Statistics for Data Science
Wintersemester 2024/25
Solutions to the 11^{th} exercise

1. (a) In this case, we have the prior density

$$f(x) \propto \exp\Big(-\frac{1}{2}(x-1)^2\Big)$$

and the measurement noise density

$$\nu(\eta) \propto \exp\Big(-\frac{1}{2}\eta^2\Big).$$

The posterior is given by Bayes' theorem:

$$\begin{aligned} f(x|y) &\propto \nu(y-x^2)f(x) \\ &\propto \exp\Big(-\frac{1}{2}(y-x^2)^2\Big)\exp\Big(-\frac{1}{2}(x-1)^2\Big) \\ &= \exp\Big(-\frac{1}{2}(y-x^2)^2 - \frac{1}{2}(x-1)^2\Big). \end{aligned}$$

(b) The posterior density $f(x|y=2)$ is illustrated in Figure 1.

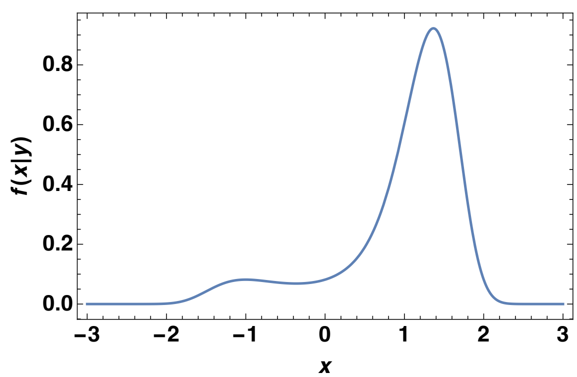

Figure 1: The posterior density $f(x|y=2)$.

We can find the MAP estimator by solving the minimum of the negative log-posterior

$$F(x) = (2-x^2)^2 + (x-1)^2.$$

The critical points can be found by solving the roots of F':

$$0 = F'(x) = -4x(2-x^2) + 2(x-1) = 4x^3 - 6x - 2 = (x+1)(4x^2-4x-2).$$

The critical points are therefore

$$x_1 = -1, \quad x_2 = \frac{1-\sqrt{3}}{2}, \quad x_3 = \frac{1+\sqrt{3}}{2}.$$

Since

$$F''(x) = 12x^2 - 6,$$

we find that $F''(x_1) > 0$, $F''(x_2) < 0$, and $F''(x_3) > 0$. Therefore x_1 and x_3 are local minima of F (and hence local maxima of $f(x|y = 2)$). Since $F(x) \xrightarrow{x\to\pm\infty} \infty$ and $F(x_3) < F(x_1)$, we conclude that

$$\hat{x}_{\text{MAP}} = x_3 = \frac{1+\sqrt{3}}{2}.$$

2. (a) The measurement noise has the density

$$\nu(\eta) \propto \exp\left(-\frac{1}{2\sigma^2}\|\eta\|^2 \right),$$

while the prior has the density

$$f(x) \propto \exp\left(-\frac{1}{2\gamma^2}\|x - x_0\|^2 \right).$$

The posterior is given by Bayes' theorem:

$$\begin{aligned} f(x|y) &\propto \nu(y - Ax)f(x) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\|y - Ax\|^2 \right) \exp\left(-\frac{1}{2\gamma^2}\|x - x_0\|^2 \right) \\ &= \exp\left(-\frac{1}{2\sigma^2}\|y - Ax\|^2 - \frac{1}{2\gamma^2}\|x - x_0\|^2 \right). \end{aligned}$$

(b) The MAP estimate clearly satisfies

$$\begin{aligned} \hat{x}_{\text{MAP}} &= \arg\max_{x\in\mathbb{R}^d} f(x|y) \\ &= \arg\max_{x\in\mathbb{R}^d} \left\{ \exp\left(-\frac{1}{2\sigma^2}\|y - Ax\|^2 - \frac{1}{2\gamma^2}\|x - x_0\|^2 \right) \right\} \\ &\overset{!}{=} \arg\min_{x\in\mathbb{R}^d} \left\{ \frac{1}{2\sigma^2}\|y - Ax\|^2 + \frac{1}{2\gamma^2}\|x - x_0\|^2 \right\} \\ &= \arg\min_{x\in\mathbb{R}^d} \left\{ \|y - Ax\|^2 + \frac{\sigma^2}{\gamma^2}\|x - x_0\|^2 \right\} \\ &= \arg\min_{x\in\mathbb{R}^d} \left\{ \left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} x - \begin{bmatrix} y \\ \lambda x_0 \end{bmatrix} \right\|^2 \right\}, \quad \text{where } \lambda = \frac{\sigma}{\gamma}. \end{aligned}$$

The minimizer can be found by solving the normal equation

$$\begin{aligned} & \begin{bmatrix} A \\ \lambda I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} A \\ \lambda I \end{bmatrix} \hat{x}_{\text{MAP}} = \begin{bmatrix} A \\ \lambda I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} y \\ \lambda x_0 \end{bmatrix} \\ \Leftrightarrow \quad & \begin{bmatrix} A^{\mathrm{T}} & \lambda I \end{bmatrix} \begin{bmatrix} A \\ \lambda I \end{bmatrix} \hat{x}_{\text{MAP}} = A^{\mathrm{T}}y + \lambda^2 x_0 \\ \Leftrightarrow \quad & (A^{\mathrm{T}}A + \lambda^2 I)\hat{x}_{\text{MAP}} = A^{\mathrm{T}}y + \lambda^2 x_0 \\ \Leftrightarrow \quad & \hat{x}_{\text{MAP}} = (A^{\mathrm{T}}A + \lambda^2 I)^{-1}(A^{\mathrm{T}}y + \lambda^2 x_0). \end{aligned}$$

Note that $A^{\mathrm{T}}A + \lambda^2 I$ is positive definite and thus invertible for all $\lambda > 0$.

3. We derive the likelihood density $f(y|x)$ for the multiplicative noise model

$$y_j = a_j x_j, \quad j = 1, \dots, n, \quad y, a, x \in \mathbb{R}^n,$$

where a_j are independent, log-normally distributed random variables, i.e., $\log a_j \sim \mathcal{N}(\log a_0, \sigma^2)$. Moreover, a is independent of x and $x_j > 0$ for $j = 1, \dots, n$.

We introduce $w_j := \log y_j = \log a_j + \log x_j$ for $j = 1, \dots, n$. By assumption, x_j and a_j are mutually independent and so fixing the realization $x_j = x_j$ does not affect the probability distribution of $\log a_j$.† We notice that w_j conditioned on x_j is distributed normally as $\log a_j$ shifted by a constant $\log x_j$. Writing this down, we get

$$\begin{aligned} f(w_j \mid x_j) = f_{\log a_j}(w_j - \log x_j) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w_j - \log x_j - \log a_0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(w_j - \log(a_0 x_j))^2 \right). \end{aligned}$$

Since $w_j = \log y_j$, we have $\mathrm{d}w_j = \frac{1}{y_j}\mathrm{d}y_j$. By change of variable, it follows that

$$\begin{aligned} f(y_j|x_j)\,\mathrm{d}y_j = f(w_j|x_j)\,\mathrm{d}w_j &= f(\log y_j|x_j)\frac{1}{y_j}\,\mathrm{d}y_j \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y_j} \exp\left(-\frac{1}{2\sigma^2}(\log y_j - \log(a_0 x_j))^2 \right) \mathrm{d}y_j \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y_j} \exp\left(-\frac{1}{2\sigma^2}\left(\log \frac{y_j}{a_0 x_j}\right)^2 \right) \mathrm{d}y_j. \end{aligned}$$

By the previous argument, we have $f(y_i|x_i) = f_{a_i}(\frac{y_i}{x_i})$ for $a_i \sim \log\mathcal{N}(\log a_0, \sigma^2)$ and these are independent, so

$$f(y|x) = \prod_{j=1}^{n} f(y_j|x_j) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \frac{1}{y_1 \cdots y_n} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} \left(\log \frac{y_j}{a_0 x_j}\right)^2 \right),$$

which is the desired result.

4. (a) We have

$$f(x, \alpha|y) \propto f(y|x, \alpha) f(x, \alpha) = f(y|x, \alpha) f(x|\alpha) f(\alpha),$$

where

$$f(y|x, \alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(y - \frac{1}{2}x\right)^2 \right), \qquad \text{(likelihood)}$$

$$f(x|\alpha) = \frac{\alpha^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\alpha x^2 \right), \qquad \text{(conditional prior)}$$

$$f(\alpha) = \mathbf{1}_{(0,\infty)}(\alpha)\sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}\alpha^2 \right). \qquad \text{(hyperprior)}$$

†It is a simple exercise to verify that the mutual independence of random variables x and y implies that the composite random variables $f(x)$ and $g(y)$, too, are mutually independent for functions f and g defined on the (Borel measurable) codomains of x and y.

Therefore

$$f(x,\alpha|y) \propto \mathbf{1}_{(0,\infty)}(\alpha)\alpha^{1/2}\exp\left(-\frac{1}{2}\left(y-\frac{1}{2}x\right)^2-\frac{1}{2}\alpha x^2-\frac{1}{2}\alpha^2\right), \quad x\in\mathbb{R},\ \alpha\in\mathbb{R}.$$

(b) We observe $y=\frac{3}{2}$. Let us write the posterior as

$$f(x,\alpha|\tfrac{3}{2}) \propto \exp(-F(x,\alpha)),$$

where

$$F(x,\alpha) := -\frac{1}{2}\log\alpha+\frac{1}{2}\left(\frac{3}{2}-\frac{1}{2}x\right)^2+\frac{1}{2}\alpha x^2+\frac{1}{2}\alpha^2, \quad x\in\mathbb{R},\ \alpha>0.$$

Let us begin by solving the zeros of the gradient. The partial derivatives of F are

$$\begin{aligned}0 &= \frac{\partial}{\partial x}\left(-\frac{1}{2}\log\alpha+\frac{1}{2}\left(\frac{3}{2}-\frac{1}{2}x\right)^2+\frac{1}{2}\alpha x^2+\frac{1}{2}\alpha^2\right)\\ &= -\frac{1}{2}\left(\frac{3}{2}-\frac{1}{2}x\right)+\alpha x\\ &= -\frac{3}{4}+\frac{1}{4}x+\alpha x\end{aligned}$$

and

$$\begin{aligned}0 &= \frac{\partial}{\partial \alpha}\left(-\frac{1}{2}\log\alpha+\frac{1}{2}\left(\frac{3}{2}-\frac{1}{2}x\right)^2+\frac{1}{2}\alpha x^2+\frac{1}{2}\alpha^2\right)\\ &= -\frac{1}{2\alpha}+\frac{1}{2}x^2+\alpha.\end{aligned}$$

The only eligible ($\alpha>0$) solution to the nonlinear system of equations

$$\begin{cases}-\frac{3}{4}+\frac{1}{4}x+\alpha x=0\\ -\frac{1}{2\alpha}+\frac{1}{2}x^2+\alpha=0\end{cases}$$

is $(x,\alpha)=(1,1/2)$.

To establish that this point is a local minimum of F (and, consequently, a local maximum of the posterior), we proceed as follows. The Hessian is given by

$$\nabla^2F(x,\alpha)=\begin{bmatrix}\frac{1}{4}+\alpha & x\\ x & 1+\frac{1}{2\alpha^2}\end{bmatrix} \quad\Rightarrow\quad \nabla^2F(1,\tfrac{1}{2})=\begin{bmatrix}\frac{3}{4} & 1\\ 1 & 3\end{bmatrix}.$$

This is positive definite since the eigenvalues $\lambda=\frac{15\pm\sqrt{145}}{8}$ are both positive. Thus $(x,\alpha)=(1,1/2)$ is a local minimum of F (and thus a local maximum of the posterior).

Since F is smooth with only one critical point $(x,\alpha)\in\mathbb{R}\times(0,\infty)$ and we easily observe that

$$\begin{aligned}&F(x,\alpha)\to+\infty \quad \text{as } (x,\alpha)\to(x^*,0) \text{ for any } x^*\in\mathbb{R}\\ &F(x,\alpha)\to+\infty \quad \text{as } |x|\to\infty \text{ or } \alpha\to\infty,\end{aligned}$$

we conclude that $(x,\alpha)=(1,1/2)$ is the global minimum of F and therefore the MAP estimator of $f(x,\alpha|\frac{3}{2})$.

Remark on numerical solution. In problems of this kind, one could consider for example the following kind of alternating minimization algorithm:

- Set $k = 0$ and choose an initial guess for α, e.g., $\alpha_0 = 1$.

  repeat

  - Find the minimizer

$$x_k = \arg\min_{x \in \mathbb{R}} F(x, \alpha_k) = \arg\min_{x \in \mathbb{R}} \left\{ \left(\frac{3}{2} - \frac{1}{2}x \right)^2 + \alpha_k x^2 \right\}.$$

  - Solve $\alpha > 0$ from $\frac{\partial J(x_k, \alpha)}{\partial \alpha} = 0$ and set $\alpha_{k+1} = \alpha$.
  - Set $k \leftarrow k + 1$.

  until convergence