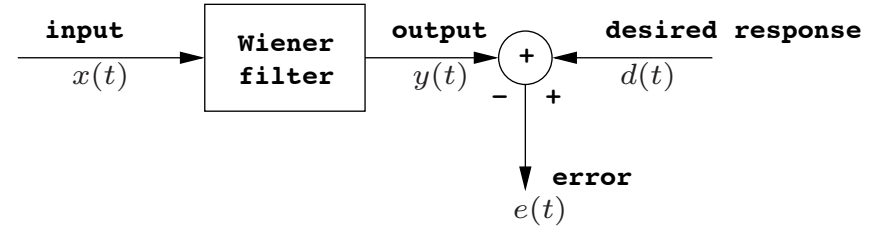


Wiener Filters and Least Mean Squares (LMS)

Seungjin Choi

Department of Computer Science
 Pohang University of Science and Technology, Korea
 seungjin@postech.ac.kr

Wiener Filter



Input-output relation: $y(t) = \sum_{\tau=0}^{L-1} w_{\tau} x(t - \tau) = \mathbf{w}^T \mathbf{x}_t$.

Error: $e(t) = d(t) - y(t)$.

Objective function: $\mathcal{J} = \langle e^2(t) \rangle$.

Wiener filter involves finding minimum mean squared error (MMSE) solution.

- ▶ Wiener filter is a linear optimal filter.
- ▶ Wiener filter is optimal when the assumption holds that the joint probability density function of the observed signal and the underlying signal is Gaussian.
- ▶ Applications of Wiener filters: denosing, echo cancelation, channel equalization, speech enhancement, image restoration, and traffic flow forecasting.

Wiener-Hopf Equation

- ▶ Assume that $x(t)$ and $y(t)$ are jointly stationary and $x(t)$ and $y(t)$ have zero means.

$$\begin{aligned} \mathcal{J} &= \langle e^2(t) \rangle \\ &= \langle (d(t) - \mathbf{w}^T \mathbf{x}_t)^2 \rangle \\ &= \sigma_d^2 - 2\mathbf{w}^T \mathbf{r}_{xd} + \mathbf{w}^T \mathbf{R}_{xx} \mathbf{w}, \end{aligned}$$

where $\sigma_d^2 = \langle d^2(t) \rangle$, $\mathbf{r}_{xd} = \langle \mathbf{x}(t)d(t) \rangle$, and $\mathbf{R}_{xx} = \langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle$.

- ▶ $\frac{\partial \mathcal{J}}{\partial \mathbf{w}} = 0$ leads to Wiener-Hopf equation: $\mathbf{R}_{xx} \mathbf{w} = \mathbf{r}_{xd}$.
- ▶ The optimal weight is given by $\mathbf{w}_* = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xd}$.
 Thus, we have $y_*(t) = (\mathbf{R}_{xx}^{-1} \mathbf{r}_{xd})^T \mathbf{x}_t$.
- ▶ $\mathcal{J}_{min} = \sigma_d^2 - \mathbf{r}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{r}_{xd}$.

1 / 12

2 / 12

Gaussian Identity

Lemma

Let $\mathcal{G}_x(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a Gaussian density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Then,

$$p(\mathbf{x}) = \mathcal{G}_{x_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) \mathcal{G}_{x_1}(\boldsymbol{\mu}_{x_1|x_2}, \boldsymbol{\Sigma}_{x_1|x_2}),$$

where

$$\begin{aligned} \boldsymbol{\mu}_{x_1|x_2} &= \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{\mu}_2 - \mathbf{x}_2), \\ \boldsymbol{\Sigma}_{x_1|x_2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}. \end{aligned}$$

3 / 12

4 / 12

Minimum Mean Squared Error

Theorem

Let the estimate of y , given \mathbf{x} , be $f(\mathbf{x})$. Then, $\langle \|y - f(\mathbf{x})\|^2 \rangle$ is minimized over all functions $f(\mathbf{x})$ by $f(\mathbf{x}) = \langle y|\mathbf{x} \rangle$.

Proof. The mean squared error function is given by

$$\begin{aligned} \mathcal{E}(f) &= \int \int \|y - f(\mathbf{x})\|^2 p(\mathbf{x}, y) d\mathbf{x} dy \\ &= \int \int \|y - f(\mathbf{x})\|^2 p(\mathbf{x}) p(y|\mathbf{x}) d\mathbf{x} dy \\ &= \int p(\mathbf{x}) \left[\int \|y - f(\mathbf{x})\|^2 p(y|\mathbf{x}) dy \right] d\mathbf{x}. \end{aligned}$$

This should be minimized for every \mathbf{x} , so we have $\frac{\partial}{\partial f(\mathbf{x})} [\int \|y - f(\mathbf{x})\|^2 p(y|\mathbf{x}) dy] = 0$. Then, $\int 2(f(\mathbf{x}) - y) p(y|\mathbf{x}) dy = 0$. Therefore,

$$f(\mathbf{x}) = \int y p(y|\mathbf{x}) dy = \langle y|\mathbf{x} \rangle.$$

Least Mean Square (LMS) Algorithm

- ▶ Proposed by B. Widrow.
- ▶ Consider the instantaneous estimate of the mean-squared error, i.e., $\mathcal{J}(t) = \frac{1}{2} e^2(t)$.
- ▶ Steepest descent algorithm,

$$\begin{aligned} \mathbf{w}_{t+1} &= \mathbf{w}_t - \eta \nabla_{\mathbf{w}} \mathcal{J}(t) \\ &= \mathbf{w}_t + \eta e(t) \mathbf{x}_t. \end{aligned}$$

Alternative View of Wiener Filter

Assume: (1) $\{x(t)\}$ and $\{d(t)\}$ are stationary; (2) $\{x(t)\}$ and $\{d(t)\}$ are jointly Gaussian. Under these assumptions, we show that the general optimal filter is Wiener filter.

Introduce an augmented random vector $\mathbf{z} = [d, \mathbf{x}^\top]^\top$. Then we have $p(d, \mathbf{x}) = \mathcal{G}_z(\mathbf{u}, \mathbf{\Sigma})$ where

$$\mathbf{u} = [\mu_d, \boldsymbol{\mu}_x^\top]^\top, \quad \mathbf{\Sigma} = \begin{bmatrix} c_{dd} & \mathbf{c}_{dx}^\top \\ \mathbf{c}_{dx} & \mathbf{C}_{xx} \end{bmatrix}.$$

The optimal filter is given by

$$\begin{aligned} y(t) &= \langle d(t) | \mathbf{x}_t \rangle \\ &= \mu_d + \mathbf{c}_{dx}^\top \mathbf{C}_{xx}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_x). \end{aligned}$$

If $\boldsymbol{\mu}_x = 0$, $\mu_d = 0$, then $y(t) = \mathbf{c}_{dx}^\top \mathbf{C}_{xx}^{-1} \mathbf{x}_t$.

5 / 12

6 / 12

Convergence in the Mean

We show that $\langle \mathbf{w}_t \rangle \rightarrow \mathbf{w}_*$ as $t \rightarrow \infty$.

Consider the LMS algorithm

$$\begin{aligned} \mathbf{w}_{t+1} &= \mathbf{w}_t + \eta (d(t) - \mathbf{w}_t^\top \mathbf{x}_t) \mathbf{x}_t \\ &= (\mathbf{I} - \eta \mathbf{x}_t \mathbf{x}_t^\top) \mathbf{w}_t + \eta \mathbf{x}_t d(t). \end{aligned} \quad (1)$$

Assume \mathbf{w} and \mathbf{x} are uncorrelated. Take the expectation in both sides of (1),

$$\langle \mathbf{w}_{t+1} \rangle = (\mathbf{I} - \eta \mathbf{R}_{xx}) \langle \mathbf{w}_t \rangle + \eta \mathbf{r}_{xd}. \quad (2)$$

The covariance matrix \mathbf{R}_{xx} is a positive definite and symmetric matrix

$$\mathbf{R}_{xx} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top. \quad (3)$$

Using the result of Wiener-Hopf equation, $\mathbf{R}_{xx} \mathbf{w}_* = \mathbf{r}_{xd}$, we rewrite (2) as

$$\langle \mathbf{w}_{t+1} \rangle = (\mathbf{I} - \eta \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top) \langle \mathbf{w}_t \rangle + \eta \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^\top \mathbf{w}_*. \quad (4)$$

7 / 12

8 / 12

Convergence in the Mean (Cont'd)

Pre-multiplying both sides of (4) by \mathbf{U}^\top , leads to

$$\mathbf{U}^\top \langle \mathbf{w}_{t+1} \rangle = (\mathbf{I} - \eta \mathbf{\Lambda}) \mathbf{U}^\top \langle \mathbf{w}_t \rangle + \eta \mathbf{\Lambda} \mathbf{U}^\top \mathbf{w}_* \quad (5)$$

Define a new vector \mathbf{v}_t by $\mathbf{v}_t = \mathbf{U}^\top (\langle \mathbf{w}_t \rangle - \mathbf{w}_*)$. Then we have

$$\langle \mathbf{w}_t \rangle = \mathbf{U} \mathbf{v}_t + \mathbf{w}_*. \quad (\text{affine transform}) \quad (6)$$

Then we rewrite (5) as $\mathbf{v}_{t+1} = (\mathbf{I} - \eta \mathbf{\Lambda}) \mathbf{v}_t$. From this relation, we have

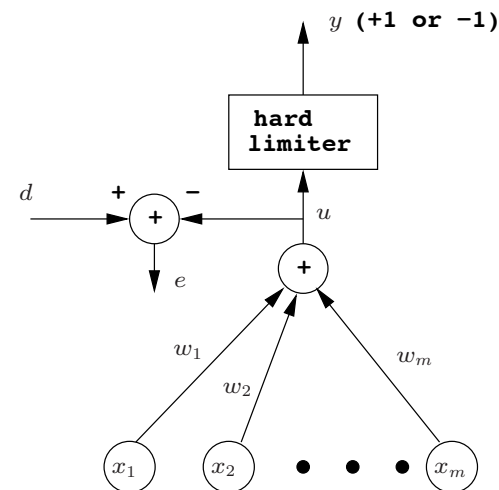
$$v_i(t+1) = (1 - \eta \lambda_i)^t v_i(0). \quad (7)$$

For the LMS algorithm to be convergent in the mean, the following condition should be satisfied:

$$|1 - \eta \lambda_i| < 1, \quad \text{for } i = 1, \dots, L. \quad (8)$$

Thus, $0 < \eta < \frac{2}{\lambda_{\max}}$.

Adaptive Linear Element (ADALINE)



9 / 12

10 / 12

Normalized LMS (NLMS)

- ▶ The NLMS algorithm is given by

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \frac{e(t)}{\|\mathbf{x}_t\|^2} \mathbf{x}_t.$$

- ▶ The NLMS algorithm is less sensitive to the variation of dynamic range of input data \mathbf{x}_t , compared to the LMS.
- ▶ The NLMS algorithm is derived in the framework of the constrained optimization where given \mathbf{x}_t and $d(t)$, \mathbf{w}_{t+1} is determined so as to minimize $\|\delta \mathbf{w}_{t+1}\|^2 = \|\mathbf{w}_{t+1} - \mathbf{w}_t\|^2$ subject to a constraint $\mathbf{w}_{t+1}^\top \mathbf{x}_t = d(t)$.

NLMS: Algorithm Derivation

The objective function is given by

$$\mathcal{J} = \|\delta \mathbf{w}_{t+1}\|^2 + \lambda (d(t) - \mathbf{w}_{t+1}^\top \mathbf{x}_t),$$

where λ is an **Lagrangian multiplier**.

Solving $\frac{\partial \mathcal{J}}{\partial \mathbf{w}_{t+1}} = 0$, leads to $\mathbf{w}_{t+1} = \mathbf{w}_t + \frac{\lambda}{2} \mathbf{x}_t$.

The optimal value of λ is found by replacing \mathbf{w}_{t+1} with $\mathbf{w}_t + \frac{\lambda}{2} \mathbf{x}_t$ in the constraint, i.e.,

$$\left(\mathbf{w}_t + \frac{\lambda}{2} \mathbf{x}_t \right)^\top \mathbf{x}_t = d(t).$$

Thus, $\lambda = \frac{2e(t)}{\|\mathbf{x}_t\|^2}$. Using this optimal value of λ , the NLMS algorithm is given by

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \frac{e(t)}{\|\mathbf{x}_t\|^2} \mathbf{x}_t.$$

11 / 12

12 / 12