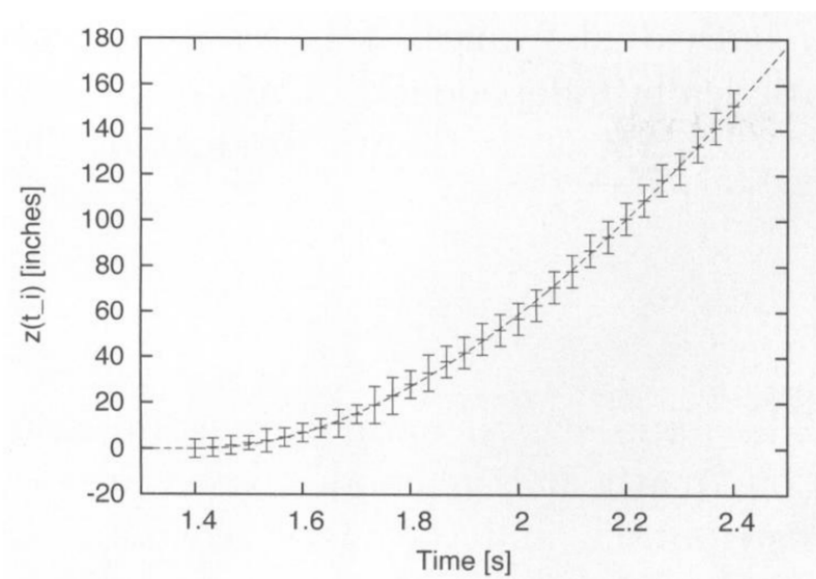
From: Allmaras et al, SIREV, 2013

Measurement data errors:

- low. res of video
- rotation of object
- blurring
- initial time (first assumed known)



$$g_i(d_i|m) = \frac{3}{4\delta_i}\left(1 - \frac{|d_i - f_i(m)|^2}{\delta_i^2}\right) \cdot \chi_{[\dots]}$$



A concrete example

From: Allmaras et al, SIREV, 2013

Want: density (posterior)

$\pi_{post}(m|d)$

Answer: Prob. that true g is in $[g_0, g_1]$ &

C is in $[C_0, C_1]$

Mean of g, C?

Bayes:

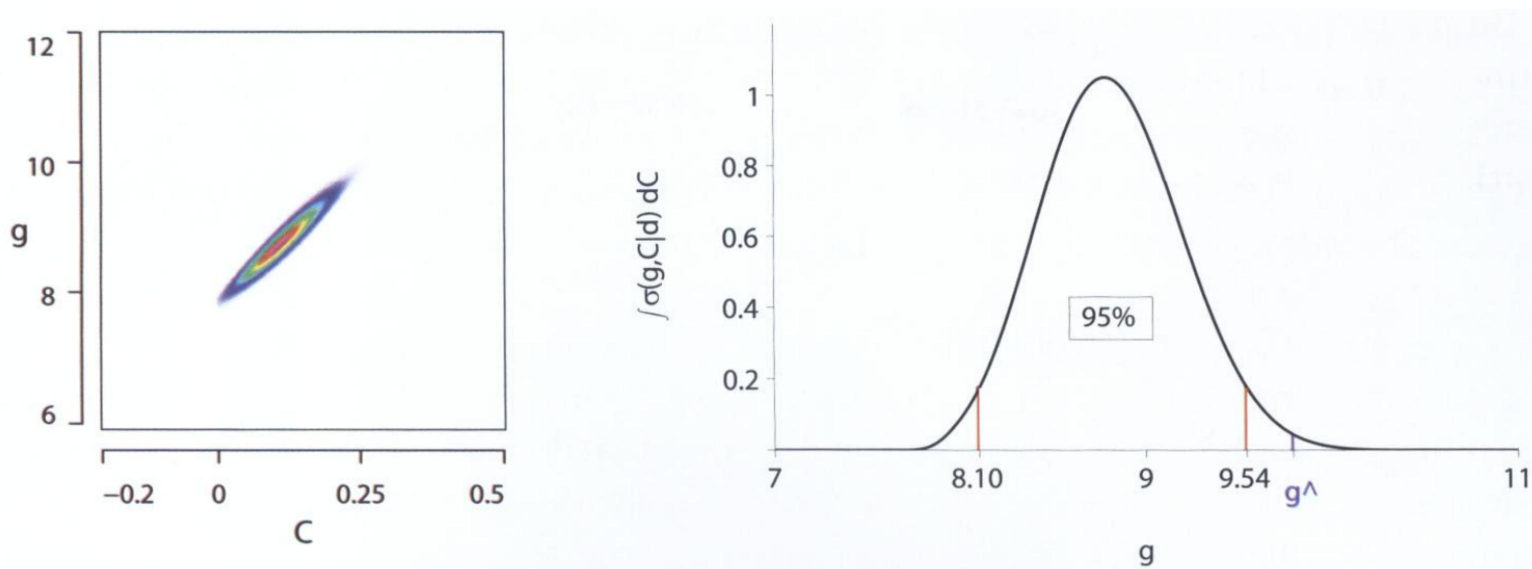
$$\pi_{post}(m|d) = k\, \pi_{pr}(m)\, \pi_{like}(d|m)$$

A concrete example

From: Allmaras et al, SIREV, 2013

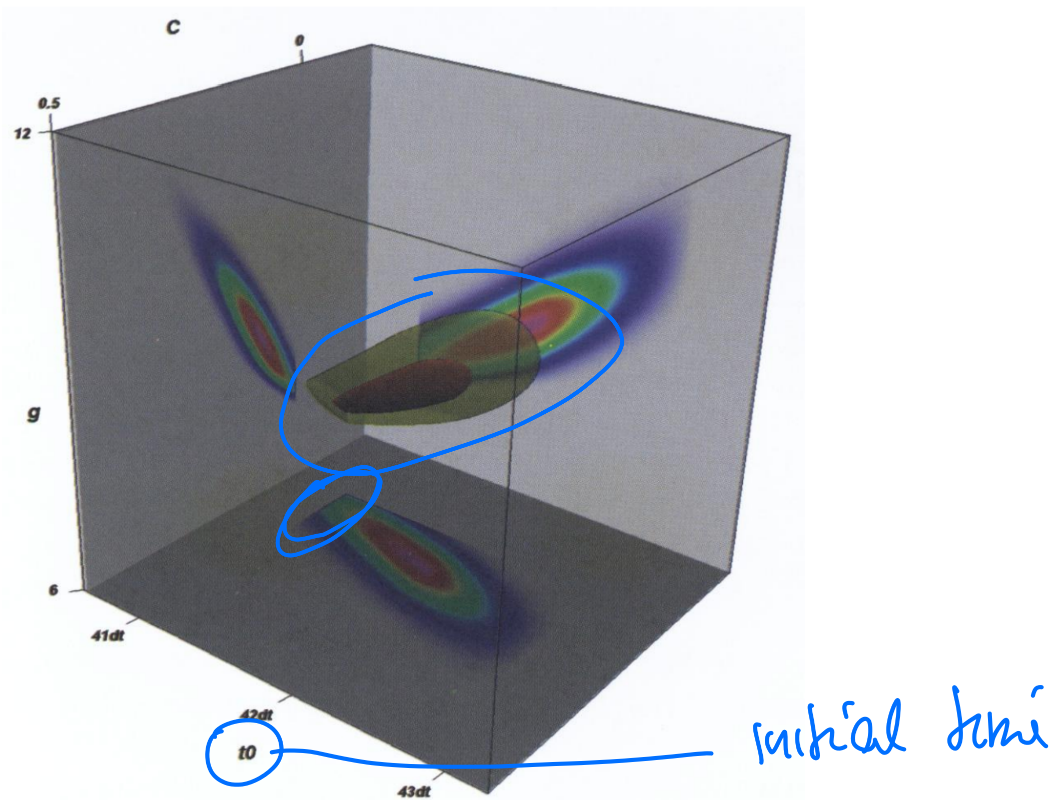
$\mathbb{E}(g|d) \approx 8.82\ m/s^2$ (truth: 9.79)

$\mathbb{E}(C|d) \approx 0.116\ 1/m$



A concrete example:towards OED

From: Allmaras et al, SIREV, 2013



A concrete example: towards OED

From: Allmaras et al, SIREV, 2013