Numerical Methods for Bayesian Inverse Problems

Lecture 2: Bayesian Approach to 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)

Brief Recap

Observational model

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

- **Inverse problems are typically ill-posed!**
- Deterministic approach and (frequentist) statistical approach yield 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
- Proper choice of the regularization term R and the regularization parameter α are crucial!
- **No quantification of the uncertainty** in the unknown u!

Bayesian Approach to Inverse Problems

The Bayesian Model

Besides random noise η we **also assume now the unknown** u **to be random!**

The Bayesian Model

Besides random noise η we **also assume now the unknown u to be random!**

Bayesian data model

$$Y = \mathcal{G}(U) + \eta,$$

where

- U is a random variable in $\mathcal{X}$ following a prior distribution π_0
  (for simplicity for the moment only finite dimensional $\mathcal{X} \subseteq \mathbb{R}^n$)
- η is a random variable in $\mathbb{R}^d$ following a distribution π_{noise}
- U and η are stochastically independent
- $\mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d$ is known (and measurable)

The Bayesian Model

Besides random noise η we **also assume now the unknown u to be random!**

Bayesian data model

$$Y = \mathcal{G}(U) + \eta,$$

where

- U is a random variable in $\mathcal{X}$ following a prior distribution π_0
  (for simplicity for the moment only finite dimensional $\mathcal{X} \subseteq \mathbb{R}^n$)
- η is a random variable in $\mathbb{R}^d$ following a distribution π_{noise}
- U and η are stochastically independent
- $\mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d$ is known (and measurable)

Bayesian statistical model

is a triple $(\mathcal{Y}, \mathcal{P}, \pi_0)$ consisting of a data space $\mathcal{Y}$, a family of probability distributions $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$, and a prior probability distribution π_0 on $\mathcal{X}$

How a Bayesian learns from data

- Before observational data is available, the Bayesian encodes prior knowledge on unknown u in the prior distribution π_0 .

How a Bayesian learns from data

- Before observational data is available, the Bayesian encodes prior knowledge on unknown u in the prior distribution π_0.

- Given observational data $y \in \mathbb{R}^d$ as realization of the observable

$$Y = \mathcal{G}(U) + \eta, \qquad U \sim \pi_0,\ \eta \sim \pi_{\text{noise}} \ \text{ independent}$$

  the Bayesian updates his prior knowledge π_0 taking into account the information provided by the occured event $Y = y$.

How a Bayesian learns from data

- Before observational data is available, the Bayesian encodes prior knowledge on unknown u in the prior distribution π_0 .

- Given observational data $y \in \mathbb{R}^d$ as realization of the observable

$$Y = \mathcal{G}(U) + \eta, \qquad U \sim \pi_0,\ \eta \sim \pi_{\text{noise}} \ \text{ independent}$$

  the Bayesian updates his prior knowledge π_0 taking into account the information provided by the occured event $Y = y$.

- This update/learning from data is done
  - by conditioning the prior $U \sim \pi_0$ on the event $Y = y$,
  - yielding a conditioned or posterior distribution $\pi_{u|y}$ for the unknown u
  - that represents current knowledge about u and is explicitly given by

Bayes' rule.

Bayes' Rule

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

  $$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.

- Formally,

$$\mathbb{P}\left(H \mid D\right) := \frac{\mathbb{P}(H \cap D)}{\mathbb{P}(D)}$$

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.
- Formally,

$$\mathbb{P}\left(H \mid D\right) := \frac{\mathbb{P}(H \cap D)}{\mathbb{P}(D)}$$

- Since analogously,

$$\mathbb{P}\left(D \mid H\right) = \frac{\mathbb{P}(D \cap H)}{\mathbb{P}(H)}$$

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.

- Formally,

$$\mathbb{P}\left(H \mid D\right) := \frac{\mathbb{P}(H \cap D)}{\mathbb{P}(D)}$$

- Since analogously,

$$\mathbb{P}\left(D \mid H\right) \mathbb{P}(H) = \mathbb{P}(H \cap D)$$

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.

- Formally,

$$\mathbb{P}(H \mid D) := \frac{\mathbb{P}(H \cap D)}{\mathbb{P}(D)}$$

- Since analogously,

$$\mathbb{P}(D \mid H)\, \mathbb{P}(H) = \mathbb{P}(H \cap D)$$

Thomas Bayes
(1701 – 1761)

Bayes' rule

$$\mathbb{P}(H \mid D) = \frac{1}{\mathbb{P}(D)}\, \mathbb{P}(D \mid H)\, \mathbb{P}(H)$$

Basics: Bayes' rule for conditional probabilities

- Consider two events H and D, e.g.,

$$H = \text{Person has a desease}, \quad D = \text{Person has a positive diagnosis test}$$

- **Task:** Find conditional probability that H is true if we know D has occured.
- Formally,

$$\mathbb{P}\left(H \mid D\right) := \frac{\mathbb{P}(H \cap D)}{\mathbb{P}(D)}$$

- Since analogously,

$$\mathbb{P}\left(D \mid H\right) \mathbb{P}(H) = \mathbb{P}(H \cap D)$$

Bayes' rule

$$\mathbb{P}\left(H \mid D\right) = \frac{1}{\mathbb{P}(D)} \mathbb{P}\left(D \mid H\right) \mathbb{P}(H)$$

Pierre-Simon Laplace
(1749 – 1827)

Why Bayes' rule is useful – A classical application

- Often conditional probability $\mathbb{P}(D \mid H)$ easier to determine than $\mathbb{P}(D \cap H)$.

Why Bayes' rule is useful – A classical application

- Often conditional probability $\mathbb{P}(D \mid H)$ easier to determine than $\mathbb{P}(D \cap H)$.
- **Classical application:** In medicine we usually know
  - *Prevalence* of a desease: $\pi_0 = \mathbb{P}(H)$
  - *Sensitivity* of the diagnosis test: $p_{\mathsf{sens}} = \mathbb{P}(D \mid H)$
  - *Specificity* of the diagnosis test: $p_{\mathsf{spec}} = \mathbb{P}(D^c \mid H^c)$

  where D^c and H^c denote the corresponding counter events

Why Bayes' rule is useful – A classical application

- Often conditional probability $\mathbb{P}(D \mid H)$ easier to determine than $\mathbb{P}(D \cap H)$.
- **Classical application:** In medicine we usually know
  - *Prevalence* of a desease: $\pi_0 = \mathbb{P}(H)$
  - *Sensitivity* of the diagnosis test: $p_{\mathsf{sens}} = \mathbb{P}(D \mid H)$
  - *Specificity* of the diagnosis test: $p_{\mathsf{spec}} = \mathbb{P}(D^c \mid H^c)$

  where D^c and H^c denote the corresponding counter events
- Thus, given a positive diagnosis test result D by the total law of probabilities

  $\mathbb{P}(D) = \mathbb{P}(D \mid H)\,\mathbb{P}(H) + \mathbb{P}(D \mid H^c)\,\mathbb{P}(H^c)$

Why Bayes' rule is useful – A classical application

- Often conditional probability $\mathbb{P}(D \mid H)$ easier to determine than $\mathbb{P}(D \cap H)$.
- **Classical application:** In medicine we usually know
  - *Prevalence* of a desease: $\pi_0 = \mathbb{P}(H)$
  - *Sensitivity* of the diagnosis test: $p_{\mathsf{sens}} = \mathbb{P}(D \mid H)$
  - *Specificity* of the diagnosis test: $p_{\mathsf{spec}} = \mathbb{P}(D^c \mid H^c)$

  where D^c and H^c denote the corresponding counter events
- Thus, given a positive diagnosis test result D by the total law of probabilities

  $\mathbb{P}(D) = \mathbb{P}(D \mid H)\,\mathbb{P}(H) + \mathbb{P}(D \mid H^c)\,\mathbb{P}(H^c)$

  it follows that
$$\mathbb{P}\left(H \mid D\right) = \frac{1}{\mathbb{P}(D)}\,\mathbb{P}\left(D \mid H\right)\,\mathbb{P}(H)$$

Why Bayes' rule is useful – A classical application

- Often conditional probability $\mathbb{P}(D \mid H)$ easier to determine than $\mathbb{P}(D \cap H)$.
- **Classical application:** In medicine we usually know
  - *Prevalence* of a desease: $\pi_0 = \mathbb{P}(H)$
  - *Sensitivity* of the diagnosis test: $p_{\text{sens}} = \mathbb{P}(D \mid H)$
  - *Specificity* of the diagnosis test: $p_{\text{spec}} = \mathbb{P}(D^c \mid H^c)$

  where D^c and H^c denote the corresponding counter events
- Thus, given a positive diagnosis test result D by the total law of probabilities

  $$\mathbb{P}(D) = \mathbb{P}(D \mid H)\,\mathbb{P}(H) + \mathbb{P}(D \mid H^c)\,\mathbb{P}(H^c) = p_{\text{sens}}\,\pi_0 + (1 - p_{\text{spec}})\,(1 - \pi_0)$$

  it follows that

  $$\mathbb{P}\,(H \mid D) = \frac{p_{\text{sens}}}{p_{\text{sens}}\,\pi_0 + (1 - p_{\text{spec}})\,(1 - \pi_0)}\,\pi_0$$

Interpretation

Bayes' rule for conditional/posterior probabilities

$$\mathbb{P}\left(H \mid D\right) = \frac{1}{\mathbb{P}(D)}\,\mathbb{P}\left(D \mid H\right)\mathbb{P}(H)$$

where

- $\mathbb{P}(H)$ is the prior probability
- $\mathbb{P}\left(D \mid H\right)$ is the likelihood for the event D given H is true
- $1/\mathbb{P}(D)$ serves as normalizing constant with $\mathbb{P}(D)$ called evidence

Interpretation

Bayes' rule for conditional/posterior probabilities

$$\mathbb{P}\left(H \mid D\right) = \frac{1}{\mathbb{P}(D)}\,\mathbb{P}\left(D \mid H\right)\mathbb{P}(H)$$

where

- $\mathbb{P}(H)$ is the prior probability
- $\mathbb{P}\left(D \mid H\right)$ is the likelihood for the event D given H is true
- $1/\mathbb{P}(D)$ serves as normalizing constant with $\mathbb{P}(D)$ called evidence

Bayes' rule updates the prior probability by reweighting it with the likelihood of the observed data.

Towards Bayes' rule for conditional probability densities

- Consider pair (U, Y) of random variables with values $(u, y) \in \mathbb{R}^n \times \mathbb{R}^d$ and
- suppose the random vector (U, Y) follows a joint distribution $(U, Y) \sim \pi$ on $\mathbb{R}^{n+d}$ with joint probability density function $\pi\colon \mathbb{R}^n \times \mathbb{R}^d \to [0, \infty)$, i.e., for subsets $H \subseteq \mathbb{R}^n$ and $D \subset \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \int_{H \times D} \pi(u, y)\ \mathrm{d}(u, y) = \int_H \int_D \pi(u, y)\ \mathrm{d}y\ \mathrm{d}u$$

Towards Bayes' rule for conditional probability densities

- Consider pair (U, Y) of random variables with values $(u, y) \in \mathbb{R}^n \times \mathbb{R}^d$ and
- suppose the random vector (U, Y) follows a joint distribution $(U, Y) \sim \pi$ on $\mathbb{R}^{n+d}$ with joint probability density function $\pi\colon \mathbb{R}^n \times \mathbb{R}^d \to [0, \infty)$, i.e., for subsets $H \subseteq \mathbb{R}^n$ and $D \subset \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \int_{H \times D} \pi(u, y) \, \mathrm{d}(u, y) = \int_H \int_D \pi(u, y) \, \mathrm{d}y \, \mathrm{d}u$$

- For U and Y we have the marginal distributions $U \sim \pi_U$, $Y \sim \pi_Y$ given by the marginal probability densities $\pi_U : \mathbb{R}^n \to [0, \infty)$, $\pi_Y : \mathbb{R}^d \to [0, \infty)$

$$\pi_U(u) := \int_{\mathbb{R}^d} \pi(u, y) \, \mathrm{d}y, \qquad \pi_Y(y) := \int_{\mathbb{R}^n} \pi(u, y) \, \mathrm{d}u.$$

Towards Bayes' rule for conditional probability densities

- Consider pair (U, Y) of random variables with values $(u, y) \in \mathbb{R}^n \times \mathbb{R}^d$ and
- suppose the random vector (U, Y) follows a joint distribution $(U, Y) \sim \pi$ on $\mathbb{R}^{n+d}$ with joint probability density function $\pi \colon \mathbb{R}^n \times \mathbb{R}^d \to [0, \infty)$, i.e., for subsets $H \subseteq \mathbb{R}^n$ and $D \subset \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \int_{H \times D} \pi(u, y)\, \mathrm{d}(u, y) = \int_H \int_D \pi(u, y)\, \mathrm{d}y\, \mathrm{d}u$$

- For U and Y we have the marginal distributions $U \sim \pi_U$, $Y \sim \pi_Y$ given by the marginal probability densities $\pi_U : \mathbb{R}^n \to [0, \infty)$, $\pi_Y : \mathbb{R}^d \to [0, \infty)$

$$\pi_U(u) := \int_{\mathbb{R}^d} \pi(u, y)\, \mathrm{d}y, \qquad \pi_Y(y) := \int_{\mathbb{R}^n} \pi(u, y)\, \mathrm{d}u.$$

- The conditional probability density $\pi_{U|Y}(\cdot; y) \colon \mathbb{R}^n \to [0, \infty)$ of U given $Y = y$ and vice versa, $\pi_{Y|U}(\cdot; u) \colon \mathbb{R}^d \to [0, \infty)$ of Y given $U = u$ are defined as

$$\pi_{U|Y}(u; y) := \frac{\pi(u, y)}{\pi_Y(y)}, \qquad \pi_{Y|U}(y; u) := \frac{\pi(u, y)}{\pi_U(u)}.$$

Remarks

- **Independence:** If U and Y are independent, then for all $H \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \mathbb{P}(U \in H)\,\mathbb{P}(Y \in D) \quad \Leftrightarrow \quad \pi(u,y) = \pi_U(u)\,\pi_Y(y)$$

We then write for the joint probability measure $\pi = \pi_U \otimes \pi_Y$.

Remarks

- **Independence:** If U and Y are independent, then for all $H \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \mathbb{P}(U \in H)\,\mathbb{P}(Y \in D) \quad \Leftrightarrow \quad \pi(u, y) = \pi_U(u)\,\pi_Y(y)$$

  We then write for the joint probability measure $\pi = \pi_U \otimes \pi_Y$.

- For any (measurable) $H \subseteq \mathbb{R}^n$ we have

$$\mathbb{P}(U \in H \mid Y = y) = \int_H \pi_{U|Y}(u; y)\,\mathrm{d}u$$

Remarks

- **Independence:** If U and Y are independent, then for all $H \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \mathbb{P}(U \in H)\,\mathbb{P}(Y \in D) \quad \Leftrightarrow \quad \pi(u,y) = \pi_U(u)\,\pi_Y(y)$$

We then write for the joint probability measure $\pi = \pi_U \otimes \pi_Y$.

- For any (measurable) $H \subseteq \mathbb{R}^n$ we have

$$\mathbb{P}(U \in H \mid Y = y) = \int_H \pi_{U|Y}(u;y)\,\mathrm{d}u$$

but for (measurable) $D \in \mathbb{R}^d$

$$\mathbb{P}(U \in H \mid Y \in D) \neq \int_D \mathbb{P}(U \in H \mid Y = y)\,\mathrm{d}y = \int_{H \times D} \pi_{U|Y}(u;y)\,\mathrm{d}(u,y)$$

Remarks

- **Independence:** If U and Y are independent, then for all $H \subseteq \mathbb{R}^n$, $D \subseteq \mathbb{R}^d$

$$\mathbb{P}(U \in H, Y \in D) = \mathbb{P}(U \in H)\,\mathbb{P}(Y \in D) \quad \Leftrightarrow \quad \pi(u,y) = \pi_U(u)\,\pi_Y(y)$$

We then write for the joint probability measure $\pi = \pi_U \otimes \pi_Y$.

- For any (measurable) $H \subseteq \mathbb{R}^n$ we have

$$\mathbb{P}(U \in H \mid Y = y) = \int_H \pi_{U|Y}(u;y)\,\mathrm{d}u$$

but for (measurable) $D \in \mathbb{R}^d$

$$\mathbb{P}(U \in H \mid Y \in D) \neq \int_D \mathbb{P}(U \in H \mid Y = y)\,\mathrm{d}y = \int_{H \times D} \pi_{U|Y}(u;y)\,\mathrm{d}(u,y)$$

- However, we can write

$$\mathbb{P}(U \in H \mid Y \in D) = \int_H \pi_{U|Y \in D}(u)\,\mathrm{d}u, \qquad \pi_{U|Y\in D}(u) := \frac{\pi(u,y)}{\int_D \pi_Y(y)\,\mathrm{d}y}$$

Bayes' rule

Given pair (U, Y) with joint probability density function $\pi\colon \mathbb{R}^n \times \mathbb{R}^d \to [0, \infty)$ and

$$\pi_{U|Y}(u; y) = \frac{\pi(u, y)}{\pi_Y(y)}, \qquad \pi(u, y) = \pi_{Y|U}(y; u)\,\pi_U(u)$$

we immediately obtain

Bayes' rule

Given pair (U, Y) with joint probability density function $\pi\colon \mathbb{R}^n \times \mathbb{R}^d \to [0, \infty)$ and

$$\pi_{U|Y}(u; y) = \frac{\pi(u, y)}{\pi_Y(y)}, \qquad \pi(u, y) = \pi_{Y|U}(y; u)\, \pi_U(u)$$

we immediately obtain

Bayes' rule for conditional probability densities

$$\pi_{U|Y}(u; y) = \frac{1}{\pi_Y(y)}\, \pi_{Y|U}(y; u)\, \pi_U(u)$$

Bayesian Inference

Bayesian inference

Bayesian statistical model (absolutely continuous case in $\mathbb{R}^n$)

is a triple $(\mathbb{R}^d, \mathcal{P}, \pi_0)$ with a family of distributions $\mathcal{P} = \{\pi_{y|u} : u \in \mathcal{X}\}$ and prior distribution π_0 on $\mathcal{X} \subseteq \mathbb{R}^n$ where each data distribution $\pi_{y|u} \in \mathcal{P}$ and π_0 have probability density functions

$$\pi_{y|u} : \mathbb{R}^d \to [0, \infty), \qquad \pi_0 : \mathcal{X} \to [0, \infty)$$

For consistent notation we write sloppily $\pi_{y|u}(y) = \pi_{Y|U}(y; u)$ and $\pi_{u|y}(u) = \pi_{U|Y}(u; y)$.

Bayesian inference

Bayesian statistical model (absolutely continuous case in $\mathbb{R}^n$)

is a triple $(\mathbb{R}^d, \mathcal{P}, \pi_0)$ with a family of distributions $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$ and prior distribution π_0 on $\mathcal{X} \subseteq \mathbb{R}^n$ where each data distribution $\pi_{y|u} \in \mathcal{P}$ and π_0 have probability density functions

$$\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty), \qquad \pi_0 \colon \mathcal{X} \to [0, \infty)$$

For consistent notation we write sloppily $\pi_{y|u}(y) = \pi_{Y|U}(y; u)$ and $\pi_{u|y}(u) = \pi_{U|Y}(u; y)$.

- This yields joint probability density

$$\pi(u, y) := \pi_{y|u}(y)\, \pi_0(u)$$

Bayesian inference

Bayesian statistical model (absolutely continuous case in $\mathbb{R}^n$)

is a triple $(\mathbb{R}^d, \mathcal{P}, \pi_0)$ with a family of distributions $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$ and prior distribution π_0 on $\mathcal{X} \subseteq \mathbb{R}^n$ where each data distribution $\pi_{y|u} \in \mathcal{P}$ and π_0 have probability density functions

$$\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty), \qquad \pi_0 \colon \mathcal{X} \to [0, \infty)$$

For consistent notation we write sloppily $\pi_{y|u}(y) = \pi_{Y|U}(y; u)$ and $\pi_{u|y}(u) = \pi_{U|Y}(u; y)$.

- This yields joint probability density

$$\pi(u, y) := \pi_{y|u}(y)\, \pi_0(u)$$

- Thus, given data $y \in \mathbb{R}^d$ we obtain the posterior density

$$\pi_{u|y}(u) = \frac{1}{\int_{\mathcal{X}} \pi_{y|u}(y)\, \pi_0(u)\, \mathrm{d}u}\, \pi_{y|u}(y)\, \pi_0(u)$$

via **Bayes' rule**, with the three ingredients prior, likelihood, and evidence.

Bayesian inference

Bayesian statistical model (absolutely continuous case in $\mathbb{R}^n$)

is a triple $(\mathbb{R}^d, \mathcal{P}, \pi_0)$ with a family of distributions $\mathcal{P} = \{\pi_{y|u} \colon u \in \mathcal{X}\}$ and prior distribution π_0 on $\mathcal{X} \subseteq \mathbb{R}^n$ where each data distribution $\pi_{y|u} \in \mathcal{P}$ and π_0 have probability density functions

$$\pi_{y|u} \colon \mathbb{R}^d \to [0, \infty), \qquad \pi_0 \colon \mathcal{X} \to [0, \infty)$$

For consistent notation we write sloppily $\pi_{y|u}(y) = \pi_{Y|U}(y; u)$ and $\pi_{u|y}(u) = \pi_{U|Y}(u; y)$.

- This yields joint probability density

$$\pi(u, y) := \pi_{y|u}(y)\, \pi_0(u)$$

- **Moreover**, given i.i.d. data $y_1, \ldots, y_m \in \mathbb{R}^d$ we obtain the posterior density

$$\pi_{u|y}(u) = \frac{1}{\int_{\mathcal{X}} \left(\prod_{j=1}^m \pi_{y|u}(y_j) \right) \pi_0(u)\, \mathrm{d}u} \left(\prod_{j=1}^m \pi_{y|u}(y_j) \right) \pi_0(u)$$

via **Bayes' rule**, with the three ingredients prior, likelihood, and evidence.

Bayesian inverse problems

Bayesian data model (in $\mathbb{R}^n$ with Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

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

Bayesian inverse problems

Bayesian data model (in $\mathbb{R}^n$ with Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

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

- The likelihood simply is (using affine invariance of Gaussian)

$$\pi_{y|u}(y) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}}\right)$$

- We assume a prior π_0 with probability density $\pi_0\colon \mathcal{X} \to [0, \infty)$

Bayesian inverse problems

Bayesian data model (in $\mathbb{R}^n$ with Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

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

- The likelihood simply is (using affine invariance of Gaussian)

$$\pi_{y|u}(y) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}}\right)$$

- We assume a prior π_0 with probability density $\pi_0 \colon \mathcal{X} \to [0, \infty)$

Bayesian inverse problem (in $\mathbb{R}^n$ with Gaussian noise)

Given data $y \in \mathbb{R}^d$ the **solution** to the Bayesian inverse problem of infering u given $y = \mathcal{G}(U) + \eta$ is the conditional or posterior distribution $\pi_{u|y}$ given by the density

$$\pi_{u|y}(u) \;\propto\; \exp\left(- \tfrac{1}{2}\|y - G(u)\|^2_{\Sigma^{-1}}\right) \pi_0(u)$$

Illustration in 1D

Example: Condition the prior $U \sim \mu_0 = \mathrm{N}(0,1)$ on the observation $y = 2$ where

$$y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}(u) = u^2 + u, \quad \eta \sim \mathrm{N}(0, \sigma^2)$$

Illustration in 1D

Example: Condition the prior $U \sim \mu_0 = \mathrm{N}(0,1)$ on the observation $y = 2$ where

$$y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}(u) = u^2 + u, \quad \eta \sim \mathrm{N}(0, \sigma^2)$$

Bayes' rule

$$\pi_{u|y}(u) \;\propto\; \exp\left(-\frac{1}{2\sigma^2}|2 - u^2 - u|^2\right) \exp\left(-\frac{1}{2}u^2\right)$$

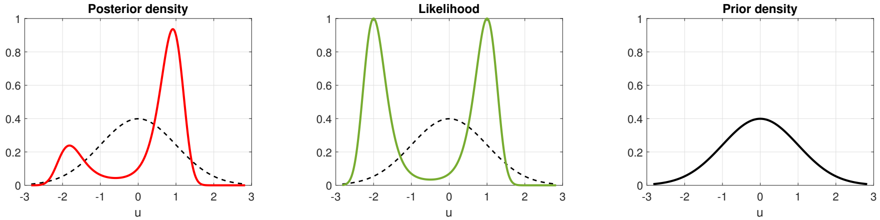


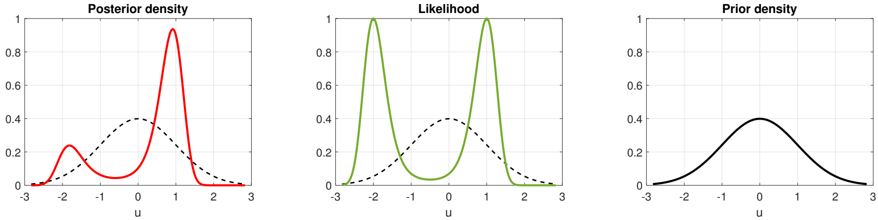
Illustration in 1D

Example: Condition the prior $U \sim \mu_0 = \mathrm{N}(0,1)$ on the observation $y = 2$ where

$$y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}(u) = u^2 + u, \quad \eta \sim \mathrm{N}(0, \sigma^2)$$

Bayes' rule

$$\pi_{u|y}(u) \;\propto\; \exp\left(-\frac{1}{2\sigma^2}|2 - u^2 - u|^2\right) \exp\left(-\frac{1}{2}u^2\right)$$



The **posterior distribution** $\pi_{u|y}$ describes our updated knowledge about u given data y and quantifies our remaining uncertainty**!**

Illustration in 2D

- We consider the following boundary value problem on $D = [0, 1]$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(u_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = f(x), \qquad p(0) = p_0,\ p(1) = u_2$$

  with unknown log-conductivity $u_1 \in \mathbb{R}$ and unknown boundary data $u_2 \in \mathbb{R}$

Illustration in 2D

- We consider the following boundary value problem on $D = [0,1]$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(u_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = f(x), \qquad p(0) = p_0,\ p(1) = u_2$$

with unknown log-conductivity $u_1 \in \mathbb{R}$ and unknown boundary data $u_2 \in \mathbb{R}$

- We are given noisy data $y = (27.5, 79.7)$ for $y = \mathcal{G}(U) + \eta$ with

$$\mathcal{G}(u) = \begin{pmatrix} p(0.25) \\ p(0.75) \end{pmatrix}, \quad \eta \sim \mathrm{N}\left(0, \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}\right).$$

For f and p_0 given, the forward map $\mathcal{G}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is in fact given explicitly here, since the solution p can be computed analytically as a fct. of $(u_1, u_2)^\top$ (**Exercise!**)

Illustration in 2D

- We consider the following boundary value problem on $D = [0,1]$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\exp(u_1)\frac{\mathrm{d}}{\mathrm{d}x}p(x)\right) = f(x), \qquad p(0) = p_0,\ p(1) = u_2$$

  with unknown log-conductivity $u_1 \in \mathbb{R}$ and unknown boundary data $u_2 \in \mathbb{R}$

- We are given noisy data $y = (27.5, 79.7)$ for $y = \mathcal{G}(U) + \eta$ with

$$\mathcal{G}(u) = \begin{pmatrix} p(0.25) \\ p(0.75) \end{pmatrix}, \quad \eta \sim \mathrm{N}\left(0, \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}\right).$$

  For f and p_0 given, the forward map $\mathcal{G}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is in fact given explicitly here, since the solution p can be computed analytically as a fct. of $(u_1, u_2)^\top$ (**Exercise!**)

- As prior we assume

$$\pi_0 = \mathrm{N}(0,1) \otimes \mathrm{U}(90, 110)$$

  and obtain as posterior measure . . .

Illustration in 2D

Posterior for data $y = (27.5, 79.7)^\top$:

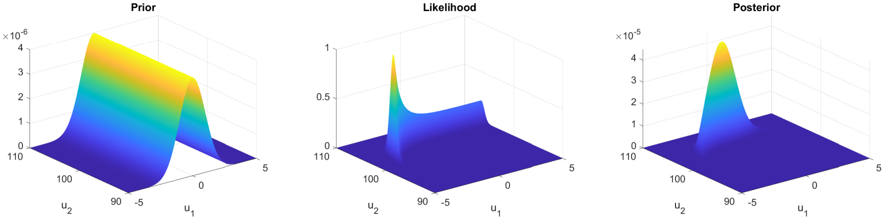


Illustration in 2D

Posterior for data $y = (27.5, 79.7)^\top$:

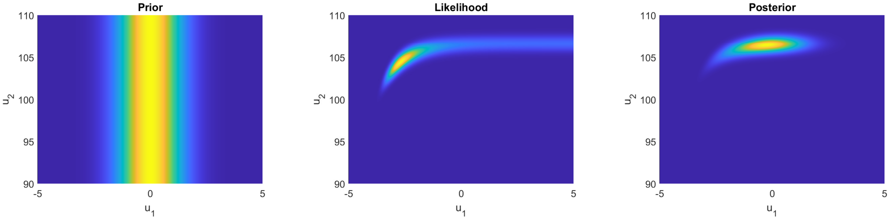
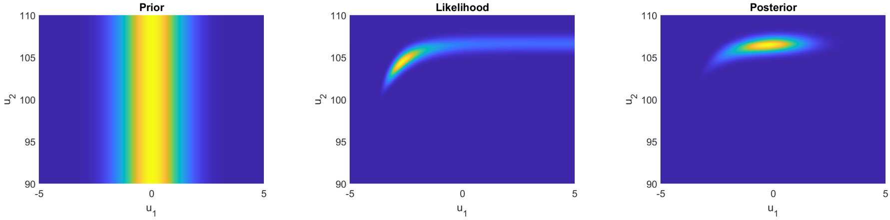
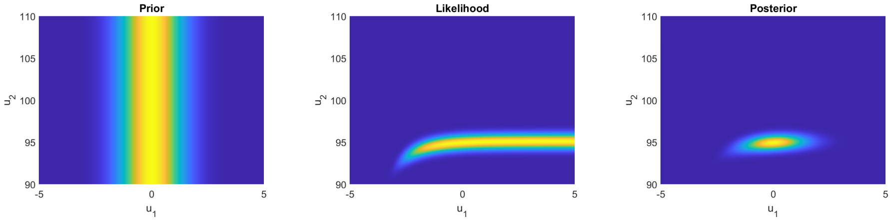


Illustration in 2D

Posterior for data $y = (27.5, 79.7)^\top$:



For different data $\tilde{y} = (23.8, 71.3)^\top$ we obtain the entirely different posterior:



Linear inverse problems

Gauss–Linear model

Given $A \in \mathbb{R}^{d \times n}$

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C), \;\; \eta \sim \mathrm{N}(0, \Sigma) \;\; \text{independent.}$$

Linear inverse problems

Gauss–Linear model

Given $A \in \mathbb{R}^{d \times n}$

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C),\ \eta \sim \mathrm{N}(0, \Sigma) \ \text{ independent.}$$

Then, for data $Y = y$ the posterior distribution is also Gaussian and given explicitly by

$$\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$$

where

$$u^y_{\mathsf{PM}} = CA^\top (ACA^\top + \Sigma)^{-1}(y - Au_0)$$
$$C^y = C - CA^\top (ACA^\top + \Sigma)^{-1} AC$$

Linear inverse problems

Gauss–Linear model

Given $A \in \mathbb{R}^{d\times n}$

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C),\ \eta \sim \mathrm{N}(0, \Sigma) \ \text{ independent.}$$

Then, for data $Y = y$ the posterior distribution is also Gaussian and given explicitly by

$$\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$$

where

$$u^y_{\mathsf{PM}} = CA^\top (ACA^\top + \Sigma)^{-1}(y - Au_0)$$
$$C^y = C - CA^\top (ACA^\top + \Sigma)^{-1} AC$$

- Proof uses affine invariance and then factorisation ('completing the square').

Linear inverse problems

Gauss–Linear model

Given $A \in \mathbb{R}^{d\times n}$

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C),\ \eta \sim \mathrm{N}(0,\Sigma) \ \text{ independent.}$$

Then, for data $Y = y$ the posterior distribution is also Gaussian and given explicitly by

$$\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$$

where

$$u^y_{\mathsf{PM}} = CA^\top(ACA^\top + \Sigma)^{-1}(y - Au_0)$$
$$C^y = C - CA^\top(ACA^\top + \Sigma)^{-1}AC$$

- Proof uses affine invariance and then factorisation ('completing the square').
- Observere the reduction in uncertainty in terms of the covariance matrix

$$C \geq C^y = C - CA^\top(ACA^\top + \Sigma)^{-1}AC \qquad \text{(independent of } y\text{!)}$$

- This formula is also the core for the famous Kalman filter.

Comparison to Tikhonov

$$y = Au + \eta$$

- We saw in Lecture 1 that the the (classical) Tikhonov-regularized solution to this problem with regulariser $R(u) := \alpha\|u\|^2$ is

$$u^\alpha = (A^\top A + 2\alpha I)^{-1} A^\top y$$

Comparison to Tikhonov

$$y = Au + \eta$$

- We saw in Lecture 1 that the the (classical) Tikhonov-regularized solution to this problem with regulariser $R(u) := \alpha\|u\|^2$ is

$$u^\alpha = (A^\top A + 2\alpha I)^{-1} A^\top y$$

- By the Sherman-Morrison-Woodbury formula we get

$$\begin{aligned} u^\alpha &= (A^\top I_d A + 2\alpha I_n)^{-1} A^\top y \\ &= \left[\frac{1}{2\alpha} I_n - \frac{1}{2\alpha} I_n A^\top \left(I_d + \frac{1}{2\alpha} A I_n A^\top\right)^{-1} \frac{1}{2\alpha} A I_n\right] A^\top y \end{aligned}$$

Comparison to Tikhonov

$$y = Au + \eta$$

- We saw in Lecture 1 that the the (classical) Tikhonov-regularized solution to this problem with regulariser $R(u) := \alpha\|u\|^2$ is
$$u^\alpha = (A^\top A + 2\alpha I)^{-1} A^\top y$$
- By the Sherman-Morrison-Woodbury formula we get
$$\begin{aligned} u^\alpha &= (A^\top I_d A + 2\alpha I_n)^{-1} A^\top y \\ &= \frac{1}{2\alpha} A^\top \left[I_d - \left(I_d + \frac{1}{2\alpha} A A^\top \right)^{-1} \frac{1}{2\alpha} A A^\top \right] y \end{aligned}$$

Comparison to Tikhonov

$$y = Au + \eta$$

- We saw in Lecture 1 that the the (classical) Tikhonov-regularized solution to this problem with regulariser $R(u) := \alpha \|u\|^2$ is
$$u^\alpha = (A^\top A + 2\alpha I)^{-1} A^\top y$$
- By the Sherman-Morrison-Woodbury formula we get
$$\begin{aligned} u^\alpha &= (A^\top I_d A + 2\alpha I_n)^{-1} A^\top y \\ &= \frac{1}{2\alpha} A^\top \left[I_d - \left(I_d + \frac{1}{2\alpha} A A^\top \right)^{-1} \frac{1}{2\alpha} A A^\top \right] y \\ &= \frac{1}{2\alpha} A^\top \left(I_d + \frac{1}{2\alpha} A A^\top \right)^{-1} y \end{aligned}$$

Comparison to Tikhonov

$$y = Au + \eta$$

- We saw in Lecture 1 that the the (classical) Tikhonov-regularized solution to this problem with regulariser $R(u) := \alpha\|u\|^2$ is

$$u^\alpha = (A^\top A + 2\alpha I)^{-1} A^\top y$$

- By the Sherman-Morrison-Woodbury formula we get

$$\begin{aligned} u^\alpha &= (A^\top I_d A + 2\alpha I_n)^{-1} A^\top y \\ &= \frac{1}{2\alpha} A^\top \left[I_d - \left(I_d + \frac{1}{2\alpha} A A^\top \right)^{-1} \frac{1}{2\alpha} A A^\top \right] y \\ &= \frac{1}{2\alpha} A^\top \left(I_d + \frac{1}{2\alpha} A A^\top \right)^{-1} y \end{aligned}$$

- For prior mean $u_0 = 0$, prior covariance $C = \frac{1}{2\alpha} I_n$, noise covariance $\Sigma = I_d$ coincides with posterior mean for Gaussian-linear Bayesian inverse problems:

$$u^y_{\mathsf{PM}} = C A^\top (A C A^\top + \Sigma)^{-1} y$$

Back to the general case

Gauss–Linear model

Given $A \in \mathbb{R}^{d\times n}$ and

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C),\ \eta \sim \mathrm{N}(0, \Sigma) \ \text{ independent,}$$

we have for data $Y = y$ as posterior distribution $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$ where

$$u^y_{\mathsf{PM}} = CA^\top (ACA^\top + \Sigma)^{-1}(y - Au_0)$$

$$C^y = C - CA^\top (ACA^\top + \Sigma)^{-1} AC$$

Back to the general case (again using the Sherman-Morrison-Woodbury formula):

Gauss–Linear model

Given $A \in \mathbb{R}^{d\times n}$ and

$$Y = AU + \eta, \qquad U \sim \mathrm{N}(u_0, C),\ \eta \sim \mathrm{N}(0,\Sigma) \ \text{ independent},$$

we have for data $Y = y$ as posterior distribution $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$ where

$$\begin{aligned} u^y_{\mathsf{PM}} &= CA^\top(ACA^\top + \Sigma)^{-1}(y - Au_0) \\ &= (A^\top\Sigma^{-1}A + C^{-1})^{-1}\left[\Sigma A^\top y + C^{-1}u_0\right] \\ C^y &= C - CA^\top(ACA^\top+\Sigma)^{-1}AC = \left(A^\top\Sigma^{-1}A + C^{-1}\right)^{-1} \end{aligned}$$

Thus, we recover **(generalized) Tikhonov solution to (linear) inverse problem**

$$\operatorname*{argmin}_u \frac{1}{2}\|y - Au\|^2_{\Sigma^{-1}} + \frac{1}{2}\|u - u_0\|^2_{C^{-1}}$$

and can associate with $\pi_{u|y}$ the **Bayesian posterior uncertainty about u^y_{PM}.**

Bayes estimates

- The **posterior distribution** $\pi_{u|y}$ is the formal solution to the Bayesian inverse problem and allows for uncertainty quantification.

Bayes estimates

- The **posterior distribution** $\pi_{u|y}$ is the formal solution to the Bayesian inverse problem and allows for uncertainty quantification.

- However, usually we also require a **"single, best guess"** for u, an estimate.

Bayes estimates

- The **posterior distribution** $\pi_{u|y}$ is the formal solution to the Bayesian inverse problem and allows for uncertainty quantification.
- However, usually we also require a **“single, best guess”** for u, an estimate.
- The two most common Bayesian (point) esimtates for u are

  1. the **posterior mean (PM)**

$$u^y_{\mathsf{PM}} := \int_{\mathcal{X}} u\, \pi_{u|y}(u)\ \mathrm{d}u = \mathbb{E}_{\pi_{u|y}}[U]$$

  Interpretation: “average guess” for u given posterior $\pi_{u|y}$

Bayes estimates

- The **posterior distribution** $\pi_{u|y}$ is the formal solution to the Bayesian inverse problem and allows for uncertainty quantification.
- However, usually we also require a **"single, best guess"** for u, an estimate.
- The two most common Bayesian (point) esimtates for u are

  1. the **posterior mean (PM)**

  $$u^y_{\mathsf{PM}} := \int_{\mathcal{X}} u\, \pi_{u|y}(u)\ \mathrm{d}u = \mathbb{E}_{\pi_{u|y}}[U]$$

  Interpretation: "average guess" for u given posterior $\pi_{u|y}$

  2. the **maximum a posteriori estiamte (MAP)**

  $$u^y_{\mathsf{MAP}} \in \operatorname*{argmax}_{u \in \mathcal{X}} \pi_{u|y}(u)$$

  Interpretation: "Most probable/likely guess" for u given posterior $\pi_{u|y}$

MAP estimate and Tikhonov

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- Assuming prior density $\pi_0 \colon \mathcal{X} \to (0, \infty)$ we have

$$u^y_{\mathsf{MAP}} \in \underset{u \in \mathcal{X}}{\operatorname{argmax}}\, \pi_{u|y}(u)$$

MAP estimate and Tikhonov

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- Assuming prior density $\pi_0 \colon \mathcal{X} \to (0, \infty)$ we have

$$u^y_{\mathsf{MAP}} \in \operatorname*{argmax}_{u \in \mathcal{X}} \left\{ \exp\left(-\tfrac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}}\right) \ \pi_0(u) \right\}$$

MAP estimate and Tikhonov

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- Assuming prior density $\pi_0 \colon \mathcal{X} \to (0, \infty)$ we have

$$u^y_{\mathsf{MAP}} \in \operatorname*{argmin}_{u \in \mathcal{X}} \left\{ \tfrac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} - \log(\pi_0(u)) \right\}$$

MAP estimate and Tikhonov

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- Assuming prior density $\pi_0 \colon \mathcal{X} \to (0, \infty)$ we have

$$u^y_{\mathsf{MAP}} \in \operatorname*{argmin}_{u \in \mathcal{X}} \left\{ \tfrac{1}{2} \| y - \mathcal{G}(u) \|^2_{\Sigma^{-1}} - \log(\pi_0(u)) \right\}$$

- Corresponds to regularized solution of the (deterministic) inverse problem with weighted least squares functional $\frac{1}{2} \| y - \mathcal{G}(u) \|^2_{\Sigma^{-1}}$ and regularizing functional

$$R(u) = -\log(\pi_0(u))$$

MAP estimate and Tikhonov

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- Assuming prior density $\pi_0 \colon \mathcal{X} \to (0, \infty)$ we have

$$u^y_{\mathsf{MAP}} \in \underset{u \in \mathcal{X}}{\operatorname{argmin}} \left\{ \tfrac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} - \log(\pi_0(u)) \right\}$$

- Corresponds to regularized solution of the (deterministic) inverse problem with weighted least squares functional $\frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}}$ and regularizing functional

$$R(u) = -\log(\pi_0(u))$$

- For Gaussian prior $\pi_0 = \mathrm{N}(u_0, C)$, recover Tikhonov–Philipps regularization

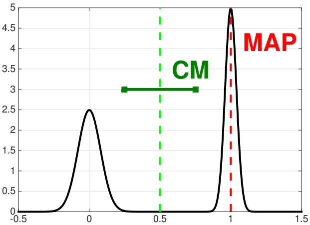
$$u^y_{\mathsf{MAP}} \in \underset{u \in \mathcal{X}}{\operatorname{argmin}} \, \tfrac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} + \tfrac{1}{2} \|(u - u_0)\|^2_{C^{-1}}$$

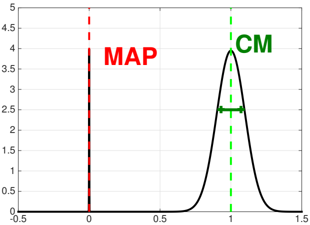
Comparison between PM and MAP

- Posterior mean $u^y_{\mathsf{PM}} = \int_{\mathcal{X}} u\pi_{u|y}(u)\,\mathrm{d}u$ computed via **numerical integration**
- MAP estimate $u^y_{\mathsf{MAP}} \in \operatorname*{argmax}_{u\in\mathcal{X}} \pi_{u|y}(u)$ computed via **numerical optimization**

Comparison between PM and MAP

- Posterior mean $u^y_{\mathsf{PM}} = \int_{\mathcal{X}} u \pi_{u|y}(u)\ \mathrm{d}u$ computed via **numerical integration**
- MAP estimate $u^y_{\mathsf{MAP}} \in \underset{u \in \mathcal{X}}{\operatorname{argmax}}\, \pi_{u|y}(u)$ computed via **numerical optimization**
- Although for Gaussian $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$, we have $u^y_{\mathsf{PM}} = u^y_{\mathsf{MAP}}$, in general, they yield very different estimates:

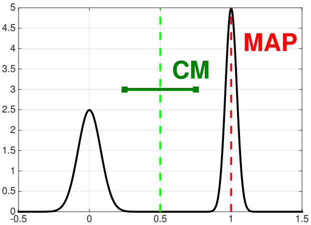


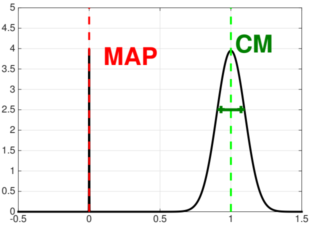


Thus, in general, neither represents the "center of mass" of $\pi_{u|y}$ ...

Comparison between PM and MAP

- Posterior mean $u^y_{\mathsf{PM}} = \int_{\mathcal{X}} u\pi_{u|y}(u)\, \mathrm{d}u$ computed via **numerical integration**
- MAP estimate $u^y_{\mathsf{MAP}} \in \underset{u\in\mathcal{X}}{\operatorname{argmax}}\, \pi_{u|y}(u)$ computed via **numerical optimization**
- Although for Gaussian $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$, we have $u^y_{\mathsf{PM}} = u^y_{\mathsf{MAP}}$, in general, they yield very different estimates:





Thus, in general, neither represents the "center of mass" of $\pi_{u|y}$ ...

... but for unimodal $\pi_{u|y}$ they are often not too different.

Choice of prior

- The choice of the prior distribution π_0 is crucial for the outcome of the resulting Bayesian data analysis or Bayesian inverse problem.
  (like the choice of the regularization functional $\mathcal{R}(u)$ in the deterministic approach)

- There are some principles for choosing the prior in the statistics literature:

Choice of prior

- The choice of the prior distribution π_0 is crucial for the outcome of the resulting Bayesian data analysis or Bayesian inverse problem.
  (like the choice of the regularization functional $\mathcal{R}(u)$ in the deterministic approach)

- There are some principles for choosing the prior in the statistics literature:

  1. **Conjugacy:** Given likelihood model $L(\cdot; u) = \pi_{y|u}(\cdot)$ choose the prior π_0 such that the posterior is analytically computable, e.g.,

     $$\text{Gaussian likelihood} + \text{Gaussian prior} \;\Rightarrow\; \text{Gaussian posterior}$$

Choice of prior

- The choice of the prior distribution π_0 is crucial for the outcome of the resulting Bayesian data analysis or Bayesian inverse problem.
  (like the choice of the regularization functional $\mathcal{R}(u)$ in the deterministic approach)

- There are some principles for choosing the prior in the statistics literature:

  1. **Conjugacy:** Given likelihood model $L(\cdot; u) = \pi_{y|u}(\cdot)$ choose the prior π_0 such that the posterior is analytically computable, e.g.,

     $$\text{Gaussian likelihood} + \text{Gaussian prior} \;\Rightarrow\; \text{Gaussian posterior}$$

  2. **Cromwells rule:** The prior should asign probability 0 only to logically or physically impossible events, i.e., the support of π_0 should always be $\mathbb{R}^n$

Choice of prior

- The choice of the prior distribution π_0 is crucial for the outcome of the resulting Bayesian data analysis or Bayesian inverse problem.
(like the choice of the regularization functional $\mathcal{R}(u)$ in the deterministic approach)

- There are some principles for choosing the prior in the statistics literature:

1. **Conjugacy:** Given likelihood model $L(\cdot;u) = \pi_{y|u}(\cdot)$ choose the prior π_0 such that the posterior is analytically computable, e.g.,

$$\text{Gaussian likelihood} + \text{Gaussian prior} \;\Rightarrow\; \text{Gaussian posterior}$$

2. **Cromwells rule:** The prior should asign probability 0 only to logically or physically impossible events, i.e., the support of π_0 should always be $\mathbb{R}^n$

3. **Uninformative/objective prior:** If we have no prior knowledge, then π_0 should be uninformative regarding the data model, which leads to so-called Jeffreys priors based on Fisher information:

Choice of prior

- The choice of the prior distribution π_0 is crucial for the outcome of the resulting Bayesian data analysis or Bayesian inverse problem.
  (like the choice of the regularization functional $\mathcal{R}(u)$ in the deterministic approach)

- There are some principles for choosing the prior in the statistics literature:

  1. **Conjugacy:** Given likelihood model $L(\cdot; u) = \pi_{y|u}(\cdot)$ choose the prior π_0 such that the posterior is analytically computable, e.g.,

     $$\text{Gaussian likelihood} + \text{Gaussian prior} \;\Rightarrow\; \text{Gaussian posterior}$$

  2. **Cromwells rule:** The prior should asign probability 0 only to logically or physically impossible events, i.e., the support of π_0 should always be $\mathbb{R}^n$

  3. **Uninformative/objective prior:** If we have no prior knowledge, then π_0 should be uninformative regarding the data model, which leads to so-called Jeffreys priors based on Fisher information:

     $$\pi_0(u) \;\propto\; \sqrt{\det \mathcal{I}^{\mathcal{F}}(u)}, \qquad \mathcal{I}^{\mathcal{F}}(u) := \mathbb{C}\mathrm{ov}_{\pi_u}\left(\nabla_u \log(\pi_u(Y))\right)$$

Asymptotics

- For more and more observed realizations $y_1, y_2, y_3, \ldots \in \mathbb{R}^d$ of an assumed true observable

$$Y = \mathcal{G}(u_{\mathsf{true}}) + \eta, \qquad \eta \sim \mathrm{N}(0, \Sigma),$$

Asymptotics

- For more and more observed realizations $y_1, y_2, y_3, \ldots \in \mathbb{R}^d$ of an assumed true observable
$$Y = \mathcal{G}(u_{\mathsf{true}}) + \eta, \qquad \eta \sim \mathrm{N}(0, \Sigma),$$
a natural question is whether posterior $\pi_{u|y_1,\ldots,y_m}$ concentrates around u_{true}.

Asymptotics

- For more and more observed realizations $y_1, y_2, y_3, \ldots \in \mathbb{R}^d$ of an assumed true observable
$$Y = \mathcal{G}(u_{\mathsf{true}}) + \eta, \qquad \eta \sim \mathrm{N}(0, \Sigma),$$
a natural question is whether posterior $\pi_{u|y_1,\ldots,y_m}$ concentrates around u_{true}.
- This requires of course, that u_{true} belongs to the support of the prior π_0, but such results do exist:

Asymptotics

- For more and more observed realizations $y_1, y_2, y_3, \ldots \in \mathbb{R}^d$ of an assumed true observable
$$Y = \mathcal{G}(u_{\mathsf{true}}) + \eta, \qquad \eta \sim \mathrm{N}(0, \Sigma),$$
a natural question is whether posterior $\pi_{u|y_1,\ldots,y_m}$ concentrates around u_{true}.
- This requires of course, that u_{true} belongs to the support of the prior π_0, but such results do exist:
- **Doob's Consistency Theorem** (informal): For $\mathcal{G}$ injective and any $r > 0$,
$$\lim_{m\to\infty} \pi_{u|y_1,\ldots,y_m}\Big(\{u\colon |u - u_{\mathsf{true}}| \le r\}\Big) = 1$$

Asymptotics

- For more and more observed realizations $y_1, y_2, y_3, \ldots \in \mathbb{R}^d$ of an assumed true observable
$$Y = \mathcal{G}(u_{\mathsf{true}}) + \eta, \qquad \eta \sim \mathrm{N}(0, \Sigma),$$
a natural question is whether posterior $\pi_{u|y_1,\ldots,y_m}$ concentrates around u_{true}.
- This requires of course, that u_{true} belongs to the support of the prior π_0, but such results do exist:
- **Doob's Consistency Theorem** (informal): For $\mathcal{G}$ injective and any $r > 0$,
$$\lim_{m\to\infty} \pi_{u|y_1,\ldots,y_m}\Big(\{u\colon |u - u_{\mathsf{true}}| \le r\}\Big) = 1$$
- **Bernstein–von Mises Theorem** (informal): Under suitable assumptions
$$\lim_{m\to\infty} d_{\mathsf{TV}}\Big(\pi_{u|y_1,\ldots,y_m}, \mathrm{N}(u_{\mathsf{PM}}^{y_1,\ldots,y_m}, \tfrac{1}{m}\mathcal{I}^{\mathcal{F}}(u_{\mathsf{true}}))\Big) = 0 \qquad \text{(asymptotically Gaussian)}$$
where d_{TV} denotes the total variation distance of probability measures
$$d_{\mathsf{TV}}(\pi, \mu) := \sup_{A \in \mathbb{R}^n} |\pi(A) - \mu(A)|.$$

Laplace Approximation

- Often expensive to compute the posterior distribution $\pi_{u|y}$ in practice and one seeks a simple approximation $\tilde{\pi}_{u|y}$.

Laplace Approximation

- Often expensive to compute the posterior distribution $\pi_{u|y}$ in practice and one seeks a simple approximation $\tilde{\pi}_{u|y}$.

- A classical approximation (again based on a Gaussian distribution) is the Laplace approximation:

$$\pi_{u|y}(u) \approx \tilde{\pi}_{u|y} = c \exp\Big(-\tfrac{1}{2}(u - u^y_{\mathsf{MAP}})^\top \nabla^2 \log \pi_{u|y}(u^y_{\mathsf{MAP}})(u - u^y_{\mathsf{MAP}})\Big)$$

Laplace Approximation

- Often expensive to compute the posterior distribution $\pi_{u|y}$ in practice and one seeks a simple approximation $\tilde{\pi}_{u|y}$.

- A classical approximation (again based on a Gaussian distribution) is the Laplace approximation:

$$\pi_{u|y}(u) \approx \tilde{\pi}_{u|y} = c \exp\Big(- \tfrac{1}{2}(u - u^y_{\mathsf{MAP}})^\top \nabla^2 \log \pi_{u|y}(u^y_{\mathsf{MAP}})(u - u^y_{\mathsf{MAP}})\Big)$$

- It is derived formally by a quadratic Taylor approximation, i.e.,

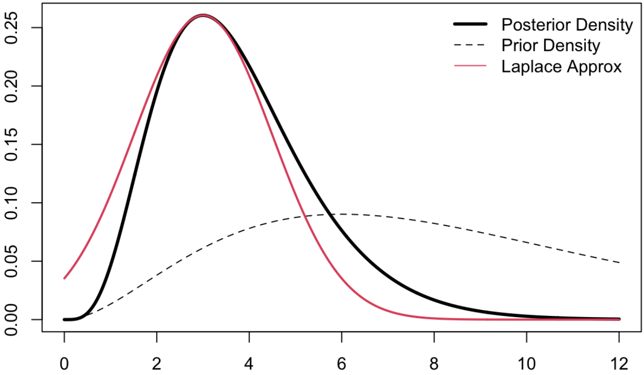
$$f(u) \approx f(u_*) - \nabla f(u_*)(u - u_*) - \frac{1}{2}(u - u_*)^\top \nabla^2 f(u_*)(u - u_*)$$

  with $\nabla^2 f(u_*)$ denoting the Hessian of f at u_*, of the negative log posterior density $f(u) := -\log \pi_{u|y}(u)$ around the MAP estimate $u_* := u^y_{\mathsf{MAP}}$.

Laplace Approximation (Illustration)

The posterior $\pi_{u|y}$ is approximated by a Gaussian distribtion $\mathrm{N}(u^y_{\mathsf{MAP}}, H^{-1})$ where

$$H := -\nabla^2 \log \pi_{u|y}(u^y_{\mathsf{MAP}})$$



Laplace approximation for inverse problems

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \mathrm{N}(0, C) \otimes \mathrm{N}(0, \Sigma)$$

Laplace approximation

$$\pi_{u|y} \approx \mathrm{N}(u^y_{\mathsf{PM}}, H^{-1})$$

where

$$u^y_{\mathsf{MAP}} = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \Phi(u) + \tfrac{1}{2} \|C^{-1/2} u\|^2 \right\}, \qquad \Phi(u) := \tfrac{1}{2} |y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

and

$$H := \nabla^2 \Phi(u^y_{\mathsf{MAP}}) + C$$

Thus, to compute the Laplace approximation relies only on numerical optimization and can be used for approximate Bayesian inference (without sampling).

Summary – Part I

- In the Bayesian approach all variables in the inverse problem are treated as random variables.
- The prior distribution π_0 for unknown u serves as probabilistic regularization.
- The solution of the Bayesian inverse problem is the posterior distribution $\pi_{u|y}$, the prior conditioned on the data y.
- The posterior describes/quantifies all remaining uncertainty about unknown u.
- MAP (maximum a posteriori) estimate is the Tikhonov-regularized solution.
- Another common point estimate for u is the posterior mean.
- Asymptotically the posterior concentrates around the ground truth and is approximately Gaussian (at least in finite dimensional Euclidean spaces).

Beyond Finite Dimensions

Infinite dimensional case

In general, we can consider Bayesian inverse problems

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

where $\mathcal{G} \colon \mathcal{X} \to \mathcal{Y}$ for separable infinite-dimensional Hilbert space $\mathcal{X}$, e.g., $L^2(D)$. For the most part $\mathcal{Y} = \mathbb{R}^d$, but also $\mathcal{Y}$ can be a separable ∞-dim'l Hilbert space.

Infinite dimensional case

In general, we can consider Bayesian inverse problems

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

where $\mathcal{G}\colon \mathcal{X} \to \mathcal{Y}$ for separable infinite-dimensional Hilbert space $\mathcal{X}$, e.g., $L^2(D)$. For the most part $\mathcal{Y} = \mathbb{R}^d$, but also $\mathcal{Y}$ can be a separable ∞-dim'l Hilbert space.

- We will not have time for a detailed treatment. It requires probability theory in Hilbert spaces. Only give a few pointers.

Infinite dimensional case

In general, we can consider Bayesian inverse problems

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

where $\mathcal{G}\colon \mathcal{X} \to \mathcal{Y}$ for separable infinite-dimensional Hilbert space $\mathcal{X}$, e.g., $L^2(D)$. For the most part $\mathcal{Y} = \mathbb{R}^d$, but also $\mathcal{Y}$ can be a separable ∞-dim'l Hilbert space.

- We will not have time for a detailed treatment. It requires probability theory in Hilbert spaces. Only give a few pointers.
- In particular, prior measures on $\mathcal{X}$ will now be random fields:

Random field (or stochastic process)

Considering domain $D \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$, a **random field or stochastic process** is a family of (real-valued) random variables $U = \{U_x \colon \Omega \to \mathbb{R} \,:\, x \in D\}$ such that $U\colon \Omega \times D \to \mathbb{R}$ is measurable. The function $u := U(\omega, \cdot)$, is called a **path** of U.

Infinite dimensional case

In general, we can consider Bayesian inverse problems

$$y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

where $\mathcal{G}\colon \mathcal{X} \to \mathcal{Y}$ for separable infinite-dimensional Hilbert space $\mathcal{X}$, e.g., $L^2(D)$. For the most part $\mathcal{Y} = \mathbb{R}^d$, but also $\mathcal{Y}$ can be a separable ∞-dim'l Hilbert space.

- We will not have time for a detailed treatment. It requires probability theory in Hilbert spaces. Only give a few pointers.
- In particular, prior measures on $\mathcal{X}$ will now be random fields:

Random field (or stochastic process)

Considering domain $D \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$, a **random field or stochastic process** is a family of (real-valued) random variables $U = \{U_x\colon \Omega \to \mathbb{R} \;:\; x \in D\}$ such that $U\colon \Omega \times D \to \mathbb{R}$ is measurable. The function $u := U(\omega, \cdot)$, is called a **path** of U.

Main workhorse: Gaussian random fields on $\mathcal{X} = L^2(D)$.

Random Fields

Gaussian random fields

Gaussian random field

A **Gaussian random field** on $D \subseteq \mathbb{R}^k$ is a stochastic process U such that for any $n \in \mathbb{N}$ and any points $x_1, \ldots, x_n \in D$ the random vector $(U(x_1), \ldots, U(x_n))^\top$ follows a multivariate normal distribution in $\mathbb{R}^n$.

- Gaussian random fields are uniquely determined by their first two moments:

mean function: $\qquad m(x) := \mathbb{E}\left[U(x)\right]$

covariance function: $\qquad c(x, x') := \mathbb{C}\mathrm{ov}(U(x), U(x'))$

where $m\colon D \to \mathbb{R}$ and $c\colon D \times D \to \mathbb{R}$ is a symmetric positive semidefinite fct.

Gaussian random fields

Gaussian random field

A **Gaussian random field** on $D \subseteq \mathbb{R}^k$ is a stochastic process U such that for any $n \in \mathbb{N}$ and any points $x_1, \ldots, x_n \in D$ the random vector $(U(x_1), \ldots, U(x_n))^\top$ follows a multivariate normal distribution in $\mathbb{R}^n$.

- Gaussian random fields are uniquely determined by their first two moments:

  mean function: $\quad m(x) := \mathbb{E}\left[U(x)\right]$

  covariance function: $\quad c(x, x') := \mathbb{C}\mathrm{ov}(U(x), U(x'))$

  where $m\colon D \to \mathbb{R}$ and $c\colon D \times D \to \mathbb{R}$ is a symmetric positive semidefinite fct.

- **Ex.:** Brownian motion on $D = [0, \infty)$ with $m \equiv 0$ and $c(x, x') = \min(x, x')$.

Gaussian random fields

Gaussian random field

A **Gaussian random field** on $D \subseteq \mathbb{R}^k$ is a stochastic process U such that for any $n \in \mathbb{N}$ and any points $x_1, \ldots, x_n \in D$ the random vector $(U(x_1), \ldots, U(x_n))^\top$ follows a multivariate normal distribution in $\mathbb{R}^n$.

- Gaussian random fields are uniquely determined by their first two moments:

$$\textbf{mean function:} \qquad m(x) := \mathbb{E}\left[U(x)\right]$$
$$\textbf{covariance function:} \qquad c(x,x') := \mathbb{C}\mathrm{ov}(U(x), U(x'))$$

  where $m\colon D \to \mathbb{R}$ and $c\colon D \times D \to \mathbb{R}$ is a symmetric positive semidefinite fct.

- **Ex.:** Brownian motion on $D = [0, \infty)$ with $m \equiv 0$ and $c(x,x') = \min(x,x')$.

- Often in UQ practice, parametric models for m and c are used, e.g.,

$$m(x) = \sum_{j=1}^{J} \beta_j f_j(x), \qquad c(x,x') = c_\theta(x,x')$$

  where parameters $\beta \in \mathbb{R}^J$ and $\theta \in \mathbb{R}^p$ are given or estimated from data.

Matérn covariance functions

- The covariance (and mean) function determine smoothness properties of the paths $u\colon D \to \mathbb{R}$ of a Gaussian random field U.

- A common parametrized class are the Matérn covariance functions

$$c_{\sigma^2,\rho,\nu}(x,x') := \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,|x-x'|}{\rho}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}\,|x-x'|}{\rho}\right)$$

with variance $\sigma^2 > 0$, correlation length $\rho > 0$, smoothness $\nu > 0$.
K_ν is the modified (2nd-kind) Bessel function of order ν and Γ is the Gamma-function.

Matérn covariance functions

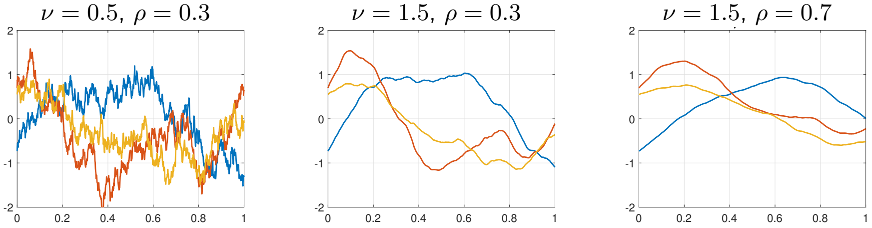
- The covariance (and mean) function determine smoothness properties of the paths $u \colon D \to \mathbb{R}$ of a Gaussian random field U.

- A common parametrized class are the Matérn covariance functions

$$c_{\sigma^2,\rho,\nu}(x,x') := \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,|x-x'|}{\rho}\right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}\,|x-x'|}{\rho}\right)$$

with variance $\sigma^2 > 0$, correlation length $\rho > 0$, smoothness $\nu > 0$.

K_ν is the modified (2nd-kind) Bessel function of order ν and Γ is the Gamma-function.



Measures and Moments in Hilbert space

- Considering a probability measure π on a separable Hilbert space $\mathcal{X}$ or its Borel-σ-algebra $\mathcal{B}(\mathcal{X})$, respectively.
- For $q \in \mathbb{N}_0$, we denote by $\mathcal{P}^q(\mathcal{X})$ all probability measures π in $\mathcal{X}$ which satisfy
$$\int_{\mathcal{X}} \|u\|^q \, \pi(\mathrm{d}u) < +\infty.$$
We say $U \sim \pi$ for $\pi \in \mathcal{P}^q(\mathcal{X})$ belongs to Lebesgue–Bochner space $L^q(\Omega; \mathcal{X})$.

Measures and Moments in Hilbert space

- Considering a probability measure π on a separable Hilbert space $\mathcal{X}$ or its Borel-σ-algebra $\mathcal{B}(\mathcal{X})$, respectively.
- For $q \in \mathbb{N}_0$, we denote by $\mathcal{P}^q(\mathcal{X})$ all probability measures π in $\mathcal{X}$ which satisfy
$$\int_{\mathcal{X}} \|u\|^q \, \pi(\mathrm{d}u) < +\infty.$$
We say $U \sim \pi$ for $\pi \in \mathcal{P}^q(\mathcal{X})$ belongs to Lebesgue–Bochner space $L^q(\Omega; \mathcal{X})$.
- The mean value $m_\pi \in \mathcal{X}$ of $\pi \in \mathcal{P}^1(\mathcal{X})$ of the r.v. $U \sim \pi$ is given by
$$m_\pi := \mathbb{E}_\pi\left[U\right] = \int_{\mathcal{X}} u \, \pi(\mathrm{d}u)$$

Measures and Moments in Hilbert space

- Considering a probability measure π on a separable Hilbert space $\mathcal{X}$ or its Borel-σ-algebra $\mathcal{B}(\mathcal{X})$, respectively.
- For $q \in \mathbb{N}_0$, we denote by $\mathcal{P}^q(\mathcal{X})$ all probability measures π in $\mathcal{X}$ which satisfy
$$\int_{\mathcal{X}} \|u\|^q \, \pi(\mathrm{d}u) < +\infty.$$
We say $U \sim \pi$ for $\pi \in \mathcal{P}^q(\mathcal{X})$ belongs to Lebesgue–Bochner space $L^q(\Omega; \mathcal{X})$.
- The mean value $m_\pi \in \mathcal{X}$ of $\pi \in \mathcal{P}^1(\mathcal{X})$ of the r.v. $U \sim \pi$ is given by
$$m_\pi := \mathbb{E}_\pi [U] = \int_{\mathcal{X}} u \, \pi(\mathrm{d}u)$$
- The covariance $C_\pi \in \mathcal{L}(\mathcal{X})$ of $\pi \in \mathcal{P}^2(\mathcal{X})$ is given by the unique linear, bounded operator satisfying
$$\langle v, C_\pi w \rangle = \int_{\mathcal{X}} \langle v, (u - m_\pi) \rangle \langle w, (u - m_\pi) \rangle \, \pi(\mathrm{d}u) \qquad \forall v, w \in \mathcal{X}.$$

Measures and Moments in Hilbert space

- Considering a probability measure π on a separable Hilbert space $\mathcal{X}$ or its Borel-σ-algebra $\mathcal{B}(\mathcal{X})$, respectively.
- For $q \in \mathbb{N}_0$, we denote by $\mathcal{P}^q(\mathcal{X})$ all probability measures π in $\mathcal{X}$ which satisfy

$$\int_{\mathcal{X}} \|u\|^q \, \pi(\mathrm{d}u) < +\infty.$$

  We say $U \sim \pi$ for $\pi \in \mathcal{P}^q(\mathcal{X})$ belongs to Lebesgue–Bochner space $L^q(\Omega; \mathcal{X})$.
- The mean value $m_\pi \in \mathcal{X}$ of $\pi \in \mathcal{P}^1(\mathcal{X})$ of the r.v. $U \sim \pi$ is given by

$$m_\pi := \mathbb{E}_\pi[U] = \int_{\mathcal{X}} u \, \pi(\mathrm{d}u)$$

- The covariance $C_\pi \in \mathcal{L}(\mathcal{X})$ of $\pi \in \mathcal{P}^2(\mathcal{X})$ is given by the unique linear, bounded operator satisfying

$$\langle v, C_\pi w \rangle = \int_{\mathcal{X}} \langle v, (u - m_\pi) \rangle \langle w, (u - m_\pi) \rangle \, \pi(\mathrm{d}u) \qquad \forall v, w \in \mathcal{X}.$$

- Considering the tensor product $\mathcal{X} \otimes \mathcal{X}$ we can identify C_π for $U \sim \pi$ with

$$\mathbb{C}\mathrm{ov}(U) = \mathbb{E}\left[(U - \mathbb{E}[U]) \otimes (U - \mathbb{E}[U])\right].$$

Densities in Hilbert space

- On infinite dimensional Hilbert spaces $\mathcal{X}$ there exists no Lebesgue measure! Hence, **cannot** work with 'simple' probability density fcts. $\pi\colon \mathcal{X} \to [0, \infty)$ s.t.

$$\pi(A) = \int_A \pi(u)\,\mathrm{d}u, \qquad A \subset \mathcal{X}.$$

Densities in Hilbert space

- On infinite dimensional Hilbert spaces $\mathcal{X}$ there exists no Lebesgue measure! Hence, **cannot** work with 'simple' probability density fcts. $\pi\colon \mathcal{X} \to [0,\infty)$ s.t.

$$\pi(A) = \int_A \pi(u)\,\mathrm{d}u, \qquad A \subset \mathcal{X}.$$

- However, given a suitable reference (probability) measure μ on $\mathcal{X}$ we may define the density of π with respect to μ.

Radon–Nikodym theorem

We say $\pi \in \mathcal{P}(\mathcal{X})$ is absolutely continuous w.r.t. or dominated by $\mu \in \mathcal{P}(\mathcal{X})$ if

$$\mu(A) = 0 \quad \Rightarrow \quad \pi(A) = 0, \qquad A \subseteq \mathcal{X}.$$

Densities in Hilbert space

- On infinite dimensional Hilbert spaces $\mathcal{X}$ there exists no Lebesgue measure! Hence, **cannot** work with 'simple' probability density fcts. $\pi\colon \mathcal{X} \to [0,\infty)$ s.t.

$$\pi(A) = \int_A \pi(u)\,\mathrm{d}u, \qquad A \subset \mathcal{X}.$$

- However, given a suitable reference (probability) measure μ on $\mathcal{X}$ we may define the density of π with respect to μ.

Radon–Nikodym theorem

We say $\pi \in \mathcal{P}(\mathcal{X})$ is absolutely continuous w.r.t. or dominated by $\mu \in \mathcal{P}(\mathcal{X})$ if

$$\mu(A) = 0 \quad \Rightarrow \quad \pi(A) = 0, \qquad A \subseteq \mathcal{X}.$$

In that case, we write $\pi \ll \mu$ and there exists a density $f\colon \mathcal{X} \to [0,\infty)$ such that

$$\pi(A) = \int_A f(u)\,\mu(\mathrm{d}u), \qquad \forall A \subseteq \mathcal{X}.$$

We denote f by $\dfrac{\mathrm{d}\pi}{\mathrm{d}\mu}$ and call it Radon–Nikodym derivative/density of π w.r.t. μ.

Conditional Distribution

- Since we cannot work with conditional probability densities anymore, need more general notions, i.e., conditional distributions (requires some technicalities).

Conditional Distribution

- Since we cannot work with conditional probability densities anymore, need more general notions, i.e., conditional distributions (requires some technicalities).

Stochastic kernel

A mapping $K\colon \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ is called a **stochastic kernel** if

- $K(y,\cdot)$ is a probabiliy measure on $\mathcal{X}$ for each $y \in \mathcal{Y}$
- $K(\cdot, A)$ is a measurable for each $A \in \mathcal{B}(\mathcal{X})$.

Conditional Distribution

- Since we cannot work with conditional probability densities anymore, need more general notions, i.e., conditional distributions (requires some technicalities).

Stochastic kernel

A mapping $K\colon \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ is called a **stochastic kernel** if

- $K(y,\cdot)$ is a probabiliy measure on $\mathcal{X}$ for each $y \in \mathcal{Y}$
- $K(\cdot, A)$ is a measurable for each $A \in \mathcal{B}(\mathcal{X})$.

Regular conditional distribution

Let $U\colon \Omega \to \mathcal{X}$ and $Y\colon \Omega \to Y$ with joint distribution $(U,Y) \sim \pi$ and marginal $Y \sim \pi_Y$, then the **regular conditional distribution** of U given Y is a stochastic kernel $\pi_{U|Y}\colon \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ such that

$$\int_B \pi_{U|Y}(y,A)\,\pi_Y(\mathrm{d}y) = \pi(A \times B) = \mathbb{P}(U \in A, Y \in B), \quad \forall A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y})$$

Bayes' rule for conditional distributions

Theorem

Let $U : \Omega \to \mathcal{X}$ and $Y : \Omega \to Y$ such that $U \sim \pi_U$, $Y \sim \pi_Y$ with joint distribution $(U, Y) \sim \pi$ written as

$$\pi(\mathrm{d}u\,\mathrm{d}y) = \pi_{Y|U}(u, \mathrm{d}y)\,\pi_U(\mathrm{d}u).$$

Bayes' rule for conditional distributions

Theorem

Let $U\colon \Omega \to \mathcal{X}$ and $Y : \Omega \to Y$ such that $U \sim \pi_U$, $Y \sim \pi_Y$ with joint distribution $(U, Y) \sim \pi$ written as

$$\pi(\mathrm{d}u\,\mathrm{d}y) = \pi_{Y|U}(u, \mathrm{d}y)\,\pi_U(\mathrm{d}u).$$

If there exists a (measurable) $L\colon \mathcal{Y} \times \mathcal{X} \to [0, \infty)$ and a measure μ on $\mathcal{Y}$ dominating each conditional distribution $\pi_{Y|U}(u, \cdot) \ll \mu$, $u \in \mathcal{X}$, such that

$$\frac{\mathrm{d}\pi_{Y|U}(u)}{\mathrm{d}\mu}(y) = L(y; u),$$

Bayes' rule for conditional distributions

Theorem

Let $U\colon \Omega \to \mathcal{X}$ and $Y\colon \Omega \to Y$ such that $U \sim \pi_U$, $Y \sim \pi_Y$ with joint distribution $(U,Y) \sim \pi$ written as

$$\pi(\mathrm{d}u\,\mathrm{d}y) = \pi_{Y|U}(u,\mathrm{d}y)\,\pi_U(\mathrm{d}u).$$

If there exists a (measurable) $L\colon \mathcal{Y} \times \mathcal{X} \to [0,\infty)$ and a measure μ on $\mathcal{Y}$ dominating each conditional distribution $\pi_{Y|U}(u,\cdot) \ll \mu$, $u \in \mathcal{X}$, such that

$$\frac{\mathrm{d}\pi_{Y|U}(u)}{\mathrm{d}\mu}(y) = L(y;u),$$

then provided $Z(y) := \int_{\mathcal{X}} L(y;u)\,\pi_U(\mathrm{d}u) < \infty$, we have

$$\pi_{U|Y}(y,\mathrm{d}u) = \frac{1}{Z(y)}\,L(y;u)\,\pi_U(\mathrm{d}u).$$

Bayes' rule for conditional distributions

Theorem

Let $U : \Omega \to \mathcal{X}$ and $Y : \Omega \to Y$ such that $U \sim \pi_U$, $Y \sim \pi_Y$ with joint distribution $(U, Y) \sim \pi$ written as

$$\pi(\mathrm{d}u\,\mathrm{d}y) = \pi_{Y|U}(u, \mathrm{d}y)\,\pi_U(\mathrm{d}u).$$

If there exists a (measurable) $L : \mathcal{Y} \times \mathcal{X} \to [0, \infty)$ and a measure μ on $\mathcal{Y}$ dominating each conditional distribution $\pi_{Y|U}(u, \cdot) \ll \mu$, $u \in \mathcal{X}$, such that

$$\frac{\mathrm{d}\pi_{Y|U}(u)}{\mathrm{d}\mu}(y) = L(y; u),$$

then provided $Z(y) := \int_{\mathcal{X}} L(y; u)\,\pi_U(\mathrm{d}u) < \infty$, we have

$$\pi_{U|Y}(y, \mathrm{d}u) = \frac{1}{Z(y)}\,L(y; u)\,\pi_U(\mathrm{d}u). \qquad \textbf{(Bayes' rule)}$$

- π_U is the prior probability measure;
- $L(y; u)$ is the likelihood function for $Y = y$ given $U = u$;
- $1/Z(y)$ is a normalizing constant with evidence $Z(y)$.

Bayesian inverse problems

Bayesian data model (for $\mathcal{Y} = \mathbb{R}^d$ and Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad (U,\eta) \sim \pi_0 \otimes \mathrm{N}(0,\Sigma)$$

Bayesian inverse problems

Bayesian data model (for $\mathcal{Y} = \mathbb{R}^d$ and Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- The data model yields as conditional distributions

$$\pi_{Y|U}(u) = \mathrm{N}(\mathcal{G}(u), \Sigma), \qquad u \in \mathcal{X},$$

- These are dominated, for any $y \in \mathbb{R}^d$, by Lebesgue measure with likelihood function

$$L(y; u) = c_\Sigma \, \exp(-\Phi(u; y)), \qquad \Phi(u; y) := \frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

where $\Phi\colon \mathcal{X} \times \mathbb{R}^d \to [0, \infty)$ is called the potential.

Bayesian inverse problems

Bayesian data model (for $\mathcal{Y} = \mathbb{R}^d$ and Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

- The data model yields as conditional distributions

$$\pi_{Y|U}(u) = \mathrm{N}(\mathcal{G}(u), \Sigma), \qquad u \in \mathcal{X},$$

- These are dominated, for any $y \in \mathbb{R}^d$, by Lebesgue measure with likelihood function

$$L(y; u) = c_\Sigma \exp(-\Phi(u; y)), \qquad \Phi(u; y) := \tfrac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

where $\Phi\colon \mathcal{X} \times \mathbb{R}^d \to [0, \infty)$ is called the potential.

Thus, by Bayes' rule, the solution to the **Bayesian inverse problem** in an **infinite-dimensional** Hilbert space $\mathcal{X}$ is the posterior distribution $\pi_{u|y}$ of U given $Y = y$

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u), \;\; Z(y) = \int_X \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u) \in (0, 1].$$

Bayesian inverse problems

Bayesian data model (for $\mathcal{Y} = \mathbb{R}^d$ and Gaussian noise)

$$Y = \mathcal{G}(U) + \eta, \qquad \mathcal{G}\colon \mathcal{X} \to \mathbb{R}^d, \quad (U,\eta) \sim \pi_0 \otimes \mathrm{N}(0,\Sigma)$$

- The data model yields as conditional distributions

$$\pi_{Y|U}(u) = \mathrm{N}(\mathcal{G}(u),\Sigma), \qquad u \in \mathcal{X},$$

- These are dominated, for any $y \in \mathbb{R}^d$, by Lebesgue measure with likelihood function

$$L(y;u) = c_\Sigma \exp(-\Phi(u;y)), \qquad \Phi(u;y) := \tfrac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$$

where $\Phi\colon \mathcal{X} \times \mathbb{R}^d \to [0,\infty)$ is called the potential.

Thus, by Bayes' rule, the solution to the **Bayesian inverse problem** in an **infinite-dimensional** Hilbert space $\mathcal{X}$ is the posterior distribution $\pi_{u|y}$ of U given $Y = y$

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u;y))\,\pi_0(\mathrm{d}u), \;\; Z(y) = \int_X \exp(-\Phi(u;y))\,\pi_0(\mathrm{d}u) \in (0,1].$$

Note. Since $0 < \exp(-\Phi(u;y)) \le 1 \quad \Rightarrow \quad 0 < \int_{\mathcal{X}} \exp(-\Phi(u;y))\,\pi_0(\mathrm{d}u) \le 1$ for any prior π_0.

Gaussian Linear Model

Given observation $y \in \mathbb{R}^d$ of

$$Y = AU + \eta, \qquad \mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \quad (U, \eta) \sim \mathrm{N}(u_0, C) \otimes \mathrm{N}(0, \Sigma)$$

where $A \in \mathcal{L}(\mathcal{X}, \mathbb{R}^d)$, the solution to the Bayesian inverse problem is the Gaussian posterior distribution $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$ where

$$u^y_{\mathsf{PM}} = (A^* \Sigma^{-1} A + C^{-1})^{-1} \left[\Sigma A^* y + C^{-1} u_0 \right]$$
$$C^y = \left(A^* \Sigma^{-1} A + C^{-1} \right)^{-1}$$

[1] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, A. Spantini. Likelihood-Informed Dimension Reduction for Nonlinear Inverse Problems. *Inverse Problems*, 30(11), 2014.

Gaussian Linear Model

Given observation $y \in \mathbb{R}^d$ of

$$Y = AU + \eta, \qquad \mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d, \quad (U, \eta) \sim \mathrm{N}(u_0, C) \otimes \mathrm{N}(0, \Sigma)$$

where $A \in \mathcal{L}(\mathcal{X}, \mathbb{R}^d)$, the solution to the Bayesian inverse problem is the Gaussian posterior distribution $\pi_{u|y} = \mathrm{N}(u^y_{\mathsf{PM}}, C^y)$ where

$$u^y_{\mathsf{PM}} = (A^*\Sigma^{-1}A + C^{-1})^{-1} \left[\Sigma A^* y + C^{-1} u_0\right]$$
$$C^y = \left(A^*\Sigma^{-1}A + C^{-1}\right)^{-1}$$

- The operators $CA^\top(ACA^* + \Sigma)^{-1}AC$ and $A^*\Sigma^{-1}A$ have finite rank $r \leq d$
- **Thus, only in the r-dimensional subspace $\mathcal{R}(A^*\Sigma^{-1}A)$ there is a change from prior to posterior!**
- The marginal of $\pi_{u|y}$ in $\mathcal{N}(A)$ coincides with the corresponding prior marginal
- Such "active" subspaces can be exploited for nonlinear forward maps $\mathcal{G}$ too[1]

[1] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, A. Spantini. Likelihood-Informed Dimension Reduction for Nonlinear Inverse Problems. *Inverse Problems*, 30(11), 2014.

Well-Posedness of Bayesian Inverse Problems

Well-posedness

So far, we have seen, that Bayesian inverse problems admit under mild assumptions a unique solution — the posterior distribution

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u), \quad Z(y) = \int_{\mathcal{X}} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u).$$

Thus, two of three conditions for well-posedness are satisfied.

How about the third, the continuous dependence on the data $y \in \mathcal{Y}$?

Well-posedness

So far, we have seen, that Bayesian inverse problems admit under mild assumptions a unique solution — the posterior distribution

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u), \quad Z(y) = \int_{\mathcal{X}} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u).$$

Thus, two of three conditions for well-posedness are satisfied.

How about the third, the continuous dependence on the data $y \in \mathcal{Y}$?

- To this end, we require suitable distances for probability measures.

Well-posedness

So far, we have seen, that Bayesian inverse problems admit under mild assumptions a unique solution — the posterior distribution

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u;y))\,\pi_0(\mathrm{d}u), \quad Z(y) = \int_{\mathcal{X}} \exp(-\Phi(u;y))\,\pi_0(\mathrm{d}u).$$

Thus, two of three conditions for well-posedness are satisfied.

How about the third, the continuous dependence on the data $y \in \mathcal{Y}$?

- To this end, we require suitable distances for probability measures.
- More than 70 metrics are known. We will use *Hellinger distance* (see below).

Well-posedness

So far, we have seen, that Bayesian inverse problems admit under mild assumptions a unique solution — the posterior distribution

$$\pi_{u|y}(\mathrm{d}u) = \frac{1}{Z(y)} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u), \quad Z(y) = \int_{\mathcal{X}} \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u).$$

Thus, two of three conditions for well-posedness are satisfied.

How about the third, the continuous dependence on the data $y \in \mathcal{Y}$?

- To this end, we require suitable distances for probability measures.
- More than 70 metrics are known. We will use *Hellinger distance* (see below).
- We focus in the following again on the case of finite-dimensional data and Gaussian noise, i.e.,

$$\mathcal{Y} = \mathcal{R}^d \quad \text{and} \quad \Phi(u;y) := \tfrac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}.$$

However, similar results can be obtained for infinite-dimensional data.

Hellinger distance – Properties

The *Hellinger distance* is given by

$$d_{\mathsf{Hell}}(\pi, \tilde{\pi}) = \sqrt{\int_{\mathcal{X}} \left| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}(u)} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}(u)} \right|^2 \mu(\mathrm{d}u)} = \left\| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}} \right\|_{L^2_\mu}$$

where μ is a common dominating measure of $\pi, \tilde{\pi}$, e.g., $\mu = \pi + \tilde{\pi}$.

Hellinger distance – Properties

The *Hellinger distance* is given by

$$d_{\mathsf{Hell}}(\pi, \tilde{\pi}) = \sqrt{\int_{\mathcal{X}} \left| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}(u)} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}(u)} \right|^2 \mu(\mathrm{d}u)} = \left\| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}} \right\|_{L^2_\mu}$$

where μ is a common dominating measure of $\pi, \tilde{\pi}$, e.g., $\mu = \pi + \tilde{\pi}$.

- Hellinger distance is topologically equivalent to total variation distance:

$$\frac{1}{2} d_{\mathsf{Hell}}(\pi, \tilde{\pi})^2 \leq d_{\mathsf{TV}}(\pi, \tilde{\pi}) \leq d_{\mathsf{Hell}}(\pi, \tilde{\pi})$$

Hellinger distance – Properties

The *Hellinger distance* is given by

$$d_{\mathsf{Hell}}(\pi, \tilde{\pi}) = \sqrt{\int_{\mathcal{X}} \left| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}(u)} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}(u)} \right|^2 \mu(\mathrm{d}u)} = \left\| \sqrt{\frac{\mathrm{d}\pi}{\mathrm{d}\mu}} - \sqrt{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\mu}} \right\|_{L^2_\mu}$$

where μ is a common dominating measure of $\pi, \tilde{\pi}$, e.g., $\mu = \pi + \tilde{\pi}$.

- Hellinger distance is topologically equivalent to total variation distance:

$$\frac{1}{2} d_{\mathsf{Hell}}(\pi, \tilde{\pi})^2 \leq d_{\mathsf{TV}}(\pi, \tilde{\pi}) \leq d_{\mathsf{Hell}}(\pi, \tilde{\pi})$$

- For any $f \in L^2_\pi(\mathcal{X}, \mathcal{Y}) \cap L^2_{\tilde{\pi}}(\mathcal{X}, \mathcal{Y})$ we have (**Exercise** – *Hint:* Cauchy-Schwarz!)

$$|\mathbb{E}_\pi [f] - \mathbb{E}_{\tilde{\pi}} [f]| \leq \sqrt{2\|f\|^2_{L^2_\pi} + 2\|f\|^2_{L^2_{\tilde{\pi}}}}\, d_{\mathsf{Hell}}(\pi, \tilde{\pi}),$$

  i.e., Hellinger distance allows to control differences in mean and covariance.

A technical lemma

Lemma

Given two potentials $\Phi_1, \Phi_2 \colon \mathcal{X} \to [0, \infty)$, let $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ be given by

$$\pi_i(\mathrm{d}u) = \frac{1}{Z_i} \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u), \qquad Z_i := \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u).$$

Then

A technical lemma

Lemma

Given two potentials $\Phi_1, \Phi_2 \colon \mathcal{X} \to [0, \infty)$, let $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ be given by

$$\pi_i(\mathrm{d}u) = \frac{1}{Z_i} \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u), \qquad Z_i := \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u).$$

Then

$$d_{\mathsf{Hell}}(\pi_1, \pi_2) \leq \frac{1}{\min(Z_1, Z_2)} \|\Phi_1 - \Phi_2\|_{L^2_{\pi_0}}.$$

A technical lemma

Lemma

Given two potentials $\Phi_1, \Phi_2 \colon \mathcal{X} \to [0, \infty)$, let $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ be given by

$$\pi_i(\mathrm{d}u) = \frac{1}{Z_i} \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u), \qquad Z_i := \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u).$$

Then

$$d_{\mathsf{Hell}}(\pi_1, \pi_2) \leq \frac{1}{\min(Z_1, Z_2)} \|\Phi_1 - \Phi_2\|_{L^2_{\pi_0}}.$$

where $\min(Z_1, Z_2) \geq \exp(-\|\Phi_1\|_{L^1_{\pi_0}} - \|\Phi_1 - \Phi_2\|_{L^1_{\pi_0}})$.

A technical lemma

Lemma

Given two potentials $\Phi_1, \Phi_2 \colon \mathcal{X} \to [0, \infty)$, let $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ be given by

$$\pi_i(\mathrm{d}u) = \frac{1}{Z_i} \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u), \qquad Z_i := \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u).$$

Then

$$d_{\mathsf{Hell}}(\pi_1, \pi_2) \leq \frac{1}{\min(Z_1, Z_2)} \|\Phi_1 - \Phi_2\|_{L^2_{\pi_0}}.$$

where $\min(Z_1, Z_2) \geq \exp(-\|\Phi_1\|_{L^1_{\pi_0}} - \|\Phi_1 - \Phi_2\|_{L^1_{\pi_0}})$.

- Consider now our Gaussian potential $\Phi(u; y) := \frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$.

A technical lemma

Lemma

Given two potentials $\Phi_1, \Phi_2 \colon \mathcal{X} \to [0,\infty)$, let $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ be given by

$$\pi_i(\mathrm{d}u) = \frac{1}{Z_i} \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u), \qquad Z_i := \exp(-\Phi_i(u))\, \pi_0(\mathrm{d}u).$$

Then

$$d_{\mathsf{Hell}}(\pi_1, \pi_2) \leq \frac{1}{\min(Z_1, Z_2)} \|\Phi_1 - \Phi_2\|_{L^2_{\pi_0}}.$$

where $\min(Z_1, Z_2) \geq \exp(-\|\Phi_1\|_{L^1_{\pi_0}} - \|\Phi_1 - \Phi_2\|_{L^1_{\pi_0}})$.

- Consider now our Gaussian potential $\Phi(u; y) := \frac{1}{2}|y - \mathcal{G}(u)|^2_{\Sigma^{-1}}$.
- For any two observations $y, \tilde{y} \in \mathbb{R}^d$ it follows easily that (**Exercise**)
  (using the identity $|a^2 - b^2| = |a+b|\,|a-b|$)

$$\|\Phi(\cdot; y) - \Phi(\cdot; \tilde{y})\|_{L^2_{\pi_0}} \leq c \sqrt{2 \max\{|y|^2, |\tilde{y}|^2\} + 2\, \|\mathcal{G}\|_{L^2_{\pi_0}}}\ |y - \tilde{y}|$$

Continuous dependence on the data

Theorem (Continuous dependence on the data)

Assume that $\mathcal{G} \in L^2_{\pi_0}(\mathcal{X};\mathbb{R}^d)$. Then, for each $r > 0$ there exists a constant c_r such that for the posterior measures resulting from the data model

$$Y = \mathcal{G}(U) + \eta, \qquad (U,\eta) \sim \pi_0 \otimes \mathrm{N}(0,\Sigma)$$

we have

$$d_{\mathsf{Hell}}(\pi_{u|y}, \pi_{u|\tilde{y}}) \leq c_r |y - \tilde{y}| \qquad \forall y, \tilde{y} \colon |y|, |\tilde{y}| \leq r$$

Continuous dependence on the data & on the forward map

Theorem (Continuous dependence on the data)

Assume that $\mathcal{G} \in L^2_{\pi_0}(\mathcal{X};\mathbb{R}^d)$. Then, for each $r > 0$ there exists a constant c_r such that for the posterior measures resulting from the data model

$$Y = \mathcal{G}(U) + \eta, \qquad (U,\eta) \sim \pi_0 \otimes \mathrm{N}(0,\Sigma)$$

we have

$$d_{\mathsf{Hell}}(\pi_{u|y}, \pi_{u|\tilde{y}}) \le c_r |y - \tilde{y}| \qquad \forall y, \tilde{y} \colon |y|, |\tilde{y}| \le r$$

Theorem (Continuous dependence on the forward map)

Consider a (numerical) approximation $\mathcal{G}_h \colon \mathcal{X} \to \mathbb{R}^d$ of $\mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d$ such that

$$|\mathcal{G}_h(u) - \mathcal{G}(u)| \le c(u)\ h^p \qquad \forall u \in \mathcal{X}$$

for $h \to 0$ and that $|\mathcal{G}_h(u)|, |\mathcal{G}(u)| \le g(u)$, $u \in \mathcal{X}$, with $c,\ g \in L^2_{\pi_0}(\mathcal{X};\mathbb{R})$, and let $\pi^h_{u|y}$ be the approximate posterior measures using $\mathcal{G}_h(U)$ instead of $\mathcal{G}(U)$.

Continuous dependence on the data & on the forward map

Theorem (Continuous dependence on the data)

Assume that $\mathcal{G} \in L^2_{\pi_0}(\mathcal{X};\mathbb{R}^d)$. Then, for each $r > 0$ there exists a constant c_r such that for the posterior measures resulting from the data model

$$Y = \mathcal{G}(U) + \eta, \qquad (U, \eta) \sim \pi_0 \otimes \mathrm{N}(0, \Sigma)$$

we have

$$d_{\mathsf{Hell}}(\pi_{u|y}, \pi_{u|\tilde{y}}) \leq c_r |y - \tilde{y}| \qquad \forall y, \tilde{y} \colon |y|, |\tilde{y}| \leq r$$

Theorem (Continuous dependence on the forward map)

Consider a (numerical) approximation $\mathcal{G}_h \colon \mathcal{X} \to \mathbb{R}^d$ of $\mathcal{G} \colon \mathcal{X} \to \mathbb{R}^d$ such that

$$|\mathcal{G}_h(u) - \mathcal{G}(u)| \leq c(u)\ h^p \qquad \forall u \in \mathcal{X}$$

for $h \to 0$ and that $|\mathcal{G}_h(u)|, |\mathcal{G}(u)| \leq g(u)$, $u \in \mathcal{X}$, with $c,\ g \in L^2_{\pi_0}(\mathcal{X};\mathbb{R})$, and let $\pi^h_{u|y}$ be the approximate posterior measures using $\mathcal{G}_h(U)$ instead of $\mathcal{G}(U)$. Then

$$d_{\mathsf{Hell}}(\pi_{u|y}, \pi^h_{u|y}) \leq c_r h^p \qquad \forall y \colon |y| \leq r$$

Bayes estimates

As in finite dimensions, we can derive Bayesian point estimates for the unknown $u \in \mathcal{X}$ based on the posterior $\pi_{u|y}$, such as the

Bayes estimates

As in finite dimensions, we can derive Bayesian point estimates for the unknown $u \in \mathcal{X}$ based on the posterior $\pi_{u|y}$, such as the

Posterior mean:

- Given $\pi_0 \in \mathcal{P}^1(\mathcal{X})$ and the above assumptions for existence of $\pi_{u|y}$. Then

$$u^y_{\mathsf{PM}} = \int_{\mathcal{X}} u \, \pi_{u|y}(\mathrm{d}u)$$

  is well-defined.

Bayes estimates

As in finite dimensions, we can derive Bayesian point estimates for the unknown $u \in \mathcal{X}$ based on the posterior $\pi_{u|y}$, such as the

Posterior mean:

- Given $\pi_0 \in \mathcal{P}^1(\mathcal{X})$ and the above assumptions for existence of $\pi_{u|y}$. Then

$$u^y_{\mathsf{PM}} = \int_{\mathcal{X}} u \, \pi_{u|y}(\mathrm{d}u)$$

  is well-defined.

- Given the assumptions for the well-posedness theorems and $\pi_0 \in \mathcal{P}^2(\mathcal{X})$,

$$\left\| u^y_{\mathsf{PM}} - u^{\tilde{y}}_{\mathsf{PM}} \right\| \le C_r d_{\mathsf{Hell}}(\pi_{u|y}, \pi_{u|\tilde{y}}) \le C_r \|y - \tilde{y}\| \qquad \forall y, \tilde{y} \colon |y|, |\tilde{y}| \le r$$

  and

$$\left\| u^y_{\mathsf{PM}} - u^y_{\mathsf{PM},h} \right\| \le C_r d_{\mathsf{Hell}}(\pi_{u|y}, \pi^h_{u|y}) \le C_r h^p \qquad \forall y \colon |y| \le r.$$

Further comments

- Again can also define the MAP estimate in infinite dimensions, although in general not by maximization of the posterior probability density function...

Further comments

- Again can also define the MAP estimate in infinite dimensions, although in general not by maximization of the posterior probability density function...
- In the case of $y \in \mathbb{R}^d$, for Gaussian prior $u \sim \mathrm{N}(0, C)$ and Gaussian noise, i.e., $\pi_{u|y}(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u)$, we recover again Tikhonov-Philipps regularization (under suitable assumptions):

$$u_{\mathsf{MAP}} \in \operatorname{argmin}_{u \in \mathcal{X}} \tfrac{1}{2} \|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} + \tfrac{1}{2} \|u\|^2_{C^{-1}}$$

with regularizing functional $R(u) = \frac{1}{2}\|u\|^2_{C^{-1}}$ determined by the prior.

Further comments

- Again can also define the MAP estimate in infinite dimensions, although in general not by maximization of the posterior probability density function...
- In the case of $y \in \mathbb{R}^d$, for Gaussian prior $u \sim \mathrm{N}(0, C)$ and Gaussian noise, i.e., $\pi_{u|y}(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u)$, we recover again Tikhonov-Philipps regularization (under suitable assumptions):

$$u_{\mathsf{MAP}} \in \mathrm{argmin}_{u \in \mathcal{X}}\, \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} + \tfrac{1}{2}\|u\|^2_{C^{-1}}$$

  with regularizing functional $R(u) = \frac{1}{2}\|u\|^2_{C^{-1}}$ determined by the prior.
- Under suitable assumptions continuous dependence on the prior π_0 can also be shown fairly straightforwardly.

Further comments

- Again can also define the MAP estimate in infinite dimensions, although in general not by maximization of the posterior probability density function...

- In the case of $y \in \mathbb{R}^d$, for Gaussian prior $u \sim \mathrm{N}(0, C)$ and Gaussian noise, i.e., $\pi_{u|y}(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u)$, we recover again Tikhonov-Philipps regularization (under suitable assumptions):

$$u_{\mathsf{MAP}} \in \operatorname{argmin}_{u \in \mathcal{X}} \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} + \tfrac{1}{2}\|u\|^2_{C^{-1}}$$

  with regularizing functional $R(u) = \frac{1}{2}\|u\|^2_{C^{-1}}$ determined by the prior.

- Under suitable assumptions continuous dependence on the prior π_0 can also be shown fairly straightforwardly.

- There are also theorems (much harder!) about the asymptotic behaviour of the posterior in infinite dimensions, i.e., **Bernstein-von Mises** type results.

Further comments

- Again can also define the MAP estimate in infinite dimensions, although in general not by maximization of the posterior probability density function...

- In the case of $y \in \mathbb{R}^d$, for Gaussian prior $u \sim \mathrm{N}(0, C)$ and Gaussian noise, i.e., $\pi_{u|y}(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \pi_0(\mathrm{d}u)$, we recover again Tikhonov-Philipps regularization (under suitable assumptions):

$$u_{\mathsf{MAP}} \in \mathrm{argmin}_{u \in \mathcal{X}}\, \tfrac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Sigma^{-1}} + \tfrac{1}{2}\|u\|^2_{C^{-1}}$$

  with regularizing functional $R(u) = \frac{1}{2}\|u\|^2_{C^{-1}}$ determined by the prior.

- Under suitable assumptions continuous dependence on the prior π_0 can also be shown fairly straightforwardly.

- There are also theorems (much harder!) about the asymptotic behaviour of the posterior in infinite dimensions, i.e., **Bernstein-von Mises** type results.

- However, in contrast to the case $\mathcal{X} = \mathbb{R}^n$, there are positive but also negative results for infinite-dimensional spaces $\mathcal{X}$. The results are highly problem dependent and still a very active field of research!

Summary – Part II

- The Bayesian approach to inverse problems is **well-posed**!
- **Unique solution:** the posterior measure $\pi_{u|y}$.
- Incorporation of prior knowledge/belief via prior measure π_0
- Conditioning / updating / learning based on measured data y.
- Point estimates available, e.g., recover Tikhonov–Philipps regularization
- **But crucially:** $\pi_{u|y}$ allows for uncertainty quantification for unknown u.

Summary – Part II

- The Bayesian approach to inverse problems is **well-posed**!
- **Unique solution:** the posterior measure $\pi_{u|y}$.
- Incorporation of prior knowledge/belief via prior measure π_0
- Conditioning / updating / learning based on measured data y.
- Point estimates available, e.g., recover Tikhonov–Philipps regularization
- **But crucially:** $\pi_{u|y}$ allows for uncertainty quantification for unknown u.

But how to actually compute Bayes estimates and quantify uncertainty?

$\longrightarrow$ **Lecture 3**