What we already discussed . . .

**Theorem 5.1 (Linear estimator for linear models)** Let \( \{y, x, v\} \) be zero-mean random variables that are related via the linear model \( y = Hx + v \), for some data matrix \( H \) of compatible dimensions. Both \( x \) and \( v \) are assumed uncorrelated with invertible covariance matrices, \( R_v = E vv^* \) and \( R_x = E xx^* \). The linear least-mean-squares estimator of \( x \) given \( y \) can be evaluated by either expression:

\[
\hat{x} = R_x H^* \left[ R_v + H R_x H^* \right]^{-1} y = \left[ R_x^{-1} + H^* R_v^{-1} H \right]^{-1} H^* R_v^{-1} y
\]

and the resulting minimum mean-square error matrix is

\[
m.m.s.e. = \left[ R_x^{-1} + H^* R_v^{-1} H \right]^{-1}
\]
APPLICATION: MULTIPLE-ANTENNA RECEIVERS

\[ N \text{ noisy measurements,} \]
\[ y(i) = x + v(i), \quad i = 0, 1, \ldots, N - 1 \]

of some zero-mean random variable \( x \). Let \( y = \text{col}\{y(0), y(1), \ldots, y(N - 1)\} \) denote the observation vector. Here we evaluate \( \hat{x} \) by showing first that \( y \) and \( x \) are related through a linear model as in (5.1), and then using (5.5). For generality, we shall assume that the variances of \( x \) and \( v(i) \) are \( \sigma_x^2 \) and \( \sigma_v^2 \), respectively. In Ex. 4.1, we used \( \sigma_x^2 = \sigma_v^2 = 1 \).

Introduce the \( N \times 1 \) column vectors:

\[ v \triangleq \text{col}\{v(0), v(1), \ldots, v(N - 1)\}, \quad h \triangleq \text{col}\{1, 1, \ldots, 1\} \]

Then \( y = hx + v \) and the covariance matrix of \( v \) is \( \sigma_v^2 I \). We now obtain from (5.5) that

\[ \hat{x} = \left[ \frac{1}{\sigma_x^2} + h^* h / \sigma_v^2 \right]^{-1} h^* y / \sigma_v^2 = \frac{1}{N + \frac{1}{\text{SNR}}} \sum_{i=0}^{N-1} y(i) \]

where \( \text{SNR} = \sigma_x^2 / \sigma_v^2 \). Observe that we are not dividing by the number of observations (which is \( N \)) but by \( N + 1 / \text{SNR} \).
APPLICATION: MULTIPLE-ANTENNA RECEIVERS

FIGURE An optimal linear receiver for recovering a symbol $x$ transmitted over additive-noise channels from multiple-antenna measurements.
But what if we consider a different model for $\mathbf{x}$, whereby it is assumed to be a constant of unknown value, say, $x$, rather than a random quantity? How will the expression for $\hat{x}$ change? The purpose of this chapter is to study such estimators. Specifically, we shall now consider linear models of the form

$$y = Hx + v$$

(6.6)
Constrained Estimation

• *What kind of constraints?*

• *Difference with the previous case?*
Constrained Estimation

Problem Formulation

We are interested in determining a linear estimator for $x$ of the form $\hat{x} = Ky$, for some $n \times N$ matrix $K$. The choice of $K$ should satisfy two conditions:

1. **Unbiasedness.** First, the estimator $\hat{x}$ should be unbiased. That is, the choice of $K$ should guarantee $E\hat{x} = x$, which is the same as $KEy = x$. But from (6.7) we have $Ey = Hx$ so that $K$ should satisfy $KHx = x$, no matter what the value of $x$ is. This condition means that $K$ should satisfy

   $KH = I$ \hspace{5cm} (6.10)

   Note that $KH$ is $n \times n$ and is therefore a square matrix.
Constrained Estimation

Problem Formulation
We are interested in determining a linear estimator for $x$ of the form $\hat{x} = Ky$, for some $n \times N$ matrix $K$. The choice of $K$ should satisfy two conditions:

2. Optimality. Second, the choice of $K$ should minimize the covariance matrix of the estimation error, $\tilde{x} = x - \hat{x}$. Using the condition $KH = I$, we find that

$$\hat{x} = Ky = K(Hx + v) = KHx + Kv = x + Kv$$

so that $\tilde{x} = -Kv$. This means that the error covariance matrix, as a function of $K$, is given by

$$E\tilde{x}\tilde{x}^* = E(Kvv^*K^*) = KR_vK^*$$

(6.11)
Constrained Estimation

Combining (6.10) and (6.11), we conclude that the desired $K$ is found by solving the following constrained optimization problem:

$$
\min_K \quad KR_u K^* \quad \text{subject to} \quad KH = I
$$

(6.12)

The estimator $\hat{x} = K_0 y$ that results from the solution of (6.12) is known as the minimum-variance-unbiased estimator, or m.v.u.e. for short. It is also sometimes called the best linear unbiased estimator (BLUE).
Constrained Estimation

**Interpretation and Solution**

Let $\mathcal{J}(K)$ denote the cost function that appears in (6.12), i.e.,

$$\mathcal{J}(K) \triangleq KR_vK^*$$

Then problem (6.12) means the following. We seek a matrix $K_o$ satisfying $K_oH = I$ such that

$$\mathcal{J}(K) - \mathcal{J}(K_o) \geq 0 \quad \text{for all } K \text{ satisfying } KH = I$$

There are several ways of determining $K_o$. We choose to use the already known solution of the linear estimation problem (cf. Sec. 5.1) in order to guess what the solution $K_o$ for (6.12) should be. Once this is done, we shall then provide an independent verification of the result.
Constrained Estimation

Thus recall, as mentioned in the introduction of this chapter, that for two zero-mean random variables \( \{x, y\} \) that are related as in (6.1), the linear least-mean-squares estimator of \( x \) given \( y \) is (cf. the second expression in (6.3)):

\[
\hat{x} = (R_x^{-1} + H^* R_v^{-1} H)^{-1} H^* R_v^{-1} y
\]

Now assume that the covariance matrix of \( x \) has the particular form \( R_x = \alpha I \), with a sufficiently large positive scalar \( \alpha \) (i.e., \( \alpha \to \infty \)). That is, assume that the variance of each of the entries of \( x \) is infinitely large. In this way, \( x \) can be “interpreted” as playing the role of some unknown constant vector, \( x \). Then the above expression for \( \hat{x} \) reduces to

\[
\hat{x} = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1} y
\]

This conclusion suggests that the choice

\[
K_o = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1}
\]

the result is known as the Gauss-Markov theorem.
Constrained Estimation

**Theorem 6.1 (Gauss-Markov Theorem)** Consider the linear model \( y = Hx + \nu \), where \( \nu \) is a zero-mean random variable with positive-definite covariance matrix \( R_\nu \), and \( x \) is an unknown constant vector. Assume further that \( H \) is a full-rank \( N \times n \) matrix with \( N \geq n \). Then the minimum-variance-unbiased linear estimator of \( x \) given \( y \) is \( \hat{x} = K_o y \), where

\[
K_o = \left( H^* R_\nu^{-1} H \right)^{-1} H^* R_\nu^{-1}
\]

Moreover, the resulting cost is m.m.s.e. \( = (H^* R_\nu^{-1} H)^{-1} \).
Constrained Estimation

**Proof:** For any matrix $K$ that satisfies $KH = I$, it is easy to verify that

$$\mathcal{J}(K) = KR_vK^* = (K - K_0)R_v(K - K_0)^* + K_0R_vK_0^* \quad (6.13)$$

This is because

$$KR_vK_0^* = KR_v[R_v^{-1}H(H^*R_v^{-1}H)^{-1}] = KH(H^*R_v^{-1}H)^{-1} = (H^*R_v^{-1}H)^{-1}$$

Likewise, $K_0R_vK_0^* = (H^*R_v^{-1}H)^{-1}$. Relation (6.13) expresses the cost $\mathcal{J}(K)$ as the sum of two nonnegative-definite terms: one is independent of $K$ and is equal to $K_0R_vK_0^*$, while the other is dependent on $K$. It is then clear, since $R_v > 0$, that the cost is minimized by choosing $K = K_0$, and that the resulting minimum cost is

$$\mathcal{J}(K_0) = (H^*R_v^{-1}H)^{-1}$$

Note further that the matrix $K_0$ in the statement of the theorem satisfies the constraint $K_0H = I$. 


Going back to examples...
APPLICATION: MULTIPLE-ANTENNA RECEIVERS

\[ y(i) = x + v(i), \quad i = 0, 1, \ldots, N - 1 \]

of some zero-mean random variable \( x \). Let \( y = \text{col}\{y(0), y(1), \ldots, y(N - 1)\} \) denote the observation vector. Here we evaluate \( \hat{x} \) by showing first that \( y \) and \( x \) are related through a linear model as in (5.1), and then using (5.5). For generality, we shall assume that the variances of \( x \) and \( v(i) \) are \( \sigma_x^2 \) and \( \sigma_v^2 \), respectively. In Ex. 4.1, we used \( \sigma_x^2 = \sigma_v^2 = 1 \).

Introduce the \( N \times 1 \) column vectors:

\[ v \triangleq \text{col}\{v(0), v(1), \ldots, v(N - 1)\}, \quad h \triangleq \text{col}\{1, 1, \ldots, 1\} \]

Then \( y = hx + v \) and the covariance matrix of \( v \) is \( \sigma_v^2 I \). We now obtain from (5.5) that

\[ \hat{x} = [1/\sigma_x^2 + h^*h/\sigma_v^2]^{-1} h^*y/\sigma_v^2 = \frac{1}{N + \frac{1}{\text{SNR}}} \sum_{i=0}^{N-1} y(i) \]

where \( \text{SNR} = \sigma_x^2/\sigma_v^2 \). Observe that we are not dividing by the number of observations (which is \( N \)) but by \( N + 1/\text{SNR} \).
APPLICATION: MULTIPLE-ANTENNA RECEIVERS

Now assume instead that we model \( x \) as an unknown constant, rather than a random variable, say,

\[
y(i) = x + v(i), \quad i = 0, 1, \ldots, N - 1
\]

(6.16)

In this case, the value of \( x \) can be regarded as the mean value of each \( y(i) \). We collect the measurements and the noises into vector form,

\[
y \triangleq \operatorname{col}\{y(0), y(1), \ldots, y(N - 1)\}, \quad v \triangleq \operatorname{col}\{v(0), v(1), \ldots, v(N - 1)\}
\]

and define the data vector

\[
h = \operatorname{col}\{1, 1, \ldots, 1\}
\]

Then

\[
y = hx + v
\]

with \( R_v = \mathbb{E}vv^* = \sigma_v^2 I \). Invoking the result of Thm. 6.1 with \( H = h \), we conclude that the optimal linear estimator, or the m.v.u.e., of \( x \) is

\[
\hat{x}_{mvue} = \frac{1}{N} \sum_{i=0}^{N-1} y(i)
\]

(6.17)
APPLICATION: MULTIPLE-ANTENNA RECEIVERS

\[ \hat{x}_{\text{mvue}} = \frac{1}{N} \sum_{i=0}^{N-1} y(i) \]  

(6.17)

This result is simply the sample-mean estimator that the reader may be familiar with from an introductory course on statistics. Comparing the expressions for \( \hat{x}_{\text{llmse}} \) and \( \hat{x}_{\text{mvue}} \) we see that we are now dividing the sum of the observations \( \{y(i)\} \) by \( N \), and not by \( N + 1/\text{SNR} \). This modification guarantees that the estimator \( \hat{x}_{\text{mvue}} \) is truly unbiased.
APPLICATION: CHANNEL ESTIMATION

Then we can write in matrix form, say, for $M = 3$ and $N = 6$,

$$
\begin{bmatrix}
y(0) \\
y(1) \\
y(2) \\
y(3) \\
y(4) \\
y(5) \\
y(6)
\end{bmatrix}
\begin{bmatrix}
s(0) & s(0) \\
s(1) & s(0) \\
s(2) & s(1) & s(0) \\
s(3) & s(2) & s(1) \\
s(4) & s(3) & s(2) \\
s(5) & s(4) & s(3) \\
s(6) & s(5) & s(4)
\end{bmatrix}
\begin{bmatrix}
c \\
v(0) \\
v(1) \\
v(2) \\
v(3) \\
v(4) \\
v(5) \\
v(6)
\end{bmatrix}
= \begin{bmatrix}
y(0) \\
y(1) \\
y(2) \\
y(3) \\
y(4) \\
y(5) \\
y(6)
\end{bmatrix}
$$

(5.7)

**FIGURE** Channel estimation in the presence of additive noise.

\[
c = R_c H^* \left[ H R_c H^* + R_v \right]^{-1} y = \left[ R_c^{-1} + H^* R_v^{-1} H \right]^{-1} H^* R_v^{-1} y
\]
APPLICATION: CHANNEL ESTIMATION

\[
\hat{c}_{\text{limse}} = \left[R_c^{-1} + H^* R_v^{-1} H \right]^{-1} H^* R_v^{-1} y
\]

\[
\hat{c}_{\text{mvue}} = (H^* R_v^{-1} H)^{-1} H^* R_v^{-1} y
\]

requires knowledge of the covariances matrix \( R_c = \mathbb{E}cc^* \)

\[
\hat{c}_{\text{mvue}} = (H^* H)^{-1} H^* y
\]

(6.19)

If the noise sequence \( \{v(i)\} \) is modeled as white with variance \( \sigma_v^2 \), then \( R_v = \sigma_v^2 I \) and \( R_v \) would end up disappearing from the expression for \( \hat{c}_{\text{mvue}} \). Specifically, (6.18) would become

It is worth remarking that expression (6.19) has the form of a least-squares solution; which we shall study in great detail in Part VII (Least-Squares Methods).