ELC 4351: Digital Signal Processing

Linear Optimal Filter

Prof. Liang Dong

Baylor University
SCHOOL OF ENGINEERING & COMPUTER SCIENCE
Department of Electrical & Computer Engineering

Linear Optimal Filter

- Linear Estimator
- Error Criterion
- Linear Minimum Mean Square Error (MMSE) Estimation
    1. Error Performance Surface
    2. Derivation of Linear MMSE Estimator
    3. Principle-Component Analysis
    4. Orthogonality Principle

Linear Estimator

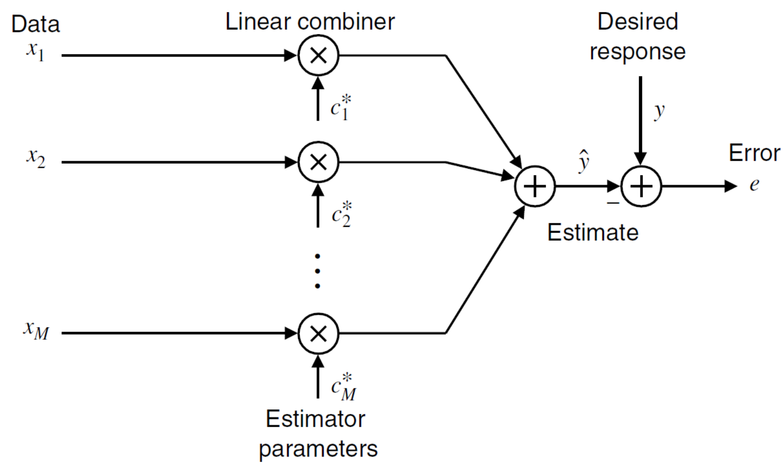


Figure: Block diagram of the linear estimator.

Linear Estimator: $\hat{y} = c_1^* x_1 + c_2^* x_2 + \cdots + c_M^* x_M = \sum_{k=1}^{M} c_k^* x_k$

Error Criterion

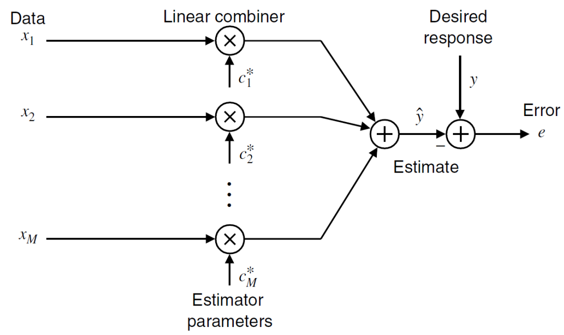
- Estimation Error: $e = \hat{y} - y$
- Error Criterion:

$$|e|, \quad \mathrm{E}[|e|] = \text{avg}[|e|]$$
$$|e|^2 = ee^*, \quad \mathrm{E}[|e|^2] = \text{avg}[|e|^2]$$

- Mean square error (MSE) Criterion:

$$P = \mathrm{E}[|e|^2]$$

Linear Mean Square Error Estimation

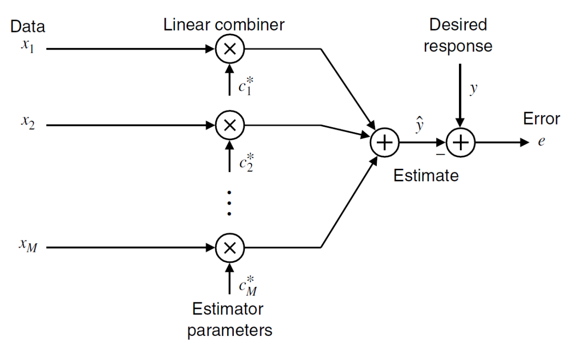


- Linear Estimator: $\hat{y} = \sum_{k=1}^{M} c_k^* x_k = \mathbf{c}^H \mathbf{x}$

  where, input data vector: $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$,
  and parameter/coefficient vector: $\mathbf{c} = [c_1, c_2, \ldots, c_M]^T$.

- Random variables are assumed to have zero-mean.

Linear Mean Square Error Estimation



- Linear Estimator: $\hat{y} = \sum_{k=1}^{M} c_k^* x_k = \mathbf{c}^H \mathbf{x}$

- Minimization of the MSE $P = \mathrm{E}[|\hat{y} - y|^2]$ with respect to parameters $\mathbf{c}$ leads to a linear estimator $\mathbf{c}_0$.

- The parameters $\mathbf{c}_0$ is the linear MMSE estimator and $\hat{y}_0$ the LMMSE estimate.

Error Performance Surface

- Express the MSE P as a function of the parameter vector $\mathbf{c}$.

$$\begin{aligned} P(\mathbf{c}) &= \mathrm{E}[|e|^2] \\ &= \mathrm{E}\left[(y - \mathbf{c}^H\mathbf{x})(y - \mathbf{c}^H\mathbf{x})^*\right] \\ &= \mathrm{E}\left[(y - \mathbf{c}^H\mathbf{x})(y^* - \mathbf{x}^H\mathbf{c})\right] \\ &= \mathrm{E}[yy^*] - \mathrm{E}[\mathbf{c}^H\mathbf{x}y^*] - \mathrm{E}[y\mathbf{x}^H\mathbf{c}] + \mathrm{E}[\mathbf{c}^H\mathbf{x}\mathbf{x}^H\mathbf{c}] \\ &= \mathrm{E}[|y|^2] - \mathbf{c}^H\mathrm{E}[\mathbf{x}y^*] - \mathrm{E}[y\mathbf{x}^H]\mathbf{c} + \mathbf{c}^H\mathrm{E}[\mathbf{x}\mathbf{x}^H]\mathbf{c} \end{aligned}$$

- Power of the desired output: $P_y = \mathrm{E}[|y|^2]$.
- Correlation matrix $\mathbf{R}$ of data vector $\mathbf{x}$ is

$$\mathbf{R} = \mathrm{E}[\mathbf{x}\mathbf{x}^H]$$

  $\mathbf{R}$ is Hermitian and nonnegtive definite. $\mathbf{R}^H = \mathbf{R}$.
- Cross-correlation vector between data vector $\mathbf{x}$ and the desired output y is

$$\mathbf{d} = \mathrm{E}[\mathbf{x}y^*]$$

Error Performance Surface

- Express the MSE P as a function of the parameter vector $\mathbf{c}$.

$$\begin{aligned} P(\mathbf{c}) &= \mathrm{E}[|y|^2] - \mathbf{c}^H\mathrm{E}[\mathbf{x}y^*] - \mathrm{E}[y\mathbf{x}^H]\mathbf{c} + \mathbf{c}^H\mathrm{E}[\mathbf{x}\mathbf{x}^H]\mathbf{c} \\ &= P_y - \mathbf{c}^H\mathbf{d} - \underbrace{\mathbf{d}^H\mathbf{c}}_{\text{linear function of } \mathbf{c}} + \underbrace{\mathbf{c}^H\mathbf{R}\mathbf{c}}_{\text{quadratic function of } \mathbf{c}} \end{aligned}$$

- If $\mathbf{R}$ is positive definite $(\mathbf{x}^H\mathbf{R}\mathbf{x} > 0, \forall \mathbf{x} \neq \mathbf{0})$, the quadratic function is bowl-shaped and has a unique minimum.

- The minimum of the error performance surface corresponds to the optimum parameters $\mathbf{c}_0$.

Error Performance Surface

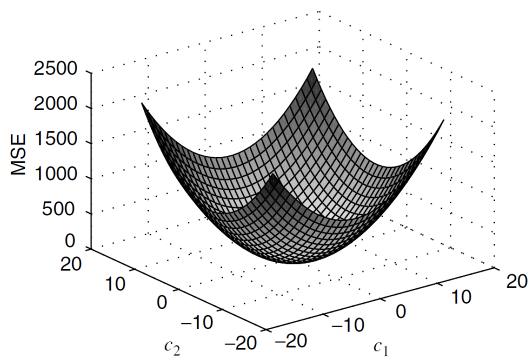


Figure: Error performance surface of quadratic function $\mathbf{c}^H\mathbf{R}\mathbf{c}$.

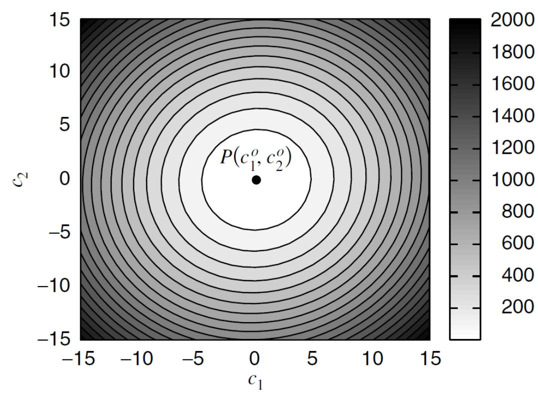


Figure: Error performance contour of quadratic function $\mathbf{c}^H\mathbf{R}\mathbf{c}$.

Derivation of the Linear MMSE Estimator

- Error performance surface (reconstruct)

$$\begin{aligned}
P(\mathbf{c}) &= P_y + (\mathbf{R}\mathbf{c}-\mathbf{d})^H\mathbf{R}^{-1}(\mathbf{R}\mathbf{c}-\mathbf{d}) - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d} \\
&= P_y + (\mathbf{c}^H\mathbf{R}^H - \mathbf{d}^H)\mathbf{R}^{-1}(\mathbf{R}\mathbf{c}-\mathbf{d}) - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d} \\
&= P_y + \mathbf{c}^H\mathbf{R}^H\mathbf{R}^{-1}\mathbf{R}\mathbf{c} - \mathbf{c}^H\mathbf{R}^H\mathbf{R}^{-1}\mathbf{d} - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{R}\mathbf{c} \\
&\quad + \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d} - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d} \\
&= P_y + \mathbf{c}^H\mathbf{R}\mathbf{c} - \mathbf{c}^H\mathbf{d} - \mathbf{d}^H\mathbf{c}
\end{aligned}$$

- Indeed, $P(\mathbf{c}) = \underbrace{P_y - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d}}_{\text{independent of } \mathbf{c}} + \underbrace{(\mathbf{R}\mathbf{c}-\mathbf{d})^H\mathbf{R}^{-1}(\mathbf{R}\mathbf{c}-\mathbf{d})}_{\text{quadratic function of } (\mathbf{R}\mathbf{c}-\mathbf{d})}$

Derivation of the Linear MMSE Estimator

- Error performance surface,

$$P(\mathbf{c}) = \underbrace{P_y - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d}}_{\text{independent of } \mathbf{c}} + \underbrace{(\mathbf{Rc} - \mathbf{d})^H\mathbf{R}^{-1}(\mathbf{Rc} - \mathbf{d})}_{\text{quadratic function of } (\mathbf{Rc}-\mathbf{d})}$$

- $\mathbf{R}^{-1}$ is also a positive definite matrix. That is,

$$\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}$$

  The minimum is achieved $\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x} = 0$ when $\mathbf{x} = \mathbf{0}$ (zero vector).

- Therefore, the minimum of the error performance surface is reached when $\mathbf{Rc} - \mathbf{d} = \mathbf{0}$.

$$\boxed{\mathbf{Rc}_0 = \mathbf{d}}$$

Derivation of the Linear MMSE Estimator

- Error performance surface,

$$P(\mathbf{c}) = \underbrace{P_y - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d}}_{\text{independent of } \mathbf{c}} + \underbrace{(\mathbf{Rc} - \mathbf{d})^H\mathbf{R}^{-1}(\mathbf{Rc} - \mathbf{d})}_{\text{quadratic function of } (\mathbf{Rc}-\mathbf{d})}$$

- The minimum of the error performance surface is reached when $\mathbf{Rc} - \mathbf{d} = \mathbf{0}$.

Normal Equation

$$\boxed{\mathbf{Rc}_0 = \mathbf{d}}$$

- The linear MMSE estimator $\mathbf{c}_0$ is

$$\mathbf{c}_0 = \mathbf{R}^{-1}\mathbf{d}$$

- The MMSE is $P(\mathbf{c}_0) = P_y - \mathbf{d}^H\mathbf{R}^{-1}\mathbf{d} = P_y - \mathbf{d}^H\mathbf{c}_0$

Excess MSE

- If $\tilde{\mathbf{c}}$ is a deviation from the optimum vector $\mathbf{c}_0$, i.e., $\mathbf{c} = \mathbf{c}_0 + \tilde{\mathbf{c}}$, we have

$$P(\mathbf{c}) = P(\mathbf{c}_0 + \tilde{\mathbf{c}}) = P(\mathbf{c}_0) + \underbrace{\tilde{\mathbf{c}}^H \mathbf{R} \tilde{\mathbf{c}}}_{\text{positive}}$$

- Excess MSE $= \tilde{\mathbf{c}}^H \mathbf{R} \tilde{\mathbf{c}}$

Principle-Component Analysis of Linear MMSE Estimator

- Eigen-decomposition of correlation matrix $\mathbf{R}$

$$\mathbf{R} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^H$$

where

$$\begin{aligned} \mathbf{Q} &= [\mathbf{q}_1 \, \mathbf{q}_2 \cdots \mathbf{q}_M] \\ \boldsymbol{\Lambda} &= \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\} \end{aligned}$$

- $\mathbf{q}_k$ and λ_k are the kth eigenvector and the corresponding eigenvalue of matrix $\mathbf{R}$.

- Decomposition:

$$\mathbf{R} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^H + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^H + \ldots + \lambda_M \mathbf{q}_M \mathbf{q}_M^H = \sum_{k=1}^{M} \lambda_k \mathbf{q}_k \mathbf{q}_k^H .$$

Principle-Component Analysis

- Each vector $\mathbf{q}_k$ has a length of one (normalized)

$$\|\mathbf{q}_k\|_2 = \sqrt{\mathbf{q}_k^H \mathbf{q}_k} = 1, \quad \forall k$$

$$\|\mathbf{q}_k\|_2^2 = \mathbf{q}_k^H \mathbf{q}_k = 1, \quad \forall k$$

- $\mathbf{q}_k$'s are orthogonal to each other

$$\mathbf{q}_k^H \mathbf{q}_l = 0, \quad k \neq l$$

- Therefore, $\mathbf{Q}$ is a unitary matrix.

$$\mathbf{Q}^H \mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}$$

$$\implies \mathbf{Q}^{-1} = \mathbf{Q}^H$$

- Correlation matrix $\mathbf{R}$ is positive definite and Hermitian $\mathbf{R}^H = \mathbf{R}$. The eigenvalues $\{\lambda_k\}_{k=1}^M$ are real and positive.

Principle-Component Analysis

- Rotation of a vector (coordinate transformation)

$$\mathbf{c}_0' = \mathbf{Q}^H \mathbf{c_0} \quad \text{or} \quad \mathbf{c}_0 = \mathbf{Q}\mathbf{c}_0'$$

- Let us check the (squared) length of the vector

$$\|c_0\|^2 = (\mathbf{Q}\mathbf{c}_0')^H \mathbf{Q}\mathbf{c}_0' = \mathbf{c}_0'^H \mathbf{Q}^H \mathbf{Q}\mathbf{c}_0' = \|\mathbf{c}_0'\|^2$$

  This means that he transformation only changes the direction of the vector but not its length.

- We can also rotate vector $\mathbf{d}$

$$\mathbf{d}' = \mathbf{Q}^H \mathbf{d} \quad \text{or} \quad \mathbf{d} = \mathbf{Q}\mathbf{d}'$$

Principle-Component Analysis

- The Normal Equation

$$\mathbf{R}\mathbf{c}_0 = \mathbf{d}$$

- Substituting $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$ in the normal equation, we have

$$\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H\mathbf{c}_0 = \mathbf{d}$$

- It follow that (left multiplying with $\mathbf{Q}^H$)

$$\mathbf{\Lambda}\mathbf{Q}^H\mathbf{c}_0 = \mathbf{Q}^H\mathbf{d}$$

$$\mathbf{\Lambda}\mathbf{c}_0' = \mathbf{d}'$$

where $\mathbf{d}' = \mathbf{Q}^H\mathbf{d}$.

Principle-Component Analysis

- This is a "decoupled" Normal Equation

$$\boxed{\mathbf{\Lambda}\mathbf{c}_0' = \mathbf{d}'}$$

- Because Λ is diagonal, it can be written into a set of M equations

$$\lambda_i c_{0,i}' = d_i', \quad 1 \leq i \leq M$$

- A set of M first-order equations. If $\lambda_i \neq 0$, we have

$$c_{0,i}' = \frac{d_i'}{\lambda_i}, \quad 1 \leq i \leq M$$

Principle-Component Analysis

- The minimum mean square error (MMSE) becomes

$$\begin{aligned} P_0 &= P_y - \mathbf{d}^H \mathbf{c}_0 \\ &= P_y - (\mathbf{Q}\mathbf{d}')^H \mathbf{Q}\mathbf{c}_0' \\ &= P_y - \mathbf{d}'^H \mathbf{c}_0' \\ &= P_y - \sum_{i=1}^{M} d_i'^* c_{0,i}' \\ &= P_y - \sum_{i=1}^{M} \frac{|d_i'|^2}{\lambda_i} \end{aligned}$$

Principle-Component Analysis

- The excess MSE becomes

$$\begin{aligned} \Delta P &= \tilde{\mathbf{c}}^H \mathbf{R} \tilde{\mathbf{c}} \\ &= \tilde{\mathbf{c}}^H \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \tilde{\mathbf{c}} \\ &= \tilde{\mathbf{v}}^H \mathbf{\Lambda} \tilde{\mathbf{v}} \\ &= \sum_{i=1}^{M} \lambda_i |\tilde{v}_i|^2 \end{aligned}$$

where $\tilde{\mathbf{v}} = \mathbf{Q}^H \tilde{\mathbf{c}}$.

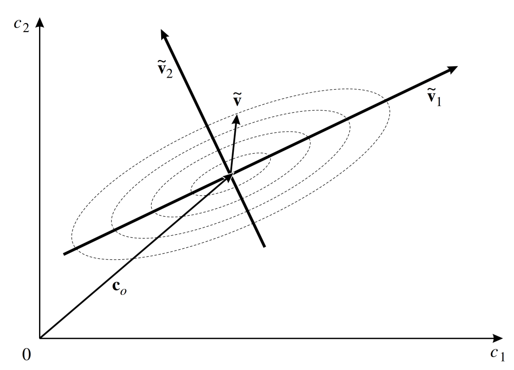


Figure: Contours of principle-component axes for excess MSE.

Principle-Component Analysis

- The MMSE estimator is

$$\begin{aligned}\mathbf{c}_0 &= \mathbf{R}^{-1}\mathbf{d} \\ &= \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^H\mathbf{d} \\ &= \sum_{i=1}^{M}\frac{\mathbf{q}_i^H\mathbf{d}}{\lambda_i}\mathbf{q}_i \\ &= \sum_{i=1}^{M}\frac{d_i'}{\lambda_i}\mathbf{q}_i\end{aligned}$$

- The MMSE estimate is

$$\begin{aligned}\hat{y}_0 &= \mathbf{c}_0^H\mathbf{x} \\ &= \sum_{i=1}^{M}\frac{d_i'}{\lambda_i}(\mathbf{q}_i^H\mathbf{x})\end{aligned}$$

Principle-Component Analysis

- The MMSE estimate is

$$\hat{y}_0 = \sum_{i=1}^{M}\frac{d_i'}{\lambda_i}(\mathbf{q}_i^H\mathbf{x})$$

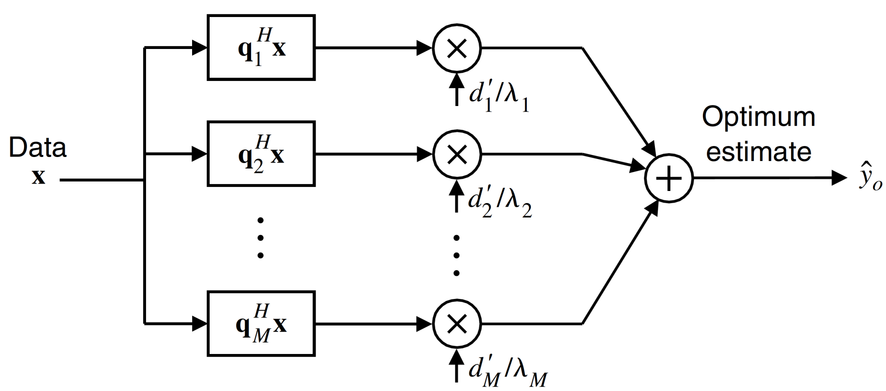


Figure: Principle-components representation of the Optimal linear estimator.

Principle of Orthogonality

- The correlation of two (zero-mean) random variables is equivalent to the inner product of two vectors in the vector space (Hilbert space).

$$\langle x, y \rangle = \mathrm{E}[xy^*]$$

- The squared length of a vector is

$$\|x\|^2 = \langle x, x \rangle = \mathrm{E}[|x|^2]$$

- Therefore, by the Cauchy-Schwartz inequality, we have

$$|\langle x, y \rangle|^2 \leq \|x\| \|y\|$$

- The two random variables are orthogonal $x \perp y$, if

$$\langle x, y \rangle = \mathrm{E}[xy^*] = 0 \Longrightarrow \text{uncorrelated}$$

Principle of Orthogonality

- Intuitive interpretation for MMSE

$$\begin{aligned}
\mathrm{E}[\mathbf{x}e_0^*] &= \mathrm{E}[\mathbf{x}(y^* - \mathbf{x}^H\mathbf{c}_0)] \\
&= \mathrm{E}[\mathbf{x}y^*] - \mathrm{E}[\mathbf{x}\mathbf{x}^H]\mathbf{c}_0 \\
&= \mathbf{d} - \mathbf{R}\mathbf{c}_0 \\
&= \mathbf{0}
\end{aligned}$$

Orthogonality Principle of MMSE Estimation

$$\boxed{\mathrm{E}[x_m e_0^*] = 0, \quad \text{for } 1 \leq m \leq M}$$

The estimation error is orthogonal to the data used for the estimation.

Principle of Orthogonality

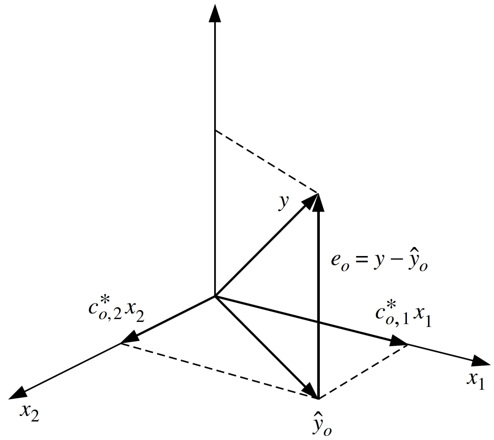


Figure: Illustration of the orthogonality principle. $x_m \perp e_0, m = 1, 2.$

- Applying the Pythagorean theorem, we have

$$\|y\|^2 = \|\hat{y}_0\|^2 + \|e_0\|^2 \quad \text{or} \quad \mathrm{E}[|y|^2] = \mathrm{E}[|\hat{y}_0|^2] + \mathrm{E}[|e_0|^2]$$