UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 13
Fall 2021

1. Orthogonal LLSE

(a) Consider zero-mean random variables X, Y, Z such that Y, Z are orthogonal. Show that $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$.

(b) Explain why for any zero-mean random variables X, Y, Z it holds that:

$$L[X \mid Y, Z] = L[X \mid Y] + L\big[X \bigm| Z - L[Z \mid Y]\big]$$

Solution:

(a) Let $U(Y) = L[X \mid Y]$, $V(Z) = L[X \mid Z]$. X, $U(Y)$, and $V(Z)$ are all zero-mean. Observe that $V(Z)$ and Y are orthogonal. To see this, observe that Y is orthogonal to 1 (this is the statement that Y is zero-mean) and to Z, and hence to any affine function of Z (in particular, Y is orthogonal to $V(Z)$). A similar argument establishes that $U(Y)$ and Z are orthogonal as well. Now,

$$\begin{aligned}
\mathbb{E}[X - U(Y) - V(Z)] &= 0, \\
\mathbb{E}\big[\big(X - U(Y) - V(Z)\big)Y\big] &= \mathbb{E}[V(Z)Y] = 0, \\
\mathbb{E}\big[\big(X - U(Y) - V(Z)\big)Z\big] &= \mathbb{E}[U(Y)Z] = 0,
\end{aligned}$$

since $X - U(Y)$ is orthogonal to Y and $X - V(Z)$ is orthogonal to Z. Therefore, $X - U(Y) - V(Z)$ is orthogonal to any linear function of 1, Y, and Z, and hence it is the LLSE of X given Y, Z.

(b) Let $W = Z - L[Z \mid Y]$, so W and Y are orthogonal. From the previous part we know $L[X \mid Y] + L[X \mid W] = L[X \mid W, Y]$, so it remains to argue that $L[X \mid W, Y] = L[X \mid Y, Z]$. This is intuitively clear since (W, Y) and (Y, Z) are linear functions of each other.

2. MMSE for Jointly Gaussian Random Variables

Provide justification for each of the following steps (1 - 5) to prove that the LLSE is equal to the MMSE estimator for jointly Gaussian random variables X and Y.
Let $g(X) = L[Y \mid X]$.

$$\begin{aligned}
& E[(Y - g(X))X] = 0 && (1) \\
\implies\ & \operatorname{cov}(Y - g(X), X) = 0 && (2) \\
\implies\ & Y - g(X) \text{ is independent of } X && (3) \\
\implies\ & E[(Y - g(X))f(X)] = 0\ \forall f(\cdot) && (4) \\
\implies\ & g(X) = E[Y \mid X] && (5)
\end{aligned}$$

Solution:

1. Since $g(X)$ is the LLSE, $Y - g(X)$ is orthogonal to all linear functions of X.
2. Since $Y - g(X)$ has 0 mean, $E[(Y - g(X))X] = E[(Y - g(X))X] - E[Y - g(X)]E[X] = \operatorname{cov}(Y - g(X), X)$.
3. Since X and Y are JG, so are all linear combinations of them, i.e. $Y - g(X)$ and X are JG. For JG random variables, uncorrelated implies independent.
4. Since $Y - g(X)$ and X are independent, so are any function of $Y - g(X)$ and any function of X. Therefore $E[(Y - g(X))f(X)] = E[Y - g(X)]E[f(X)] = 0$.
5. $E[X \mid Y]$ is the one and only function $g(X)$ that satisfies $E[(Y - g(X))f(X)] = 0$ for any function $f(X)$ of X. Since $g(X)$ satisfies this property, it must be $E[X \mid Y]$.

3. Stochastic Linear System MMSE

Let $(V_n,\ n \in \mathbb{N})$ be i.i.d. $\mathcal{N}(0, \sigma^2)$ and independent of $X_0 = \mathcal{N}(0, u^2)$. Let $|a| < 1$. Define

$$X_{n+1} = aX_n + V_n, \qquad n \in \mathbb{N}.$$

(a) What is the distribution of X_n, where n is a positive integer?

(b) Find $\mathbb{E}[X_{n+m} \mid X_n]$ for $m, n \in \mathbb{N}$, $m \geq 1$.

(c) Find u so that the distribution of X_n is the same for all $n \in \mathbb{N}$.

Solution:

(a) First, we find X_n as a function of X_0 and $(V_n)_{n \in \mathbb{N}}$.

$$\begin{aligned} X_1 &= aX_0 + V_0 \\ X_2 &= aX_1 + V_1 = a^2 X_0 + aV_0 + V_1 \\ X_3 &= aX_2 + V_2 = a^3 X_0 + a^2 V_0 + aV_1 + V_2. \end{aligned}$$

Thus, if we proceed doing this recursively, we find that

$$X_n = a^n X_0 + \sum_{i=0}^{n-1} a^i V_{n-1-i}.$$

Since X_0 and $(V_n)_{n \in \mathbb{N}}$ are independent Gaussian random variables, X_n is also Gaussian, so we need to find the mean and variance. X_0 and $(V_n)_{n \in \mathbb{N}}$ are zero-mean so

$$\mathbb{E}(X_n) = 0.$$

We know that

$$\sum_{i=0}^{n-1} a^i = \frac{1 - a^n}{1 - a}.$$

Thus,

$$\operatorname{var} X_n = a^{2n} \operatorname{var} X_0 + \sum_{i=0}^{n-1} a^{2i} \operatorname{var} V_{n-1-i} = a^{2n} u^2 + \frac{1 - a^{2n}}{1 - a^2} \sigma^2.$$

Hence,

$$X_n \sim \mathcal{N}\Big(0, a^{2n} u^2 + \frac{1 - a^{2n}}{1 - a^2} \sigma^2\Big).$$

(b) Similarly, by a shift of index

$$X_{n+m} = a^m X_n + \sum_{i=0}^{m-1} a^i V_{n+m-1-i}.$$

Now suppose that we have zero-mean random variables X, Y, and Z where $X = aY + Z$ and Y and Z are independent, then

$$\text{LLSE}[X \mid Y] = aY.$$

(Why?) Now since the random variables are jointly Gaussian, the MMSE is actually linear. Furthermore, X_n is independent of $\sum_{i=0}^{m-1} a^i V_{n+m-1-i}$. Thus,

$$\mathbb{E}(X_{n+m} \mid X_n) = a^m X_n.$$

(c) This is equivalent to X_1 having the same variance as X_0. Thus,

$$a^2 u^2 + \sigma^2 = u^2.$$

Thus,

$$u^2 = \frac{\sigma^2}{1 - a^2}.$$

4. **(Optional, included for practice) Random Walk with Unknown Drift**

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

$$\begin{aligned} X_1(n+1) &= X_1(n) + X_2(n) + V(n), \\ X_2(n+1) &= X_2(n), \\ Y(n) &= X_1(n) + W(n). \end{aligned}$$

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 $Y_0, \dots, Y_n$, 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 equation.

Hint: Use the law of total expectation

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

as well as the *innovation*

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

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

$$\begin{aligned}
\hat{X}(n) &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \hat{X}(n-1) + K_n\big(Y(n) - \begin{bmatrix} 1 & 1 \end{bmatrix} \hat{X}(n-1)\big), \\
K_n &= S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \end{bmatrix} S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sigma_W^2 \right)^{-1}, \\
S_n &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Sigma_{n-1} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & 0 \end{bmatrix}, \\
\Sigma_n &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - K_n \begin{bmatrix} 1 & 0 \end{bmatrix} \right) S_n.
\end{aligned}$$

(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

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

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

$$\begin{aligned}\mathbb{E}\big(X(n) \bigm| X(n+1), Y^{(n)}\big) &= L[X(n) \mid X(n+1), Y^{(n)}] \\ &= L[X(n) \mid Y^{(n)}] + L[X(n) \mid \tilde{X}(n+1)]\end{aligned}$$

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

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