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:

\[ \hat{y} = c_1^* x_1 + c_2^* x_2 + \cdots + c_M^* x_M = \sum_{k=1}^{M} c_k^* x_k \]

Error Criterion

- Estimation Error: \( e = \hat{y} - y \)
- Error Criterion:
  \[ |e|, \quad E[|e|] = \text{avg}[|e|] \]
  \[ |e|^2 = ee^*, \quad E[|e|^2] = \text{avg}[|e|^2] \]
- Mean square error (MSE) Criterion:
  \[ P = E[|e|^2] \]
Linear Mean Square Error Estimation

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

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

Random variables are assumed to have zero-mean.

Minimization of the MSE $P = E[|\hat{y} - y|^2]$ with respect to parameters $c$ leads to a linear estimator $c_0$.

The parameters $c_0$ is the linear MMSE estimator and $\hat{y}_0$ the LMMSE estimate.
Express the MSE $P$ as a function of the parameter vector $c$.

\[
P(c) = E[|e|^2]
= E[(y - c^H x)(y - c^H x)^*]
= E[(y - c^H x)(y^* - x^H c)]
= E[yy^*] - E[c^H xy^*] - E[yx^H c] + E[c^H xx^H c]
= E[|y|^2] - c^H E[xy^*] - E[yx^H]c + c^H E[xx^H]c
\]

Power of the desired output: $P_y = E[|y|^2]$.

Correlation matrix $R$ of data vector $x$ is

\[
R = E[xx^H]
\]

$R$ is Hermitian and nonnegative definite. $R^H = R$.

Cross-correlation vector between data vector $x$ and the desired output $y$ is

\[
d = E[xy^*]
\]

Express the MSE $P$ as a function of the parameter vector $c$.

\[
P(c) = E[|y|^2] - c^H E[xy^*] - E[yx^H]c + c^H E[xx^H]c
= P_y - c^H d - d^H c + c^H Rc
\]

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

The minimum of the error performance surface corresponds to the optimum parameters $c_0$. 

Error Performance Surface
Error Performance Surface

Figure: Error performance surface of quadratic function $c^H R c$.

Figure: Error performance contour of quadratic function $c^H R c$.

Derivation of the Linear MMSE Estimator

- Error performance surface (reconstruct)

$$P(c) = P_y + (Rc - d)^H R^{-1} (Rc - d) - d^H R^{-1} d$$
$$= P_y + (c^H R^H - d^H) R^{-1} (Rc - d) - d^H R^{-1} d$$
$$= P_y + c^H R^H R^{-1} Rc - c^H R^H R^{-1} d - d^H R^{-1} Rc$$
$$+ d^H R^{-1} d - d^H R^{-1} d$$
$$= P_y + c^H Rc - c^H d - d^H c$$

- Indeed, $P(c) = P_y - d^H R^{-1} d + (Rc - d)^H R^{-1} (Rc - d)$

  independent of $c$  quadratic function of $(Rc - d)$
Derivation of the Linear MMSE Estimator

- Error performance surface,

\[ P(c) = P_y - d^H R^{-1} d + (Rc - d)^H R^{-1} (Rc - d) \]

\begin{align*}
\text{independent of } c & \quad \text{quadratic function of } (Rc - d)
\end{align*}

\[ R^{-1} \text{ is also a positive definite matrix. That is,} \]

\[ x^H R^{-1} x > 0, \quad \forall x \neq 0 \]

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

- Therefore, the minimum of the error performance surface is reached when \( Rc - d = 0 \).

\[ Rc_0 = d \]

Normal Equation

\[ Rc_0 = d \]

- The linear MMSE estimator \( c_0 \) is

\[ c_0 = R^{-1} d \]

- The MMSE is \( P(c_0) = P_y - d^H R^{-1} d = P_y - d^H c_0 \)
If \( \tilde{c} \) is a deviation from the optimum vector \( c_0 \), i.e., \( c = c_0 + \tilde{c} \), we have

\[
P(c) = P(c_0 + \tilde{c}) = P(c_0) + \tilde{c}^H R \tilde{c}
\]

where \( R = Q \Lambda Q^H \)

with

\[
Q = [q_1 \quad q_2 \cdots \quad q_M]
\]

\[
\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\}
\]

\( q_k \) and \( \lambda_k \) are the \( k \)th eigenvector and the corresponding eigenvalue of matrix \( R \).

Decomposition:

\[
R = \lambda_1 q_1 q_1^H + \lambda_2 q_2 q_2^H + \ldots + \lambda_M q_M q_M^H = \sum_{k=1}^{M} \lambda_k q_k q_k^H.
\]
Principle-Component Analysis

- Each vector $q_k$ has a length of one (normalized)
  \[
  \|q_k\|_2 = \sqrt{q_k^H q_k} = 1, \quad \forall k
  \]
  \[
  \|q_k\|_2 = q_k^H q_k = 1, \quad \forall k
  \]
- $q_k$'s are orthogonal to each other
  \[
  q_k^H q_l = 0, \quad k \neq l
  \]
- Therefore, $Q$ is a unitary matrix.
  \[
  Q^H Q = QQ^H = I
  \]
  \[
  \implies Q^{-1} = Q^H
  \]
- Correlation matrix $R$ is positive definite and Hermitian
  $R^H = R$. The eigenvalues $\{\lambda_k\}_{k=1}^M$ are real and positive.

Principle-Component Analysis

- Rotation of a vector (coordinate transformation)
  \[
  c_0' = Q^H c_0 \quad \text{or} \quad c_0 = Qc_0'
  \]
- Let us check the (squared) length of the vector
  \[
  \|c_0\|^2 = (Qc_0')^H Qc_0 = c_0'^H Q^H Qc_0 = \|c_0'\|^2
  \]
  This means that the transformation only changes the direction of the vector but not its length.
- We can also rotate vector $d$
  \[
  d' = Q^H d \quad \text{or} \quad d = Qd'
  \]
Principle-Component Analysis

- The Normal Equation
  \[ \mathbf{Rc}_0 = \mathbf{d} \]

- Substituting \( \mathbf{R} = \mathbf{Q}\Lambda\mathbf{Q}^H \) in the normal equation, we have
  \[ \mathbf{Q}\Lambda\mathbf{Q}^H \mathbf{c}_0 = \mathbf{d} \]

- It follows that (left multiplying with \( \mathbf{Q}^H \))
  \[ \Lambda\mathbf{Q}^H \mathbf{c}_0 = \mathbf{Q}^H \mathbf{d} \]
  \[ \Lambda\mathbf{c}'_0 = \mathbf{d}' \]
  where \( \mathbf{d}' = \mathbf{Q}^H \mathbf{d} \).

Principle-Component Analysis

- This is a “decoupled” Normal Equation
  \[ \boxed{\Lambda\mathbf{c}'_0 = \mathbf{d}'} \]

- Because \( \Lambda \) 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

\[
P_0 = P_y - d^H c_0
= P_y - (Qd')^H Qc'_0
= P_y - d'^H c'_0
= P_y - \sum_{i=1}^{M} d'^*_i c'_{0,i}
= P_y - \sum_{i=1}^{M} |d'_i|^2 \frac{\lambda_i}{\lambda_i}
\]

Principle-Component Analysis

- The excess MSE becomes

\[
\Delta P = \bar{c}^H R \bar{c}
= \bar{c}^H Q \Lambda Q^H \bar{c}
= \tilde{\nu}^H \Lambda \tilde{\nu}
= \sum_{i=1}^{M} \lambda_i |\tilde{\nu}_i|^2
\]

where \( \tilde{\nu} = Q^H \bar{c} \).

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

- The MMSE estimator is

\[ c_0 = R^{-1}d \]
\[ = QA^{-1}Q^H d \]
\[ = \sum_{i=1}^{M} \frac{q_i^H d}{\lambda_i} q_i \]
\[ = \sum_{i=1}^{M} \frac{d'_i}{\lambda_i} q_i \]

- The MMSE estimate is

\[ \hat{y}_0 = c_0^H x \]
\[ = \sum_{i=1}^{M} \frac{d'_i}{\lambda_i} (q_i^H x) \]

Figure: Principle-components representation of the Optimal linear estimator.
The correlation of two (zero-mean) random variables is equivalent to the inner product of two vectors in the vector space (Hilbert space).

\[ \langle x, y \rangle = E[xy^*] \]

The squared length of a vector is

\[ \|x\|^2 = \langle x, x \rangle = E[|x|^2] \]

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

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

Intuitive interpretation for MMSE

\[
E[xe_0^*] = E[x(y^* - x^Hc_0)] \\
= E[xy^*] - E[xx^H]c_0 \\
= d - Rc_0 \\
= 0
\]

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 

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