Exercise Round 2.

Exercise 1. (Linear Bayesian Estimation)

The priors for parameters a_1 and a_2 are independent Gaussian as follows:

$$a_1 \sim \mathrm{N}(0, \sigma^2)$$
$$a_2 \sim \mathrm{N}(0, \sigma^2),$$

where the variance σ^2 is known. The measurements y_k are modeled as

$$y_k = a_1\, x_k + a_2 + e_k, \qquad k = 1, \ldots, n$$

where e_k's are independent Gaussian error terms with mean 0 and variance 1, that is, $e_k \sim \mathrm{N}(0, 1)$. The values x_k are fixed and known. The posterior distribution can be now written as

$$p(\mathbf{a} \,|\, y_1, \ldots, y_n) = Z \exp(-\frac{1}{2} \sum_{k=1}^{n} (a_1\, x_k + a_2 - y_k)^2) \; \exp(-\frac{1}{2\sigma^2} a_1^2) \; \exp(-\frac{1}{2\sigma^2} a_2^2)$$

where Z is the normalization term, which is independent of the parameter $\mathbf{a} = (a_1\; a_2)^T$. The posterior distribution can be seen to be Gaussian and your task is to derive its mean and covariance.

A) Write the exponent of the posterior distribution in matrix form as in Exercise 1 of Round 1 (in terms of $\mathbf{y}$, $\mathbf{X}$, $\mathbf{a}$ and σ^2).

B) Because Gaussian distribution is always symmetric, its mean $\mathbf{m}$ is at the maximum of the distribution. Solve the mean by computing the gradient of the exponent and finding where it vanishes.

C) Find the covariance of the distribution by computing the second derivative matrix (Hessian matrix) $\mathbf{H}$ of the exponent. The covariance is then $\Sigma = -\mathbf{H}^{-1}$ (why?).

D) What is the resulting posterior distribution? What is the relationship with the LS-estimate in Exercise 1 of round 1?

Exercise 2. (Linear Regression with Kalman Filter)

The model in Exercise 1 can be written as a linear state space model as follows:

$$\begin{aligned} \mathbf{a}_k &= \mathbf{a}_{k-1} \\ y_k &= \mathbf{H}_k \, \mathbf{a}_k + \epsilon_k, \end{aligned}$$

where $\mathbf{H}_k = (x_k \; 1)$, $\mathbf{a}_0 \sim \mathrm{N}(0, \sigma^2 \mathbf{I})$ and $\epsilon_k \sim \mathrm{N}(0, 1)$. The state in the model is now $\mathbf{a}_k = (a_{1,k} \; a_{2,k})^T$ and the measurements are y_k for $k = 1, \ldots, n$. Assume that Kalman filter is used for processing the measurements $y_1, \ldots, y_n$. Your task is to prove that at time step n, the mean and covariance of $\mathbf{a}_n$ computed by the Kalman filter are the same as the mean and covariance of the posterior distribution computed in the previous exercise.

The Kalman filter equations for the above model can be written as:

$$\begin{aligned} S_k &= \mathbf{H}_k \, \mathbf{P}_{k-1} \, \mathbf{H}_k^T + 1 \\ \mathbf{K}_k &= \mathbf{P}_{k-1} \, \mathbf{H}_k^T \, S_k^{-1} \\ \mathbf{m}_k &= \mathbf{m}_{k-1} + \mathbf{K}_k \, (y_k - \mathbf{H}_k \, \mathbf{m}_{k-1}) \\ \mathbf{P}_k &= \mathbf{P}_{k-1} - \mathbf{K}_k \, S_k \, \mathbf{K}_k^T. \end{aligned}$$

A) Write formulas for the posterior mean $\mathbf{m}_{k-1}$ and covariance $\mathbf{P}_{k-1}$ assuming that they are the same as what would be obtained if the pairs $\{(y_i, x_i) \; : \; i = 1, \ldots, k-1\}$ were (batch) processed as in the previous exercise. Write similar equations for the mean $\mathbf{m}_k$ and covariance $\mathbf{P}_k$. Show that the posterior means can be expressed in form

$$\begin{aligned} \mathbf{m}_{k-1} &= \mathbf{P}_{k-1} \, \mathbf{X}_{k-1}^T \, \mathbf{y}_{k-1} \\ \mathbf{m}_k &= \mathbf{P}_k \, \mathbf{X}_k^T \, \mathbf{y}_k, \end{aligned}$$

where $\mathbf{X}_{k-1}$ and $\mathbf{y}_{k-1}$ have been constructed as $\mathbf{X}$ and $\mathbf{y}$ in the previous exercise, except that only the pairs $\{(y_i, x_i) \; : \; i = 1, \ldots, k-1\}$ have been used. $\mathbf{X}_k$ and $\mathbf{y}_k$ have been constructed similarly from pairs up to step k.

B) Rewrite the expressions $\mathbf{X}_k^T \, \mathbf{X}_k$ and $\mathbf{X}_k^T \, \mathbf{y}_k$ in terms of $\mathbf{X}_{k-1}$, $\mathbf{y}_{k-1}$, $\mathbf{H}_k$ and y_k. Substitute these into the expressions of $\mathbf{m}_k$ and $\mathbf{P}_k$ obtained in A).

C) Expand the expression of the covariance $\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{K}_k \, S_k \, \mathbf{K}_k^T$ by substituting the expressions for $\mathbf{K}_k$ and S_k. Convert it to simpler form by applying the matrix inversion lemma:

$$\mathbf{P}_{k-1} - \mathbf{P}_{k-1} \, \mathbf{H}_k^T \, (\mathbf{H}_k \, \mathbf{P}_{k-1} \, \mathbf{H}_k^T + 1)^{-1} \, \mathbf{H}_k \, \mathbf{P}_{k-1} = (\mathbf{P}_{k-1}^{-1} + \mathbf{H}_k^T \, \mathbf{H}_k)^{-1}.$$

Show that this expression for $\mathbf{P}_k$ is equivalent to the expression in A).

D) Expand the expression of the mean $\mathbf{m}_k = \mathbf{m}_{k-1} + \mathbf{K}_k \, (y_k - \mathbf{H}_k \, \mathbf{m}_{k-1})$ and show that the result is equivalent to the expression obtained in A). Hint: The Kalman gain can also be written as $\mathbf{K}_k = \mathbf{P}_k \, \mathbf{H}_k^T$.

E) Prove by induction argument that the mean and covariance computed by the Kalman filter at step n is the same as the posterior mean and covariance obtained in the previous exercise.

Exercise 3. (Gaussian Identities)

Recall that the Gaussian probability density is defined as

$$\mathrm{N}(\mathbf{x} \,|\, \mathbf{m}, \mathbf{P}) = \frac{1}{(2\,\pi)^{n/2} \, |\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \, \mathbf{P}^{-1} \, (\mathbf{x} - \mathbf{m}) \right)$$

Derive the following Gaussian identities:

A) Let $\mathbf{x}$ and $\mathbf{y}$ have the Gaussian densities

$$p(\mathbf{x}) = \mathrm{N}(\mathbf{x} \,|\, \mathbf{m}, \mathbf{P}), \qquad p(\mathbf{y} \,|\, \mathbf{x}) = \mathrm{N}(\mathbf{y} \,|\, \mathbf{H}\,\mathbf{x}, \mathbf{R}),$$

then the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{m} \\ \mathbf{H}\,\mathbf{m} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{P}\,\mathbf{H}^T \\ \mathbf{H}\,\mathbf{P} & \mathbf{H}\,\mathbf{P}\,\mathbf{H}^T + \mathbf{R} \end{pmatrix} \right)$$

and the marginal distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim \mathrm{N}(\mathbf{H}\,\mathbf{m}, \mathbf{H}\,\mathbf{P}\,\mathbf{H}^T + \mathbf{R}).$$

Hint: Use the properties of expectation $\mathrm{E}[\mathbf{H}\,\mathbf{x} + \mathbf{r}] = \mathbf{H}\ \mathrm{E}[\mathbf{x}] + \mathrm{E}[\mathbf{r}]$ and $\mathrm{Cov}[\mathbf{H}\,\mathbf{x} + \mathbf{r}] = \mathbf{H}\ \mathrm{Cov}[\mathbf{x}]\,\mathbf{H}^T + \mathrm{Cov}[\mathbf{r}]$ (if $\mathbf{x}$ and $\mathbf{r}$ independent).

B) Write down the explicit expression for the joint and marginal probability densities above:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \,|\, \mathbf{x}) \, p(\mathbf{x}) = ?$$
$$p(\mathbf{y}) = \int p(\mathbf{y} \,|\, \mathbf{x}) \, p(\mathbf{x}) \, d\mathbf{x} = ?$$

C) If the random variables $\mathbf{x}$ and $\mathbf{y}$ have the joint Gaussian probability density

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{pmatrix} \right),$$

then the conditional density of $\mathbf{x}$ given $\mathbf{y}$ is given as follows:

$$\mathbf{x} \,|\, \mathbf{y} \sim \mathrm{N}(\mathbf{a} + \mathbf{C}\,\mathbf{B}^{-1}\,(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\,\mathbf{B}^{-1}\mathbf{C}^T)$$

Hints:

- Denote inverse covariance as $\mathbf{D} = [\mathbf{D}_{11}\ \mathbf{D}_{12}; \mathbf{D}_{12}^T\ \mathbf{D}_{22}]$ and expand the quadratic form in the Gaussian exponent.
- Compute the derivative with respect to $\mathbf{x}$ and set it to zero. Conclude that due to symmetry the point where the derivative vanishes is the mean.
- Check from a linear algebra book that the inverse of $\mathbf{D}_{11}$ is given by the Schur complement:
$$\mathbf{D}_{11}^{-1} = \mathbf{A} - \mathbf{C}\,\mathbf{B}^{-1}\,\mathbf{C}^T$$
and that $\mathbf{D}_{12}$ can be then written as
$$\mathbf{D}_{12} = -\mathbf{D}_{11}\,\mathbf{C}\,\mathbf{B}^{-1}.$$
- Find the simplified expression for the mean by applying the identities above.
- Find the second derivative of the negative Gaussian exponent with respect to $\mathbf{x}$. Conclude that it must be the inverse conditional covariance of $\mathbf{x}$.
- Use the Schur complement expression above for computing the conditional covariance.