Bayesian Inverse Problems and Data Assimilation

Naoki Sakai

The University of Tokyo

December 16, 2024

Based on the work in [1] [2].

Outline

1. Bayesian Inverse Problems
   - Formulation
   - Well-Posedness
   - Optimization view
   - Remarks
2. Data Assimilation
   - Smoothing Problem and Filtering Problem
   - Kalman Filter
   - Extended Kalman Filter
   - Ensemble Kalman Filter
   - Particle Filter
   - Remarks

Outline

Bayesian Inverse Problems

Let $G : \mathbb{R}^d \to \mathbb{R}^k, u \longmapsto G(u)$ and assume that observation $y \in \mathbb{R}^k$ is perturbed by additive noise η with pdf ν.

$$y = G(u) + \eta$$

With the prior distribution $\rho(u)$,by Bayes formula, the posterior distribution is

$$\pi^y(u) = \frac{1}{Z(y)}\nu(y - G(u))\rho(u)$$

where $Z(y) = \int \nu(y - G(u))\rho(u)$

Note that when $Z(y) \in \{0, \infty\}$ the posterior is not defined but this event is of zero probability with respect to the joint distribution of (u, y)

Distance between probability measures

Let μ, ν be two probability measures on $\mathbb{R}^d$

- Total Variation distance

$$d_{TV}(\mu,\nu) = \frac{1}{2}\int |\mu(u) - \nu(u)|du = \frac{1}{2}||\mu - \nu||_{L^1(\mathbb{R}^d)}$$

- Hellinger Distance

$$d_{Hell}(\mu,\nu) = (\frac{1}{2}\int |\sqrt{\mu(u)} - \sqrt{\nu(u)}|^2 du)^{1/2} = \frac{1}{\sqrt{2}}||\sqrt{\mu} - \sqrt{\nu}||_{L^2(\mathbb{R}^d)}$$

Lemma

1. $0 \leq d_{TV}(\mu,\nu), d_{Hell}(\mu,\nu) \leq 1$
2. $\frac{1}{\sqrt{2}}d_{TV}(\mu,\nu) \leq d_{Hell}(\mu,\nu) \leq \sqrt{d_{TV}(\mu,\nu)}$

Well-Posedness

In addition to the well-defined posterior distribution, we need the stabilities of the posterior distribution.

- perturbed data and forward model
  $\pi^y(u) = \frac{1}{Z(y)}\nu(y - G(u))\rho(u)$ and $\pi_h^{\tilde{y}}(u) = \frac{1}{\tilde{Z}_h(y)}\nu(\tilde{y} - G_h(u))\rho(u)$

Lemma

Assume

- **Boundedness:** $\nu(y) \leq k_1, \quad \forall y \in \mathbb{R}^k.$
- **Lipschitz Continuity:** $|\sqrt{\nu(y_1)} - \sqrt{\nu(y_2)}| \leq k_2|y_1 - y_2|.$
- $G, G_h \in L_p^2(\mathbb{R}^d)$
- $Z(y), \tilde{Z}_h(y) > 0$

Then $d_{Hell}(\pi^y, \pi_h^{\tilde{y}}) \leq C(||y - \tilde{y}|| + ||G - G_h||_{L_\rho^2})$
where C *depends on* $k_1, k_2, Z(y), \tilde{Z}_h(y).$

Proof

$$d_{Hell}(\pi^{y}, \pi_h^{\tilde{y}})^2 = \frac{1}{2}\int_{\mathbb{R}^d} \left(\sqrt{\pi^{y}} - \sqrt{\pi_h^{\tilde{y}}}\right)^2 du$$

$$= \frac{1}{2}\int \left(\sqrt{\frac{J(y - G(u))}{Z(y)}} - \sqrt{\frac{J(\tilde{y} - G_h(u))}{\tilde{Z}_h(y)}}\right)^2 \rho(u)\, du$$

$$\leq \underbrace{\int_{\mathbb{R}^d} \left(\sqrt{\frac{J(y - G(u))}{Z(y)}} - \sqrt{\frac{J(\tilde{y} - G_h(u))}{Z(y)}}\right)^2 \rho(u)\, du}_{A} + \underbrace{\int \left(\sqrt{\frac{J(\tilde{y} - G_h(u))}{Z(y)}} - \sqrt{\frac{J(\tilde{y} - G_h(u))}{\tilde{Z}_h(y)}}\right)^2 \rho(u)\, du}_{B}$$

$$A \leq \frac{1}{Z(y)} K_2^2 \int_{\mathbb{R}^d} \left(\|y - \hat{y}\| + \|G(u) - G_h(u)\|\right)^2 \rho(u)\, du$$

$$\leq \frac{2K_2^2}{Z(y)} \left(\|y - \hat{y}\|^2 + \int_{\mathbb{R}^d} \|G(u) - G_h(u)\|^2 \rho(u)\, du\right)$$

$$B = \left(\frac{1}{\sqrt{Z(y)}} - \frac{1}{\sqrt{\tilde{Z}_h(y)}}\right)^2 \underbrace{\int_{\mathbb{R}^d} J(\hat{y} - G_h(u))\, \rho(u)\, du}_{\tilde{Z}_h(y)}$$

$$= \frac{\left(\sqrt{\tilde{Z}_h(y)} - \sqrt{Z(y)}\right)^2}{Z(y)}$$

$$\leq \frac{1}{Z(y)} \left(\frac{1}{2\sqrt{\min(Z(y), \tilde{Z}_h(y))}} \left|\hat{Z}_h(y) - Z(y)\right|\right)^2$$

$$\leq \frac{1}{4 \min\{Z(y), \tilde{Z}_h(y)\}^2} \left(\hat{Z}_h(y) - Z(y)\right)^2$$

and

$$(\tilde{Z}_n(z) - Z(z))^2 = \left(\int_{\mathbb{R}^d} (J(\tilde{z} - G_n(u)) - J(z - G(u))) \rho(u)\, du \right)^2$$

$$= \left(\int_{\mathbb{R}^d} \left(\sqrt{J(\tilde{z} - G_n)} - \sqrt{J(z - G)}\right) \underbrace{\left(\sqrt{J(\tilde{z} - G_n)} + \sqrt{J(z - G)}\right)}_{\le 2\sqrt{k_1}} \rho(u)\, du \right)^2$$

$$\le 4k_1 \int_{\mathbb{R}^d} \left(\sqrt{J(\tilde{z} - G_n)} - \sqrt{J(z - G)}\right)^2 \rho(u)\, du$$

$$\le 4k_1 k_2 \int_{\mathbb{R}^d} (\|z - \tilde{z}\| + \|G - G_n\|)^2 \rho(u)\, du$$

$$\le 8 k_1 k_2^2 \left(\|z - \tilde{z}\|^2 + \|G - G_n\|^2_{L^2_\rho} \right)$$

Thus,

$$d_{Hell}(\pi^z, \pi_n^{\tilde{z}})^2 \le \left(\frac{2k_2^2}{Z(z)} + \frac{8k_1 k_2^2}{4\min\{Z(z), Z_n(\tilde{z})\}} \right) \left(\|z - \tilde{z}\|^2 + \|G - G_n\|^2_{L^2_\rho} \right)$$

$\blacksquare$

Well-Posedness

- perturbed prior
  $\pi^y(u) = \frac{1}{Z(y)}\nu(y - G(u))\rho(u)$ and $\pi^{\tilde{y}}(u) = \frac{1}{\tilde{Z}(y)}\nu(y - G(u))\tilde{\rho}(u)$

Lemma

Assume

- **Boundedness:** $\nu(y) \leq k_1, \quad \forall y \in \mathbb{R}^k$.
- $Z(y), \tilde{Z}(y) > 0$

Then $d_{Hell}(\pi^y, \pi^{\tilde{y}}) \leq C d_{Hell}(\rho, \tilde{\rho})$
where C *depends on* $k_1, Z(y), \tilde{Z}(y)$

proof is similar to the previous one.

Optimization view

$$y = Au + \eta$$

where $A \in \mathbb{R}^{k \times d}$ is linear operator(matrix) and η has mean zero and symmmetric, positive-definite covariance matrix Q

Theorem (Gauss-Markov)

In addition to the assumption above, suppose that $A^T QA$ is invertible,then, Among all linear unbiased estimators $\hat{u} = Ky$ of u, the estimator:

$$\tilde{K} = (A^\top Q^{-1} A)^{-1} A^\top Q^{-1}$$

minimizes both the mean-squared error: $\mathbb{E}[\|\hat{u} - u\|^2] = \mathbb{E}[(\hat{u} - u)^\top (\hat{u} - u)]$, and the covariance matrix: $\mathbb{E}[(\hat{u} - u)(\hat{u} - u)^\top]$.

Note that this theorem is true even for separable Hilbert Spaces with some additonal assumptions.
$\tilde{K}y$ is the solution of the weighted least square problem,

$$\tilde{u} = \arg\min_{u \in \mathbb{R}^d} J(u), \quad J(u) := \frac{1}{2} ||Au - y||^2_{Q^{-1}}$$

Bayesian Interpretation of Regularization

We further assume that noise $\eta \sim \mathcal{N}(0, Q)$ and prior distribution is $\mathcal{N}(\bar{u}, R)$.

$$J(u) := \frac{1}{2}||Au - y||^2_{Q^{-1}} + \frac{1}{2}||u - \bar{u}||^2_{R^{-1}}$$

The solution of the minimizer $J(u)$ is the maximum a posteriori estimator (MAP estimator).

Recovery of Sparse Signals

Sometimes you want to recover sparse signals ,i.e.,

$$J(u) := \frac{1}{2}||Au - y||^2_{Q^{-1}} + \lambda||u||_0$$

where $||u||_0 := \{i \in \{1,\cdots,d\}|u_i \neq 0\}$.
This is not convex but the minimizer often coincides with the minimizer of

$$J(u) := \frac{1}{2}||Au - y||^2_{Q^{-1}} + \lambda||u||_1$$

where $||u||_1 := \sum_{i=1}^{d} |u_i|$.

The topics not mentioned in this slide

- Accessing the Bayesian Posterior Measure(ex. Markov chain Monte Carlo)
- Formulation and Well-Posedness on Banach spaces
- Frequentist Consistency
- Linear Gaussian Setting and Small noise limit and Consistency

Outline

Markov Model

Data Assimilation:

- Identify the states of a dynamical system from the observations

Markov Model:

- Random initial state:

$$v_0 \sim \rho_0.$$

- Markov transition probabilities for $j = 0, \ldots, J-1$:

$$v_{j+1} \sim P(\cdot | v_j).$$

- Observations:

$$y_j = h(v_j) + \eta_j, \quad \eta_j \overset{\text{iid}}{\sim} \nu, \quad \eta_j \perp (v_0, \ldots, v_J), \quad h \in C^0(\mathbb{R}^d, \mathbb{R}^k).$$

Example: Additive Noise Model

- Transition model:

$$v_{j+1} = \psi(v_j) + \xi_j, \quad \psi \in C^0(\mathbb{R}^d, \mathbb{R}^d), \quad \xi_j \overset{\text{iid}}{\sim} \mu$$

- Which implies:

$$P(v_{j+1} | v_j) = \mu(v_{j+1} - \psi(v_j)).$$

Notation

Notation:

- State (signal):
$$V = (v_0, v_1, \ldots, v_J).$$
- Data:
$$Y = (y_1, \ldots, y_J).$$
- Data up to time j:
$$Y_j = (y_1, \ldots, y_j).$$

Smoothing Problem

Goal: Find the distribution of the state given all the data:

$$\pi(V) = P(V|Y) = P(v_0, \ldots, v_J | y_1, \ldots, y_J).$$

Prior Distribution:

- The prior distribution is given by:
  $\rho(V) = P(v_0, \ldots, v_J) = \rho_0(v_0) \prod_{j=0}^{J-1} P(v_{j+1}|v_j)$.

Likelihood:

- Let $Y = G(V) + \eta$, where $G(V) = (h(v_1), \ldots, h(v_J))$ and $\eta = (\eta_1, \ldots, \eta_J)$.
- The likelihood is:
  $\mathcal{L}(V|Y) = P(Y|V) = \prod_{j=1}^{J} P(y_j|v_j) = \prod_{j=1}^{J} \nu(y_j - h(v_j))$.

Bayes' Formula:

- The posterior distribution is:

$$P(V) = \frac{1}{Z}\mathcal{L}(V|Y)\rho(V),$$

  where Z is the normalizing constant.

Well-posedness of the Smoothing Problem

Well-posedness of the Smoothing Problem:

- Can be derived from results on Bayesian inverse problems (BIPs).

Assumption 1:

- $\nu(y) \leq k_1, \quad \forall y \in \mathbb{R}^k$.
- $|\sqrt{\nu(y_1)} - \sqrt{\nu(y_2)}| \leq k_2 |y_1 - y_2|, \quad \forall y_1, y_2 \in \mathbb{R}^k$.
- $\mathbb{E}_\rho \left(\sum_{j=1}^{J} |h(v_j)|^2 \right) < \infty$.

Consider two sets of data $Y, \tilde{Y}$ and the two posteriors:

$$\pi^Y(V) = \frac{1}{Z} \mathcal{L}(V|Y)\rho(V), \quad \pi^{\tilde{Y}}(V) = \frac{1}{\tilde{Z}} \mathcal{L}(V|\tilde{Y})\rho(V).$$

Theorem: Under Assumption 1, there exists $C > 0$, independent of Y and $\tilde{Y}$, such that:

$$d_{\mathrm{Hell}}(\pi^Y, \pi^{\tilde{Y}}) \leq C|Y - \tilde{Y}|.$$

Filtering Problem

Goal: Sequentially update the distribution of the state v_j given all observations up to time j:

$$Y_j = (y_1, \ldots, y_j).$$

Filtering Distribution: The filtering distribution is given as:

$$\pi_j(v_j) = P(v_j | Y_j).$$

Two-step Recursive Procedure:

- Prediction Step:

$$\hat{\pi}_{j+1}(v_{j+1}) := P(v_{j+1}|Y_j) = \int P(v_{j+1}|v_j)\pi_j(v_j)dv_j.$$

- Analysis Step:

$$\pi_{j+1}(v_{j+1}) = \frac{1}{Z}\nu(y_{j+1} - h(v_{j+1}))\hat{\pi}_{j+1}(v_{j+1}),$$

  where Z is the normalization constant.

Given $\pi_0 = \rho_0$, we recursively compute:

Relation between Filtering and Smoothing

Relation Between Filtering and Smoothing:

- Smoothing distribution:

$$\pi^Y(V) = P(v_0, \ldots, v_J | y_1, \ldots, y_J).$$

- Filtering distribution at the final time J:

$$\pi_J(v_J) = P(v_J | y_1, \ldots, y_J) = \int P(v_0, \ldots, v_J | y_1, \ldots, y_J) dv_0 \ldots dv_{J-1}.$$

The filtering distribution is the marginal of the smoothing distribution at the last observation time J.

Well-posedness of the Filtering Distribution

Well-posedness of the Filtering Distribution:

- We consider two sets of data Y and $\tilde{Y}$ and the corresponding filtering distributions π_J and $\tilde{\pi}_J$.
- By exploiting the relation between the filtering and smoothing distributions, and the well-posedness of the smoothing distribution, we have:

Theorem: Under Assumption 1, there exists a constant $\hat{C} > 0$, independent of Y and $\tilde{Y}$, such that:

$$d_{TV}(\pi_J, \tilde{\pi}_J) \leq \hat{C}|Y - \tilde{Y}|.$$

Kalman Filter: Linear Gaussian Case

Assumptions:

- Linear dynamics with Gaussian noise:

$$v_{j+1} = Av_j + b + \xi_j, \quad \xi_j \sim \mathcal{N}(0, \Sigma).$$

- Linear observations with Gaussian noise:

$$y_j = Hv_j + \eta_j, \quad \eta_j \sim \mathcal{N}(0, \Gamma).$$

- Gaussian initial condition:

$$u_0 \sim \mathcal{N}(m_0, C_0).$$

Prediction Step:

- Predict prior mean and covariance:

$$\hat{m}_{j+1} = Am_j + b,$$
$$\hat{C}_{j+1} = AC_jA^T + \Sigma.$$

Kalman Filter: Analysis Step

Analysis Step:

- Update posterior mean and covariance:

$$m_{j+1} = \hat{m}_{j+1} + K_{j+1} d_{j+1}$$
$$C_{j+1} = (I - K_{j+1} H)\hat{C}_{j+1}$$

where,
$d_{j+1} = y_{j+1} - H\hat{m}_{j+1}$ (Innovation)
$K_{j+1} = \hat{C}_{j+1} H^T (H\hat{C}_{j+1} H^T + \Gamma)^{-1}$ (Kalman Gain)

Extended Kalman Filter (ExKF)

Assumption: Observations are linear

1. **Input:** Initial mean $m_0 \in \mathbb{R}^d$ and covariance $C_0 \in \mathbb{R}^{d \times d}$.
2. **For** $j = 0, 1, \dots, J-1$, perform prediction and analysis:
   - **Prediction:**
   $$\begin{aligned} \hat{m}_{j+1} &= \Psi(m_j), \\ \hat{C}_{j+1} &= D\Psi(m_j) C_j D\Psi(m_j)^T + \Sigma. \end{aligned}$$
   - **Analysis:**
   $$\begin{aligned} K_{j+1} &= C_{j+1} H^T (H C_{j+1} H^T + \Gamma)^{-1}, \\ m_{j+1} &= (I - K_{j+1} H) \hat{m}_{j+1} + K_{j+1} y_{j+1}, \\ C_{j+1} &= (I - K_{j+1} H) \hat{C}_{j+1}. \end{aligned}$$
3. **Output:** Predictive means $\{m_j\}$ and covariances $\{C_j\}$.

Ensemble Kalman Filter (EnKF)

Assumption: Observations are linear

1. **Input:** Ensemble size N, initial ensemble $\{v_0^{(n)}\}_{n=1}^N$, parameter $s \in \{0, 1\}$.
2. **For** $j = 0, 1, \ldots, J-1$, perform prediction and analysis:
   - **Prediction:**

$$
\begin{aligned}
\xi_j^{(n)} &\sim \mathcal{N}(0, \Sigma), \quad n = 1, \ldots, N, \\
\hat{v}_{j+1}^{(n)} &= \Psi(v_j^{(n)}) + \xi_j^{(n)}, \\
\hat{m}_{j+1} &= \frac{1}{N} \sum_{n=1}^{N} \hat{v}_{j+1}^{(n)}, \\
\hat{C}_{j+1} &= \frac{1}{N} \sum_{n=1}^{N} (\hat{v}_{j+1}^{(n)} - \hat{m}_{j+1}) \otimes (\hat{v}_{j+1}^{(n)} - \hat{m}_{j+1}).
\end{aligned}
$$

   - **Analysis:**

$$
\begin{aligned}
\eta_j^{(n)} &\sim \mathcal{N}(0, \Gamma), \quad n = 1, \ldots, N, \\
y_{j+1}^{(n)} &= y_{j+1} + s\eta_j^{(n)}, \\
v_{j+1}^{(n)} &= (I - K_{j+1}H)\hat{v}_{j+1}^{(n)} + K_{j+1}y_{j+1}^{(n)}.
\end{aligned}
$$

3. **Output:** Ensembles $\{v_j^{(n)}\}_{n=1}^N$ for $j = 0, \ldots, J$.

Bootstrap Particle Filter

Algorithm

1. **Input:** Initial distribution π_0, number of particles N.
2. **For** $j = 0, 1, \ldots, J-1$, perform:
   - **Particle Generation:**
   $$v_j^{(n)} \sim \pi_j, \quad n = 1, \ldots, N,$$
   $$\hat{v}_{j+1}^{(n)} = \Psi(v_j^{(n)}) + \xi_j^{(n)}, \quad \xi_j^{(n)} \sim \mathcal{N}(0, \Sigma).$$
   - **Weight Update:**
   $$w_{j+1}^{(n)} \propto \exp\left(-\frac{1}{2}\|y_{j+1} - H\hat{v}_{j+1}^{(n)}\|_\Gamma^2\right),$$
   $$w_{j+1}^{(n)} \leftarrow \frac{w_{j+1}^{(n)}}{\sum_{k=1}^N w_{j+1}^{(k)}}.$$
   - **Resampling:**
   $$\pi_{j+1}(u) = \sum_{n=1}^N w_{j+1}^{(n)} \delta(u - \hat{v}_{j+1}^{(n)}).$$
3. **Output:** Particle approximations $\pi_j^N \approx \pi_j$ for $j = 1, \ldots, J$.

The topics not mentioned in this slide

- variational method in derivation of Kalman Filter
- Time continuous case
- Optimal Particle filter

References

Daniel Sanz-Alonso, Andrew M Stuart, and Armeen Taeb.
Inverse problems and data assimilation.
arXiv preprint arXiv:1810.06191, 2018.

Timothy John Sullivan.
Introduction to uncertainty quantification, volume 63.
Springer, 2015.

Thanks for your attention!