Inverse problem from a Bayesian perspective - Review 1

XXY (algebraicfrost@mail.ustc.edu.cn)

1 Problem Background

1.1 General Problem

To find μ, an input to a mathematical model, given y an observation of (some components of, or functions of) the solution of the model. We have an equation of the form

$$y = \mathcal{G}(\mu) \tag{1}$$

to solve for $\mu \in X$, given $y \in Y$, where X, Y are Banach spaces. μ can be viewed as the initial solution $\mu(t=0) = \mu_0$ of a pde, and y can be viewed as μ_t, which is observable. We can solve the problem following least-squares optimization problem of finding the norm $||\cdot||_Y$ on Y:

$$\text{argmin}_{u \in X} \frac{1}{2}\|y - \mathcal{G}(u)\|_Y^2$$

This problem, too, may be difficult to solve as it may possess minimizing sequences $u^{(n)}$ which do not converge to a limit in X, or it may possess multiple minima and sensitive dependence on the data y. These issues can be somewhat ameliorated by solving a regularized minimization problem of the form, for some Banach space $(E, \|\cdot\|_E)$ contained in X, and point $m_0 \in E$,

$$\text{argmin}_{u \in E} \left(\frac{1}{2}\|y - \mathcal{G}(u)\|_Y^2 + \frac{1}{2}\left\|u - m_0\right\|_E^2 \right). \tag{2}$$

To solve the above problem, the Bayesian approach is introduced to firstly introduce a posterior distribution of μ^y given y, and a prior distribution of $\mu^0 \sim N(m_o, \Gamma)$. The posterior distribution μ^y will have probability density of

$$\pi^y(u) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|_Y^2 - \frac{1}{2}\left\|u - m_0\right\|_E^2\right). \tag{3}$$

By solving the least-square problem in (2), we can find the maximum a posterior(MAP) estimate for the posterior distribution in (3).

1.2 Probabilistic Framework

The source of data often reveals that the observations y are subject to noise and that a more appropriate model equation is often of the form

$$y = \mathcal{G}(\mu) + \eta. \tag{4}$$

η is the observable noise which follows Gaussian distribution $N(0, B)$. Since the noise η and the observation y are both sampled from the same Banach space, we assume that they have the same probability density, that is

$$\rho(y|\mu) = \rho(y - \mathcal{G}(\mu)). \tag{5}$$

Following (5), the density of posterior distribution μ^y given y can be reconstructed as This is often referred to as the data likelihood. By Bayes' formula (6.24) we obtain

$$\pi^y(u) = \frac{\rho(y - \mathcal{G}(u))\pi_0(u)}{\int_{\mathbb{R}^n} \rho(y - \mathcal{G}(u))\pi_0(u)\mathrm{d}u}$$

Thus

$$\pi^y(u) \propto \rho(y - \mathcal{G}(u))\pi_0(u)$$

with constant of proportionality depending only on y(which means $\int_{\mathbb{R}^n} \rho(y - \mathcal{G}(u))\pi_0(u)\mathrm{d}u$ is a constant). Abstractly (2.7) expresses the fact that the posterior measure μ^y (with density π^y) and prior measure μ_0 (with density π_0) are related through the Radon-Nikodym derivative (see Theorem 6.2)

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \rho(y - \mathcal{G}(\mu))$$

Since ρ is a density and thus non-negative, without loss of generality we may write the right-hand side as the exponential of the negative of a potential $\Phi(\mu; y)$, to obtain

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(\mu) \propto \exp(-\Phi(\mu; y)).$$

Next, we assume that the noise η and prior distribution μ^0 are Gaussian, thus

$$\begin{aligned}\pi^y(u) &\propto \exp\left(-\frac{1}{2}\left|B^{-1/2}(y - \mathcal{G}(u))\right|^2 - \frac{1}{2}\left|\Sigma_0^{-1/2}(u - m_0)\right|^2\right)\\ &= \exp\left(-\frac{1}{2}|(y - \mathcal{G}(u))|_B^2 - \frac{1}{2}|(u - m_0)|_{\Sigma_0}^2\right)\end{aligned}$$

In terms of measures this is the statement that

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{1}{2}|(y - \mathcal{G}(u))|_B^2\right)$$

The maximum a posteriori estimator, or MAP estimator, is then

$$\operatorname{argmin}_{u \in \mathbb{R}^n}\left(\frac{1}{2}|y - \mathcal{G}(u)|_B^2 + \frac{1}{2}|u - m_0|_{\Sigma_0}^2\right).$$

This is a specific instance of the regularized minimization problem (3).

2 Algorithm

For common situations:

1. Set the Gaussian distribution of prior μ^0, such as $\mu^0 \sim N(0, \sigma)$.
2. Set the Gaussian distribution of observed noise η, that is $\eta \sim N(0, \Gamma)$
3. Set the linear/nonlinear system $y = \mathcal{G}(\mu) + \eta$. The linear & non-linear of the system depend on the operator mapping $\mathcal{G}$.
4. Set the Gaussian distribution of posterior μ^y, such as $\mu^y \sim N(m, \Sigma)$.

5. Compute the posterior mean and variance m, Σ using Theorem 6.20.

If the system $y = \mathcal{G}(\mu) + \eta$ is linear with the linear operator $\mathcal{G}$ being $A : X \longrightarrow \mathbb{R}^{lq}$, the posterior measure μ^y is also Gaussian with $\mu^y \sim N(m_0, \Sigma_0)$ where

$$m = m_0 + \Sigma_0 A^*(\Gamma + A\Sigma_0 A*)^{-1}(y - Am_0),$$
$$\Sigma = \Sigma_0 - \Sigma_0 A * (\Gamma + A\Sigma_0 A*)^{-1} A\Sigma_0.$$