Note Title 05-09-2014

Steepest Descent Algorithm

(A gradient descent technique)

Recall LMMSE

X is Random Variable

$\underline{Y}$ is random vector

$$\hat{X}_{LMMSE} = \underline{w}_{opt}^{*} \underline{Y} \quad \text{where}$$

$$R_Y \underline{w}_{opt} = R_{YX}$$

We will always assume R_Y is invertible (unless otherwise stated)

$$\underline{w}_{opt} = R_Y^{-1} R_{YX}$$

Steepest Descent Algorithm

* An iterative approach to find optimal LMMSE weight $\underline{w}_{opt}$
* Applies to more general cost functions (other than MSE)
* leads to adaptive filters

Algorithm

Start with weight vector $\underline{w}(0)$

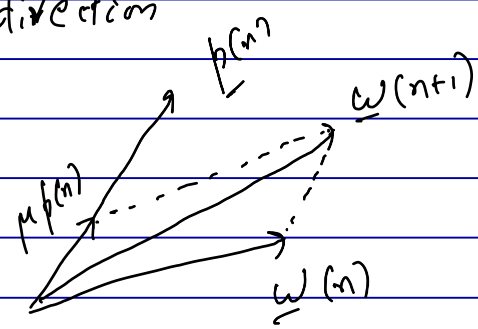
time index denoted within braces

Update the vector as follows

$$\underline{w}(n+1) = \underline{w}(n) + \mu \, \underline{p}(n)$$

$\mu \rightarrow$ step size (real positive constant)

$\underline{p}(n) \longrightarrow$ update direction



How should I choose μ & $\underline{p}(n)$

such that $\underline{w}(n) \longrightarrow \underline{w}_{opt}$ as $n \rightarrow \infty$

Recall

$\hat{X}(n) = \underline{w}^*(n)\,\underline{Y}$

$$\text{MSE Cost} = E\left(X - \hat{X}\right)^2$$

$$\| $$

$$J\left(\underline{w}(n)\right) = \sigma_X^2 - R_{XY}\,\underline{w}(n) - \underline{w}^*(n)\,R_{YX} + \underline{w}^*(n)\,R_Y\,\underline{w}(n)$$

$$J(\underline{w}(n+1)) = \sigma_X^2 - R_{XY}\,\underline{w}(n+1) - \underline{w}^*(n+1) R_{YX} + \underline{w}^*(n+1) R_Y \underbrace{\underline{w}(n+1)}_{\underline{w}(n) + \mu\,\underline{p}(n)}$$

after some algebra

$$J(\underline{w}(n+1)) = J(\underline{w}(n)) + \mu^2 \underline{p}^*(n) R_Y \underline{p}(n) + \mu\, \underline{p}^*(n)\left(R_Y \underline{w}(n) - R_{YX}\right) + \mu\left(\underline{w}^*(n) R_Y - R_{XY}\right)\underline{p}(n)$$

$$J(\underline{w}(n+1)) = J(\underline{w}(n)) + \mu^2 \underline{p}^*(n) R_Y \underline{p}(n) + 2\mu\, \mathrm{Re}\left\{ \underbrace{\left(\underline{w}^*(n) R_Y - R_{XY}\right)}_{\nabla_{\underline{w}(n)} J(\underline{w}(n))} \underline{p}(n) \right\}$$

We want $J(\underline{w}(n+1)) < J(\underline{w}(n))$

Since $\mu^2 \underline{p}^*(n) R_Y \underline{p}(n) \geq 0$

we need $\mathrm{Re}\left\{ \nabla_{\underline{w}(n)} J(\underline{w}(n))\, \underline{p}(n) \right\} < 0$

first of all Note

$\nabla_{\underline{w}} J(\underline{w}(n))$ is a row vector

(Size same as $\underline{w}(n)$)

Take $\underline{p}(n) = -\left[\nabla_{w} (J(\underline{w}(n))\right]^*$

choice for steepest descent algorithm.

$$\underline{w}(n+1) = \underline{w}(n) + \mu\left[-\nabla_{\underline{w}(n)} J(\underline{w}(n))\right]^*$$

$$= \underline{w}(n) + \mu(-1)\left(\underline{w}^*(n) R_Y - R_{XY}\right)^*$$

$$\boxed{\underline{w}(n+1) = \underline{w}(n) - \mu\left[R_Y \underline{w}(n) - R_{YX}\right]}$$

Steepest descent algorithm

$$\underbrace{\underline{w}(n+1)}_{M\times 1 \text{ Vector}} = \underbrace{(I - \mu R_Y)}_{M\times M}\underline{w}(n) + \mu \underbrace{R_{YX}}_{M\times 1}$$

8/9 Will $\underline{w}(n) \rightarrow \underline{w}_{opt}$ as $n \rightarrow \infty$

Answer will depend on the value of μ.

$(*) \longrightarrow \underline{w}(n+1) = (\underline{I} - \mu R_Y)\,\underline{w}(n) + \mu R_{YX}$

Note $\quad R_Y \underline{w}_{opt} = R_{YX}$

Subtracting $\underline{w}_{opt}$ from both sides of $(*)$

$$\underline{w}(n+1) - \underline{w}_{opt} = (\underline{I} - \mu R_Y)\,\underline{w}(n) + \mu (R_Y \underline{w}_{opt}) - \underline{w}_{opt}$$

$$\underbrace{\underline{w}(n+1) - \underline{w}_{opt}}_{\tilde{\underline{w}}(n+1)} = (\underline{I} - \mu R_Y)\underbrace{(\underline{w}(n) - \underline{w}_{opt})}_{\tilde{\underline{w}}(n)}$$

$\tilde{w}(n)$ $\Downarrow$ error between $\underline{w}(n)$ and $\underline{w}_{opt}$ at n^{th} iteration

$$\tilde{\underline{w}}(n+1) = (\underline{I} - \mu R_Y)\,\tilde{\underline{w}}(n)$$

$$= (\underline{I} - \mu R_Y)\left[(\underline{I} - \mu R_Y)\,\tilde{\underline{w}}(n-1)\right]$$

Repeating this

$$\tilde{\underline{w}}(n+1) = (\underline{I} - \mu R_Y)^{n+1}\,\tilde{\underline{w}}(0)$$

Will $(\underline{I} - \mu R_Y)^n$ converge to zero matrix as $n \to \infty$.

Let the eigen decomposition of R_r be

$$R_r = Q \Lambda Q^*$$

$Q \to$ is orthonormal eigen vector matrix

$$QQ^* = Q^*Q = I$$

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_M \end{bmatrix} = \Lambda \to \text{is diagonal matrix of eigenvalues}$$

$$I - \mu R_r = I - \mu Q \Lambda Q^*$$

$$= QQ^* - \mu Q \Lambda Q^*$$

$$= Q [I - \mu \Lambda] Q^*$$

$$(I - \mu R_r)^2 = Q[I - \mu \Lambda] \underbrace{Q^* \; Q}_{I} [I - \mu \Lambda] Q^*$$

$$= Q [I - \mu \Lambda]^2 Q^*$$

$$(I - \mu R_r)^n = Q [I - \mu \Lambda]^n Q^*$$

$$(I - \mu R_y)^n \rightarrow \text{zero matrix if}$$

$$(I - \mu \Lambda)^n \rightarrow \text{zero matrix}$$

$$\begin{bmatrix} (1-\mu\lambda_1)^n & & & \\ & (1-\mu\lambda_2)^n & & \\ & & \ddots & \\ & & & (1-\mu\lambda_M)^n \end{bmatrix}$$

will converge to zero (matrix)

if $(1 - \mu\lambda_i)^n \rightarrow 0$ for $i = 1$ to M

$|1 - \mu\lambda_i| < 1$ for all $i = 1$ to M

$\Rightarrow \quad -1 < 1 - \mu\lambda_i < 1 \quad \forall i$

$\Rightarrow \quad 0 < \mu\lambda_i < 2 \quad \forall i$

$\Rightarrow \quad \mu < \frac{2}{\lambda_i} \quad \forall i$

equivalently $\boxed{0 < \mu < \frac{2}{\lambda_{max}}}$ $\rightarrow$ Condition for convergence of SDA

λ_{max} is the largest eigen value of R_Y

Other types of cost functions

$$\tilde{X} = \cancel{X} - \underline{w}^* \underline{Y}$$

MSE : $E\{|\tilde{x}|^2\}$

Absolute Error : $E\{|\tilde{X}|\}$

Fourth power : $E\{|\tilde{X}|^4\}$

Mixed : $E\{|\tilde{x}|^2 + |\tilde{x}|^4\}$

For all these criteria,

write cost function as a function of $\underline{w}(n)$

$J(\underline{w}(n))$

Gradient descent algorithm

$$\underline{w}(n+1) = \underline{w}(n) + \mu\left[-\nabla_{\underline{w}(n)} J(\underline{w}(n))\right]^*.$$

Convergence may be difficult to analyse

due to existence of local minima

for some types of cost functions.

Modes of convergence

Recall :

$$\underline{\tilde{w}}(n) = (I - \mu R_r)^n \, \underline{\tilde{w}}(0)$$

$$= Q \, [I - \mu \Lambda]^n \, Q^* \, \underline{\tilde{w}}(0)$$

$$\underbrace{Q^* \, \underline{\tilde{w}}(n)}_{|} = \underbrace{Q^* Q}_{I} \, [I - \mu \Lambda]^n \, \underbrace{Q^* \underline{\tilde{w}}(0)}$$

$$\underline{\tilde{u}}(n) = (I - \mu \Lambda)^n \, \underline{\tilde{u}}(0)$$

Note $\| \underline{\tilde{u}}(n) \|^2 = \underline{\tilde{u}}^*(n) \, \underline{\tilde{u}}(n)$

$$= (Q^* \underline{\tilde{w}}(n))^* \, Q^* \, \underline{\tilde{w}}(n)$$

$$= \underline{\tilde{w}}^*(n) \, \underbrace{Q \, Q^*}_{I} \, \underline{\tilde{w}}(n)$$

$$= \underline{\tilde{w}}^*(n) \, \underline{\tilde{w}}(n)$$

$$= \| \underline{\tilde{w}}(n) \|^2$$

$\underline{\tilde{u}}(n)$ & $\underline{\tilde{w}}(n)$ have same norm

$\underline{\tilde{u}}(n) \to \underline{0}$ then $\underline{\tilde{w}}(n) \to 0$ (and vice versa)

$$\underline{\tilde{u}}(n) = \begin{bmatrix} \tilde{u}_1(n) \\ \tilde{u}_2(n) \\ \vdots \\ \tilde{u}_M(n) \end{bmatrix}$$

$$\underline{\tilde{u}}(n) = (\underline{I} - \mu \Lambda)^n \, \underline{\tilde{u}}(0)$$

ith coordinate

$$\tilde{u}_i(n) = (1 - \mu \lambda_i)^n \, \tilde{u}_i(0)$$

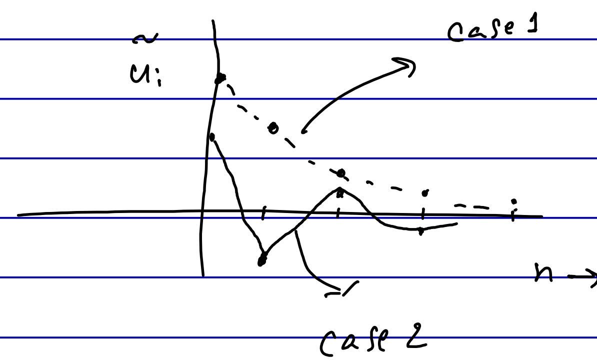
ith mode of convergence is determined by $(1 - \mu \lambda_i)$

We need $|1 - \mu \lambda_i| < 1$ for convergence

Case 1

$$0 < 1 - \mu \lambda_i < 1$$

Convergence is monotonic



Case 2

$$-1 < 1 - \mu \lambda_i < 0$$

Convergence is oscillating

Rate of convergence is determined by
$|1-\mu\lambda_i|$

fastest mode $= \min_i |1-\mu\lambda_i|$

slowest mode $= \max_i |1-\mu\lambda_i|$

Optimal step size

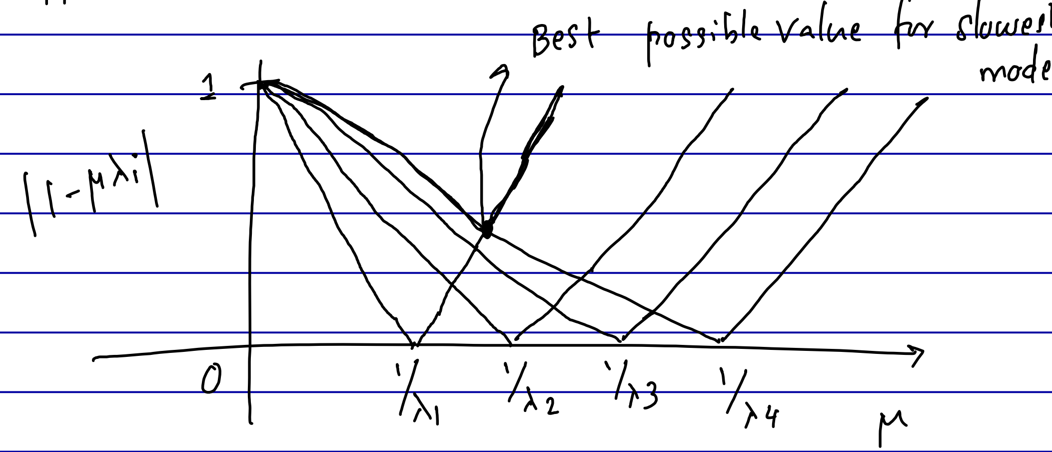
In asymptotic regime (large n)
slowest mode determines overall behaviour.

Choose step size μ such that
we get the best possible slowest mode.

$$\mu_{opt} = \min_{\mu} \max_{i} |1-\mu\lambda_i|$$

Consider 4 modes

Suppose $\lambda_{max} = \lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 = \lambda_{min}$



Slowest mode takes the minimum value when mode 1 and mode 4 intersect.

At the intersection point

$1-\mu\lambda_{max}$ will be negative

$1-\mu\lambda_{min}$ will be positive

$$|1-\mu\lambda_{max}| = |1-\mu\lambda_{min}| \rightarrow \text{for optimal step size}$$

$$-(1-\mu\lambda_{max}) = (1-\mu\lambda_{min})$$

$$\boxed{\mu_{opt} = \frac{2}{\lambda_{max}+\lambda_{min}}}$$

Corresponding slowest mode

$$= (1-\mu_{opt}\lambda_{max}) \;\&\; (1-\mu_{opt}\lambda_{min})$$

$$= \pm\frac{\lambda_{max}-\lambda_{min}}{\lambda_{max}+\lambda_{min}}$$

Time constant of convergence (of ith mode) is defined as

$$\tau_i = \frac{-1}{2\ln|1-\mu\lambda_i|}$$

MSE Behavior [Learning Curve]

So far, we studied how $\underline{w}(n) \rightarrow \underline{w}_{opt}$

Now, let us see how mse $J(\underline{w}(n))$ converge to minimum?

$$J(\underline{w}(n)) = E\{ |X - \underline{w}^*(n)\underline{Y}|^2 \}$$

$$= E\{ |\underbrace{X - \underline{w}_{opt}^*\underline{Y}}_{\mathrm{I}} + \underbrace{\underline{w}_{opt}^*\underline{Y} - \underline{w}^*(n)\underline{Y}}_{\mathrm{II}}|^2 \}$$

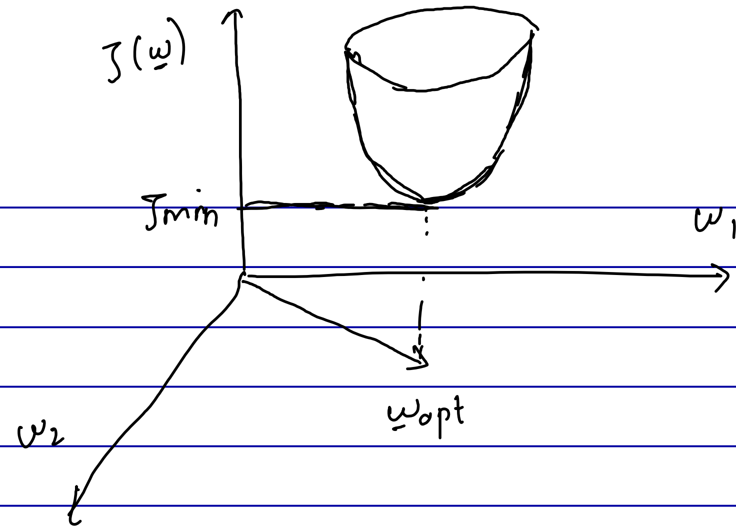
[I & II are orthogonal/uncorrelated from Orthogonality principle]

$$= \underbrace{E|X - \underline{w}_{opt}^*\underline{Y}|^2}_{\substack{J_{min} \\ \text{(MSE of optimal filter)}}} + E\{ |(\underbrace{\underline{w}_{opt} - \underline{w}(n)}_{\underline{\tilde{w}}(n)})^* \underline{Y}|^2 \}$$

$$J(\underline{w}(n)) = J_{min} + \underbrace{\underline{\tilde{w}}^*(n) R_Y \underline{\tilde{w}}(n)}_{\geq 0}$$

2D example

$$\underline{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$



$$J(\underline{w}(n)) = J_{min} + \underbrace{\underline{\tilde{w}}^*(n)\, R_y\, \underline{\tilde{w}}(n)}_{\text{excess MSE}}$$

$\xi(n) \rightarrow$ denote excess MSE at n^{th} iteration

$$\xi(n) = \underline{\tilde{w}}^*(n)\, R_y\, \underline{\tilde{w}}(n)$$

$R_y = Q \Lambda Q^*$

$\underline{\tilde{u}}(n) = Q^* \underline{\tilde{w}}(n)$

$$= \tilde{u}(n)\, \Lambda\, \tilde{u}(n)$$

$$(*) \rightarrow \xi(n) = \sum_{i=1}^{M} |\tilde{u}_i(n)|^2 \lambda_i \qquad \underline{\tilde{u}}(n) = \begin{bmatrix} \tilde{u}_1(n) \\ \vdots \\ \tilde{u}_M(n) \end{bmatrix}$$

we know $\tilde{u}_i(n) \rightarrow 0$ as $n \rightarrow \infty$ if $\mu < \frac{2}{\lambda_{max}}$

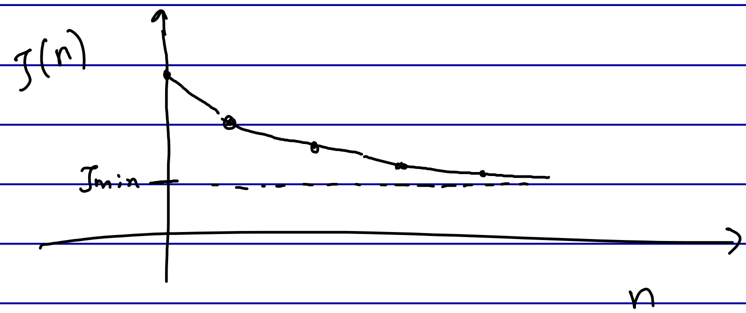
i^{th} mode convergence $\rightarrow$ $\tilde{u}_i(n) = (1-\mu\lambda_i)^n\, \tilde{u}_i(0)$

$\rightarrow$ Put this in $(*)$

$$\xi(n) = \sum_{i=1}^{M} \lambda_i (1-\mu\lambda_i)^{2n} |\tilde{u}_i(0)|^2$$

$$\lambda_i (1-\mu\lambda_i)^{2n} \longrightarrow 0 \quad \text{monotonically}$$

$$\xi(n) \longrightarrow 0 \quad \text{monotonically} \quad \text{as } n \to \infty.$$



Contours of Error Surface

$J(\underline{w}) = c$ (c is a constant)

set of all $\underline{w}$ which give same cost

what is the shape of this contour?

$$J(\underline{w}) = J_{min} + \underline{\tilde{w}}^{*} R_Y \underline{\tilde{w}}$$

$$\underline{\tilde{w}} = \underline{w} - \underline{w}_{opt}$$

$$J(\underline{w}) = c \Rightarrow$$

$$\underline{\tilde{w}}^{*} R_Y \underline{\tilde{w}} = c - J_{min} = c_1$$

change of coordinates

$$R_Y = Q \Lambda Q^{*}$$

$$\underline{\tilde{u}} = Q^{*} \underline{\tilde{w}}$$

$$\underline{\tilde{u}} = \begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \vdots \\ \tilde{u}_M \end{bmatrix}$$

$$\underline{\tilde{u}}^* \Lambda \, \underline{\tilde{u}} = c_1$$

$$\lambda_1 |\tilde{u}_1|^2 + \lambda_2 |\tilde{u}_2|^2 + \dots + \lambda_M |\tilde{u}_M|^2 = c_1$$

$\Downarrow$

Ellipsoid in M dim space.

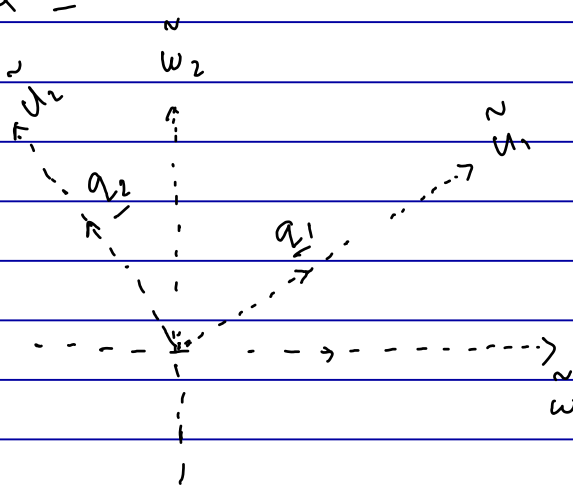
Take 2D example

$$Q = \begin{bmatrix} \underline{q}_1 & \underline{q}_2 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

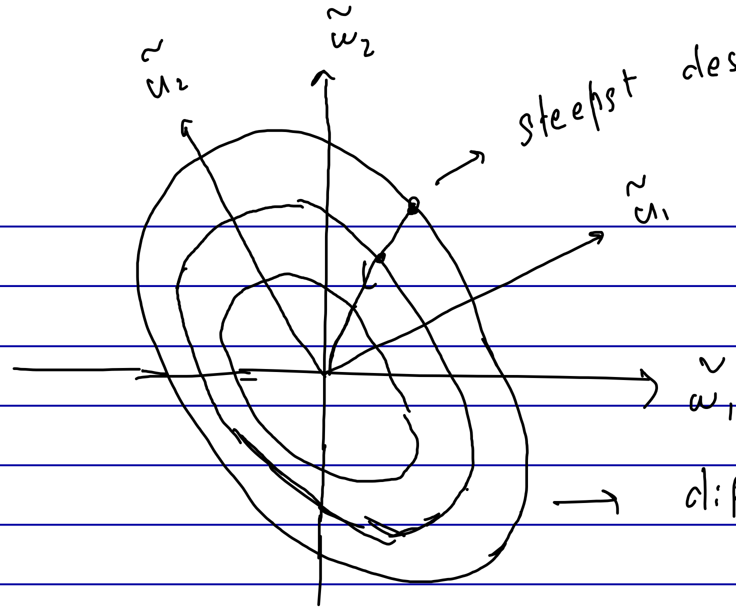
$$\underline{\tilde{u}} = Q^* \underline{\tilde{w}}$$

$$\underline{\tilde{w}} = \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \end{bmatrix}$$

$$\underline{\tilde{u}} = \begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix}$$



Note $\quad Q^* \underline{q}_1 = \begin{bmatrix} \underline{q}_1^* \\ \underline{q}_2^* \end{bmatrix} \begin{bmatrix} \underline{q}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



steepest descent algorithm starts at some point and move inwards and reach origin eventually

different ellipses correspond to different constants

Modifications to $\underline{\underline{\text{SDA}}}$

SDA : $\underline{w}(n+1) = \underline{w}(n) + \mu \left[-\nabla_w J(\underline{w}(n)) \right]^*$

Iteration dependent step size

$$\underline{w}(n+1) = \underline{w}(n) + \mu(n) \left[-\nabla_w J(\underline{w}(n)) \right]^*$$

choose $\mu(n)$ appropriately at each iteration to have faster convergence.

12/9 <u>Iteration dependent step size</u>

$$w(n+1) = w(n) + \mu(n) \left[R_{yx} - R_y \underline{w}(n) \right]$$

Error $\quad \tilde{\underline{w}}(n+1) = \underline{w}(n+1) - \underline{w}_{opt}$

$$= \left[I - \mu(n) R_Y \right] \tilde{\underline{w}}(n)$$

Change of coordinates

$$\tilde{\underline{u}}(n) = Q^* \tilde{\underline{w}}(n)$$

$$\tilde{\underline{u}}(n+1) = \left[I - \mu(n) \Lambda \right] \tilde{\underline{u}}(n)$$

i^{th} mode

$$\tilde{u}_i(n+1) = \left(\prod_{k=0}^{n} \left[1 - \mu(k) \lambda_i \right] \right) \tilde{u}_i(0)$$

want $\quad \prod_{k=0}^{n} \left[1 - \mu(k) \lambda_i \right] \to 0 \quad$ as $\quad n \to \infty$

Note that $\quad | 1 - \mu(k) \lambda_i | < 1 \quad$ for all k

is not sufficient ensure convergence.

Example.

$$a(k) = e^{-\frac{1}{(k+1)^2}}$$

$$|a(k)| < 1 \quad \text{for all } k$$

$$\prod_{k=1}^{n} a(k) \to e^{-\pi^2/6}$$

Following are some sufficient conditions (but not necessary)

① Choose $\mu(n) \to 0$ as $n \to \infty$

<u>and</u> $\sum \mu(n) = \infty$

For instance $\mu(n) = \frac{\alpha}{n+\beta}$, $\alpha > 0$, $\beta > 0$

or any $\mu(n)$ such that

$\sum \mu(n) = \infty$ & $\sum \mu^2(n) < \infty$

② choose $\mu(n)$ such that

$0 < |\mu(n)| < \alpha < 1$

× ———— ×

<u>Optimal choice of $\mu(n)$ for fastest convergence</u>

$$\underline{\tilde{w}}(n+1) = \left(\underline{I} - \mu(n) R_y\right) \underline{\tilde{w}}(n)$$

$$J(\underline{w}(n+1)) = J_{min} + \underline{\tilde{w}}(n+1)^* R_y \, \underline{\tilde{w}}(n+1)$$

$$= J_{min} + \left\{\left[\left(\underline{I} - \mu(n) R_y\right) \tilde{\underline{w}}(n)\right]^* R_y \left[\quad \cdot \quad \right]\right.$$

$$= \underbrace{J_{min} + \tilde{\underline{w}}^{*}(n) R_Y \tilde{\underline{w}}(n)} - \Big[2\mu(n) \tilde{\underline{w}}^{*}(n) R_Y^2 \tilde{\underline{w}}(n)$$

$$- \mu^2(n) \tilde{\underline{w}}^{*}(n) R_Y^3 \tilde{\underline{w}}(n) \Big]$$

$$J(\underline{w}(n+1)) = J(\underline{w}(n)) - \underbrace{\big[\quad A \quad \big]}$$

Choose $\mu(n)$ so that term A is maximized

$$\frac{dA}{d\mu} = 0$$

$$\boxed{\mu^{opt}(n) = \frac{\tilde{\underline{w}}^{*}(n) R_Y^2 \tilde{\underline{w}}(n)}{\tilde{\underline{w}}^{*}(n) R_Y^3 \tilde{\underline{w}}(n)}}$$

For this $\mu^{opt}(n)$, will this algorithm converge?

Excess MSE $\quad \xi(n) = J(\underline{w}(n)) - J_{min}$

$$\xi(n+1) = \xi(n) - A \Big|_{\text{evaluated at } \mu_{opt}}$$

$$= \xi(n) - \frac{\left[\tilde{\underline{w}}^{*}(n) R_Y^2 \tilde{\underline{w}}(n) \right]^2}{\tilde{\underline{w}}^{*}(n) R_Y^3 \tilde{\underline{w}}(n)}$$

If R_Y is positive definite,

R_Y^2, R_Y^3 are also positive definite

$$0 \le \xi(n+1) < \xi(n)$$

$\xi(n)$ is monotonically decreasing and bounded

$\Rightarrow$ $\xi(n)$ Converges

To show it converges to zero

Start with some initial condition $\underline{w}(0)$

Run fixed μ SDA

$$\underline{w}^{\mu}(n+1) = \underline{w}^{\mu}(n) + \mu\left[R_{Yx} - R_Y \underline{w}^{\mu}(n)\right]$$

We know that

$$\xi^{\mu}(n) \longrightarrow 0 \qquad \text{if } 0 < \mu < 2/\lambda_{max}$$

excess MSE for fixed μ algo

For optimal iteration dependent step size

$$\xi(n) \le \xi^{\mu}(n) \quad \forall n$$

$$\xi(n) \to 0 \quad \text{as} \quad n \to \infty.$$

Newton's Method

SDA : $\underline{w}(n+1) = \underline{w}(n) + \mu \underbrace{\left[- \nabla_w \, J(w(n)) \right]^*}$

$J(\underline{w})$ is $\overset{(MSE)}{\text{cost}}$ as function of $\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix}$

$$\nabla_{\underline{w}} \left(J(\underline{w}) \right) = \left[\frac{\partial J}{\partial w_1} \quad \frac{\partial J}{\partial w_2} \quad \cdots \quad \frac{\partial J}{\partial w_M} \right]$$

Hessian Matrix is defined as

$$\nabla^2_{\underline{w}} (J) = \nabla_{\underline{w}^*} \left[\nabla_{\underline{w}} J(\underline{w}) \right]$$

$$\nabla^2_{\underline{w}} (J(\underline{w})) = \begin{bmatrix} \frac{\partial^2 J}{\partial w_1^* \partial w_1} & \frac{\partial^2 J}{\partial w_1^* \partial w_2} & \cdots & \frac{\partial^2 J}{\partial w_1^* \partial w_M} \\ \vdots & & & \\ \frac{\partial^2 J}{\partial w_M^* \partial w_1} & & \cdots & \frac{\partial^2 J}{\partial w_M^* \partial w_M} \end{bmatrix}$$

For newton algorithm

$$\underline{w}(n+1) = \underline{w}(n) + \mu (-) \left[\nabla^2_{\underline{w}} (J(\underline{w})) \right]^{-1} \left[\nabla_{\underline{w}} (J(\underline{w})) \right]^*$$

For MSE cost

$$\left[\nabla_{\underline{w}} (J(\underline{w}))\right] = \underline{w}^* R_Y - R_{xy}$$

$$\nabla^2_{\underline{w}} (J) = \nabla_{\underline{w}^*} \left[\nabla_{\underline{w}} J\right] = R_Y$$

$$\underline{w}(n+1) = \underline{w}(n) + \mu R_Y^{-1} \left(R_{Yx} - R_Y \underline{w}(n) \right)$$

$$= \underline{w}(n) + \mu \left(\underbrace{R_Y^{-1} R_{Yx}}_{\underline{w}_{opt}} - \underline{w}(n) \right)$$

15/9

$$= \underline{w}(n) + \mu \left(\underline{w}_{opt} - \underline{w}(n) \right)$$

Error vector

$$\tilde{\underline{w}}(n+1) = (1-\mu)\, \tilde{\underline{w}}(n)$$

Convergence when $|1-\mu| < 1$

For $\mu = 1$, convergence happens in one iteration

ie) $\underline{w}(n+1) = \underline{w}(n) + 1 \cdot R_Y^{-1} \left(R_{Yx} - R_Y \underline{w}(n) \right]$

$$= R_Y^{-1} R_{Yx} = \underline{w}_{opt}.$$

equivalent to finding optimal filter directly

Learning Curve (for $\mu \neq 1$)

$$J(\underline{w}(n)) = J_{min} + \underline{\tilde{w}}^{H}(n)\, R_{\gamma}\, \underline{\tilde{w}}(n)$$

Coordinate transformation

$$\underline{\tilde{u}}(n) = Q^{H}\, \underline{\tilde{w}}(n)$$

$$= J_{min} + \underline{\tilde{u}}^{H}(n)\, \Lambda\, \underline{\tilde{u}}(n)$$

$$= J_{min} + \sum_{i=1}^{M} \lambda_i\, |\tilde{u}_i(n)|^2$$

$$= J_{min} + \sum_{i=1}^{M} \lambda_i\, |1-\mu|^{2n}\, |\tilde{u}_i(0)|^2$$

$\Downarrow$

Monontonically converge to zero if $|1-\mu| < 1$

* Contour Curves (set of $\underline{w}$ with same cost) are same as SDA

* Convergence does not depend on λ_i's.

* Time constant for rate of convergence is $\dfrac{-1}{2\ln|1-\mu|}$