Exercise 1

Given:

$$\boldsymbol{\theta} \sim \mathcal{N}_2(\boldsymbol{m_0}, P_0), \quad p(\boldsymbol{y}|\boldsymbol{\theta}) \sim \mathcal{N}_n(H\boldsymbol{\theta}, \sigma^2 I_N) \tag{1}$$

To be Proved:

$$\boldsymbol{\theta}|\boldsymbol{y} \sim N_2(\nu, C) \tag{2}$$

where

$$C^{-1} = H'\sigma^{-2}H + P_0^{-1}$$

and

$$\nu = C(H'\sigma^{-2}\boldsymbol{y} + P_0^{-1}\boldsymbol{m_0})$$

Using Bayes' Rule, we can write the posterior distribution from 2 as:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

We can rewrite the right-hand side of the above proportionality as a product of two Gaussians with mean and co-variance from 1:

$$\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{y}) &\propto \frac{1}{\sqrt{2\pi}|\sigma^2\boldsymbol{I}|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{y} - H\boldsymbol{\theta})'(\sigma^2\boldsymbol{I_n})^{-1}(\boldsymbol{y} - H\boldsymbol{\theta})) \\
&\quad \times \frac{1}{\sqrt{2\pi}|P_0|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m_0})'P_0^{-1}(\boldsymbol{\theta} - \boldsymbol{m_0})) \\
&\propto \exp(-\frac{1}{2\sigma^2}(\boldsymbol{y} - H\boldsymbol{\theta})'(\boldsymbol{y} - H\boldsymbol{\theta})) \times \exp(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m_0})'P_0^{-1}(\boldsymbol{\theta} - \boldsymbol{m_0})) \\
&\propto \exp(-\frac{1}{2}(\frac{1}{\sigma^2}(\boldsymbol{y} - H\boldsymbol{\theta})'(\boldsymbol{y} - H\boldsymbol{\theta}) + (\boldsymbol{\theta} - \boldsymbol{m_0})'P_0^{-1}(\boldsymbol{\theta} - \boldsymbol{m_0}))) \\
&\propto \exp(-\frac{1}{2}(-\frac{2}{\sigma^2}\boldsymbol{y}'H\boldsymbol{\theta} + \frac{1}{\sigma^2}\boldsymbol{\theta}'H'H\boldsymbol{\theta} + \boldsymbol{\theta}'P_0^{-1}\boldsymbol{\theta} - 2\boldsymbol{m_0'}P_0^{-1}\boldsymbol{\theta})) \qquad (3)
\end{aligned}$$

In the above process, we simplify by removing terms that do not have $\boldsymbol{\theta}$. Comparing 3 to a generic normal distribution $\mathcal{N}(\boldsymbol{\nu}, \boldsymbol{C})$:

$$\begin{aligned}
\mathcal{N}(\boldsymbol{\nu}, \boldsymbol{C}) &\propto \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\nu})'\boldsymbol{C}^{-1}(\boldsymbol{x} - \boldsymbol{\nu})) \\
&\propto \exp(-\frac{1}{2}(\boldsymbol{x}'\boldsymbol{C}^{-1}\boldsymbol{x} - 2\boldsymbol{\nu}'\boldsymbol{C}^{-1}\boldsymbol{x})) \qquad (4)
\end{aligned}$$

We can rewrite equation 3 in the form of a generic normal distribution as given in equation 4. To do this, we equate the quadratic terms to get the mean and the linear terms to get the covariance.

Equating the quadratic terms:

$$\begin{aligned}
\boldsymbol{x}'\boldsymbol{C}^{-1}\boldsymbol{x} &= \frac{1}{\sigma^2}\boldsymbol{\theta}'H'H\boldsymbol{\theta} + \boldsymbol{\theta}'P_0^{-1}\boldsymbol{\theta} \\
&= \boldsymbol{\theta}'(\frac{1}{\sigma^2}H'H + P_0^{-1})\boldsymbol{\theta}
\end{aligned}$$

Hence, the covariance in the new distribution is:

$$\boldsymbol{C} = (\frac{1}{\sigma^2}H'H + P_0^{-1})^{-1} \tag{5}$$

Now, we equate the linear terms to get $\boldsymbol{\nu}$:

$$\begin{aligned}
-2\boldsymbol{\nu}'\boldsymbol{C}^{-1}\boldsymbol{x} &= -\frac{2}{\sigma^2}\boldsymbol{y}'H\boldsymbol{\theta} - 2\boldsymbol{m_0}'P_0^{-1}\boldsymbol{\theta} \\
\boldsymbol{\nu}'\boldsymbol{C}^{-1} &= \frac{1}{\sigma^2}\boldsymbol{y}'H + \boldsymbol{m_0}'P_0^{-1} \\
\boldsymbol{\nu}'\boldsymbol{C}^{-1}\boldsymbol{C} &= (\frac{1}{\sigma^2}\boldsymbol{y}'H + \boldsymbol{m_0}'P_0^{-1})\boldsymbol{C} \\
\boldsymbol{\nu}' &= (\frac{1}{\sigma^2}\boldsymbol{y}'H + \boldsymbol{m_0}'P_0^{-1})\boldsymbol{C} \\
\boldsymbol{\nu} &= \boldsymbol{C}(\frac{1}{\sigma^2}H'\boldsymbol{y} + P_0^{-1}\boldsymbol{m_0}) \qquad (6)
\end{aligned}$$

Hence, we can write the distribution of $\boldsymbol{\theta}|\boldsymbol{y}$ as a Gaussian with mean and covariance from 6 and 5:

$$\boldsymbol{\theta}|\boldsymbol{y} \sim \mathcal{N}_2(\boldsymbol{\nu}, \boldsymbol{C})$$

where

$$\boldsymbol{\nu} = \boldsymbol{C}(\frac{1}{\sigma^2}H'\boldsymbol{y} + P_0^{-1}\boldsymbol{m_0})$$

and

$$\boldsymbol{C^{-1}} = \frac{1}{\sigma^2}H'H + P_0^{-1}$$

which proves 2.

Exercise 2

Given:

$$P_k^{-1} = H_k'\sigma^{-2}H_k + P_{k-1}^{-1} \tag{7}$$

Woodbury Matrix Identity:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(c^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{8}$$

To be Proved:

$$P_k = P_{k-1} - P_{k-1}H_k'(\sigma^2 + H_kP_{k-1}H_k')^{-1}H_kP_{k-1} \tag{9}$$

Substituting $A = P_{k-1}^{-1}, C = \sigma^{-2}, U = H_k', V = H_k$ in 8, we get:

$$(H_k'\sigma^{-2}H_k + P_{k-1}^{-1})^{-1} = P_{k-1} - P_{k-1}H_k'(\sigma^2 + H_kP_{k-1}H_k')^{-1}H_kP_{k-1}$$

Comparing the above equation to 7, the LHS is just P_k^{-1}, so we get:

$$\begin{aligned}(P_k^{-1})^{-1} &= P_{k-1} - P_{k-1}H_k'(\sigma^2 + H_kP_{k-1}H_k')^{-1}H_kP_{k-1} \\ P_k &= P_{k-1} - P_{k-1}H_k'(\sigma^2 + H_kP_{k-1}H_k')^{-1}H_kP_{k-1}\end{aligned}$$

which proves 9.

- -

Using Woodbury Matrix Identity is numerically advantageous to compute P_k because the inverse term is a scalar, which is trivial to invert(compared to the inverse in the original equation for P_k). Let us verify this by checking the dimensions of each term of the inverse in equation 9:

$$H_k \in \mathbb{R}^{1\times k}, P_{k-1} \in \mathbb{R}^{k\times k}, \sigma^2 \in \mathbb{R}$$

$$\therefore (\sigma^2 + H_kP_{k-1}H_k') \in \mathbb{R}^1$$

Comparing this to the inverse term in the original equation for P_k, as described in 7:

$$H_k'\sigma^{-2}H_k + P_{k-1}^{-1} \in \mathbb{R}^{k\times k}$$

Computing this can be non-trivial for large values of k. Hence, it would be numerically advantageous to use the Woodbury Matrix Identity to compute the updated co-variance matrix.

Exercise 3

1. The code that computes the posterior distribution is in the Appendix at the end of the report.

2. Posterior Means and Co-Variance Matrices:

   After 10 observations:

$$m_{10} = \begin{bmatrix} 40067.19 \\ 233.92 \\ 663.49 \\ 23.29 \end{bmatrix}, P_{10} = \begin{bmatrix} 98.95 & -2.11 & -4.00 & 0.31 \\ -2.11 & 94.71 & -12.06 & 0.05 \\ -4.00 & -12.06 & 67.48 & -7.80 \\ 0.31 & 0.05 & -7.80 & 1.02 \end{bmatrix}$$

   After full dataset:

$$m_{30} = \begin{bmatrix} 40072.17 \\ 355.67 \\ 1324.71 \\ -53.82 \end{bmatrix} P_{30} = \begin{bmatrix} 98.74 & -2.73 & -5.02 & 0.43 \\ -2.73 & 91.88 & -21.13 & 1.17 \\ -5.02 & -21.13 & 26.78 & -2.63 \\ 0.43 & 1.17 & -2.63 & 0.28 \end{bmatrix}$$

3. Figure 1 shows the fitted curve in red, superimposed on the observed data (shown as a scatter plot). As we can see, the curve fits the observed data quite well.

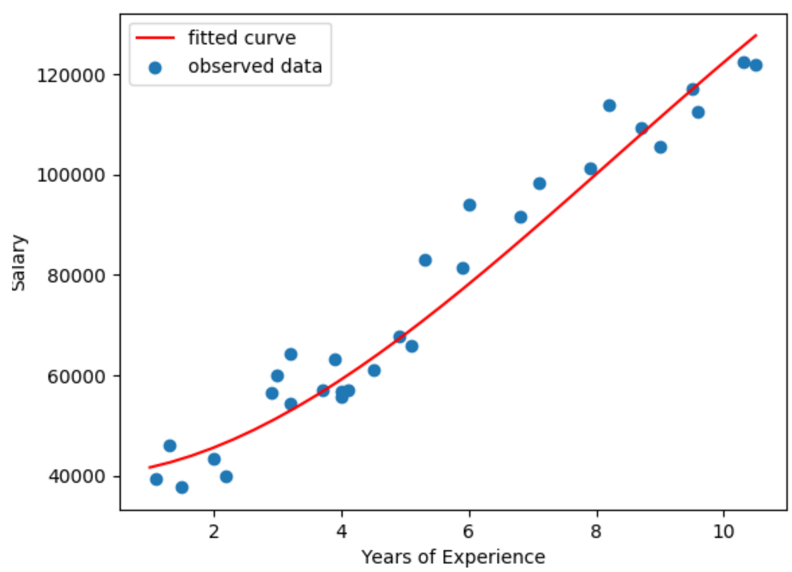

Figure 1: Fitted curve superimposed on the observed data

Exercise 4

Given:

$$(\theta_{k-1}|y_{1:k-1}) \sim \mathcal{N}(m_{k-1}, P_{k-1}), \quad (\theta_k|\theta_{k-1}) \sim \mathcal{N}(A\theta_{k-1}, \theta) \tag{10}$$

and **Lemma 1:** If the random vectors $\boldsymbol{X} \in \mathbb{R}^n$ and $\boldsymbol{Y} \in \mathbb{R}^m$ satisfy

$$X \sim \mathcal{N}(m, P)$$
$$Y|X \sim \mathcal{N}(Hx + u, R)$$

then

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} m \\ Hm + u \end{bmatrix}, \begin{bmatrix} P & PH' \\ HP & HPH' + R \end{bmatrix}) \tag{11}$$

To be Proved:

$$\theta_k|y_{1:k-1} \sim \mathcal{N}(m_k^-, P_k^-)$$

where

$$P_k^- = AP_{k-1}A' + Q, \quad m_k^- = Am_{k-1} \tag{12}$$

The distribution $P(\theta_k|y_{1:k-1})$ can be written as a marginalization over $\theta_{k-1}|y_{1:k}$:

$$P(\theta_k|y_{1:k-1}) = \int P(\theta_k, \theta_{k-1}|y_{1:k-1}) d(\theta_{k-1}|y_{1:k-1})$$

Since θ_k is independent of θ_{k-1} given previous observations $y_{1:k-1}$:

$$P(\theta_k, \theta_{k-1}|y_{1:k-1}) = P(\theta_k|\theta_{k-1}, y_{1:k-1})P(\theta_{k-1}|y_{1:k-1})$$

Since θ_k is independent of previous observations $y_{1:k-1}$ given θ_{k-1}:

$$P(\theta_k, \theta_{k-1}|y_{1:k-1}) = P(\theta_k|\theta_{k-1})P(\theta_{k-1}|y_{1:k-1})$$

Now, to represent the joint distribution of $(\theta_k, \theta_{k-1})|y_{1:k-1}$ as a Multivariate Gaussian, we use Lemma 1.

We can see clearly that $(\theta_{k-1}|y_{1:k-1})$ and $(\theta_k|\theta_{k-1})$ from 10 have a distribution of the same form as X and Y from Lemma 1.

Rewriting 11 by substituting from 10, we get:

$$\begin{bmatrix} \theta_{k-1}|y_{1:k-1} \\ \theta_k|\theta_{k-1} \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} m_{k-1} \\ Am_{k-1} \end{bmatrix}, \begin{bmatrix} P_{k-1} & P_{k-1}A' \\ AP_{k-1} & AP_{k-1}A' + Q \end{bmatrix})$$

Using the property that *The marginal of a Joint-Gaussian is a Gaussian*, we can marginalize the above joint Gaussian into its component distribution to get $\theta_k|\theta_{k-1}$:

$$\theta_k|\theta_{k-1} \sim \mathcal{N}(Am_{k-1}, AP_{k-1}A' + Q)$$

which proves 12.