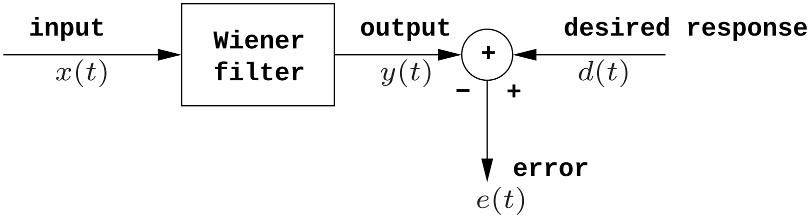
Wiener Filters and Least Mean Squares (LMS)

Seungjin Choi

Department of Computer Science and Engineering
Pohang University of Science and Technology
77 Cheongam-ro, Nam-gu, Pohang 37673, Korea
seungjin@postech.ac.kr

Wiener Filter



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

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

  where $\sigma_d^2 = \left\langle d^2(t) \right\rangle$, $\boldsymbol{r}_{xd} = \langle \boldsymbol{x}(t) d(t) \rangle$, and $\boldsymbol{R}_{xx} = \left\langle \boldsymbol{x}(t) \boldsymbol{x}^\top(t) \right\rangle$.
- $\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}} = 0$ leads to Wiener-Hopf equation: $\boxed{\boldsymbol{R}_{xx} \boldsymbol{w} = \boldsymbol{r}_{xd}}$.
- The optimal weight is given by $\boldsymbol{w}_* = \boldsymbol{R}_{xx}^{-1} \boldsymbol{r}_{xd}$.
  Thus, we have $y_*(t) = \left(\boldsymbol{R}_{xx}^{-1} \boldsymbol{r}_{xd} \right)^\top \boldsymbol{x}_t$.
- $\mathcal{J}_{min} = \sigma_d^2 - \boldsymbol{r}_{xd}^\top \boldsymbol{R}_{xx}^{-1} \boldsymbol{r}_{xd}$.

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

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Then,

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

where

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

Minimum Mean Squared Error

Theorem

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

Proof. The mean squared error function is given by

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

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

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

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 $\boldsymbol{z} = \left[d, \boldsymbol{x}^\top\right]^\top$. Then we have $p(d, \boldsymbol{x}) = \mathcal{G}_z(\boldsymbol{u}, \boldsymbol{\Sigma})$ where

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

The optimal filter is given by

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

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

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,

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

Convergence in the Mean

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

Consider the LMS algorithm

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

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

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

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

$$\boldsymbol{R}_{xx} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top. \tag{3}$$

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

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

Convergence in the Mean (Cont'd)

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

$$\boldsymbol{U}^\top \langle \boldsymbol{w}_{t+1} \rangle = (\boldsymbol{I} - \eta\boldsymbol{\Lambda})\, \boldsymbol{U}^T \langle \boldsymbol{w}_t \rangle + \eta\boldsymbol{\Lambda}\boldsymbol{U}^\top \boldsymbol{w}_*. \tag{5}$$

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

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

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

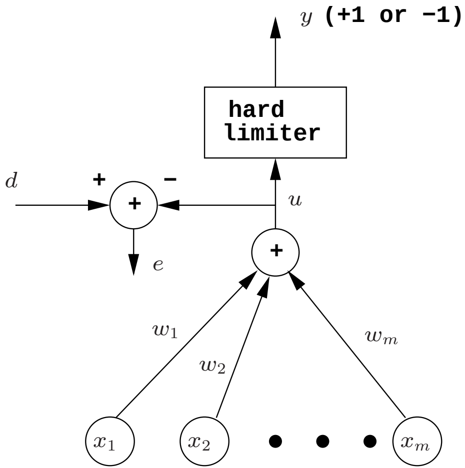
$$v_i(t+1) = (1 - \eta\lambda_i)^t\, v_i(0). \tag{7}$$

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

$$|1 - \eta\lambda_i| < 1, \quad \text{for } i = 1, \ldots, L. \tag{8}$$

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

Adaptive Linear Element (ADALINE)



Normalized LMS (NLMS)

- The NLMS algorithm is given by

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

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

NLMS: Algorithm Derivation

The objective function is given by

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

where λ is an Lagrangian multiplier.

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

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

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

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

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