Uncertainty quantification for waves: inverse UQ

Laura Scarabosio (l.scarabosio@math.ru.nl)

Summer school "Waves: Modeling, Analysis, and Numerics"
Nijmegen, 25–29 August 2025

Radboud University

Introduction

The inverse problem

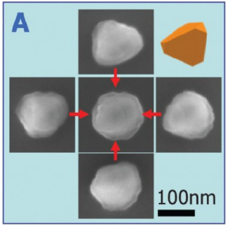
? → (Physics) model → Measurements

Goal: Infer the cause (=value of parameters) by observing effects (=indirect and noisy measurements).

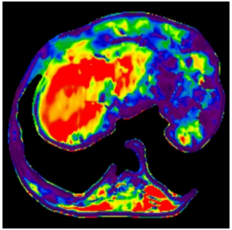
Why? Gain knowledge, make more accurate predictions of quantities of interest.

Why physics model? adds expert knowledge. Allows inference with scarce data and *interpretability*.

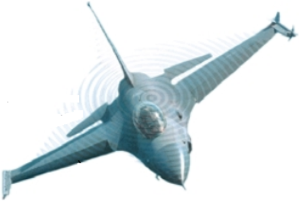
Inverse wave scattering



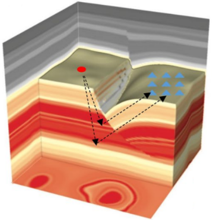
Sannomiya, Diss. ETHZ 18747



siemens-healthineers.com



Borden, Inverse Problems (2002)



Yu, Ma, Rev. Geo. (2021)

Plan for inverse UQ

Inverse problems and Bayesian approach

Bayesian inversion in time-harmonic scattering

Computational realization

Inverse problems and Bayesian approach

Setting

$$\text{measurements} = \text{model}(\vartheta) + \text{noise}$$
$$y = \mathcal{G}(\vartheta) + \varepsilon,$$

where $y \in \mathbb{R}^N$, $\vartheta \in \mathbb{R}^d$, ε realization of N-dimensional random variable, e.g. $\varepsilon \sim \mathcal{N}(0, \Sigma)$.

A possible solution: **least-squares** solution (MLE when $\varepsilon \sim \mathcal{N}(0, \Sigma)$)

$$\vartheta^* = \text{argmin}_\vartheta \|y - \mathcal{G}(\vartheta)\|^2.$$

Remarks:

- Other noise models are possible (e.g., multiplicative noise)
- Modeling error can be taken into account in ε.

Inverse problems are usually ill-posed

Well-posed problem

Existence: a solution exists

Uniqueness: the solution is unique

Stability: the solution depends continuously on data

J.S. Hadamard (1865–1963)

This does not usually hold in inverse problems:
even if we can find a solution, a slight variation in data (e.g., measurement noise) may change the solution drastically.

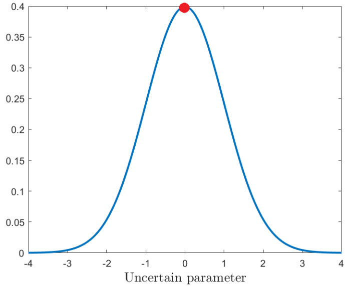
Regularization in inverse problems

Deterministic approach: optimization–based.

Slightly modify the problem or algorithm to achieve stable solution.

Tikhonov regularization: $\vartheta^* = \text{argmin}_\vartheta \|y - \mathcal{G}(\vartheta)\|^2 + \alpha\|\vartheta\|^2$.
Landweber iteration: early stopping in gradient descent.

No quantification of uncertainty.



Statistical approach: sampling–based.

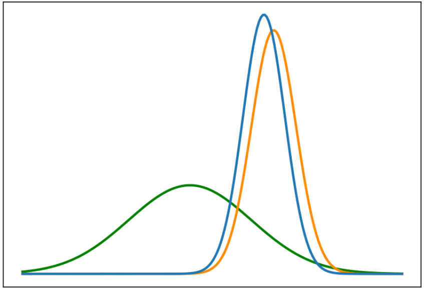
Shift of focus:
probability distribution rather than point estimate.

We model ϑ as a **random variable**.
Its distribution is now the solution to the Bayesian inverse problem.

The Bayesian inverse problem

$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{model evidence}}$$

$$\pi(\vartheta|y) = \frac{\pi(y|\vartheta)\pi_0(\vartheta)}{Z}$$



Prior distribution: distribution of ϑ before any data (expert knowledge)

Likelihood: probability of data conditioned on ϑ (data knowledge)

Posterior distribution: distribution of ϑ *conditioned on data* (data+expert knowledge)

Model evidence: normalization constant, computed for model comparison.

Well-posedness of Bayesian inverse problems [Stuart, Acta Numerica, 2010]

$$y = \mathcal{G}(\vartheta) + \varepsilon,$$

where $y \in \mathbb{R}^N$, $\mathcal{G} : X \to \mathbb{R}^N$, ε with density ρ.

Existence and uniqueness

Assume $\mathcal{G} : X \to \mathbb{R}^N$ continuous, ρ has support equal to $\mathbb{R}^N$ and $\mu_0(X) = 1$. Then the posterior measure $\mu^y(\mathrm{d}\vartheta)$ is absolutely continuous with respect to the prior $\mu_0(\mathrm{d}\vartheta)$ and has Radon-Nikodym derivative given by

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0} = \exp(-\Phi(\vartheta; y)), \quad \Phi(\vartheta; y) = -\log(\rho(y - \mathcal{G}(\vartheta))).$$

Stability

Assume $\mathcal{G} \in L^2_{\mu_0}(X; \mathbb{R}^N)$. Then, for each $r > 0$ there exists $c_r > 0$:

$$d_{Hell}(\mu^y, \mu^{\tilde{y}}) \leq c_r |y - \tilde{y}|, \quad \text{for all } y, \tilde{y} : |y|, |\tilde{y}| \leq r.$$

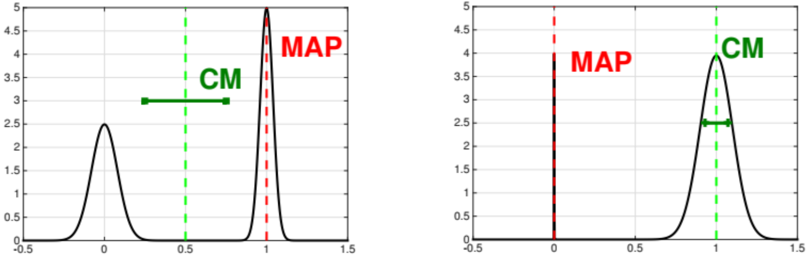
Bayesian estimators

Also in the Bayesian approach, it may be useful to retrieve some point estimates.

Posterior mean (PM or CM): $\vartheta_{PM} = \int_{\mathbb{R}^d} \vartheta \, \mathrm{d}\pi(\vartheta|y)$ "average guess"

Maximum A Posteriori (MAP): $\vartheta_{MAP} = \mathrm{argmax}_\vartheta \pi(\vartheta|y)$ "most likely guess"

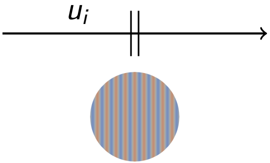
$\vartheta_{PM} = \vartheta_{MAP}$ for Gaussian distributions, but in general they can be very different:

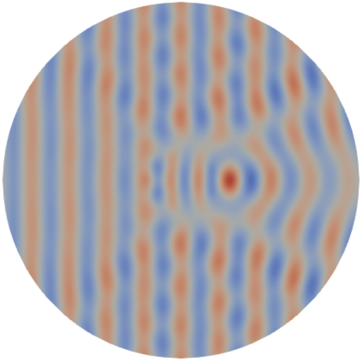


Source: Björn Sprungk, Radboud Summer School 2022.

Bayesian inversion in time-harmonic scattering

The setting





Goal: infer scatterer's shape from measurements of the scattered field.
Focus: effect of the frequency on the inversion result.

Bayesian shape inverse problem

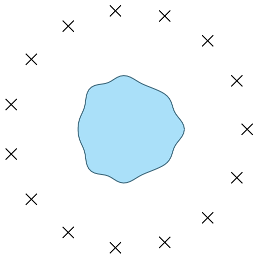
Assumptions

star-shaped scatterer

non-trapping regime [Moiola, Spence 2017]

finite dimensional measurements

additive noise $\varepsilon \sim \mathcal{N}(0, \Sigma)$



Given a prior measure μ_0 on r, find the posterior μ^y given the observations

$$y = \mathcal{G}(r) + \varepsilon,$$

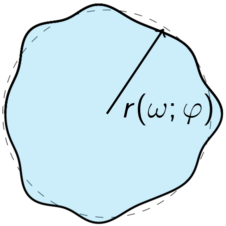
Prior for the shape

$$r(\omega;\varphi) = r_0(\varphi) + \sum_{j=1}^{d} \beta_j Y_j(\omega)\psi_j(\varphi), \quad Y_j \sim \mathcal{U}([-1,1]) \text{ independent}$$

Choices for $\{\psi_j\}_j$:

Laplace-Beltrami eigenfunctions [Church et al. 2020]

Localized supports - wavelets [van Harten, S. 2024]



Coefficients $\{\beta_j\}_j$:

asymptotic decay $\leftrightsquigarrow$ smoothness

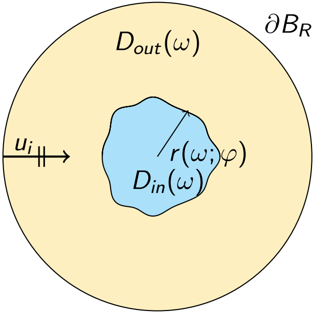
preasymptotic decay $\leftrightsquigarrow$ correlation length

Example:

$$r(\omega;\varphi) = r_0 + \frac{r_0}{4} \sum_{j=1}^{d/2} \frac{1}{1+\ell j^p} \left(Y_{2j-1}(\omega)\cos(j\varphi) + Y_{2j}(\omega)\sin(j\varphi)\right)$$

Helmholtz transmission problem

$$\begin{cases} -\alpha_{in}\Delta(u+u_i) - \kappa_0^2 n_{in}(u+u_i) = 0 \text{ in } D_{in}(\omega) \\ -\alpha_{out}\Delta(u+u_i) - \kappa_0^2 n_{out}(u+u_i) = 0 \text{ in } D_{out}(\omega) \\ + \text{ continuity conditions at interface} \\ + \text{ radiation condition on } u \text{ at } \partial B_R \end{cases}$$



Non-trapping assumption [Moiola, Spence 2019]: $\frac{n_{in}}{n_{out}} \leq \frac{\alpha_{in}}{\alpha_{out}}$

$$\boxed{\mathcal{G}(r) = \mathcal{O} \circ G(r), \text{ where } G(r) = u}$$

Frequency-explicit well-posedness

Theorem (Kuijpers, S. 2023)

Case 1: *r is μ_0-a.s. Lipschitz, $\alpha_{in} = \alpha_{out}$, $\frac{n_{in}}{n_{out}} < 1$ and $V = H^1(B_R)$*

Case 2: *r is μ_0-a.s. of class $C^{2,1}$, $\frac{n_{in}}{n_{out}} < 1 < \frac{\alpha_{in}}{\alpha_{out}}$ and $V = H^1(B_R \setminus U)$.*

Then:

(i) $\mu^y \ll \mu_0$ *with likelihood* $\propto \exp\left(-\frac{1}{2}|y - \mathcal{G}(r)|_\Sigma^2\right)$

(ii) for each $\gamma > 0$ *s.t.* $|y|, |\tilde{y}| \leq \gamma$,

$$d_{\mathrm{Hell}}(\mu^y, \mu^{\tilde{y}}) \leq C\|u_i\|_{H^1_{\kappa_0,\alpha,n}(B_R)}|y - \tilde{y}| \sim (\kappa_0 R)|y - \tilde{y}|$$

Main tools: shape calculus (i) and estimates from [Moiola, Spence 2019] (ii).

Frequency-explicit well-posedness: remarks

Theorem (Kuijpers, S. 2023)

Case 1: *r is μ_0-a.s. Lipschitz, $\alpha_{in} = \alpha_{out}$, $\frac{n_{in}}{n_{out}} < 1$ and $V = H^1(B_R)$*

Case 2: *r is μ_0-a.s. of class $C^{2,1}$, $\frac{n_{in}}{n_{out}} < 1 < \frac{\alpha_{in}}{\alpha_{out}}$ and $V = H^1(B_R \setminus U)$.*

Then:

(i) $\mu^y \ll \mu_0$ with likelihood $L(y|r) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(r)\|_\Sigma^2\right)$

(ii) for each $\gamma > 0$ s.t. $|y|, |\tilde{y}| \leq \gamma$,

$$d_{\mathrm{Hell}}(\mu^y, \mu^{\tilde{y}}) \leq C\|u_i\|_{H^1_{\kappa_0,\alpha,n}(B_R)}|y - \tilde{y}| \sim (\kappa_0 R)|y - \tilde{y}|$$

Frequency dictates the lengthscale

Constants depend on the inverse problem setting (prior, measurements)

High frequency and/or *high contrast* worsen stability

For $\alpha_{in} = \alpha_{out}$, wider class of measurements allowed

Computational realization

From vanilla Monte Carlo to MCMC

To visualize posterior, we need to **sample** (also if we have analytical expression!)

The posterior is in general a *non*-standard distribution on a high-dimensional space: **i.i.d. sampling not feasible**.

Markov Chain Monte Carlo (MCMC)

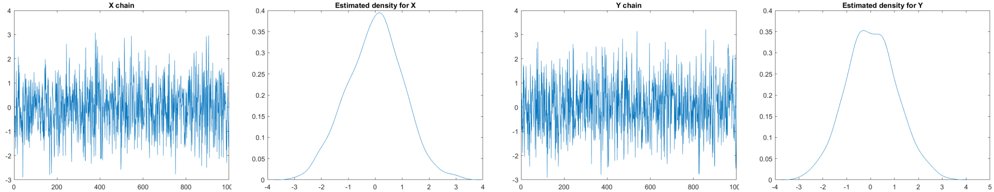
Give up on independence: generate a chain (=sequence) of correlated samples that, after a transient, follow the posterior distribution.

Markov chains: definition

Markov chain: stochastic process where, at time t, the distribution of $\vartheta^{(t)}$ depends on $\vartheta^{(t-1)} = \bar{\vartheta}$ only and not on all previous values (memoryless process).

$$\vartheta^{(0)} \longrightarrow \vartheta^{(1)} \longrightarrow \cdots \longrightarrow \vartheta^{(t-1)} \longrightarrow \vartheta^{(t)}$$

Example: Markov chain sampling for $\vartheta = (X, Y)$



MCMC: Metropolis-Hastings algorithm

Algorithm Metropolis-Hastings algorithm.

1: Select a starting value $\vartheta^{(0)} = (\vartheta_1^{(0)}, \ldots, \vartheta_d^{(0)})$.
2: **for** $t = 1, 2, \ldots \ldots$ **do**
3: Draw $\vartheta_{prop} \sim q(\cdot|\vartheta^{(t-1)})$
4: Compute

$$\alpha(\vartheta_{prop}|\vartheta^{(t-1)}) = \min\left\{1, \frac{f(\vartheta_{prop})q(\vartheta^{(t-1)}|\vartheta_{prop})}{f(\vartheta^{(t-1)})q(\vartheta_{prop}|\vartheta^{(t-1)})}\right\}$$

5: With probability $\alpha(\vartheta_{prop}|\vartheta^{(t-1)})$ set $\vartheta^{(t)} = \vartheta_{prop}$, otherwise $\vartheta^{(t)} = \vartheta^{(t-1)}$.
6: **end for**

where:

f **target density**, for us it's the posterior: $f(\vartheta) = \pi(\vartheta|y)$

q **proposal density**, proposes local moves

we can draw $\vartheta^{(0)} \sim \pi_0(\vartheta)$

MCMC: random walk Metropolis-Hastings algorithm

We propose the new value as

$$\vartheta_{prop} = \vartheta^{(t-1)} + s\xi,$$

where $\xi \sim \mathcal{N}(0, C)$ (or another symmetric distribution)
$\Rightarrow q(\cdot|\vartheta^{(t-1)})$ density of $\mathcal{N}(\vartheta^{(t-1)}, s^2 C)$.

Acceptance probability simplifies to

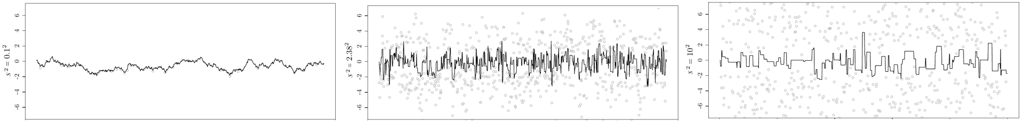
$$\alpha(\vartheta_{prop}, \vartheta^{(t-1)}) = \min \left\{ 1, \frac{f(\vartheta_{prop})}{f(\vartheta^{(t-1)})} \right\}$$

(why?)

Random walk Metropolis-Hastings: choice of step size s

Sources of correlation between samples:

1. ϑ_{prop} too close to $\vartheta^{(t-1)}$ (chains moves slowly)
2. ϑ_{prop} too far from $\vartheta^{(t-1)}$ (high chance of rejection)



Source: A. M. Johansen, L. Evers, and N. Whiteley, Monte carlo methods, Lecture notes (2010).

"Rule of thumb" [Roberts, Rosenthal, 2001]:

$$\text{Choose } s \text{ such that } \mathbb{E}[\alpha(\cdot,\cdot)] \approx 0.234 \text{ when } d > 2,$$
$$\mathbb{E}[\alpha(\cdot,\cdot)] \approx 0.5 \text{ when } d = 1, 2.$$

MCMC: efficiency

For Q with finite variance and $S_N := \frac{1}{N}\sum_{t=1}^{N} Q(\vartheta^{(t)})$:

$$\lim_{N\to\infty} N\mathbb{E}\left[(S_N - \mathbb{E}[Q(\vartheta)])^2\right] = Var[Q(\vartheta)] \underbrace{\left[1 + 2\sum_{j=1}^{\infty} \rho(Q(\vartheta^{(0)}), Q(\vartheta^{(j)}))\right]}_{:=\text{integrated autocorrelation time } (IACT_Q)}$$

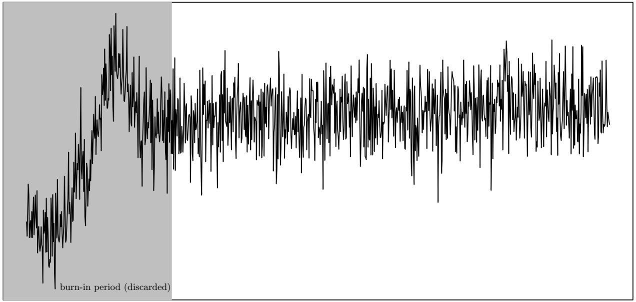
Remarks:

Same convergence rate as vanilla Monte Carlo.
The sum is the price to pay for correlation in samples.

MCMC: practical considerations – burn-in

Depending on $\vartheta^{(0)}$, the distribution of $(\vartheta^{(t)})_t$ for small t might be far from the target distribution (posterior).

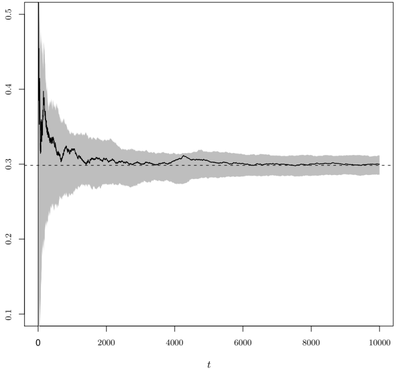
Remedy: discard the first iterations, how many depends on how fast mixing the Markov chain is.



Source: A. M. Johansen et al., Monte carlo methods, Lecture notes (2010)

MCMC: practical considerations – basic plots

1. Plot the chain! Possibly for more runs.
2. Plot cumulative averages $\left(\sum_{\tau=1}^{t} \varphi(\vartheta^{(\tau)})/t\right)_t$ and standard deviation or variance. **Desired**: average converges to a value, variance decreases.



Source: A. M. Johansen et al., Monte carlo methods, Lecture notes (2010)

Bayesian inverse problems and sampling: literature

M. Dashti, M. and A. M. Stuart, *The Bayesian approach to inverse problems*, arXiv preprint arXiv:1302.6989 (2013) and Handbook of Uncertainty Quantification (2016).

A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta numerica (2010), 19, pp. 451–559.

J. Kaipio and E. Somersalo, *Statistical and computational inverse problems* (2006) Springer Science & Business Media.

T. J. Sullivan, *Introduction to uncertainty quantification* (2015), Springer.

C. P. Robert and G. Casella, *Monte Carlo statistical methods* (1999), New York: Springer.

J. S. Liu, *Sequential Monte Carlo methods in practice* (2001), New York: springer.