Adaptive Filter

\[ u(n) \rightarrow D \rightarrow u(n-1) \rightarrow D \rightarrow u(n-2) \]

\[ D \times w_0 \]

\[ D \times w_1 \]

\[ D \times w_2 \]

\[ y(n) = \sum \]

\[ e(n) = y(n) - \sum d(n) \]
Model Assumptions and Parameters

- Training examples \((x(k),d(k))\) drawn randomly, second order zero mean sequences.

- Parameters
  - Inputs: \(u(k) \in \mathbb{R}^n\)
  - Weights: \(w(k) \in \mathbb{R}^n\)
  - Outputs: \(y(k) = w(k)^T x(k)\)
  - Desired outputs: \(d(k)\)
  - Error: \(e(k) = d(k) - y(k)\)

- Error criterion (MSE)
  \[
  \min J(w) = E [0.5(e(k))^2]
  \]
Wiener solution

Define $P = \mathbb{E}(u(k)d(k))$ and $R = \mathbb{E}(u(k)u(k)^T)$.

$J(w) = .5 \mathbb{E}[(d(k)-y(k))^2]$  
$= .5\mathbb{E}(d(k)^2) - \mathbb{E}(u(k)d(k))^T w + w^T \mathbb{E}(u(k)u(k)^T) w$  
$= .5\mathbb{E}[d(k)^2] - P^T w + .5w^T R w$

Note $J(w)$ is a quadratic function of $w$. To minimize $J(w)$ find gradient, $\nabla J(w)$ and set to 0.

$\nabla J(w) = -P + Rw = 0$

$Rw = P$ (Wiener solution)

If $R$ is nonsingular, then $w = R^{-1} P$.

Resulting $MSE = .5\mathbb{E}[d(k)^2] - .5P^T R^{-1} P$
Gradient based iterative algorithms

- Steepest descent algorithm (move in direction of negative gradient)
  \[ w(k+1) = w(k) - \mu \nabla J(w(k)) = w(k) + \mu (P - Rw(k)) \]
- Least mean square algorithm (approximate gradient from training example)
  \[ \nabla J(w(k)) = -e(k)u(k) \]
  \[ w(k+1) = w(k) + \mu e(k)u(k) \]
Steepest Descent Convergence

- \( w(k+1) = w(k) + \mu (P-Rw(k)) \); Let \( w^* \) be solution.
  Center weight vector \( c = w - w^* \)

- \( c(k+1) = (I - \mu R)c(k) \); Assume \( R \) is nonsingular.
  Decorrelate weight vector \( v = Q^{-1}c \) where \( R = Q\Lambda Q^{-1} \) is the transformation that diagonalizes \( R \).

- \( v(k+1) = (I - \mu \Lambda)v(k), \ v(k) = (I - \mu \Lambda)^k v(0) \).
  Conditions for convergence \( 0 < \mu < 2/\lambda_{\text{max}} \).
Step Size $\mu$

$\mu$ too large

$\mu$ too small
Rate of Convergence

- Rate of convergence depends on eigenvalues, $\lambda_i$ as convergence rate for this eigenvalue is $(1 - \mu \lambda_i)$. Key eigenvalues are $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$.
- Fastest rate of convergence achieved when setting $\mu = 2 / (\lambda_{\text{min}} + \lambda_{\text{max}})$. This results in smallest and largest eigenvalue having same convergence rate.
- Convergence of SD depends on condition number matrix of $\lambda_{\text{max}} / \lambda_{\text{min}}$. 
Energy Function

- Energy Function:
  \[ J(w) = 0.5 \sigma_d^2 - P^T w + 0.5 w^T R w \]
  For optimal weight \( R w^* = P \) and
  \[ J_{\text{min}} = J(w^*) = 0.5 \sigma_d^2 - 0.5 P^T w^* \]

- SD energy function behavior
  \[ J(w(k)) = J_{\text{min}} + 0.5 (w(k)-w^*)^T R (w(k)-w^*) \]
  \[ = J_{\text{min}} + 0.5 v(k)^T \Lambda v(k) \]
  \[ = J_{\text{min}} + 0.5 \sum_i (I - \mu \lambda_i)^{2k} v_i(0)^2 \]
SD Algorithm Summary

- Algorithm is an iterative deterministic algorithm that minimizes energy function.
- For filtering problem (finding Wiener filter solution), the energy surface is a quadratic function of weights.
- Behavior depends on initial weight, \( w(0) \) and correlation matrix \( R \) (eigenvalues \( \lambda_i \)). For convergence step size must satisfy \( 0 < \mu < 2/\lambda_{\text{max}} \) (discuss prediction example).
- Overdamped versus underdamped behavior: optimal step size \( \mu = 2 / (\lambda_{\text{min}} + \lambda_{\text{max}}) \).
LMS Algorithm

- SD requires knowledge of R and P. In many applications these second order statistics are unknown.
- Least mean square (LMS) algorithm
  - Make estimate of gradient (noisy gradient descent algorithm)
  - Estimate based on one observation \((u(k), d(k))\)
    \[
    \nabla \hat{J}(w(k)) = -e(k)u(k)
    \]
    \[
    w(k+1) = w(k) + \mu e(k)u(k)
    \]
LMS Convergence Behavior

- Assumptions: \( u(n) \) iid sequence, \( u(n) \) independent of \( d(n-k) \), \( k > 0 \), \( d(n) \) independent of \( y(n-k) \), \( k > 0 \), \( u(n) \) and \( d(n) \) are jointly Gaussian.

- Mean convergence analysis: Let \( e^*(k) = d(k) - w^*^T u(k) \), denote error from optimal weight at time \( k \).
  - \( E(c(k+1)) = (I - \mu R) E(c(k)) + \mu E(u(k)e^*(k)) \)
  - Asymptotically assuming step size is chosen correctly, then \( \lim_{k} E(c(k)) = 0 \) and \( E(w(k)) \) converges to \( w^* \)

- Mean squared analysis studies cost function \( J(w(k)) \). Note \( \text{tr}(R) > \lambda_{\text{max}} \) and more conservative bound given by \( 0 < \mu < 2 / \text{tr}(R) \).