1.1. **Linear Prediction.** Henceforth we will assume a *scalar valued* state-space model.

Definition. A scalar valued state-space model is

$$Y_n = g_n X_n + V_n, \tag{1.4}$$

$$X_{n+1} = f_n X_n + W_n, \tag{1.5}$$

for $n = 1, 2, \ldots$ where $\{V_n\}_n \sim \mathrm{WN}\left(0, \{r_n\}_n\right)$, $\{W_n\}_n \sim \mathrm{WN}\left(0, \{q_n\}_n\right)$ while f_n and g_n are constants. Furthermore, X_1, $\{V_n\}_n$ and $\{W_n\}_n$ are all mutually uncorrelated,

$$\mathrm{Cov}(X_1, W_n) = \mathrm{Cov}(X_1, V_n) = \mathrm{Cov}(W_m, V_n) = 0$$

for all $n \geq 1$ and $m \geq 1$.

This section reviews basic theory of best linear prediction which will then be used to construct the Kalman filter.

Definition. Given a sequence of random variables $\{Y_1, \dots, Y_n\}$ and a random variable X, a linear predictor of X using $\{Y_1, \dots, Y_n\}$ is

$$h_0 + h_1 Y_1 + \cdots + h_n Y_n$$

for some constants $(h_0, \dots, h_n)$.

How do we choose the values of the constants $h_0, \dots, h_n$?

Definition. The best linear predictor by X using $\{Y_1, \dots, Y_n\}$ is the predictor with coefficients $(h_0, \dots, h_n)$ that solves

$$\min_{(h_0,\dots,h_n)} \mathbb{E}\left\{(h_0 + h_1 Y_1 + \cdots + h_n Y_n - X)^2\right\}. \tag{1.6}$$

We denote this best linear predictor by $K[X \mid Y_{1:n}]$. (Memory aid: K for Kalman.)

Important points to note:

- We regard $K[\cdot \mid \cdot]$ as a function with two arguments, which is not specific to the collection X and $\{Y_1, \dots, Y_n\}$.
- So for some other collection of random variables, $K[U \mid W_{1:n}]$ is the best linear predictor of U using $\{W_1, \dots, W_n\}$.

The minimization problem (1.6) has a unique solution which satisfies the equations

$$\frac{\partial}{\partial h_i} \mathbb{E}\left\{(h_0 + h_1 Y_1 + \cdots + h_n Y_n - X)^2\right\} = 0 \tag{1.7}$$

for $i = 0, 1, \dots, n$. Interchanging the order of integration and differentiation

$$\frac{\partial}{\partial h_i} \mathbb{E}\left\{(\cdots)^2\right\} = \mathbb{E}\left\{\frac{\partial}{\partial h_i}(\cdots)^2\right\}$$

yields

$$\mathbb{E}\left\{(h_0 + h_1 Y_1 + \cdots + h_n Y_n - X)\right\} = 0, \tag{1.8}$$

$$\mathbb{E}\left\{(h_0 + h_1 Y_1 + \cdots + h_n Y_n - X)\, Y_i\right\} = 0, \quad i = 1, \ldots, n. \tag{1.9}$$

Solving for h_0 using (1.8) gives

$$h_0 = \mathbb{E}X - h_1 \mathbb{E}Y_1 - \cdots - h_n \mathbb{E}Y_n$$

which ensures unbiasedness. Substituting h_0 into (1.9) gives

(1.10)

$$\mathbb{E}\left\{[h_1(Y_1 - \mathbb{E}Y_1) + \cdots + h_n(Y_n - \mathbb{E}Y_n) - (X - \mathbb{E}X)]\,(Y_i - \mathbb{E}Y_i)\right\} = 0$$

for $i = 1, \ldots, n$.

Fact. *We can express the solution (1.10) in vector form using* $\mathbf{m} = (\mathbb{E}(Y_1), \ldots, \mathbb{E}(Y_n))^T$, $\mathbf{p} = (Cov(X, Y_1), \ldots, Cov(X, Y_n))^T$ *and*

$\Sigma = \mathbb{E}\left\{(Y_1, \ldots, Y_n)^T (Y_1, \ldots, Y_n)\right\} - \mathbf{m}\mathbf{m}^T$. *Line* i *of (1.10) is the* i*-th row of* $\Sigma\mathbf{h} = \mathbf{p}$ *which gives* $\mathbf{h} = \Sigma^{-1}\mathbf{p}$.

Remark. For random variables U and V, we say U is orthogonal to V if $\mathrm{Cov}(U, V) = 0$ or $\mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V)$. Note that if either U or V has zero mean, then orthogonality implies $\mathbb{E}(UV) = 0$.

We can deduce the following useful properties just from (1.8)-(1.9).

Fact 1.1. *Consider the collection of random variables* X*,* $(Y_1, \ldots, Y_n)$ *and* U*. Let* $\mathbf{p} = (Cov(X, Y_1), \ldots, Cov(X, Y_n))^T$*, let* Σ *be the square matrix with elements* $[\Sigma]_{i,j} = Cov(Y_i, Y_j)$ *and* $\mathbf{h} = (h_1, \ldots, h_n)^T$*.*

(1) Let $(h_1, \ldots, h_n)^{\mathrm{T}}$ satisfy $\Sigma\mathbf{h} = \mathbf{p}$ then

$$\hat{X} = K\left[X \mid Y_{1:n}\right] = \mathbb{E}(X) + h_1(Y_1 - \mathbb{E}Y_1) + \ldots + h_n(Y_n - \mathbb{E}Y_n).$$

(2) The error of the estimate $X-\hat{X}$ has zero mean and is orthogonal to all Y_i, that is

$$\mathbb{E}\left[\left(X - K\left[X \mid Y_{1:n}\right]\right) Y_i\right] = 0.$$

(3) If $\operatorname{Cov}(Y_i, Y_n) = 0$ for $i < n$ then

$$K\left[X \mid Y_{1:n}\right] = K\left[X \mid Y_{1:n-1}\right] + K\left[X \mid Y_n\right] - \mathbb{E}(X).$$

(4) For constants a, b, c,

$$K\left[aX + bU + c \mid Y_{1:n}\right] = aK\left[X \mid Y_{1:n}\right] + bK\left[U \mid Y_{1:n}\right] + c.$$

(5) Let C be an $n \times n$ invertible matrix and $\mathbf{b}$ a vector and let

$$\left(Y_1', \ldots, Y_n'\right)^{\mathrm{T}} = C\left(Y_1, \ldots, Y_n\right)^{\mathrm{T}} + \mathbf{b}.$$

Then $K[X \mid Y_{1:n}] = K[X \mid Y'_{1:n}]$, i.e. using $Y_{1:n}$ or the transformed data set $Y'_{1:n}$ to predict X gives the same result (as the matrix C is invertible.)

These properties will be proved in the examples paper. We show the final one here though, ignoring $\mathbf{b}$, but do comment on the impact $\mathbf{b}$ has in the proof.

The covariance matrix of the transformed data set $Y'_1, \ldots, Y'_n$ is

$$\mathbb{E}\left\{(Y'_1, \ldots, Y'_n)^{\mathrm{T}}(Y'_1, \ldots, Y'_n)\right\} - \mathbb{E}\left\{(Y'_1, \ldots, Y'_n)^{\mathrm{T}}\right\}\mathbb{E}\left\{(Y'_1, \ldots, Y'_n)\right\}$$

or, using $\mathbb{E}\left\{(Y'_1, \ldots, Y'_n)^{\mathrm{T}}\right\} = C\mathbb{E}\left\{(Y_1, \ldots, Y_n)^{\mathrm{T}}\right\} = C\mathbf{m}$, is

$$\mathbb{E}\left\{C(Y_1, \ldots, Y_n)^{\mathrm{T}}(Y_1, \ldots, Y_n)C^T\right\} - C\mathbf{m}\mathbf{m}^{\mathrm{T}}C^{\mathrm{T}} = C\Sigma C^T.$$

Similarly for the cross-covariance $(\mathrm{Cov}(X, Y_1'), \ldots, \mathrm{Cov}(X, Y_n'))^{\mathrm{T}}$,

$$\mathbb{E}\left\{X\left(Y_1', \ldots, Y_n'\right)^{\mathrm{T}}\right\} - \mathbb{E}\left\{X\right\}\mathbb{E}\left\{\left(Y_1', \ldots, Y_n'\right)^{\mathrm{T}}\right\}$$

$$= \mathbb{E}\left\{XC\left(Y_1, \ldots, Y_n\right)^{\mathrm{T}}\right\} - \mathbb{E}\left\{X\right\}\mathbb{E}\left\{C\left(Y_1, \ldots, Y_n\right)^{\mathrm{T}}\right\}$$

$$= C\mathbf{p}.$$

(The calculation of the covariance and cross-covariance is unchanged when $\mathbf{b}$ is included.) Now use the explicit solution to verify $K\left[X \mid Y_{1:n}\right] = K\left[X \mid Y_{1:n}'\right]$. Let

$$(h_1', \ldots, h_n')^{\mathrm{T}} = \left(C\Sigma C^{\mathrm{T}}\right)^{-1} C\mathbf{p}$$

then

$$
\begin{aligned}
K\left[X \mid Y'_{1:n}\right] &= \mathbb{E}(X) + h'_1(Y'_1 - \mathbb{E}Y'_1) + \ldots + h'_n(Y'_n - \mathbb{E}Y'_n) \\
&= \mathbb{E}(X) + (Y'_1 - \mathbb{E}Y'_1, \ldots, Y'_n - \mathbb{E}Y'_n)\left(C\Sigma C^{\mathrm{T}}\right)^{-1} C\mathbf{p} \\
&= \mathbb{E}(X) + (Y_1 - \mathbb{E}Y_1, \ldots, Y_n - \mathbb{E}Y_n)\, C^{\mathrm{T}}\left(C\Sigma C^{\mathrm{T}}\right)^{-1} C\mathbf{p} \\
&= \mathbb{E}(X) + (Y_1 - \mathbb{E}Y_1, \ldots, Y_n - \mathbb{E}Y_n)\, \Sigma^{-1}\mathbf{p} \\
&= K\left[X \mid Y_{1:n}\right].
\end{aligned}
$$

An important special cases states that if $\mathrm{Cov}(U, Y_i) = 0$ then Y_i is not useful for predicting U.

Exercise. Show $K\left[U \mid Y_{1:n}\right] = \mathbb{E}(U)$ if $\mathrm{Cov}(U, Y_i) = 0$ for all i.

The next special case states that using other random variables to predict one that is already observed is not useful.

Exercise. (Sanity check.) Show $K\left[Y_i \mid Y_{1:n}\right] = Y_i$.