A short intro to Tikhonov regularisation under a Bayesian View

Shitong Dai

Complementary to Homework2-Problem1

April 7, 2022

Introduction

I try to demonstrate the sound connection between the Bayesian approach and the Tikhonov regularisation, as proposed in lecture 5.

Consider a linear system of equations with additive noise e in the observed data, we assume it submits the Gaussian distribution. Then we have:

$$\begin{aligned} & d = Am + e \\ & m \in \mathcal{R}^n, \quad d, e \in \mathcal{R}^m, \quad A \in \mathcal{R}^{m \times n} \end{aligned}$$

and let the parameters m and the observed data d be the random variables. We assume that m and e are independent and identical Gaussian distribution with zero means,

$$m \sim \mathcal{N}\left(0, \gamma^2 C_m\right), \quad e \sim \mathcal{N}\left(0, \sigma^2 I\right).$$

The covariance of the noise, $\sigma^2 I$ indicates that each component of D is contaminated by independent and identically distributed random noise. Now, we define the a priori density of the form

$$\pi_{\text{prior}}(m) = \exp\left(-\frac{1}{2\gamma^2} m^T C_m^{-1} m\right),$$

and the likelihood density by assuming that the noise covariance is known is

$$\pi_{like}(d \mid m) = \exp\left(-\frac{1}{2\sigma^2}\|d - Am\|^2\right).$$

By Bayes theorem, we can define our a posterior density by the product of the likelihood and the a priori densities. The a posteriori density defines as

$$\begin{aligned} \pi_{\text{post}}(m \mid d) & \propto \pi_{\text{like}}(d \mid m) \pi_{\text{prior}}(m) \\ & \propto \exp\left(-\frac{1}{2\sigma^2}\|d - Am\|^2 - \frac{1}{2\gamma^2} m^T C_m^{-1} m\right) \\ & = \exp(-V(m \mid d)) \end{aligned}$$

where

$$V(m \mid d) = \frac{1}{2\sigma^2}\|d - Am\|^2 + \frac{1}{2\gamma^2} m^T C_m^{-1} m.$$

Based on the Gaussian likelihood and a priori densities, the a posteriori density $\pi_{\text{post}}(m \mid d)$ is also Gaussian. This condition is only valid for the linear case, and for the non-linear case, this might not hold.

Next, we demonstrate the sound connection between the Bayesian approach and the Tikhonov regularisation. Consider the matrix C_m is symmetric positive definite, so is its inverse; thus, it admits Cholesky factorisation of the form

$$C_m^{-1} = L^T L.$$

With this notation, we can reformulate our $V(m \mid d)$ into the Tikhonov regularisation functional, defines as

$$T(m) = 2\sigma^2 V(m \mid d) = \|d - Am\|^2 + \lambda^2 \|Lm\|^2, \quad \lambda = \frac{\sigma}{\gamma},$$

where λ is the ratio of the noise and the a priori variances. It also acts as a parameter which balancing the information contributions between the likelihood and the a priori densities.

The Tikhonov regularisation functional describes in the lecture 5 slides, plays a crucial role in the classical regularisation theory. From the Tikhonov regularisation perspectives,

The MAP estimate is the point in the parameter space that maximises the a posteriori probability density function (PDF). The problems of estimating the MAP can be simplified into minimising a misfit function of the negative logarithm of the a posteriori PDF. The negative logarithm of a posteriori PDF resembles the general Tikhonov regularisation misfit function, as shown in our Homework 2 in the previous section. The Maximum A Posteriori (MAP) estimator defines as:

$$\begin{aligned} m_{MAP} &= \arg\max_{m} \pi_{\text{post}}\left(m \mid d\right) \\ &= \arg\min_{m} -\log \pi_{\text{post}}\left(m \mid d\right) \\ &= \arg\min_{m} V(m \mid d). \end{aligned}$$

In this particular case, the MAP estimator is better describes as

$$m_{MAP} = \arg\min_{m} \left(\|d - Am\|^2 + \lambda^2 \|Lm\|^2\right).$$