Home assignment 1

Consider the one-dimensional deconvolution problem of estimating the input signal $f = f(t)$ from the noisy data

$$g(s_j) = \int_{-1}^{1} a(s_j - t) f(t) dt + e_j, \quad 1 \leq j \leq n, \tag{1}$$

where the convolution kernel is a Gaussian,

$$a(t) = \exp\left(-\frac{1}{2w^2} t^2\right),$$

and the sampling points s_j are evenly distributed in the interval $[0, 1]$,

$$s_j = \frac{j-1}{n-1}, \quad 1 \leq j \leq n.$$

Further, we assume that the signal f can be extended by zero outside the interval $[0, 1]$. The additive noise e_j is Gaussian white noise, that is, the components e_j are mutually independent, zero mean and variance σ^2 Gaussian random variables, denoted as

$$\mathbf{e} \sim \mathcal{N}(0, \sigma^2 I).$$

We discretize the equation (1), for simplicity, by using n evenly distributed discretization points,

$$g(s_j) \approx \frac{1}{n} \sum_{k=1}^{n} a(s_j - t_k) f(t_k) + e_j, \quad t_j = \frac{j-1}{n-1}, \quad 1 \leq j \leq n,$$

leading to a matrix equation

$$\mathbf{y} = A\mathbf{x} + \mathbf{e},$$

where $x_j = f(t_j)$, $y_j = g(s_j)$ and $A_{jk} = (1/n) a(s_j - t_k)$.

The likelihood function based on the model is

$$\pi(\mathbf{x} \mid \mathbf{y}) \propto \pi_{\text{noise}}(\mathbf{y} - A\mathbf{x}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - A\mathbf{x}\|^2\right).$$

For the prior, we use the Gaussian *white noise prior*,

$$\pi_{\text{prior}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2\gamma^2} \|\mathbf{x}\|^2\right),$$

where γ^2 is the prior variance. By Bayes' theorem, the posterior density is then

$$\begin{aligned}\pi_{\text{post}}(\mathbf{x}) &= \pi_{\text{prior}}(\mathbf{x})\pi(\mathbf{y} \mid \mathbf{x}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - A\mathbf{x}\|^2 - \frac{1}{2\gamma^2}\|\mathbf{x}\|^2\right).\end{aligned}$$

The *Maximum A Posteriori* (MAP) estimate is then

$$\begin{aligned}\mathbf{x}_{\text{MAP}} &= \arg\min\left(\frac{1}{2\sigma^2}\|\mathbf{y} - A\mathbf{x}\|^2 + \frac{1}{2\gamma^2}\|\mathbf{x}\|^2\right) \\ &= \arg\min\left(\|\mathbf{y} - A\mathbf{x}\|^2 + \delta\|\mathbf{x}\|^2\right), \quad \delta = \frac{\sigma^2}{\gamma^2}.\end{aligned}$$

The classical (non-statistical) regularization literature calls the above MAP estimate the *Tikhonov regularized solution* of the inverse problem, and the parameter δ is referred to as *regularization parameter.*

There is a huge litearature about how to choose the regularization parameter, one of the most common ways being the *Morozov discrepancy principle*, which is usually explained as follows: assume that we have an estimate of the noise level,

$$\|\mathbf{e}\| \approx \varepsilon.$$

It means that any vector $\mathbf{x}$ that satisfies the condition

$$\|\mathbf{y} - A\mathbf{x}\| \leq \varepsilon \tag{2}$$

can be considered as an acceptable solution, since the data is satisfied within the noise level. The Morozov discrepancy principle is to find the *largest* δ *so that the discrepancy condition (2) is satisfied.*

In practice, let $\mathbf{x}_\delta$ denote the solution

$$\begin{aligned}\mathbf{x}_\delta &= \arg\min\left(\|\mathbf{y} - A\mathbf{x}\|^2 + \delta\|\mathbf{x}\|^2\right) \\ &= \arg\min\left\|\begin{bmatrix} A \\ \sqrt{\delta}I \end{bmatrix}\mathbf{x} - \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix}\right\|.\end{aligned}$$

We solve δ from the condition

$$d(\delta) = \|\mathbf{y} - A\mathbf{x}_\delta\| = \varepsilon \Rightarrow \delta = \delta(\varepsilon).$$

The first assignment is to investigate this method.

1. Write a program that plots the curve $\delta \mapsto d(\delta)$. What are the limits $\delta \to 0+$ and $\delta \to \infty$? Clearly, these are the limits for ε where the Morozov discrepancy principle is operational.

2. Write a program that solves $\delta = \delta(\varepsilon)$. You can use, e.g., a binary search algorithm.

3. Analyze the noise level condition: if $\mathbf{e} \sim \mathcal{N}(0, \sigma^2 I)$, draw a large sample of realizations with different variances σ^2, and find an empirical "rule of thumb", $\varepsilon = \varepsilon(\sigma)$ for how the noise level ε should be chosen when σ^2 is given and vice versa.

4. Generate data using a known input signal $\mathbf{x}$. Solve the regularization parameter δ with different values of ε, and check what the prior variance $\gamma^2 = \sigma^2/\delta$ is, when δ is computed using the Morozov discrepancy principle. Does this standard deviation correspond to the dynamical range of your true signal?

5. With your input signal, plot the so called *L–curve*, that is, plot the parametrized curve

$$\delta \mapsto (\log \|\mathbf{x}_\delta\|, \log \|\mathbf{y} - A\mathbf{x}_\delta\|).$$

   The folklore has that this curve should have a shape of the letter "L", and the optimal value should be somewhere in the corner of the curve. try with different noise levels and see if you are able to reporduce this phenomenon. The method of choosing the regularization parameter with this technique is called the *L–curve method*. Any comments? Can you argue why this method might work?

Attach the Matlab code and plots to your report. Write out the details and your conclusions, not just minimal answers to the problems.

You can use the Matlab code posted on the course web page as a starting point.