Aalto University

ELEC-E8107 Stochastics models, estimation and control
Lecture 2b: Linear Estimation in Static Systems

- Minimum Mean Square Error (MMSE)
- MMSE estimation of Gaussian random vectors
- Linear MMSE estimator for arbitrarily distributed random vectors
- LS estimation of unknown constant vectors from linear observations, batch form, recursive form.
- Apply the LS technique

ESTIMATION OF GAUSSIAN RANDOM VECTORS

The Conditional Mean and Covariance for Gaussian Random Vectors

Two random vectors x and z that are **jointly normally (Gaussian)** distributed

The estimate of the random variable x in terms of

z according to the **minimum mean square error (MMSE) criterion** —

the **MMSE estimator** — is **the conditional mean of x given z**.

$$y \triangleq \begin{bmatrix} x \\ z \end{bmatrix}$$

$$\hat{x} \triangleq E[x|z] = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})$$

$$P_{xx|z} \triangleq E[(x - \hat{x})(x - \hat{x})'|z] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}$$

Joint and Conditional Gaussian random variables, Conditional pdf of x given z.

$$y = \begin{bmatrix} x \\ z \end{bmatrix} \qquad p(x,z) = p(y) = \mathcal{N}(y; \bar{y}, P_{yy}) \qquad \bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix}$$

$$P_{xx} = \text{cov}(x) = E[(x-\bar{x})(x-\bar{x})']$$
$$P_{zz} = \text{cov}(z) = E[(z-\bar{z})(z-\bar{z})']$$
$$P_{xz} = \text{cov}(x,z) = E[(x-\bar{x})(z-\bar{z})'] = P'_{zx}$$

$$P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}$$

$$p(x|z) = \frac{p(x,z)}{p(z)} = \frac{|2\pi P_{yy}|^{-1/2}\, e^{-\frac{1}{2}(y-\bar{y})' P_{yy}^{-1}(y-\bar{y})}}{|2\pi P_{zz}|^{-1/2}\, e^{-\frac{1}{2}(z-\bar{z})' P_{zz}^{-1}(z-\bar{z})}}$$

New with zero mean random variables, the exponent becomes

$$\xi \triangleq x - \bar{x}$$
$$\zeta \triangleq z - \bar{z}$$

$$\begin{aligned} q &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}^{-1} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \\ &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} T_{xx} & T_{xz} \\ T_{zx} & T_{zz} \end{bmatrix} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \end{aligned}$$

continues

$$T_{xx}^{-1} = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}$$

$$P_{zz}^{-1} = T_{zz} - T_{zx}T_{xx}^{-1}T_{xz}$$

$$T_{xx}^{-1}T_{xz} = -P_{xz}P_{zz}^{-1}$$

$$\begin{aligned} q &= \xi' T_{xx}\xi + \xi' T_{xz}\zeta + \zeta' T_{zx}\xi + \zeta' T_{zz}\zeta - \zeta' P_{zz}^{-1}\zeta \\ &= (\xi + T_{xx}^{-1}T_{xz}\zeta)' T_{xx}(\xi + T_{xx}^{-1}T_{xz}\zeta) + \zeta'(T_{zz} - T_{zx}T_{xx}^{-1}T_{xz})\zeta - \zeta' P_{zz}^{-1}\zeta \\ &= (\xi + T_{xx}^{-1}T_{xz}\zeta)' T_{xx}(\xi + T_{xx}^{-1}T_{xz}\zeta) \end{aligned} \quad (1.4.14\text{-}15)$$

completion of the squares $\xi + T_{xx}^{-1}T_{xz}\zeta = x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})$

$$\boxed{E(x|z) \triangleq \hat{x} = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})}$$

$$\boxed{\text{cov}(x|z) \triangleq P_{xx|z} = T_{xx}^{-1} = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}}$$

Fundamental equations of linear estimation

Estimation of Gaussian random vectors

x and z jointly Gaussian

z is the measurement

x random variable to be estimated

$$y \triangleq \begin{bmatrix} x \\ z \end{bmatrix} \qquad \bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} \qquad P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}$$

$$y \sim \mathcal{N}[\bar{y}, P_{yy}] \qquad P_{xx} = E[(x-\bar{x})(x-\bar{x})'] \qquad P_{xz} = E[(x-\bar{x})(z-\bar{z})'] = P'_{zx}$$

The **MMSE Minimum Mean Square Error** –estimator is the conditional mean of x given z, for linear Gaussian case,

also **Maximum a Posteriori MAP** -estimator

$$\boxed{\hat{x} \triangleq E[x|z] = \bar{x} + P_{xz} P_{zz}^{-1} (z - \bar{z})}$$

$$\boxed{P_{xx|z} \triangleq E[(x-\hat{x})(x-\hat{x})'|z] = P_{xx} - P_{xz} P_{zz}^{-1} P_{zx}}$$

- The optimal estimator (in the MMSE sense) of x in terms of z is a linear function of z. This is a consequence of the Gaussian assumption
- conditional covariance, measures the “quality” of the estimate, is independent of the observation z.

The MMSE estimate — the conditional mean — of a Gaussian random vector in terms of another Gaussian random vector (the measurement) is a linear combination of

- The prior (unconditional) mean of the variable to be estimated;
- The difference between the measurement and its prior mean.

The Principle of Orthogonality

(MMSE) estimate of a random variable *x* in terms of another random variable *z* is the conditional mean *E[x|z]*

- In many problems **the distributional information** needed for the evaluation of the conditional mean **is not available**.
- Furthermore, even if it were available, **the evaluation** of the **conditional** mean could be **prohibitively complicated**

A method that

(1) is simple — yields **the estimate as a linear function of the observation(s)** and

(2) requires little information —**only first and second moments**, is highly desirable.

Such a method, called **linear MMSE estimation**, relies on **the principle of orthogonality**

The best linear estimate (in the sense of MMSE) of a random variable in terms of another random variable — the observation(s) — is such that

1. The estimate is unbiased — the estimation error has mean zero, and

2. The estimation error is **uncorrelated** from the observation(s);

that is, they are **orthogonal**.

Linear MMSE Estimation for *Zero-Mean* Random Variables

in terms of a (normed linear) space of random variables

The set of real-valued scalar **zero-mean random variables** z_i, $i = 1, \ldots, n$, can be considered as **vectors in an abstract vector spac**e **or linear space**

A (complete) vector space, in which one defines **an inner product,** is a Hilbert space

$$\langle z_i, z_k \rangle = E[z_i z_k]$$ **(correlation !)**

Random variables under consideration are zero mean

$$\langle z_i, z_i \rangle = E[z_i^2] = \|z_i\|^2$$

satisfies the properties of **a norm** and can be taken as such.

With this definition of the norm, **linear dependence** is defined by stating that the norm of a linear combination of vectors is zero

$$E\left[\left(\sum_{i=1}^{m} \alpha_i z_i\right)^2\right] = 0$$

If, $\alpha_1 \neq 0$ then z_1 is a linear combination of $z_2, \ldots, z_m$

$$z_1 = -\frac{1}{\alpha_1} \sum_{i=2}^{m} \alpha_i z_i$$

that is, it is an element of the **subspace** spanned by $z_2, \ldots, z_m$

Two vectors are **orthogonal**, denoted as $z_i \perp z_k$, if and only if

$$\langle z_i, z_k \rangle = 0$$

which is equivalent to these zero-mean random variables being **uncorrelated**

The **linear MMSE estimator** of a **zero-mean random variable** x in terms of z_i, $i = 1, \ldots, n$, is given by

$$\hat{x} = \sum_{i=1}^{n} \beta_i z_i$$

and has to be such that the **norm of the estimation error** is minimum

$$\tilde{x} \triangleq x - \hat{x}$$

The linear MMSE estimate is denoted also by a circumflex (“hat”), even though it is **not** the conditional mean.

Thus the norm of the estimation error

$$\|\tilde{x}\|^2 = E[(x - \hat{x})^2] = E\Big[(x - \sum_{i=1}^{n} \beta_i z_i)^2\Big]$$

will have to be minimized with respect to β_i, i = 1, . . . , n.

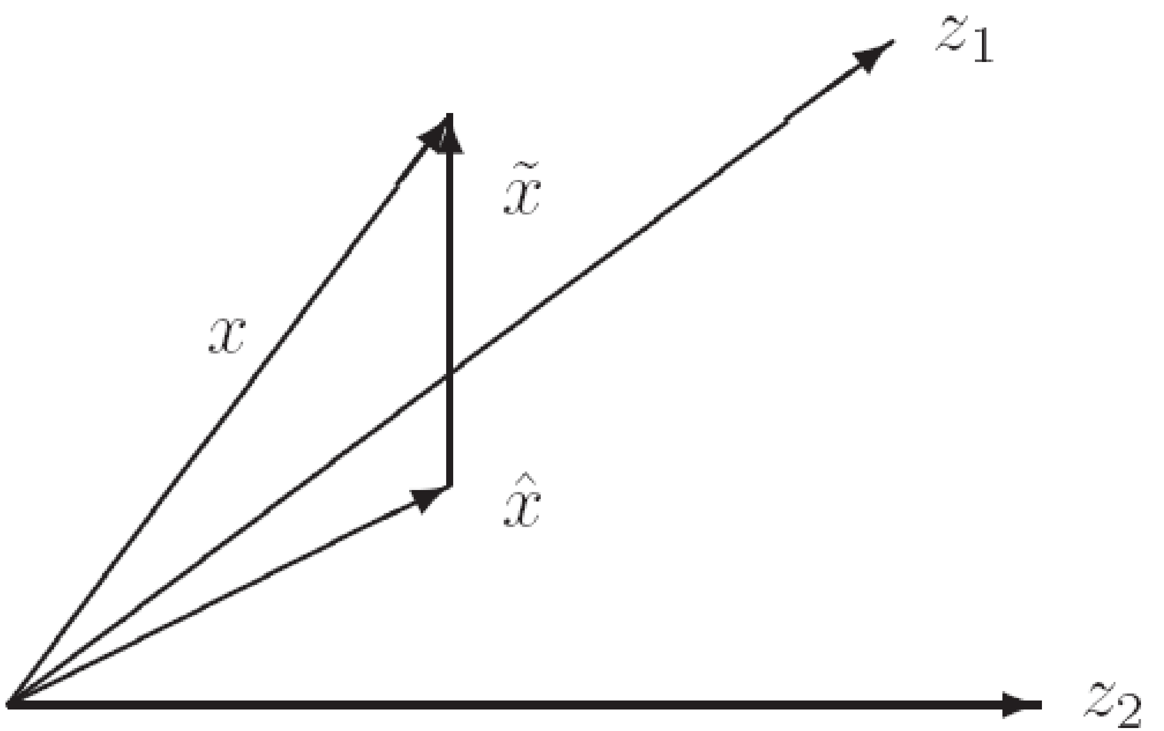
$$-\frac{1}{2}\frac{\partial}{\partial \beta_k}\|\tilde{x}\|^2 = E\Big[(x - \sum_{i=1}^{n} \beta_i z_i) z_k\Big] = E[\tilde{x} z_k] = \langle \tilde{x}, z_k \rangle = 0 \qquad k = 1, \ldots, n$$

is seen to be equivalent to requiring the following **orthogonality property**:

$$\tilde{x} \perp z_k \qquad \forall k$$

This is the **principle of orthogonality**:

In order for the error to have minimum norm, it has to be orthogonal to the observations. *This is equivalent to stating that the estimate of x has to be the orthogonal projection of x into the space spanned by the observations*



Orthogonal projection of random variable x into the subspace of $\{z_1, z_2\}$.

Linear MMSE Estimation for *Nonzero-Mean* Random Variables

For a random variable x with nonzero mean $\bar{x}$, the **best linear estimator** is of the form

$$\hat{x} = \beta_0 + \sum_{i=1}^{n} \beta_i z_i$$

Since the MSE is the sum of the square of the mean and the variance

$$E[\tilde{x}^2] = \left(E[\tilde{x}]\right)^2 + \mathrm{var}(\tilde{x})$$

in order to minimize it, the estimate should have the **unbiasedness** property

$$E[\tilde{x}] = 0 \qquad \beta_0 = \bar{x} - \sum_{i=1}^{n} \beta_i \bar{z}_i \qquad \bar{z}_i = E[z_i]$$

$$\hat{x} = \bar{x} + \sum_{i=1}^{n} \beta_i (z_i - \bar{z}_i)$$

The error corresponding to this estimate is

$$\begin{aligned} \tilde{x} &\triangleq x - \hat{x} \\ &= x - \bar{x} - \sum_{i=1}^{n} \beta_i (z_i - \bar{z}_i) \end{aligned}$$

This has transformed the nonzero-mean case into the zero-mean case.

The orthogonality principle then yields the coefficients β_i from

$$\langle \tilde{x}, z_k \rangle = E[\tilde{x} z_k] = E\Big[[x - \bar{x} - \sum_{i=1}^{n} \beta_i (z_i - \bar{z}_i)] z_k\Big] = 0 \qquad k = 1, \ldots, n$$

The estimator is also known as the **Best Linear Unbiased Estimator BLUE**

Linear MMSE Estimation for Vector Random Variables

Vector-valued random variables x and z, which are **not necessarily Gaussian or zero-mean**.

The "best linear" estimate of x in terms of z $\hat{x} = Az + b$

The criterion for "best" is the MMSE: find the estimator that minimizes the **scalar MSE criterion**, the expected value of the squared norm of the estimation error

$$J \triangleq E[(x-\hat{x})'(x-\hat{x})]$$

The linear MMSE estimator is such that the estimation error

$$\tilde{x} = x - \hat{x}$$

is zero-mean (the estimate is **unbiased**) and orthogonal to the observation z

$$E[\tilde{x}] = \bar{x} - (A\bar{z} + b) = 0 \qquad b = \bar{x} - A\bar{z}$$

The estimate is the orthogonal projection of the vector x into the space spanned by the observation vector z

$$\tilde{x} = x - \bar{x} - A(z - \bar{z})$$

The orthogonality requirement is, in the multidimensional case, that each component of $\tilde{x}$ be orthogonal to each component of z.

$$\begin{aligned} E[\tilde{x}z'] &= E\{[x - \bar{x} - A(z - \bar{z})]z'\} \\ &= E\{[x - \bar{x} - A(z - \bar{z})](z - \bar{z})'\} = P_{xz} - AP_{zz} = 0 \end{aligned}$$

the weighting matrix A

$$A = P_{xz}P_{zz}^{-1}$$

the linear MMSE estimator for the multidimensional case is identical to the conditional mean from the Gaussian case

$$\boxed{\hat{x} = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})}$$

The matrix MSE corresponding is given by

$$E[\tilde{x}\tilde{x}'] = E[[x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})][x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})]']$$

$$\boxed{E[\tilde{x}\tilde{x}'] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx} = P_{xx|z}}$$

And is identical expression to the conditional covariance in the Gaussian case

(strictly speaking, the matrix MSE is not a covariance matrix since the estimate is not the conditional mean)

Equations above are the fundamental equations of linear estimation

Remarks

Note the distinction between the scalar MSE criterion, an inner product, and the matrix MSE, an outer product. The matrix MSE is sometimes called, with abuse of language, a covariance matrix.

From the above derivations it follows that

• the best estimator (in the MMSE sense) for Gaussian random variables

is identical to

• the best **linear** estimator for arbitrarily distributed random variables with the same first- and second-order moments.

The linear estimator is the **overall best** if the random variables are

Gaussian; otherwise, it is only the **best within the class of linear estimators**

Linear MMSE Estimation — Summary

The linear MMSE estimator of one random vector x in terms of another random vector z is such that the **estimation error** is

1. Zero-mean (the estimate is unbiased)
2. Uncorrelated from the measurements

These two properties imply that the error is orthogonal to the measurements. **The principle of orthogonality.**

The expression of the linear MMSE estimator is identical to the expression of the conditional mean of Gaussian random vectors if they have the same first two moments.

Similarly, the matrix MSE associated with the LMMSE estimator has the same expression as the conditional covariance in the Gaussian case. The linear MMSE estimator is

1. The overall best if the random variables are Gaussian
2. The best within the class of linear estimators otherwise

The Batch LS Estimation

In the linear least squares (LS) problem it is desired to estimate the n_x vector x, modeled as an unknown constant, from the linear observations (n_z-vectors)

$$z(i) = H(i)x + w(i) \qquad i = 1, \ldots, k$$

to minimize the quadratic error

$$J(k) = \sum_{i=1}^{k} [z(i) - H(i)x]' R(i)^{-1} [z(i) - H(i)x]$$

$$z^k = \begin{bmatrix} z(1) \\ \vdots \\ z(k) \end{bmatrix}$$

$$J(k) = [z^k - H^k x]' (R^k)^{-1} [z^k - H^k x]$$

$$H^k = \begin{bmatrix} H(1) \\ \vdots \\ H(k) \end{bmatrix} \qquad w^k = \begin{bmatrix} w(1) \\ \vdots \\ w(k) \end{bmatrix} \qquad R^k = \begin{bmatrix} R(1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R(k) \end{bmatrix} = \mathrm{diag}[R(i)]$$

The L S estimator that minimizes J is obtained by setting its gradient with respect to x to zero.

$$\nabla_x J(k) = -2H^{k'}(R^k)^{-1}[z^k - H^k x] = 0$$

$$\boxed{\hat{x}(k) = [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}z^k}$$

assuming the required inverse exists

- It can be easily shown that since R^k is positive definite, the Hessian with respect to x is positive definite, and consequently the extremum point is a minimum.
- **A batch estimator** — the entire data have to be processed simultaneously for every k.

LS estimator is unbiased, because

$$E[\hat{x}(k)] = [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}E[H^k x + w^k] = x$$

The estimation error is

$$\tilde{x}(k) = x - \hat{x}(k) = -[H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}w^k$$

the covariance matrix of the LS estimator

$$\begin{aligned} P(k) &\triangleq E[\{\hat{x}(k) - E[\hat{x}(k)]\}\{\hat{x}(k) - E[\hat{x}(k)]\}'] \\ &= E[[\hat{x}(k) - x][\hat{x}(k) - x]'] \\ &= E[\tilde{x}(k)\tilde{x}(k)'] \\ &= [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}R^k(R^k)^{-1}H^k[H^{k'}(R^k)^{-1}H^k]^{-1} \end{aligned}$$

$$E[w^k w^{k'}] = R^k$$

$$\boxed{P(k) = [H^{k'}(R^k)^{-1}H^k]^{-1}}$$

The existence of the inverse of $H'R^{-1}H$ required is equivalent to having the covariance of the error finite. This amounts to requiring the parameter x to be observable

Relationship to the Maximum Likelihood (ML) Estimator

If the measurement errors w(i) are independent Gaussian random variables with mean zero and covariance R(i), then minimizing the LS criterion is equivalent to maximizing the likelihood function

$$\begin{aligned}\Lambda_k(x) &= p[z^k|x] = \prod_{i=1}^{k} p[z(i)|x] \\ &= c\, e^{-\frac{1}{2}\sum_{i=1}^{k}[z(i)-H(i)x]'R(i)^{-1}[z(i)-H(i)x]}\end{aligned}$$

the LS and ML estimators coincide, LS is clearly a “disguised” ML technique

The Recursive LS Estimator

In this case, k is interpreted as "discrete time."

$$z^{k+1} = \begin{bmatrix} z^k \\ z(k+1) \end{bmatrix} \qquad w^{k+1} = \begin{bmatrix} w^k \\ w(k+1) \end{bmatrix}$$

$$H^{k+1} = \begin{bmatrix} H^k \\ H(k+1) \end{bmatrix} \qquad R^{k+1} = \begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix}$$

$$\begin{aligned} P(k+1)^{-1} &= H^{k+1'}(R^{k+1})^{-1}H^{k+1} \\ &= \begin{bmatrix} H^{k'} & H(k+1)' \end{bmatrix} \begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix}^{-1} \begin{bmatrix} H^k \\ H(k+1) \end{bmatrix} \\ &= H^{k'}(R^k)^{-1}H^k + H(k+1)'R(k+1)^{-1}H(k+1) \end{aligned}$$

$$\boxed{P(k+1)^{-1} = P(k)^{-1} + H(k+1)'R(k+1)^{-1}H(k+1)}$$

The information is additive here because of the following:

1. The problem is static — the parameter is fixed.
2. The observations are modeled as independent.

$$
\begin{aligned}
P(k+1) &= [P(k)^{-1} + H(k+1)'R(k+1)^{-1}H(k+1)]^{-1} \\
&= P(k) - P(k)H(k+1)'[H(k+1)P(k)H(k+1)' \\
&\quad +R(k+1)]^{-1}H(k+1)P(k)
\end{aligned}
$$

$$\boxed{S(k+1) \triangleq H(k+1)P(k)H(k+1)' + R(k+1)}$$

$$\boxed{W(k+1) \triangleq P(k)H(k+1)'S(k+1)^{-1}}$$

$$\boxed{P(k+1) = [I - W(k+1)H(k+1)]P(k)}$$

$$\boxed{P(k+1) = P(k) - W(k+1)S(k+1)W(k+1)'}$$

Alternative Expression for the Gain

$$
\begin{aligned}
P(k+1)H(k+1)'R(k+1)^{-1} &= \{P(k)H(k+1)' - P(k)H(k+1)' \\
&\quad \cdot [H(k+1)P(k)H(k+1)' + R(k+1)]^{-1} \\
&\quad \cdot H(k+1)P(k)H(k+1)'\}R(k+1)^{-1} \\
&= P(k)H(k+1)' \\
&\quad \cdot [H(k+1)P(k)H(k+1)' + R(k+1)]^{-1} \\
&\quad \cdot \{H(k+1)P(k)H(k+1)' + R(k+1) \\
&\quad - H(k+1)P(k)H(k+1)'\}R(k+1)^{-1} \\
&= P(k)H(k+1)'S(k+1)^{-1} \\
&= W(k+1)
\end{aligned}
$$

$$
\boxed{W(k+1) = P(k+1)H(k+1)'R(k+1)^{-1}}
$$

The Recursion for the Estimate

$$
\begin{aligned}
\hat{x}(k+1) &= P(k+1)H^{k+1'}(R^{k+1})^{-1}z^{k+1} \\
&= P(k+1)\left[H^{k'} \;\; H(k+1)'\right]\begin{bmatrix} R^k & 0 \\ 0 & R(k+1) \end{bmatrix}^{-1}\begin{bmatrix} z^k \\ z(k+1) \end{bmatrix} \\
&= P(k+1)H^{k'}(R^k)^{-1}z^k + P(k+1)H(k+1)'R(k+1)^{-1}z(k+1) \\
&= [I - W(k+1)H(k+1)]P(k)H^{k'}(R^k)^{-1}z^k + W(k+1)z(k+1) \\
&= [I - W(k+1)H(k+1)]\hat{x}(k) + W(k+1)z(k+1)
\end{aligned}
$$

The above is the recursive parameter estimate updating equation — the recursive LS estimator, written as

$$
\boxed{\hat{x}(k+1) = \hat{x}(k) + W(k+1)[z(k+1) - H(k+1)\hat{x}(k)]}
$$

$$\hat{x}(k+1) = \hat{x}(k) + W(k+1)[z(k+1) - H(k+1)\hat{x}(k)]$$

The new (updated) estimate is therefore equal to the previous one plus a correction term. This correction term consists of the gain W(k+1) multiplying **the residual — the difference between the observation z(k+1) and the predicted value of this observation.**

Since this is a recursive scheme, initialization is required, for example by using a batch technique on a small number of initial measurements or by using an "a priori" initial estimate and an associated covariance.

The Residual Covariance, S

Covariance of the residual, the difference (zero mean) between the observation (noise R) and predicted observation based on estimate of x (covariance P), which are independent.

$$E[[z(k+1) - H(k+1)\hat{x}(k)][z(k+1) - H(k+1)\hat{x}(k)]'] = S(k+1)$$

$$\triangleq H(k+1)P(k)H(k+1)' + R(k+1)$$

Example of Prior Information, The Sample Mean

Noisy observations on a constant scalar x, w(i) independent and identically distributed random variables, zero mean, variance σ2

$$z(i) = x + w(i) \qquad i = 1, \ldots, k$$

$$R^k = I\sigma^2 \qquad H^k = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\begin{aligned} \hat{x}(k) &= [H^{k'}(R^k)^{-1}H^k]^{-1}H^{k'}(R^k)^{-1}z^k \\ &= \left\{ [1 \ \cdots \ 1]\,(I\sigma^2)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}^{-1} [1 \ \cdots \ 1]\,(I\sigma^2)^{-1} \begin{bmatrix} z(1) \\ \vdots \\ z(k) \end{bmatrix} \\ &= \frac{1}{k}\sum_{i=1}^{k} z(i) \end{aligned}$$

$$P(k) = [H^{k'}(R^k)^{-1}H^k]^{-1} = \frac{\sigma^2}{k}$$

The recursive form of the LS estimation

$$S(k+1) = H(k+1)P(k)H(k+1)' + R(k+1) = \frac{\sigma^2}{k} + \sigma^2 = \frac{k+1}{k}\sigma^2$$

$$W(k+1) = P(k)H(k+1)'S(k+1)^{-1} = \frac{\sigma^2}{k}\left(\frac{k+1}{k}\sigma^2\right)^{-1} = \frac{1}{k+1}$$

$$\boxed{\hat{x}(k+1) = \hat{x}(k) + \frac{1}{k+1}[z(k+1) - \hat{x}(k)]} \qquad \hat{x}(1) = z(1)$$

Or directly from the batch expression

$$\begin{aligned}
\hat{x}(k+1) &= \frac{1}{k+1}\sum_{i=1}^{k+1} z(i) \\
&= \frac{1}{k+1}\left[\sum_{i=1}^{k} z(i) + z(k+1)\right] \\
&= \frac{1}{k+1}[k\hat{x}(k) + z(k+1) - \hat{x}(k) + \hat{x}(k)] \\
&= \hat{x}(k) + \frac{1}{k+1}[z(k+1) - \hat{x}(k)] \qquad k = 1, \ldots
\end{aligned}$$