UC Berkeley
Department of Electrical Engineering and Computer Sciences

Electrical Engineering 126: Probability and Random Processes

Problem Set 10
Fall 2017

1. **Improving Estimation Error**

   Show that

   $$\mathbb{E}[(X - \mathbb{E}[X \mid Y, Z])^2] \leq \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2].$$

   In other words, more information yields better estimation error.

   Solution:

   Let $f(Y) = \mathbb{E}[X \mid Y]$ and $g(Y, Z) = \mathbb{E}[X \mid Y, Z]$. One has

   $$\mathbb{E}\big[(X - g(Y, Z))^2\big] \leq \mathbb{E}\big[(X - g(Y, Z))^2\big] + \mathbb{E}\big[(g(Y, Z) - f(Y))^2\big]$$
   $$= \mathbb{E}\big[(X - g(Y, Z) + g(Y, Z) - f(Y))^2\big]$$

   $(g(Y, Z) - f(Y)$ is a function of (Y, Z), so it is orthogonal to $X - g(Y, Z))$

   $$= \mathbb{E}\big[(X - f(Y))^2\big].$$

2. **Jointly Gaussian Decomposition**

   Let U and V be jointly Gaussian random variables with means $\mu_U = 1$, $\mu_V = 4$, respectively, with variances $\sigma_U^2 = 2.5$, $\sigma_V^2 = 2$, respectively, and with covariance $\rho = 1$. Can we write U as $U = aV + Z$, where a is a scalar and Z is independent of V? If you think we can, find the value of a and the distribution of Z; otherwise please explain the reason.

   Solution:

   The LLSE of U given V is

   $$L[U \mid V] = \mu_U + \frac{\rho}{\sigma_V^2}(V - \mu_V).$$

   For jointly Gaussian random variables, $Z' = U - L[U \mid V]$ is independent of V, so we have

   $$U = \mu_U + \frac{\rho}{\sigma_V^2}(V - \mu_V) + Z'.$$

Let

$$a = \frac{\rho}{\sigma_V^2},$$

$$Z = \mu_U - \frac{\rho}{\sigma_V^2}\mu_V + Z'.$$

Then $U = aV + Z$ and Z is independent of V. We can find that

$$\mathbb{E}[Z] = \mu_U - \frac{\rho}{\sigma_V^2}\mu_V,$$

and

$$\text{var } Z = \sigma_U^2 - \frac{\rho^2}{\sigma_V^2}.$$

Then we know that

$$Z \sim \mathcal{N}\left(\mu_U - \frac{\rho}{\sigma_V^2}\mu_V, \sigma_U^2 - \frac{\rho^2}{\sigma_V^2}\right).$$

Then we know that $a = 0.5$ and $Z \sim \mathcal{N}(-1, 2)$.

3. **Gaussian Estimation**

Let $Y = X + Z$ and $U = X - Z$, where X and Z are i.i.d. $\mathcal{N}(0, 1)$.

(a) Find the joint distribution of U and Y.

(b) Find the MMSE of X given the observation Y, call this $\hat{X}(Y)$.

(c) Let the estimation error $E = X - \hat{X}(Y)$. Find the conditional distribution of E given Y.

Solution:

(a) Note that U and Y are jointly Gaussian. Also, $\mathbb{E}[UY] = 0$, so they are uncorrelated. Since they are uncorrelated and jointly Gaussian, U and Y are independent. We can see that both U and Y are marginally Gaussian with the distribution $\mathcal{N}(0, 2)$, and thus, the joint distribution

$$f_{U,Y}(u, y) = \frac{1}{4\pi}e^{-(u^2+y^2)/4}.$$

Geometric Viewpoint

There is another, perhaps more intuitive way to see that U and Y are orthogonal. We see the following picture: One can see from Figure 1 that since $\|X\|_2 = \|Z\|_2$ and X and Z are orthogonal, the angle between Y and X is 45 degrees and similarly, the angle between X and $-Z$ is 45 degrees. Thus, we can see that Y and U are orthogonal.

(b) We are looking for $\hat{X}(Y) = \mathbb{E}[X \mid Y]$. Note that X and Y are jointly Gaussian, so

$$\mathbb{E}[X \mid Y] = L[X \mid Y] = \frac{1}{2}Y.$$

2
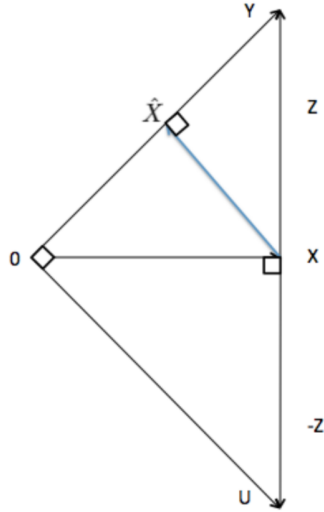
Figure 1: Geometry of U and Y.

Geometric Solution

We can also attack this part from a geometric point of view. Referring to Figure 1, we can see that $\triangle 0X\hat{X} \sim \triangle 0YX$. Thus, we have

$$\frac{\|\hat{X}\|_2}{\|X\|_2} = \frac{\|X\|_2}{\|Y\|_2}.$$

Thus, we can see that

$$\|\hat{X}\|_2 = \frac{1}{\|Y\|_2} = \frac{\|Y\|_2}{2}$$

Thus,

$$\hat{X} = \frac{Y}{2}.$$

(c) We see that:

$$E = X - \hat{X}(Y) = X - \frac{1}{2}Y = X - \frac{X+Z}{2} = \frac{U}{2}$$

Now, note that by Part (a), U and Y are independent, and thus E, the estimation error is independent of Y! Now, note that since E is a linear transformation of U, it has density

$$f_{E|Y}(\epsilon \mid y) = f_E(\epsilon) = \frac{1}{\sqrt{\pi}}e^{-\epsilon^2}.$$

4. **Successive Conditioning**

X, Y, and Z are jointly Gaussian, each having zero mean and strictly positive variance. Define $T = \mathbb{E}(Z \mid Y)$, $U = \mathbb{E}(T \mid X)$, and $V = \mathbb{E}(U \mid Y)$. Is $T = V$? Either argue or give a counterexample.

3

Solution:

The conclusion is incorrect.

Since X, Y, Z are jointly Gaussian, $T = \mathbb{E}(Z \mid Y) = aY$, where

$$a = \frac{\text{cov}(Y, Z)}{\text{var}\, Y}.$$

Similarly, $U = \mathbb{E}(T \mid X) = \mathbb{E}(aY \mid X) = abX$, where

$$b = \frac{\text{cov}(X, Y)}{\text{var}\, X}.$$

Similarly, $V = \mathbb{E}(U \mid Y) = \mathbb{E}(abX \mid Y) = abcY$, where

$$c = \frac{\text{cov}(X, Y)}{\text{var}\, Y}.$$

The condition $T = V$ implies $bc = 1$ (when $a \neq 0$), or

$$\mathbb{E}(XY)^2 = (\text{var}\, X)(\text{var}\, Y).$$

But this is not true in general. Consider the following example. One can have $\text{cov}(XY) = \mathbb{E}(XY) = 0$, with all other entries non-zero in the covariance matrix. Clearly $T \neq V$ in this case.

5. **Forrest Gump**

Forrest Gump is running across the United States, and we would like to track his progress. Assume that on day $n \in \mathbb{N}$ he runs $X(n)$ miles, and the amount he runs each day is determined by the amount he ran on the previous day with some random noise in the following manner: $X(n) = \alpha X(n-1) + V(n)$. Unfortunately, the measurements of the distance he traveled on each day are also subject to some noise. Assume that $Y(n)$ gives the measured number of miles Forrest Gump traveled on day n and that $Y(n) = \beta X(n) + W(n)$. For this problem, assume that $X(0) \sim \mathcal{N}(0, \sigma_X^2), W(n) \sim \mathcal{N}(0, \sigma_W^2), V(n) \sim \mathcal{N}(0, \sigma_V^2)$ are independent.

(a) Suppose that you observe $Y(0)$. Find the MMSE of $X(0)$ given this observation.

(b) Express both $\mathbb{E}[Y(n) \mid Y(0), \ldots, Y(n-1)]$ and $\mathbb{E}[X(n) \mid Y(0), \ldots, Y(n-1)]$ in terms of $\hat{X}(n-1)$, where $\hat{X}(n-1)$ is the MMSE of $X(n-1)$ given the observations $Y(0), Y(1), \ldots, Y(n-1)$.

(c) Show that:

$$\hat{X}(n) = \alpha \hat{X}(n-1) + k_n[Y(n) - \alpha\beta \hat{X}(n-1)]$$

where

$$k_n = \frac{\text{cov}(X(n), \tilde{Y}(n))}{\text{var}\, \tilde{Y}(n)}$$

and $\tilde{Y}(n) = Y(n) - L[Y(n) \mid Y(0), Y(1), \ldots, Y(n-1)]$.

Hint: Think geometrically.

4

Solution:

(a) We can see that $X(0), Y(0)$ are jointly Gaussian random variables, so:

$$\mathbb{E}[X(0) \mid Y(0)] = L[X(0) \mid Y(0)]$$
$$= \mathbb{E}[X(0)] + \frac{\mathrm{cov}(X(0), Y(0))}{\mathrm{var}\, Y(0)}\big(Y(0) - \mathbb{E}[Y(0)]\big)$$

We find:

$$\mathrm{cov}\big(X(0), Y(0)\big) = \mathbb{E}[X(0)Y(0)] = \mathbb{E}[\beta X(0)^2 + X(0)W(0)] = \beta\sigma_X^2$$

Additionally,

$$\mathrm{var}\, Y(0) = \mathbb{E}\big[\big(\beta X(0) + W(0)\big)^2\big] = \beta^2\sigma_X^2 + \sigma_W^2.$$

Thus, we have

$$\mathbb{E}[X(0) \mid Y(0)] = \frac{\beta\sigma_X^2}{\beta^2\sigma_X^2 + \sigma_W^2} Y(0).$$

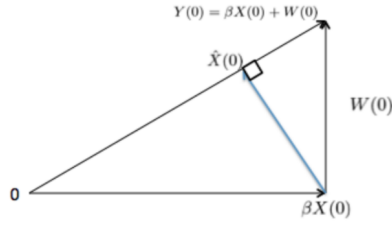**Geometric Solution**
Consider the following diagram:

Figure 2: Geometry of the 0th observation.

(b) We have the following:

$$\mathbb{E}[Y(n) \mid Y^{(n-1)}] = \mathbb{E}[\beta X(n) + W(n) \mid Y^{(n-1)}]$$
$$= \beta \, \mathbb{E}[X(n) \mid Y^{(n-1)}]$$
$$= \beta \, \mathbb{E}[\alpha X(n-1) + V(n) \mid Y^{(n-1)}]$$
$$= \alpha\beta\hat{X}(n-1)$$

Likewise, we see that:

$$\mathbb{E}[X(n) \mid Y^{(n-1)}] = \mathbb{E}[\alpha X(n-1) + V(n) \mid Y^{(n-1)}] = \alpha\hat{X}(n-1)$$

(c) We are interested in the quantity $\hat{X}(n) = \mathbb{E}[X(n) \mid Y^{(n)}]$. In other words, we would like the best estimator after n measurements have been taken. In an online setting, many of these measurements come in sequential fashion, so we would ideally like to have an estimate at time $n - 1$ and simply update the estimate when a new measurement comes at time n. Now, we can see that $\hat{X}(n) = \mathbb{E}[X(n) \mid Y^{(n)}] = L[X(n) \mid Y^{(n)}]$. How does this play into updating our observation? We can equivalently write this as $L[X(n) \mid Y(n), Y^{(n-1)}]$. We now state the following Theorem (8.1 in the book):

Theorem 1. *If X, Y, Z are 0 mean random variables such that Y and Z are orthogonal, $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$.*

The proof is given in the book and we will not repeat it here, but we will give the corresponding diagram which gives the intuition in Figure 3.
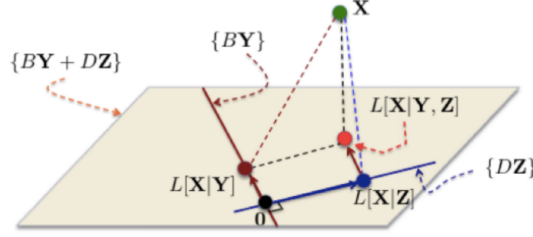

Figure 3: Updating the LLSE

Figure 3: Updating the LLSE.

We are not quite done, however, as $Y(n)$ and $Y^{(n-1)}$ are not necessarily orthogonal. To deal with this, we note that $Y(n) - L[Y(n) \mid Y(n-1)]$ and $Y^{(n-1)}$ are orthogonal. Additionally, we can see that any linear combination of $Y(n), Y^{(n-1)}$ is a linear combination of $(Y(n) - L[Y(n) \mid Y^{(n-1)}], Y^{(n-1)})$ (going forward, we let $\tilde{Y}(n) = Y(n) - L[Y(n) \mid Y^{(n-1)}]$). Thus, we have:

$$L[X(n) \mid Y(n), Y^{(n-1)}] = L[X(n) \mid Y^{(n-1)}, \tilde{Y}(n)]$$

Additionally, by the theorem above, we have:

$$L[X(n) \mid Y^{(n-1)}, \tilde{Y}(n)] = L[X(n) \mid Y^{(n-1)}] + L[X(n) \mid \tilde{Y}(n)]$$

We tackle these parts separately. Note that we determined $L[X(n) \mid Y^{(n-1)}]$ in Part (b). Now, we look at $L[X(n) \mid \tilde{Y}(n)]$. Recall that from part b, we have $L[Y(n) \mid Y^{(n-1)}] = \alpha\beta\hat{X}(n-1)$. Thus, we have the following:

$$L[X(n) \mid Y^{(n-1)}] + L[X(n) \mid \tilde{Y}(n)] \tag{1}$$

$$= \alpha\hat{X}(n-1) + k_n\big(Y(n) - L[Y(n) \mid Y^{(n-1)}]\big) \tag{2}$$

$$= \alpha\hat{X}(n-1) + k_n\big(Y(n) - \alpha\beta\hat{X}(n-1)\big) \tag{3}$$

where (2) follows from the known slope for the LLSE.

Congratulations! You have just derived the recursive structure of the scalar Kalman filter. We refer to k_n as the gain of the filter, and it turns out that this can be precomputed. Thus, as you have shown, the Kalman filter can be used to compute the MMSE at time n as a linear function of the MMSE at time $n-1$ and the newest observation.

6. **Random Walk with Unknown Drift**

Consider a random walk with unknown drift. The dynamics are given, for $n \in \mathbb{N}$, as

$$X_1(n+1) = X_1(n) + X_2(n) + V(n),$$
$$X_2(n+1) = X_2(n),$$
$$Y(n) = X_1(n) + W(n).$$

Here, X_1 represents the position of the particle and X_2 represents the velocity of the particle (which is unknown but constant throughout time). Y is the observation. V and W are independent Gaussian noise variables with mean zero and variance σ_V^2 and σ_W^2 respectively.

(a) Write down the dynamics of the system in matrix-vector form and write down the Kalman filter recursive equations for this system.

(b) Let k be a positive integer. Compute the prediction $\mathbb{E}(X(n+k) \mid Y^{(n)})$, where $Y^{(n)}$ is the history of the observations $Y_0, \ldots, Y_n$, in terms of the estimate $\hat{X}(n) := \mathbb{E}(X(n) \mid Y^{(n)})$.

(c) Now let $k = 1$ and compute the smoothing estimate $\mathbb{E}(X(n) \mid Y^{(n+1)})$ in terms of the quantities that appear in the Kalman filter equations.

Solution:

(a) In matrix form, the dynamics are

$$\begin{bmatrix} X_1(n+1) \\ X_2(n+1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1(n) \\ X_2(n) \end{bmatrix} + \underbrace{\begin{bmatrix} V(n) \\ 0 \end{bmatrix}}_{\tilde{V}(n)},$$

$$Y(n) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C} \begin{bmatrix} X_1(n) \\ X_2(n) \end{bmatrix} + W(n).$$

The Kalman filter equations are

$$\hat{X}(n) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \hat{X}(n-1) + K_n \big(Y(n) - \begin{bmatrix} 1 & 1 \end{bmatrix} \hat{X}(n-1)\big),$$

$$K_n = S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Big(\begin{bmatrix} 1 & 0 \end{bmatrix} S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sigma_W^2\Big)^{-1},$$

$$S_n = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Sigma_{n-1} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\Sigma_n = \Big(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - K_n \begin{bmatrix} 1 & 0 \end{bmatrix}\Big) S_n.$$

(b) First suppose that $k = 1$ and note that

$$\mathbb{E}\big(X(n+1) \mid Y^{(n)}\big) = \mathbb{E}\big(AX(n) + \tilde{V}(n) \mid Y^{(n)}\big)$$

and by independence of the noise and linearity of expectation,

$$\mathbb{E}\big(X(n+1) \mid Y^{(n)}\big) = A\,\mathbb{E}\big(X(n) \mid Y^{(n)}\big) = A\hat{X}(n).$$

The interpretation is quite simple: we take our estimate at time n, $\hat{X}(n)$, and then move it forwards one time step via the transition dynamics A. It is then easy to see that

$$\mathbb{E}\big(X(n+k) \mid Y^{(n)}\big) = A^k \hat{X}(n).$$

By computing

$$A^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

then one has

$$\mathbb{E}\big(X(n+k) \mid Y^{(n)}\big) = \begin{bmatrix} \hat{X}_1(n) + k\hat{X}_2(n) \\ \hat{X}_2(n) \end{bmatrix},$$

that is, your predicted velocity at time $n+k$ is still $\hat{X}_2(n)$ (makes sense; the velocity is not changing with time) and your predicted position at time $n+k$ is $\hat{X}_1(n)$, plus the velocity $\hat{X}_2(n)$ added k times.

(c) The first step is to recognize that

$$\mathbb{E}\big(X(n) \mid X(n+1), Y^{(n+1)}\big) = \mathbb{E}\big(X(n) \mid X(n+1), Y^{(n)}, Y(n+1)\big)$$
$$= \mathbb{E}\big(X(n) \mid X(n+1), Y^{(n)}\big)$$

since $Y(n+1) = CX(n+1) + W(n+1)$ and $W(n+1)$ is independent of everything else, so conditioned on $X(n+1)$, $Y(n+1)$ does not tell you anything new about $X(n)$. Now, observe that

$$\mathbb{E}\big(X(n) \mid X(n+1), Y^{(n)}\big) = L[X(n) \mid X(n+1), Y^{(n)}]$$
$$= L[X(n) \mid Y^{(n)}] + L[X(n) \mid \tilde{X}(n+1)]$$

8

where $\tilde{X}(n+1) := X(n+1) - L[X(n+1) \mid Y^{(n)}]$ is the **innovation**. By the previous part, $L[X(n+1) \mid Y^{(n)}] = A\hat{X}(n)$. So,

$$\tilde{X}(n+1) = X(n+1) - A\hat{X}(n).$$

Also,

$$\begin{aligned}
\operatorname{cov}\big(X(n), \tilde{X}(n+1)\big) &= \operatorname{cov}\big(X(n), A[X(n) - \hat{X}(n)] + \tilde{V}(n)\big) \\
&= \operatorname{cov}\big(X(n), X(n) - \hat{X}(n)\big)A^{\mathsf{T}} \\
&= \operatorname{cov}\big(X(n) - \hat{X}(n)\big)A^{\mathsf{T}}
\end{aligned}$$

since the error $X(n) - \hat{X}(n)$ is uncorrelated with the estimate $\hat{X}(n)$. We are in good shape since $\operatorname{cov}(X(n) - \hat{X}(n)) = \Sigma_n$ by definition. Also, $\operatorname{cov} \tilde{X}(n+1) = S_{n+1}$ by definition. Thus,

$$L[X(n) \mid \tilde{X}(n+1)] = \Sigma_n A^{\mathsf{T}} S_{n+1}^{-1}\big(X(n+1) - A\hat{X}(n)\big)$$

and

$$\begin{aligned}
\mathbb{E}\big(X(n) \mid Y^{(n+1)}\big) &= \mathbb{E}(\mathbb{E}\{X(n) \mid X(n+1), Y^{(n+1)}\} \mid Y^{(n+1)}) \\
&= \mathbb{E}\big(\hat{X}(n) + \Sigma_n A^{\mathsf{T}} S_{n+1}^{-1} \tilde{X}(n+1) \mid Y^{(n+1)}\big) \\
&= \hat{X}(n) + \Sigma_n A^{\mathsf{T}} S_{n+1}^{-1} \mathbb{E}\big(X(n+1) - A\hat{X}(n) \mid Y^{(n+1)}\big) \\
&= \hat{X}(n) + \Sigma_n A^{\mathsf{T}} S_{n+1}^{-1}\big(\hat{X}(n+1) - A\hat{X}(n)\big).
\end{aligned}$$