Mathematics and numerics for data assimilation and state estimation – Lecture 10

Mathematics for Uncertainty Quantification

RWTH AACHEN UNIVERSITY

Summer semester 2020

Overview

1. Bayesian inversion in different problem setting
2. Weak convergence of distributions
3. Linear-Gaussian setting
4. Posterior measure in the small-noise limit

Summary of lecture 9

Considered inverse problem

$$Y = G(U) + \eta \tag{1}$$

with assumptions: $\eta \sim \pi_\eta$, $U \sim \pi_U$ and $\eta \perp U$.

and solution:

$$\pi_{U|Y}(u|y) = \frac{\pi_\eta(y - G(u))\pi_U(u)}{Z}.$$

Stability: under some assumptions, small perturbations in input leads to small perturbations in output:

$$|G_\delta - G| = \mathcal{O}(\delta) \implies d(\pi^\delta(\cdot|y), \pi(\cdot|y)) = \mathcal{O}(\delta^p) \quad \text{for some} \quad p > 0,$$

and metrics

$$d_{TV}(\pi, \bar{\pi}) = \frac{1}{2}\|\pi - \bar{\pi}\|_{L^1(\mathbb{R}^d)} \quad \text{and} \quad d_H(\pi, \bar{\pi}) = \frac{1}{\sqrt{2}}\|\sqrt{\pi} - \sqrt{\bar{\pi}}\|_{L^2(\mathbb{R}^d)}$$

Overview

1. Bayesian inversion in different problem setting
2. Weak convergence of distributions
3. Linear-Gaussian setting
4. Posterior measure in the small-noise limit

Inverse problem with random model and exact observations

Let us consider a different type of inverse problem

$$Y = G(U)$$

with prior $U \sim U(0,1)$ and, for any $u \in (0,1)$, $G(u) \sim \mathit{Bernoulli}(u)$.

In other words U is a continuous rv, while $Y|(U = u) \sim \mathit{Bernoulli}(u)$ is discrete.

Given $Y = y$, we may formally proceed as before

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

Problem: $Y|(U = u)$ is a discrete rv!

Alternative measures-based approach:

For $y \in \{0,1\}$,

$$\mathbb{P}(Y=y, U \in du) = \mathbb{P}(Y=y|U \in du)\mathbb{P}(U \in du)$$
$$\mathbb{P}(Y=y, U \in du) = \mathbb{P}(U \in du|Y=y)\mathbb{P}(Y=y)$$

we derive by Bayes' rule the posterior measure

$$\mathbb{P}(U \in du|Y=y) = \frac{\mathbb{P}(Y=y|U \in du)\mathbb{P}(U \in du)}{\mathbb{P}(Y=y)}$$

By $Y=y \mid U=u$, it follows that

$$\mathbb{P}(Y=y \mid U \in du) = (1-u)^{1-y}u^y$$

and thus

$$\mathbb{P}(U \in du|Y=y) = \frac{(1-u)^y u^y du}{Z}.$$

With density form

$$\pi_{U|Y}(u|y) = \qquad = \frac{(1-u)^{1-y}u^y}{Z}. \tag{2}$$

Is the coin fair?

Consider an inverse problem with a sequence of **exact** observations of coin tosses

$$Y_k = G_k(U), \quad \text{for} \quad k = 1, 2, \dots$$

with $G_k(U)|U = u \sim \mathit{Bernoulli}(u)$, where for any fixed $\tilde{u} \in (0, 1)$ $(G_1(\tilde{u}), G_2(\tilde{u}), \dots)$ is an iid sequence. Hence

$$(Y_1, Y_2, \dots)|(U = u) = (G_1(u), G_2(u), \dots)$$

is a (conditionally $U = u$) iid sequence.
Input: Coin-bias prior $U \in U(0, 1)$ and flipping coin results $Y = (Y_1, \dots, Y_n) = (y_1, \dots, y_n)$.

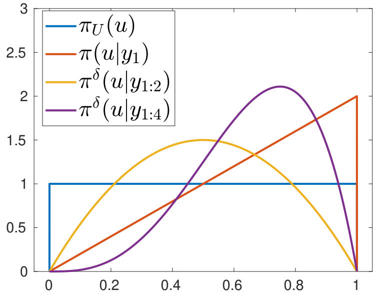
Direct extension of (2) yields

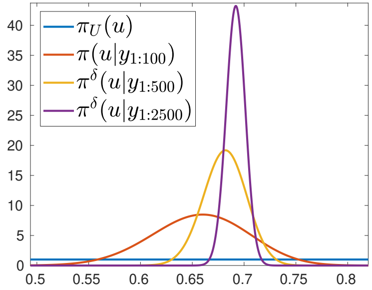
$$\pi_{U|Y}(u|y_{1:n}) = \frac{\prod_{k=1}^{n}(1-u)^{1-y_k} u^{y_k} \mathbb{1}_{(0,)]}(u)}{Z} = \frac{(1-u)^{n-\bar{y}_n} u^{\bar{y}_n} \mathbb{1}_{(0,1)}(u)}{Z}$$

where $\bar{y}_n = \sum_{k=1}^{n} y_k$.

Computational result given

$$y = (1, 0, 1, 1, \dots) \quad \text{with} \quad \bar{y}_{100} = 66,\ \bar{y}_{500} = 341,\ \bar{y}_{2500} = 1730$$



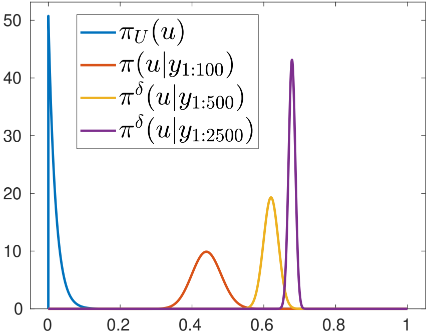


Numerical integration gives

$$\mathbb{P}(|U - 0.7| < 0.05 | Y_{1:500} = Y_{1:500}) = 0.9320$$

Sensitivity to the prior

Computational result given same y measurement sequence but now with the **very poor prior** $\pi_U(u) \propto (1-u)^{50}\mathbb{1}_{[0,1]}(u)$.



Numerical integration gives

$$\mathbb{P}(|U-0.7|<0.05|Y_{1:500}=Y_{1:500}) = 0.0695$$

See [“Data analysis” by D.S. Sivia section 2.1] for more on this example.

Overview

Definition 1 (Weak convergence of probability measures)

A sequence of distributions $\mathbb{P}_k$ on $(\mathbb{R}^d, \mathcal{B}^d)$ is said to converge weakly towards $\mathbb{P}$ if it holds for any globally bounded and continuous function $g : \mathbb{R}^d \to \mathbb{R}$ that

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} g(x)\mathbb{P}_k(dx) = \int_{\mathbb{R}^d} g(x)\mathbb{P}(dx).$$

We write $\mathbb{P}_k \Rightarrow \mathbb{P}$.

As an extension of the above, a family of distribution $\{\mathbb{P}_\gamma\}_{\gamma>0}$ converges weakly towards $\mathbb{P}$ as $\gamma \downarrow 0$ provided

$$\lim_{\gamma\downarrow 0} \int_{\mathbb{R}^d} g(x)\mathbb{P}_\gamma(dx) = \int_{\mathbb{R}^d} g(x)\mathbb{P}(dx).$$

Example 2 (Weak convergence of distributions)

For any $C \in \mathcal{B}$, let

$$\mathbb{P}_k(C) = \int_C (1 - k^{-1})\mathbb{1}_{(0,1)} + k^{-1}\mathbb{1}_{(1,2)} dx$$

Then it holds $\mathbb{P}_k \Rightarrow \mathbb{P} = U(0,1)$.

Verification: For a given $g \in C_b(\mathbb{R})$, we must show that for any $\epsilon > 0$, $\exists K > 0$ such that

$$\left| \int_{\mathbb{R}^d} g(x)\mathbb{P}_k(dx) - \int_{\mathbb{R}^d} g(x)\mathbb{P}(dx) \right| \leq \epsilon \quad \forall k > K.$$

Note that $\mathbb{P}_k = (1 - k^{-1})\mathbb{P} + k^{-1}U(1,2)$ and let $K = 2\left\lceil \frac{\max(\|g\|_\infty, 1)}{\epsilon} \right\rceil$.

Then for $\tilde{\mathbb{P}} := U(1,2)$ and $k > K$,

$$\begin{aligned} \left| \int_{\mathbb{R}^d} g(x)\mathbb{P}_k(dx) - \int_{\mathbb{R}^d} g(x)\mathbb{P}(dx) \right| &\leq k^{-1} \int_{\mathbb{R}^d} |g(x)|(\mathbb{P} + \tilde{\mathbb{P}})(dx) \\ &\leq 2k^{-1}\|g\|_\infty \leq \epsilon. \end{aligned}$$

Exercise 1: For $\mathbb{P}_\gamma = N(\mu, \gamma^2)$, show that $\mathbb{P}_\gamma \Rightarrow \delta_\mu$ as $\gamma \downarrow 0$.

Exerecise 2: For $\mathbb{P}_\gamma = N(\mu + \gamma\eta_0, \gamma^2\Gamma_0)$ with fixed $\mu, \eta_0 \in \mathbb{R}^d$ and positive definite $\Gamma_0 \in \mathbb{R}^{d\times d}$ show that $\mathbb{P}_\gamma \Rightarrow \delta_\mu$ as $\gamma \downarrow 0$.

Overview

Linear-Gaussian setting

We consider the inverse problem

$$Y = G(U) + \eta \tag{3}$$

with

Assumption 1

- linear forward model $G(u) = Au$ where $A \in \mathbb{R}^{k\times d}$
- and $\eta \sim N(0, \Gamma)$, $U \sim N(\hat{m}, \hat{C})$ where both Γ and $\hat{C}$ are positive definite and $\eta \perp U$.

Given on observation $Y = y$, Bayesian inversion yields

$$\pi(u|y) = \qquad\qquad = \frac{\pi_\eta(y - Au)\pi_U(u)}{Z}$$

where we have used that

$$Y|(U = u) = G(u) + \eta \sim N(G(u), \Gamma).$$

Recall that for $X \sim N(\mu, \Sigma)$,

$$\pi_X(x) = \frac{\exp\Big(-\frac{1}{2}|x-\mu|_\Sigma^2\Big)}{Z}$$

with

$$|x-\mu|_\Sigma := |\Sigma^{-1/2}(x-\mu)|.$$

So we may write (for a different normalizing constant Z),

$$\begin{aligned}
\pi(u|y) &= \frac{\pi_\eta(y-Au)\pi_U(u)}{Z} \\
&= \frac{\exp\Big(-\frac{1}{2}|y-Au|_\Gamma^2 - \frac{1}{2}|u-\hat{m}|_{\hat{C}}^2\Big)}{Z} \\
&= \frac{\exp(-\mathrm{J}(u))}{Z}
\end{aligned}$$

with

$$\mathrm{J}(u) := \frac{1}{2}|y-Au|_\Gamma^2 + \frac{1}{2}|u-\hat{m}|_{\hat{C}}^2 \tag{4}$$

Objective: Verify that $U|Y=y$ is Gaussian, and find its density.

On the one hand:

$$\pi(u|y) = \frac{\exp(-\mathrm{J}(u))}{Z}.$$

On the other hand, let us make the ansatz that for some $m \in \mathbb{R}^d$ and pos. def. C,

$$\pi(u|y) = \frac{\exp\Big(-\frac{1}{2}|u-m|_C^2\Big)}{Z}$$

For this to hold, we must find m and C s.t.,

$$|u-m|_C^2 = 2\mathrm{J}(u).$$

We write out these terms in sums of their polynomial parts:

$$|u-m|_C^2 = (u-m)^T C^{-1}(u-m) = u^T {\color{red}C^{-1}} u - 2u^T {\color{blue}C^{-1}m} + q$$

and

$$\begin{aligned} 2\mathrm{J}(u) &= |y-Au|_\Gamma^2 + |u-\hat{m}|_{\hat{C}}^2 \\ &= (y-Au)^T\Gamma^{-1}(y-Au) + (u-\hat{m})^T\hat{C}^{-1}(u-\hat{m}) \\ &= u^T{\color{red}(A^T\Gamma^{-1}A + \hat{C}^{-1})}u - 2u^T{\color{blue}(A^T\Gamma^{-1}y + \hat{C}^{-1}\hat{m})} + \hat{q} \end{aligned}$$

Enforcing equality for same-order-term coefficients yields

$$u^T C^{-1} u = u^T (A^T \Gamma^{-1} A + \hat{C}^{-1}) u \quad \forall u \in \mathbb{R}^d \implies C = (A^T \Gamma^{-1} A + \hat{C}^{-1})^{-1}$$

and

$$u^T C^{-1} m = u^T (A^T \Gamma^{-1} y + \hat{C}^{-1} \hat{m}) \quad \forall u \in \mathbb{R}^d \implies m = C(A^T \Gamma^{-1} y + \hat{C}^{-1} \hat{m}).$$

Theorem 3

If Assumption 1 holds, then

$$\pi(u|y) = \frac{\exp\left(-\frac{1}{2} |u - m|_C^2 \right)}{2} \tag{5}$$

with

$$C = (A^T \Gamma^{-1} A + \hat{C}^{-1})^{-1} \quad \textit{and} \quad m = C(A^T \Gamma^{-1} y + \hat{C}^{-1} \hat{m}).$$

MAP of a Gaussian posterior vs deterministic inv. methods

Consider initially ill-posed inverse problem: given y and A, find x s.t.

$$Au = y,$$

(assume either no or many solutions).

Form of Tikhonov regularization: For some $\lambda > 0$, define solution as

$$u = \arg\min_{x \in \mathbb{R}^d} \underbrace{|y - Ax|^2}_{\text{Loss term}} + \underbrace{\lambda |x|^2}_{\text{Regularizing term}}$$

Bayesian inversion of

$$Y = AU$$

for $U \sim N(0, \sigma^2 I)$ and $\eta \sim N(0, \gamma^2 I)$ and $Y = y$ yields, cf (4),

$$\pi(u|y) \propto \exp\left(- \frac{\gamma^{-2}|y - Au|^2 + \sigma^{-2}|u|^2}{2} \right)$$

Hence

$$u_{MAP}[\pi(\cdot|y)] = \arg\min_{u \in \mathbb{R}^d} |y - Au|^2 + \frac{\gamma^2}{\sigma^2}|u|^2.$$

Overview

Small-noise limit and multivariate normals

We consider the inverse problem

$$Y = AU + \eta,$$

with $U \sim N(\hat{m}, \hat{C})$ and $\eta \sim N(0, \gamma^2 \Gamma_0)$ for some positive definite Γ_0 and parameterized in $\gamma > 0$.

Theorem 3 yields that $U|(Y = y) \sim N(m, C)$ with

$$C(\gamma) = \gamma^2 (A^T \Gamma_0^{-1} A + \gamma^2 \hat{C}^{-1})^{-1}$$

and

$$m(\gamma) = (A^T \Gamma_0^{-1} A + \gamma^2 \hat{C}^{-1})^{-1} (A^T \Gamma_0^{-1} y + \gamma^2 \hat{C}^{-1} \hat{m})$$

Questions:

- What happens to the posterior density as $\gamma \downarrow 0$?
- How does $\lim_{\gamma \to 0} \pi(\cdot|y)$ depend on the prior, A and y?
- If $y_\gamma = Au^\dagger + \gamma \eta^\dagger$, for some deterministic $u^\dagger, \eta^\dagger$, will then asymptotically $\pi(\cdot|y_\gamma)$ concentrate around $u^\dagger$?

Speculations

It seems reasonable to expect that $U|Y = y \sim N(m(\gamma), C(\gamma))$ will converge in some sense to $N(m^*, C^*)$, where

$$\begin{aligned} m^* &= \lim_{\gamma \to 0} (A^T \Gamma_0^{-1} A + \gamma^2 \hat{C}^{-1})^{-1} (A^T \Gamma_0^{-1} y + \gamma^2 \hat{C}^{-1} \hat{m}) \\ &\stackrel{?}{=} (A^T \Gamma_0^{-1} A)^{-1} A^T \Gamma_0^{-1} y \end{aligned}$$

and

$$C^* = \lim_{\gamma \to 0} C(\gamma) = \lim_{\gamma \to 0} \gamma^2 (A^T \Gamma_0^{-1} A + \gamma^2 \hat{C}^{-1})^{-1} \stackrel{?}{=} 0.$$

The argument hinges on whether $A^T \Gamma_0^{-1} A$ is invertible or not.

Need to consider two cases for $A \in \mathbb{R}^{k \times d}$:

- overdetermined/determined: $k \geq d$ and $Null(A) = \{0\}$,
- underdetermined: $k < d$ and $Rank(A) = k$.

Overdetermined and determined settings

For the case $A \in \mathbb{R}^{k \times d}$, $k \geq d$ and $Null(A) = \{0\}$, it is clear that

$$Ax = 0 \iff x = 0$$

which implies that for all $x \in \mathbb{R}^d \setminus \{0\}$,

$$x^T A^T \Gamma_0^{-1} A x > 0$$

so $A^T \Gamma_0^{-1} A$ is invertible.
For the sequence of distributions $U|(Y = y) \sim N(m(\gamma), C(\gamma))$ with a **fixed** $y \in \mathbb{R}^k$, we have that

$$m^* = \lim_{\gamma \to 0} m(\gamma) = (A^T \Gamma_0^{-1} A)^{-1} A^T \Gamma_0^{-1} y \quad \text{and} \quad C^* = \lim_{\gamma \to 0} C(\gamma) = 0.$$

This yields the small-noise limit, as $\gamma \downarrow 0$,

$$N(m(\gamma), C(\gamma)) \Rightarrow \delta_{m^*} = \begin{cases} \delta_{A^{-1} y} & \text{if } k = d \\ \delta_{(A^T \Gamma_0^{-1} A)^{-1} A^T \Gamma_0^{-1} y} & \text{if } k > d \end{cases}$$

Note above: If $k = d$ then A is invertible.

Interpretation of m^* and C^*

From (4) we have

$$\pi(u|y;\gamma) = \frac{\exp(-\mathrm{J}(u,\gamma))}{Z(\gamma)}$$

with

$$\mathrm{J}(u,\gamma) := \underbrace{\frac{1}{2}\gamma^{-2}|\Gamma_0^{-1/2}(y - Au)|^2}_{\text{log likelihood – loss}} \quad + \quad \underbrace{\frac{1}{2}|u - \hat{m}|_{\hat{C}}^2}_{\text{log prior – vanishing regularizer}} \quad . \tag{6}$$

Interpretation

$$m^* = (A^T\Gamma_0^{-1}A)^{-1}A^T\Gamma_0^{-1}y$$

is mean-square minimizer of the **log likelihood** term,

$$m^* = \arg\min_{u\in\mathbb{R}^d} |\Gamma_0^{-1/2}(Au - y)|^2 = \lim_{\gamma\to 0} \arg\min_{u\in\mathbb{R}^d} J(u,\gamma)$$

Moreover, influence from prior on $\pi(u|y;\gamma)$ vanishes asymptotically since

$$C^* = \lim_{\gamma\to 0} \gamma^2(A^T\Gamma_0^{-1}A + \gamma^2\hat{C}^{-1})^{-1} = 0$$

Consistency of the estimator – overdetermined setting

Consider again the inverse problem

$$Y = AU + \eta,$$

with $U \sim N(\hat{m}, \hat{C})$ and $\eta \sim N(0, \gamma^2\Gamma_0)$, but assume now that

$$Y = y(\gamma) = Au^\dagger + \gamma\eta^\dagger \quad \text{for fixed } u^\dagger, \eta^\dagger$$

This yields the posterior distribution $U|Y = y(\gamma) \sim N(m(\gamma), C(\gamma))$ where

$$m(\gamma) = (A^T\Gamma_0^{-1}A + \gamma^2\hat{C}^{-1})^{-1}(A^T\Gamma_0^{-1}y(\gamma) + \gamma^2\hat{C}^{-1}\hat{m})$$

and $C(\gamma) =$ as earlier. Consequently,

$$m^* = \lim_{\gamma\to 0} m(\gamma) = (A^T\Gamma_0^{-1}A)^{-1}A^T\Gamma_0^{-1}Au^\dagger = u^\dagger$$

and we obtain the consistency result

$$N(m(\gamma), C(\gamma)) \Rightarrow \delta_{m^*} = \delta_{u^\dagger} \quad \text{as } \gamma \to 0.$$

Underdetermined setting

We consider the simplified inverse problem

$$Y = AU + \eta = A_0 U_1 + \eta,$$

on $\mathbb{R}^d = \mathbb{R}^k \times \mathbb{R}^{d-k}$ where

- $U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N(\hat{m}_1, I_k) \times N(\hat{m}_2, I_{d-k})$ with $\hat{m}_1, U_1 \in \mathbb{R}^k$ and $\hat{m}_2, U_2 \in \mathbb{R}^{d-k}$,
- $N(\hat{m}_1, I_k) \times N(\hat{m}_2, I_{d-k})$ is a measure on $(\mathbb{R}^k \times \mathbb{R}^{d-k}, \mathcal{B}^k \times \mathcal{B}^{d-k})$.
- $A = [A_0 \;\; 0] \in \mathbb{R}^{k \times d}$ with non-singular $A_0 = \mathbb{R}^{k \times k}$
- $\eta \sim N(0, \gamma^2 \Gamma_0)$ with positive definite $\Gamma_0 \in \mathbb{R}^{k \times k}$.

Observations only of the first k components yields

$$\pi(u_1, u_2 | y) \propto \frac{\exp\left(-\frac{1}{2}\gamma^{-2}|y - A_0 u_1|_{\Gamma_0}^2 - \frac{1}{2}|u_1 - \hat{m}_1|^2 - \frac{1}{2}|u_2 - \hat{m}_2|^2\right)}{Z}$$

Equivalently,
$U|(Y=y) \sim N(m_1(\gamma), C_1(\gamma)) \times N(\hat{m}_2, I_{d-k})$ with

$$m_1 = (A_0^T \Gamma_0^{-1} A_0 + \gamma^2 I_k)^{-1}(A_0^T \Gamma_0^{-1} y + \gamma^2 \hat{m}_1)$$

and

$$C_1 = \gamma^2 (A_0^T \Gamma_0^{-1} A_0 + \gamma^2 I_k)^{-1}$$

Restricted to the measure on $(\mathbb{R}^k, \mathcal{B}^k)$,

$$N(m_1(\gamma), C_1(\gamma)) \Rightarrow \delta_{A_0^{-1} y} \quad \text{as} \quad \delta \to 0,$$

and thus

$$N(m_1(\gamma), C_1(\gamma)) \times N(\hat{m}_2, I_k) \Rightarrow \delta_{A_0^{-1} y} \times N(\hat{m}_2, I_{d-k}) \quad \text{as} \quad \delta \to 0.$$

Observation: Asymptotically perfect "correction" in observed subspace (prior is near-irrelevant for posterior), no correction in unobserved subspace (posterior equals prior in these components).

Summary small-noise limit

For linear-Gaussian inverse problem

$$Y = AU + \eta,$$

with $U \sim N(\hat{m}, \hat{C})$ and $\eta \sim N(0, \gamma^2 \Gamma_0)$ for some positive definite Γ_0 and parameterized in $\gamma > 0$.

We obtained $U|(Y = y) \sim N(m(\gamma), C(\gamma))$, and in the small-noise limit $\gamma \to 0$

- $N(m(\gamma), C(\gamma)) \Rightarrow \delta_{A^{-1}y}$ when A is invertible,
- $N(m(\gamma), C(\gamma)) \Rightarrow \delta_{(A^T\Gamma_0^{-1}A)^{-1}A^T\Gamma_0^{-1}y}$ in the overdetermined setting
- Underdetermined setting, see [SST Theorem 2.12],

$$N(m(\gamma), C(\gamma)) \Rightarrow \text{correction in observed-subspace measure}$$
$$\times \text{no correction in unobserved-subspace measure (it remains the prior)}$$