Performance of Adaptive Filters

- Will focus on LMS algorithm
- Will study
  - → transient behaviour
  - → Steady state performance.
- For steepest descent algorithm,

  If $0 < \mu < 2/\lambda_{max}$ then

  SDA converges to $\underline{w}_{opt}$.
- Analysis of LMS is not straightforward.

  Usually requires lot of simplifying assumptions.

29/9

Steady State Performance.

- LMS Update $\underline{w}(n+1) = \underline{w}(n) + \mu\, \underline{y}(n)\left[x^*(n) - \underline{y}^*(n)\,\underline{w}(n)\right]$

$x(n)$, $\underline{y}(n)$ are realizations of random variables X and $\underline{Y}$

will use $\underbrace{X(n) \text{ \& } \underline{Y}(n)}_{\text{capital letters}}$ to denote the random quantities

Assume $X(n)$, $\underline{Y}(n) \to M\times 1$ vector are zero mean

$X(n) \downarrow$ scalar

$$E\{|X(n)|^2\} = \sigma_x^2$$

$$E\{\underline{Y}(n)\,\underline{Y}^*(n)\} = R_Y$$

$$R_{XY}^* = E\{\underline{Y}(n)\,X^*(n)\} = R_{YX}$$

Under these assumptions,

$\underline{w}(n+1)$ is also a random vector

depends on $\{X(n), X(n-1), \ldots, X(0)$
$\underline{Y}(n), \underline{Y}(n-1), \ldots, \underline{Y}(0)\}$

Will assume that initial condition $\underline{w}(-1)$ is also a random vector

Error at time n

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

Define steady state MSE as

$$\lim_{n\to\infty} E\{|D(n)|^2\}$$

Suppose $\underline{w}(n) = \underline{w}_{opt}$, then

$$E\{|D(n)|^2\} = J_{min} = \text{min. MSE.}$$

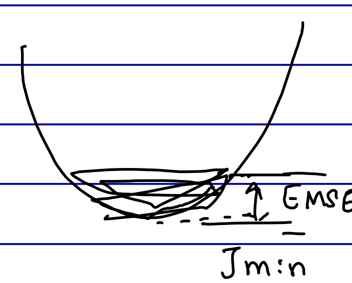
If $\underline{w}(n) \neq \underline{w}_{opt}$, then

$$E\{|D(n)|^2\} > J_{min}$$

We define excess meansquared error as

$$EMSE = \lim_{n\to\infty} |D(n)|^2 - J_{min}$$

SDA: $\underline{w}(n+1) = \underline{w}(n) + \mu\left(R_{YX} - R_Y\,\underline{w}(n)\right)$



If $\underline{w}(n) = \underline{w}_{opt} = R_Y^{-1} R_{YX}$

Then SDA will not update the weights.

LMS: $\underline{w}(n) = \underline{w}(n) + \mu\left(\underline{Y}(n)\,X^*(n) - \underline{Y}(n)\,\underline{Y}^*(n)\,\underline{w}(n)\right)$

Even if $\underline{w}(n) = \underline{w}_{opt}$, we will still update $\underline{w}(n+1)$ and $\underline{w}(n+1) \neq \underline{w}_{opt}$.

μ, R_Y, σ_X^2, R_{YX} will play a role in the value of EMSE.

x———————x

- First we will derive two useful results (which are true for many adaptive filters)

  * Energy Conservation Relation
  * Fundamental Variance Relation

* Using these results, we derive EMSE by using some approximations.

x———————x

Energy Conservation Relation

Define two more error quantities

Let $\underline{\tilde{w}}(n) = \underline{w}_{opt} - \underline{w}(n)$

apriori error $D_a(n) = \underline{\tilde{w}}^*(n)\, \underline{Y}(n)$

posteriori error $D_p(n) = \underline{\tilde{w}}^*(n+1)\, \underline{Y}(n)$

Will consider general filter update equation

(1) — $\underline{w}(n+1) = \underline{w}(n) + \mu\, \underline{Y}(n) \cdot g(D(n))$

$g(\cdot)$ is some arbitrary function

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

$g(.)$ is any function.

Note: For LMS $g(D(n)) = D^*(n)$

LMF : $g(D(n)) = D^*(n)\,|D(n)|^2$

ε-NLMS : $g(D(n)) = \dfrac{D^*(n)}{\varepsilon + \|\underline{Y}(n)\|^2}$

Subtracting $\underline{w}_{opt}$ in (1)

$$\underline{\tilde{w}}(n+1) = \underline{\tilde{w}}(n) - \mu\, \underline{Y}(n)\, g(D(n)) \quad \text{—— (2)}$$

Multiply (2) with $\underline{Y}^*(n)$

$$\underbrace{\underline{Y}^*(n)\,\underline{\tilde{w}}(n+1)}_{D_p^*(n)} = \underbrace{\underline{Y}^*(n)\,\underline{\tilde{w}}(n)}_{D_a^*(n)} - \mu\, \|\underline{Y}(n)\|^2\, g(D(n))$$

So,

$$g(D(n)) = \frac{D_a^*(n) - D_p^*(n)}{\mu\, \|\underline{Y}(n)\|^2} \quad \text{—— (3)}$$

Substitute (3) in (2), we have

$$\underline{\tilde{w}}(n+1) = \underline{\tilde{w}}(n) - \frac{\underline{Y}(n)}{\|\underline{Y}(n)\|^2}\left(D_a^*(n) - D_p^*(n)\right)$$

$$\underline{\tilde{w}}(n+1) + \frac{D_a^*(n)}{\|\underline{Y}(n)\|^2}\underline{Y}(n) = \underline{\tilde{w}}(n) + \frac{D_p^*(n)}{\|\underline{Y}(n)\|^2}\underline{Y}(n)$$

Energy (norm squared) of both sides are equal

Take LHS

$$\left(\underline{\tilde{w}}(n+1) + \frac{D_a^*(n)}{\|\underline{Y}(n)\|^2}\underline{Y}(n)\right)^* \left(\underline{\tilde{w}}(n+1) + \frac{D_a(n)}{\|\underline{Y}(n)\|^2}\underline{Y}(n)\right)$$

$$= \underline{\tilde{w}}^*(n+1)\,\underline{\tilde{w}}(n+1) + \frac{|D_a(n)|^2}{\|\underline{Y}(n)\|^4}\underline{Y}^*(n)\,\underline{Y}(n)$$

$$+ \frac{D_a(n)}{\|\underline{Y}(n)\|^2}\underbrace{\underline{Y}^*(n)\,\underline{\tilde{w}}(n+1)}_{D_p^*(n)} + \frac{\overbrace{\underline{\tilde{w}}^*(n+1)\,\underline{Y}(n)}^{D_p(n)}\,D_a^*(n)}{\|\underline{Y}(n)\|^2}$$

$$\text{Norm}^2\ \text{LHS} = \|\underline{\tilde{w}}(n+1)\|^2 + \frac{|D_a(n)|^2}{\|\underline{Y}(n)\|^2} + \frac{D_a(n)\,D_p^*(n) + D_p(n)\,D_a^*(n)}{\|\underline{Y}(n)\|^2}$$

$$\text{Norm}^2\ \text{RHS} = \|\underline{\tilde{w}}(n)\|^2 + \frac{|D_p(n)|^2}{\|\underline{Y}(n)\|^2} + \frac{D_a(n)\,D_p^*(n) + D_p(n)\,D_a^*(n)}{\|\underline{Y}(n)\|^2}$$

Equating the norms,

$$\|\underline{\tilde{w}}(n+1)\|^2 + \frac{|D_a(n)|^2}{\|\underline{Y}(n)\|^2} = \|\underline{\tilde{w}}(n)\|^2 + \frac{|D_p(n)|^2}{\|\underline{Y}(n)\|^2}$$

for $\|\underline{Y}(n)\| \neq 0$

Note, if $\|\underline{Y}(n)\| = 0$, then

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

so $\|\underline{\tilde{w}}(n+1)\|^2 = \|\underline{\tilde{w}}(n)\|^2$ if $\underline{Y}(n) = \underline{0}$

This can be written compactly.

$$\bar{\mu}(n) = \begin{cases} 0 & \text{if } \|\underline{Y}(n)\| = 0 \\ \dfrac{1}{\|\underline{Y}(n)\|^2} & \text{if } \|\underline{Y}(n)\| \neq 0 \end{cases}$$

Then

$$\|\underline{\tilde{w}}(n+1)\|^2 + \bar{\mu}(n)\,|D_a(n)|^2 = \|\underline{\tilde{w}}(n)\|^2 + \bar{\mu}(n)\,|D_p(n)|^2$$

This is energy conservation relation

Holds true for any adaptive filter update of form (1)

x ——— x

20/9 <u>Steady state Conditions</u>

We say that adaptive filter has reached steady state if both the following conditions are true.

$\underline{s}$ is a constant vector

① $E\{\underline{\tilde{w}}(n)\} \longrightarrow \underline{s}$ as $n \to \infty$

C is a constant matrix.

② $E\{\underline{\tilde{w}}(n)\,\underline{\tilde{w}}^*(n)\} \longrightarrow \underline{\underline{C}}$ as $n \to \infty$

Note that steady state conditions imply that

$$E\,\|\underline{\tilde{w}}(n)\|^2 \longrightarrow \alpha$$ where α is a constant as $n \to \infty$.

because

$E\{\|\underline{\tilde{w}}(n)\|^2\}$ = Sum of diagonal entries of $E\{\underline{\tilde{w}}(n)\,\underline{\tilde{w}}^*(n)\}$

So α = sum of diagonal entries of C.

= trace(C)

x ———————— x

<u>Fundamental Variance relation.</u>

From Energy conservation, we have

$$\|\underline{\tilde{w}}(n+1)\|^2 + \bar{\mu}(n)\,|D_a(n)|^2 = \|\underline{\tilde{w}}(n)\| + \bar{\mu}(n)\,|D_p(n)|^2$$

Taking Expectation on both sides,

$$E\{\|\underline{\tilde{w}}(n+1)\|^2\} + E\{\bar{\mu}(n)\,|D_a(n)|^2\} =$$

$$E\{\|\tilde{\underline{w}}(n)\|^2\} + E\{\bar{\mu}(n)\,|D_p(n)|^2\}$$

where $\bar{\mu}(n) = \begin{cases} 0 & \text{if } \|\underline{Y}(n)\|^2 = 0 \\ \dfrac{1}{\|\underline{Y}(n)\|^2} & \text{else.} \end{cases}$

As $n \to \infty$, (in steady state)

$$E\{\|\tilde{\underline{w}}(n+1)\|^2\} = E\{\|\tilde{\underline{w}}(n)\|^2\} = \text{trace}(C)$$

So, $$E\{\bar{\mu}(n)\,|D_a(n)|^2\} = E\{\bar{\mu}(n)\,|D_p(n)|^2\}$$

Using (3) from previous class,

$$D_p^*(n) = D_a^*(n) - \mu\,\|\underline{Y}(n)\|^2\, g(D(n))$$

Recall $D(n) = X(n) - \underline{w}^b(n)\,\underline{Y}(n)$

$$E\{\bar{\mu}(n)\,|D_a(n)|^2\} = E\{\bar{\mu}(n)\,|D_a^*(n) - \mu\,\|\underline{Y}(n)\|^2 g(D(n))|^2\}$$

$$= E\{\bar{\mu}(n)\,(D_a^*(n) - \mu\,\|\underline{Y}(n)\|^2 g(D(n)))^*.$$
$$(D_a(n) - \mu\|\underline{Y}(n)\|^2 g(D(n)))\}$$

after some algebra

using the fact that $\bar{\mu}(n)\cdot\|\underline{Y}(n)\|^4 = \|\underline{Y}(n)\|^2$

$$= E\{\bar{\mu}(n)\,|D_a(n)|^2 + \mu^2\,\|\underline{Y}(n)\|^2\,|g^*(D(n))|^2 - 2\mu\,\text{Re}\{g^*(D(n))\,D_a^*(n)\}\}$$

So we have

$$\boxed{E\left\{\mu \, \|\underline{Y}(n)\|^2 \, |g^*(D(n))|^2\right\} = 2\,\mathrm{Re}\left\{E\left(D_a^*(n)\, g^*(D(n))\right)\right\}}$$

$\Updownarrow$

Fundamental variance relation

will use this to compute Excess MSE

Need some assumptions on model (to proceed further)

1) There exist a (constant) vector $\underline{w}_o$ such that

$$X(n) = \underline{w}_o^* \underline{Y}(n) + V(n)$$

$V(n)$: additive noise

(We encountered this type of model in system identification problem)

2) The noise sequence $\{V(n)\}$ is i.i.d. (independent & identically distributed) with zero-mean and variance $\sigma_V^2 = E\{|V(n)|^2\}$

3) The noise sequence $\{V(n)\}$ is independent of $\{\underline{Y}(k)\}$ for all n, k

4) The weight $\underline{w}_o$ is independent of $\{\underline{Y}(n), V(n)\}$

5) $X(n)$, $\underline{Y}(n)$ are zero mean

$$E\{|X(n)|^2\} = \sigma_x^2$$

$$E\{\underline{Y}(n)\,\underline{Y}^*(n)\} = R_Y$$

This is a stationary model.

(Later we will consider non-stationary models)

<u>Some Consequences of our model</u>

$$\underline{w}(n+1) = \underline{w}(n) + \mu\,\underline{Y}(n)\,g(D(n))$$

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

1. $\underline{V}(n+1)$ is independent of $\underline{w}(k)$ for all $k \le n+1$

(Since $\underline{w}(n+1)$ depends on $\{X(n), \dots, X(0), Y(n), \dots, Y(0)\}$ $\Rightarrow$ indep. of $V(n+1)$)

2. $V(n+1)$ is independent of $\tilde{w}(k)$ for all $k \le n+1$

3) $Y(n)$ is independent of $\tilde{\underline{w}}^*(n)\,\underline{Y}(n) = D_a(n)$ apriori estimation error.

4) (Recall system identification problem)

$$\underline{w}_{opt} = \underline{w}_0$$

Optimal estimator

$$\hat{X}(n)\Big|_{LMMSE} = \underline{w}_0^* \underline{Y}(n).$$

$$\text{and error} = X(n) - \underline{w}_0^* (\underline{Y}(n))$$

$$= V(n)$$

$$J_{min} = E|V(n)|^2 = \sigma_v^2$$

5) $$D(n) = D_a(n) + V(n)$$

$$D(n) = X(n) - \underline{W}^*(n)\,\underline{Y}(n)$$

$$= X(n) + (\tilde{\underline{W}}(n) - \underline{w}_{opt})^* \underline{Y}(n)$$

$$= \underbrace{X(n) - \underline{w}_{opt}^* \underline{Y}(n)} + \underbrace{\tilde{\underline{w}}^*(n)\,\underline{Y}(n)}$$

$$= V(n) + D_a(n).$$

6) Since $V(n)$ is independent of $D_a(n)$

$$\underbrace{E\{|D(n)|^2\}}_{\text{MSE}} = \underbrace{E\{|D_a(n)|^2\}}_{\text{EMSE}} + \underbrace{E\{|V(n)|^2\}}_{J_{min}}$$

x ———————————— x

Characterization of Excess MSE for LMS

LMS update $\rightarrow$ $$\underline{w}(n+1) = \underline{w}(n) + \mu\{\underline{Y}(n)\, D^*(n)\}$$

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

$$g(D(n)) = D^*(n)$$

Recall Variance relation (which holds true in steady state)

$$① \quad \mu\, E\{\|\underline{Y}(n)\|^2\, |g^*(D(n))|^2\} = 2\,\mathrm{Re}\left[E\{D_a^*(n)\, g^*(D(n))\}\right]$$

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

$$\underline{\tilde{w}}(n) = \underline{w}_{opt} - \underline{w}(n)$$

From our model,

$$D(n) = D_a(n) + V(n)$$

Also $D_a(n)$ is ind. of $V(n)$

$$\underbrace{E\{|D(n)|^2\}}_{MSE} = E\{|D_a(n)|^2\} + \underbrace{E\{|V(n)|^2\}}_{J_{min}}$$

Recall, excess MSE (denoted by η^{LMS})

is defined as

$$\eta^{LMS} = \lim_{n\to\infty} E\,|D(n)|^2 - J_{min}$$

$$= \lim_{n\to\infty} E\,|D_a(n)|^2$$

Term in LHS of (1)

$g^*(D(n)) = D(n)$ for LMS

$$\mu\, E\left\{ \|\underline{Y}(n)\|^2 \, |g^*(D(n))|^2 \right\}$$

$$= \mu\, E\left\{ \|\underline{Y}(n)\|^2 \, |D(n)|^2 \right\}$$

$$= \mu\, E\left\{ \|\underline{Y}(n)\|^2 \cdot |D_a(n) + V(n)|^2 \right\}$$

($V(n)$ is indep. of $\underline{Y}(n)$, $D_a(n)$

$$= \mu\, E\left\{ \|\underline{Y}(n)\|^2 \, |D_a(n)|^2 \right\}$$

$$+ \mu\, \underbrace{E\left\{ \|\underline{Y}(n)\|^2 \right\}}_{\text{trace}(R_Y)} \, \underbrace{E\left\{ |V(n)|^2 \right\}}_{\sigma_v^2}$$

$$= \mu\, E\left\{ \|\underline{Y}(n)\|^2 \, |D_a(n)|^2 \right\} + \mu\, \sigma_v^2 \, \text{trace}(R_Y)$$

Term in RHS of 1

$$E\left\{ D_a^*(n)\, g^*(D(n)) \right\} = E\left\{ D_a^*(n)\, D(n) \right\}$$

$$= E\left\{ D_a^*(n) \left(D_a(n) + V(n) \right) \right\}$$

$$= E\left\{ |D_a(n)|^2 \right\} + \underbrace{E\left\{ D_a^*(n)\, V(n) \right\}}_{0}$$

$$= E\left\{ |D_a(n)|^2 \right\}$$

$\Rightarrow$ RHS of 1 is $2E|D_a(n)|^2$

So we have

$$\mu E\left\{ \|\underline{Y}(n)\|^2 |D_a(n)|^2 \right\} + \mu \sigma_v^2 \operatorname{trace}(R_Y) = 2E|D_a(n)|^2$$

Taking limit as $n \to \infty$,

$$\lim_{n\to\infty} \mu E\left\{ \|\underline{Y}(n)\|^2 |D_a(n)|^2 \right\} + \mu \sigma_v^2 \operatorname{trace}(R_Y) = 2 \underbrace{\lim_{n\to\infty} E|D_a(n)|^2}_{\eta^{LMS}} \quad (2)$$

Still need to compute $\lim_{n\to\infty} E\left\{ \|\underline{Y}(n)\|^2 |D_a(n)|^2 \right\}$

. We need further assumptions to proceed

- Will consider two scenarios

Scenario 1 (Small step size)

When step size is small,

$|D_a(n)|^2$ is expected to be small

then $E\left\{ \|\underline{Y}(n)\|^2 \, |D_a(n)|^2 \right\}$ is small

compared to $\sigma_v^2 \, \mathrm{trace}\,(R_Y)$

Then $$\boxed{\eta^{LMS} \approx \frac{\mu \sigma_v^2}{2} \, \mathrm{trace}\,(R_Y)}$$

for sufficiently small step size

- η^{LMS} is proportional to

$\mu, \; \sigma_v^2, \; \mathrm{trace}\,(R_Y)$

- Intuitively satisfying

Scenario 2 (Seperation principle)

Assume that $\|\underline{Y}(n)\|^2$ & $D_a(n)$

are independent at steady state

(Some simulation results are given in book to justify this)

Seems to hold true for larger range of μ than Scenario 1

In this case

$$\lim_{n\to\infty} E\left\{\|\underline{Y}(n)\|^2 |D_a(n)|^2\right\}$$

$$= \lim_{n\to\infty} E\left\{\|\underline{Y}(n)\|^2\right\} E\left\{|D_a(n)|^2\right\}$$

$$= \text{trace}(R_Y) \cdot \eta^{LMS}$$

substituting this in (2)

$$\boxed{\eta^{LMS} = \frac{\mu \sigma_v^2 \,\text{trace}(R_Y)}{2 - \mu \,\text{trace}(R_Y)}}$$

Performance of ϵ- NLMS

$$\underline{w}(n+1) = \underline{w}(n) + \mu \, \underline{Y}(n) \frac{D^{*}(n)}{\|\underline{Y}(n)\|^2 + \epsilon}$$

$$g(D(n)) = \frac{D^{*}(n)}{\|\underline{Y}(n)\|^2 + \epsilon}$$

Proceeding from Variance relation (similar to LMS)

& using separation principle (Scenario 2)

$\|\underline{Y}(n)\|^2$ is indep. of $D_a(n)$ as $n \to \infty$

we get

$$\eta^{N}_{\epsilon\text{-LMS}} = \frac{\mu \sigma_v^2 \alpha_Y}{2\beta_Y - \mu \alpha_Y}$$

$$\alpha_Y = E\left\{ \frac{\|\underline{Y}(n)\|^2}{\left(\epsilon + \|\underline{Y}(n)\|^2\right)^2} \right\}$$

$$\beta_Y = E\left\{ \frac{1}{\|\underline{Y}(n)\|^2} \right\}$$

Suppose ϵ is very small, $\alpha_y \approx \beta_y$

$$\Rightarrow \boxed{\eta^{\varepsilon\text{-NLMS}} = \frac{\mu \sigma_v^2}{2-\mu}}$$

- This EMSE expression does not depend on statistics of $\underline{Y}(n)$
- ε - NLMS is less sensitive to the statistics of input $\underline{Y}(n)$ compared to LMS.

* ——— *

13/10

Tracking Performance of adaptive filters

- Now, we consider non-stationary environments
- R_Y, R_{YX} changes with time (n)
- In this case, will the adaptive filter track be able to track the optimal filter? If so, under what conditions?

- Suppose $\sigma_X^2(n)$, $R_Y(n)$, $R_{XY}(n)$ depend on time 'n'.

Optimal filter

$$\underline{w}_{opt}(n) = R_Y^{-1}(n) \, R_{YX}(n)$$

$$J_{min}(n) = \sigma_X^2(n) - R_{YX}^*(n) \, R_Y^{-1}(n) R_{YX}(n)$$

Consider LMS update:

$$\underline{w}(n+1) = \underline{w}(n) + \mu \, \underline{Y}(n) \, D^*(n)$$

where $D(n) = X(n) - \underline{w}^*(n) \underline{Y}(n)$

Will adaptive filter $\underline{w}(n)$ be able to "closely" follow the optimal filter $\underline{w}_{opt}(n)$?

Model (time-varying system)

$$X(n) = \underline{w}_0^*(n) \, \underline{Y}(n) + V(n)$$

$\underline{w}_0(n) \rightarrow$ time varying filter

$V(n) \rightarrow$ i.i.d. noise with variance σ_V^2

$V(n) \rightarrow$ is independent of $\underline{Y}(k)$, $\underline{w}_0(k)$ for all n, k.

$X, \underline{Y}, V$ are zero mean

$E\{\underline{Y}(n) \underline{Y}^*(n)\} = R_Y$ (doesn't change with time)

$$\underline{w}_0(n) = \underline{w}_0(n-1) + \underline{q}(n) \qquad n \geq 1$$

with $\underline{w}_0 = \underline{w}_0(0)$ being the initial condition.

$\underline{q}(n) \rightarrow$ is zero mean with covariance Q.

$\rightarrow$ $q(n)$ is i.i.d,

$\underline{q}(n)$ is independent of $\underline{Y}(k)$, $V(k)$ and $\underline{w}_0$

<u>Goal:</u> Study if & when the adaptive filter (LMS) be able to track the time varying system $\underline{w}_0(n)$

Consider adaptive filter update equation:

$$\underline{w}(n) = \underline{w}(n-1) + \mu\, \underline{Y}(n)\, g\left(D(n)\right)$$

$$\text{where} \quad D(n) = x(n) - \underline{w}^{*}(n-1)\, \underline{Y}(n)$$

Note that $\underline{w}_0(n) \rightarrow$ true system

$\underline{w}(n) \rightarrow$ adaptive system which is trying to track the true system.

$$D(n) = X(n) - \underline{w}^{*}(n-1)\, \underline{Y}(n)$$

$$= \left(\underline{w}_0^*(n)\,\underline{Y}(n) + V(n)\right) - \underline{w}^*(n-1)\,\underline{Y}(n)$$

$$= \underbrace{\left(\underline{w}_0^*(n) - \underline{w}^*(n-1)\right)\underline{Y}(n)}_{D_a(n)} + V(n)$$

$$D(n) = D_a(n) + V(n).$$

From the assumptions in our model

$D_a(n)$ & $V(n)$ are independent (as before)

(Since $V(n)$ is indep. of $\underline{Y}(n)$, $\underline{w}_0(n)$, $\underline{w}(n-1)$)

$$E\,|D(n)|^2 = E\,|D_a(n)|^2 + \sigma_V^2$$

Verify yourself for our model $J_{min}(n) = \sigma_V^2$ $\quad \underline{w}_{opt}(n) = \underline{w}_0(n)$

Excess MSE $\quad \eta = \lim_{n\to\infty} E\,|D_a(n)|^2$

Lower bound on EMSE.

$$D_a(n) = \left(\underline{w}_0(n) - \underline{w}(n-1)\right)^*\underline{Y}(n)$$

$$= \left(\underline{w}_0(n-1) + \underline{q}(n) - \underline{w}(n-1)\right)^*\underline{Y}(n)$$

$$= \underbrace{\left(\underline{w}_0(n-1) - \underline{w}(n-1)\right)}_{\underline{\tilde{w}}(n-1)}^{*} \underline{Y}(n) + \underline{q}^{*}(n)\, \underline{Y}(n).$$

$$= \underline{\tilde{w}}^{*}(n-1)\, \underline{Y}(n) + \underline{q}^{*}(n)\, \underline{Y}(n)$$

Since $\underline{q}(n)$ is independent of $\underline{\tilde{w}}(n-1)$, $\underline{Y}(n)$

we have (verify)

$$E\,|D_a(n)|^2 = E\left\{ |\underline{\tilde{w}}^{*}(n-1)\, \underline{Y}(n)|^2 \right\} + E\left\{ |\underline{q}^{*}(n)\, \underline{Y}(n)|^2 \right\}$$

$$\geq E\left\{ |\underline{q}^{*}(n)\, \underline{Y}(n)|^2 \right\}$$

$\Downarrow$

lower bound on EMSE

$$E\,|\underline{q}^{*}(n)\, \underline{Y}(n)|^2 = E\left\{ \left(\underline{q}(n)\, \underline{Y}(n)\right)^{*} \left(\underline{q}^{*}(n)\, \underline{Y}(n)\right) \right\}$$

$$= \mathrm{trace}\left(E\left\{ \underline{q}^{*}(n)\, \underline{Y}(n) \left(\underline{q}(n)\, \underline{Y}(n)\right)^{*} \right\} \right)$$

$E(X) = E(E(X|Y))$

$$= \mathrm{trace}\; E\left\{ E\left(\underline{q}^{*}(n)\, \underline{Y}(n)\, Y^{*}(n)\, \underline{q}(n) \,\middle|\, \underline{q}(n) \right) \right\}$$

$$= \mathrm{trace}\; E\left\{ \underline{q}^{*}(n)\, \underbrace{E\left(\underline{Y}(n)\, \underline{Y}^{*}(n) \,\middle|\, \underline{q}(n)\right)}_{R_Y}\, \underline{q}(n) \right\}$$

$$= \operatorname{trace}\left(E\left\{ \underline{q}^{*}(n)\, R_Y\, \underline{q}(n) \right\} \right)$$

$\operatorname{trace}(ABC)$
$= \operatorname{trace}(CAB)$
$= \operatorname{trace}(BCA)$

$$= E\left(\operatorname{trace}\left(\underline{q}^{*}(n)\, R_Y\, \underline{q}(n) \right) \right)$$

$$= E\left(\operatorname{trace}\left(R_Y\, \underline{q}(n)\, \underline{q}^{*}(n) \right) \right)$$

$$= \operatorname{trace}\left\{ E\left(R_Y\, \underline{q}(n)\, \underline{q}^{*}(n) \right) \right\}$$

$$= \operatorname{trace}(R_Y Q)$$

So $\quad E\left(D_a(n)\right)^2 \geq \operatorname{trace}(R_Y Q)$

Degree of Non-stationarity (DN)

$$DN \triangleq \sqrt{\frac{\operatorname{trace}(R_Y Q)}{J_{min}}}$$

$$= \sqrt{\frac{\operatorname{trace}(R_Y Q)}{\sigma_v^2}}$$

$DN \gg 1 \Rightarrow$ Statistics are changing too fast
Adaptive filter will not be able to track optimal filter

$DN \ll 1 \Rightarrow$ Adaptive filter will be able to track optimal filter.

14/10

Computation of Excess MSE

* Energy Conservation Relation

update: $\underline{w}(n) = \underline{w}(n-1) + \mu\, \underline{Y}(n)\, g\big(D(n)\big)$

where $D(n) = x(n) - \underline{w}^*(n-1)\,\underline{Y}(n)$

Now,

$$\underline{w}_0(n) - \underline{w}(n) = \underline{w}_0(n) - \underline{w}(n-1) - \mu\, \underline{Y}(n)\, g\big(D(n)\big)$$

$(\cdot)^*\, \underline{Y}(n)$

$$\underbrace{\big(\underline{w}_0(n) - \underline{w}(n)\big)^* \underline{Y}(n)}_{D_p(n)} = \overbrace{\big(\underline{w}_0(n) - \underline{w}(n-1)\big)^* \underline{Y}(n)}^{D_a(n)} - \mu\, g^*\big(D(n)\big) \underbrace{\underline{Y}^*(n)\,\underline{Y}(n)}_{\|\underline{Y}(n)\|^2}$$

$$D_p(n) = D_a(n) - \mu\, \|\underline{Y}(n)\|^2\, g^*\big(D(n)\big)$$

Proceeding same way as before

$$\|\underline{w}_0(n) - \underline{w}(n)\|^2 + \bar{\mu}(n)\, |D_a(n)|^2$$

$$= \|\underline{w}_0(n) - \underline{w}(n-1)\|^2 + \bar{\mu}(n)\, |D_p(n)|^2$$

where $\bar{\mu}(n) = \begin{cases} 0 & \text{if } \|\underline{Y}(n)\| = 0 \\ 1/\|\underline{Y}(n)\|^2 & \text{else} \end{cases}$

* Variance Relation.

Our model $\underline{w}_o(n) = \underline{w}_0(n-1) + \underline{q}(n)$

$\underline{q}(n)$ is i.i.d with covariance Q

$\underline{q}(n)$ is independent of $\underline{Y}(k)$, $X(k)$, $\underline{w}_0(0)$ for all k

Consider $E\|\underline{w}_o(n) - \underline{w}(n-1)\|^2$

$$= E\|\underline{w}_0(n-1) + \underline{q}(n) - \underline{w}(n-1)\|^2$$

$$= E\|\underbrace{\underline{w}_0(n-1) - \underline{w}(n-1)}_{\underline{\tilde{w}}(n-1)} + \underline{q}(n)\|^2$$

based on assumptions on $\underline{q}(n)$ $\rightarrow$

$$= E\|\underline{\tilde{w}}(n-1)\|^2 + \underbrace{E\|\underline{q}(n)\|^2}_{\text{trace}(Q)}$$

Substituting this in energy conservation relation,

$$E\|\tilde{w}(n)\|^2 + E\{\bar{\mu}(n)|D_a(n)|^2\}$$
$$= E\|\tilde{w}(n-1)\|^2 + \underbrace{tr(Q)}_{\text{only extra term for our non-stationary model.}} + E\{\bar{\mu}(n)|D_p(n)|^2\}$$

In steady state, as $n \to \infty$

$$E\|\tilde{w}(n)\|^2 = E\{\|\tilde{w}(n-1)\|^2\}$$

Assuming steady state,

$$E\{\bar{\mu}(n)|D_a(n)|^2\} = E\{\bar{\mu}(n)|D_p(n)|^2\} + tr(Q)$$

Recall $D_p(n) = D_a(n) - \mu\|\underline{Y}(n)\|^2 g^*(D(n))$

Substituting & simplifying

$$\mu E\{\|\underline{Y}(n)\|^2 |g(D(n))|^2\} + tr(Q)$$

$$= 2\,Re\{E\{D_a^*(n)\, g(D(n))\}\}$$

Tracking Performance of LMS

EMSE $= \lim_{n\to\infty} E|D_a(n)|^2$

$$g(D(n)) = D^*(n)$$

$$EMSE = \eta^{LMS} = \frac{1}{2}\Big[\mu E\{\|\underline{Y}(n)\|^2 |D_a(n)|^2\} + \mu\sigma_v^2 tr(R_Y) + \frac{1}{\mu} tr(Q)\Big]$$

(As before)

Approximation 1 (small step size)

neglect the first term

$$\boxed{\eta^{LMS} = \frac{\mu \sigma_v^2 \, tr(R_Y) + \frac{1}{\mu} tr(Q)}{2}}$$

Choice of μ:

As μ increases term 1 increases

but term 2 decreases

Take derivative of η^{LMS} wrt μ &

equate to zero

$$\mu_{opt} = \sqrt{\frac{tr(Q)}{\sigma_v^2 \, tr(R_Y)}}$$

Corresponding EMSE $\eta^{LMS}_{opt} = \sqrt{\frac{tr(Q)}{\sigma_v^2 \, tr(R_Y)}}$

Approximation 2 (Separation principle)

$\|\underline{Y}(n)\|^2$ is independent of $|D_a(n)|^2$

$$\eta^{LMS} = \frac{\mu \sigma_v^2 \, tr(R_Y) + \frac{1}{\mu} tr(Q)}{2 - \mu \, tr(R_Y)}$$

Optimal μ to minimize η^{LMS} is

$$\mu_{opt} = \sqrt{\frac{tr(Q)}{\sigma_v^2 \, tr(R_Y)} + \left(\frac{tr(Q)}{2\sigma_v^2}\right)^2}$$

$$- \frac{tr(Q)}{2\sigma_v^2}$$

Tracking performance of other algorithms

ε-NLMS

$$\eta^{\varepsilon\text{-NLMS}} \approx \frac{\mu \sigma_v^2}{2-\mu} + \frac{1}{2-\mu} \frac{tr(Q)}{E\left\{\frac{1}{\|\underline{Y}(n)\|^2}\right\}}$$

(using separation principle)

RLS

$$\underline{w}(n) = \underline{w}(n-1) + P_n \underline{Y}(n) D^*(n)$$

$$P_n = \frac{1}{\lambda}\left[P_{n-1} - \frac{\lambda^{-1} P_{n-1} \underline{Y}_n \underline{Y}_n^* P_{n-1}}{1+\lambda^{-1} \underline{Y}_n^* P_{n-1} \underline{Y}_n}\right]$$

$$P_{-1} = \frac{I}{\varepsilon}, \quad 0 << \lambda < 1$$

Using our non-stationary model and some approximations

$$\eta^{RLS} = \frac{\sigma_v^2 (1-\lambda) M + \frac{1}{1-\lambda} tr(Q R_Y)}{2-(1-\lambda)M}$$

$M \rightarrow$ Size of $\underline{Y}(n)$

Find optimal value of λ

$$\lambda_{opt}^{RLS} = 1 - \frac{1}{\sigma_v}\sqrt{\frac{tr(Q R_Y)}{M}}$$

Corresponding EMSE

$$\eta_{opt}^{RLS} = \sigma_v \sqrt{M\, tr(Q R_Y)}$$

* ______ *

15/10

Transcient Analysis of adaptive filters.

$X(n)$, $\underline{Y}(n)$ are zero-mean stationary (wide sense)

$E|X(n)|^2 = \sigma_x^2$ $\quad E\{\underline{Y}(n)\underline{Y}^\dagger(n)\} = R_Y$;

$E\{\underline{Y}(n)X^*(n)\} = R_{YX}$

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

Focus on LMS

$$\underline{w}(n+1) = \underline{w}(n) + \mu\, \underline{Y}(n)\left[X(n) - \underline{w}^\dagger(n)\underline{Y}(n)\right]^*$$

Note that $\{\underline{w}(n)\}$ is a sequence of random vectors.

Define $\underline{\tilde{w}}(n) = \underline{w}_{opt} - \underline{w}(n)$

error

Will study how

(a) $E\{\underline{\tilde{w}}(n)\}$ behaves with n

(Mean convergence)

(b) $E\{\|\underline{\tilde{w}}(n)\|^2\}$ behaves with n

(Mean square convergence)

Mean Convergence. (of LMS)

Consider

$$\underline{w}_{opt} - \underline{w}(n+1) = \underline{w}_{opt} - \underline{w}(n) - \mu\,\underline{Y}(n)\,x^*(n) + \mu\,\underline{Y}(n)\,\underline{Y}^*(n)\,\underline{w}(n)$$

$$\underline{\tilde{w}}(n+1) = \underline{\tilde{w}}(n) - \mu\,\underline{Y}(n)\,x^*(n) + \mu\,\underline{Y}(n)\,\underline{Y}^*(n)\big(\underbrace{\underline{w}(n) - \underline{w}_{opt}}_{-\underline{\tilde{w}}(n)} + \underline{w}_{opt}\big)$$

$$= \left(I - \mu\,\underline{Y}(n)\underline{Y}^*(n)\right)\underline{\tilde{w}}(n) - \mu\,\underline{Y}(n)\,x^*(n) + \mu\,\underline{Y}(n)\,\underline{Y}^*(n)\,\underline{w}_{opt}$$

$$\underline{\tilde{w}}(n+1) = \left(I - \mu\,\underline{Y}(n)\,\underline{Y}^*(n)\right)\underline{\tilde{w}}(n) - \mu\,\underline{Y}(n)\left[x(n) - \underline{w}^*_{opt}\,\underline{Y}(n)\right]^*$$

Taking expectation on both sides

$$E\{\tilde{\underline{w}}(n+1)\} = E\{(\underline{I} - \mu \underline{Y}(n)\underline{Y}^*(n))\,\tilde{\underline{w}}(n)\}$$
$$- \mu\, E\{\underline{Y}(n)\,[x(n) - \underline{w}_{opt}^*\underline{Y}(n)]^*\}$$

From orthogonality principle,

$(x(n) - \underline{w}_{opt}^*\underline{Y}(n)) \rightarrow$ error corresponding to optimal lmmse estimator

is orthogonal to any linear function of $\underline{Y}(n)$

So $E\{\underline{Y}(n)\,[x(n) - \underline{w}_{opt}^*\underline{Y}(n)]^*\} = 0$

Consider $E\{(\underline{I} - \mu\, \underline{Y}(n)\underline{Y}^*(n))\,\tilde{\underline{w}}(n)\}$

$\Downarrow$ not easy to evaluate.

Typically μ (step size is small)

$\tilde{\underline{w}}(n)$ changes very slowly compared to input $\underline{Y}(n)$

two time scales

$\underline{Y}(n) \rightarrow$ fast

$\tilde{\underline{w}}(n) \rightarrow$ slow

$$E\left\{ \left(\underline{I} - \mu \underline{Y}(n)\underline{Y}^*(n)\right) \underline{\tilde{w}}(n) \right\}$$

$$\approx E\left(\underline{I} - \mu \underline{Y}(n)\underline{Y}^*(n)\right) E\left(\underline{\tilde{w}}(n)\right)$$

Proceeding with this approximation

$$E\left\{\underline{\tilde{w}}(n+1)\right\} = E\left\{\underline{I} - \mu \underline{Y}(n)\underline{Y}^*(n)\right\} E\left\{\underline{\tilde{w}}(n)\right\}$$

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

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

$\left(\underline{I} - \mu R_Y\right)^{n+1}$ converges if

all the eigen values of $\underline{I} - \mu R_Y$

have absolute value < 1

Say $R_Y = Q \Lambda Q^* \rightarrow$ eigen decomposition

$$\underline{I} - \mu R_Y = Q\left(\underline{I} - \mu \Lambda\right) Q^*$$

eigen values of $\underline{I} - \mu R_Y$ = eigen values of $\underline{I} - \mu \Lambda$

So $|1 - \mu \lambda_i| < 1$ for all i

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

$$\Rightarrow \quad 0 < \mu < \frac{2}{\lambda_i} \quad \forall i$$

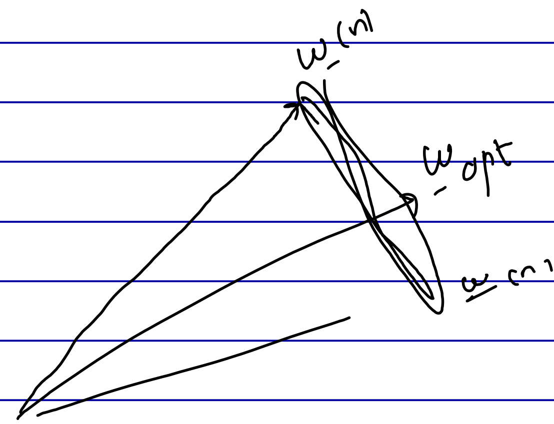
$$\Rightarrow \quad \boxed{0 < \mu < \frac{2}{\lambda_{max}}}$$

Same as steepest descent convergence.

Under this condition $E\{\underline{\tilde{w}}(n)\} \rightarrow \underline{0}$.

$$E\{\underline{w}_{opt} - \underline{w}(n)\} \rightarrow \underline{0}$$

$$\Rightarrow \quad \boxed{E\{\underline{w}(n)\} \rightarrow \underline{w}_{opt}}$$



Mean square Convergence of LMS

Update equation:

$$\underline{w}(n+1) = \underline{w}(n) + \mu \underline{Y}(n) [X^*(n) - \underline{Y}^*(n) \underline{w}(n)]$$

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

$$\underline{\tilde{w}}(n) = \underline{w}_{opt} - \underline{w}(n)$$

we have

$$\underline{\tilde{w}}(n+1) = [\underline{I} - \mu \underline{Y}(n) \underline{Y}^*(n)] \underline{\tilde{w}}(n)$$

obtained this in last lecture

$$- \mu \underline{Y}(n) \underbrace{[x(n) - \underline{w}^*_{opt} \underline{Y}(n)]^*}_{D_0(n)}$$

$D_0(n) \rightarrow$ estimation error corresponding to optimal filter

$$(1) — \underline{\tilde{w}}(n+1) = [\underline{I} - \mu \underline{Y}(n) \underline{Y}^*(n)] \underline{\tilde{w}}(n) - \mu \underline{Y}(n) D_0(n)$$

Again using small step size assumption

($\underline{\tilde{w}}(n)$ changes slow compared to $\underline{Y}(n)$)

we approximate (1) as

$$(2) — \underline{\tilde{w}}(n+1) = (\underline{I} - \mu R_Y) \underline{\tilde{w}}(n) - \mu \underline{Y}(n) D_0(n)$$

Will study the behaviour of

$$E\{\|\underline{\tilde{w}}(n)\|^2\} \quad \text{with } n.$$

Define the weight error covariance matrix

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

Clearly, $E\{\|\underline{\tilde{w}}(n)\|^2\} = tr(K(n))$

MSE at time n

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

Behaviour of $J(n)$ and $K(n)$ are closely related.

Will study them together.

Model Assumptions (Independence theory)

- Input vectors $\underline{Y}(1), \underline{Y}(2), \ldots, Y(n), \ldots$ is a sequence of i.i.d. random vectors
- $\underline{Y}(n)$ at time n is independent of past desired response $X(n-1), X(n-2), \ldots, X(1)$
- At time n, $X(n)$ depends on $\underline{Y}(n)$ but it is independent of $X(1), \ldots, X(n-1)$
- $X(n)$ and $\underline{Y}(n)$ are jointly Gaussian.

Model Implications.

- $\underline{w}(n+1)$ depends on $\{\underline{Y}(1), \ldots, \underline{Y}(n)\}$
  $\{x(1), \ldots, x(n)\}$ & $\underline{w}(0)$

- $\underline{w}(n+1)$ is independent of $\underline{Y}(n+1), x(n+1)$

- $\underline{\tilde{w}}(n+1)$ is independent of $\underline{Y}(n+1), x(n+1)$

Definition:

LMS is said to converge in mean square sense if $E\{\|\underline{\tilde{w}}(n)\|^2\} \longrightarrow$ constant as $n \to \infty$

Will see that this mean square convergence is equivalent to $J(n) \longrightarrow$ constant as $n \to \infty$.

Recursion for weight error vector covariance

$$\underline{\tilde{w}}(n+1) = (\underline{I} - \mu R_Y)\underline{\tilde{w}}(n) - \mu \underline{Y}(n) D_0^*(n)$$

$$K(n+1) = E\{\underline{\tilde{w}}(n+1)\, \underline{\tilde{w}}^*(n+1)\}$$

Consider:

$$E\{(\underline{I} - \mu R_Y)\underline{\tilde{w}}(n) \; [(\underline{I} - \mu R_Y)\underline{\tilde{w}}(n)]^*\}$$

$$= E\{(\underline{I} - \mu R_Y)\, \underline{\tilde{w}}(n)\, \underline{\tilde{w}}^*(n)\, (\underline{I} - \mu R_Y)^*\}$$

$$= (I - \mu R_Y)\ \underbrace{E\{\tilde{\underline{w}}(n)\,\tilde{\underline{w}}^*(n)\}}_{K(n)}\ (I - \mu R_Y)$$

$$= (I - \mu R_Y)\ K(n)\ (I - \mu R_Y)$$

Now, $E\{(\mu\, D_o^*(n)\, \underline{Y}(n))\ (\underline{Y}(n)^*\, D_o(n)\, \mu)\}$

$$= E\{\mu^2\ |D_o(n)|^2\ \underline{Y}(n)\,\underline{Y}^*(n)\}$$

using iterative expectation identity $E(x) = E(E(x|Y))$

$$= \mu^2\ E\{E\{|D_o(n)|^2\ \underline{Y}(n)\,\underline{Y}^*(n)\ |\ D_o(n)\}\}$$

$$= \mu^2\ E\{|D_o(n)|^2\ E\{\underline{Y}(n)\,\underline{Y}^*(n)\ |\ D_o(n)\}\}$$

From orthogonality principle $D_o(n)$ is uncorrelated with $\underline{Y}(n)$

Since $X(n)$ & $\underline{Y}(n)$ are jointly Gaussian

$D_o(n) = X(n) - \underline{w}_{opt}^*\ \underline{Y}(n)$

$D_o(n)$ & $\underline{Y}(n)$ are jointly Gaussian

$D_o(n)$ is independent of $\underline{Y}(n)$

$$= \mu^2\ E\{|D_o(n)|^2 \cdot \underbrace{E\{\underline{Y}(n)\,\underline{Y}^*(n)\}}_{R_Y}\}$$

$$= \mu^2\ R_Y\ J_{min}$$

Considering the cross terms,

$$E\left\{ (I - \mu R_Y)\, \underline{\tilde{w}}(n) \quad \mu\, \underline{Y}^{*}(n)\, D_0(n) \right\}$$

$\underline{\tilde{w}}(n)$ is independent of $\underline{Y}(n)$ & $D_0(n)$

$$= E\left\{ \left(I - \mu R_Y\, \tilde{w}(n) \right) \right\} \underbrace{E\left\{ \mu\, \underline{Y}^{*}(n)\, D_0(n) \right\}}_{=0}$$

$$= 0$$

Putting these together, we have

$$K(n+1) = (I - \mu R_Y)\, K(n)\, (I - \mu R_Y) + \mu^2 J_{min} R_Y. \quad \text{— (3)}$$

due to the presence of this term

$K(n)$ doesnot converge to 0.

Now consider MSE

$$J(n) = E\left\{ |D(n)|^2 \right\}$$

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

$$= X(n) - \left(\underline{w}_{opt} - \underline{\tilde{w}}(n) \right)^{*} \underline{Y}(n)$$

$$= x(n) - \underline{w}_{opt}^* \underline{Y}(n) + \underline{\tilde{w}}^*(n)\, \underline{Y}(n)$$

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

$$E\, |D(n)|^2 = E \Big\{ |D_o(n)|^2 + \underline{\tilde{w}}^*(n)\, \underline{Y}(n)\, \underline{Y}^*(n)\, \underline{\tilde{w}}(n)$$

$$+ D_o(n)\, \underline{Y}^*(n)\, \underline{\tilde{w}}(n)$$

$$+ D_o^*(n)\, \underline{\tilde{w}}^*(n)\, \underline{Y}(n) \Big\}$$

Since $\underline{\tilde{w}}(n)$ is indep. of $\underline{Y}(n)$ & $D_o(n)$

the cross terms vanish.

$$J(n) = E\{|D_o(n)|^2\} + E\{ \underline{\tilde{w}}^*(n)\, \underline{Y}(n)\, \underline{Y}^*(n)\, \underline{\tilde{w}}(n) \}$$

$$= J_{min} + E\, \text{trace}\{ \underline{\tilde{w}}^*(n)\, \underline{Y}(n)\, \underline{Y}^*(n)\, \underline{\tilde{w}}(n) \}$$

$$= J_{min} + E\, \text{trace}\{ \underline{Y}(n)\, \underline{Y}^*(n)\, \underline{\tilde{w}}(n)\, \underline{\tilde{w}}^*(n) \}$$

$$= J_{min} + \text{trace}\Big(E\{ \underline{Y}(n)\, \underline{Y}^*(n) \}\, E\{ \underline{\tilde{w}}(n)\, \underline{\tilde{w}}^*(n) \} \Big)$$

$$J(n) = J_{min} + \text{trace}\{ R_Y\, K(n) \}$$

We can define $J_{emse}(n) = J(n) - J_{min}$

$$= \text{tr}\{ R_Y\, K(n) \}$$

Also, $E\{\|\tilde{\underline{w}}(n)\|^2\} = tr\{K(n)\}$

Behaviour of $J(n)$ and $E\{\|\tilde{w}(n)\|^2\}$ can be studied using $K(n)$ behaviour.

Will work with change of coordinates.

Recall $\tilde{\underline{w}}(n+1) = (I - \mu R_Y)\tilde{\underline{w}}(n) - \mu D_o^*(n)\underline{Y}(n)$

Let $R_Y = Q \Lambda Q^*$

$\Lambda \rightarrow$ diagonal matrix containing the eigen values

$Q \rightarrow$ orthonormal matrix of eigen vector.

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

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

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

$$= Q(I - \mu\Lambda)Q^*$$

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

& $\underline{\phi}(n) = Q^*(-\mu D_o^*(n)\underline{Y}(n))$

$$Q^* \tilde{\underline{w}}(n+1) = Q^* Q (I - \mu\Lambda) Q^* \tilde{\underline{w}}(n)$$
$$Q^*(-\mu D_o^*(n)\underline{Y}(n))$$

Multiplying with Q^* on both sides.

$$\tilde{\underline{u}}(n+1) = (I - \mu\Lambda)\,\tilde{\underline{u}}(n) + \underline{\phi}(n) \longrightarrow$$

$$\mathrm{Cov}(\underline{\phi}(n)) = \mathrm{Cov}\left\{ Q^*(-\mu D_o^*(n)\,\underline{Y}(n)) \right\}$$

$\mathrm{cov}(A\underline{x})$

$= A\,\mathrm{Cov}(\underline{x})\,A^*$

$$= Q^*\,\underbrace{\mathrm{Cov}(-\mu D_o^*(n)\,\underline{Y}(n))}_{\mu^2 J_{min} R_Y}\,Q$$

$$= \mu^2 J_{min}\, Q^*(Q\Lambda Q^*)\,Q$$

$$= \mu^2 J_{min}\,\Lambda$$

Define $P(n) = E\{\tilde{\underline{u}}(n)\,\tilde{\underline{u}}^*(n)\}$

Note that

$$P(n) = \mathrm{cov}(\tilde{\underline{u}}(n))$$

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

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

$$= Q^*\,K(n)\,Q$$

Also, $E\|\tilde{\underline{u}}(n)\|^2 = \mathrm{tr}(P(n))$

$$= tr\left(Q^* K(n) Q\right)$$

$$= tr\left(K(n) \underbrace{Q Q^*}_{I}\right)$$

$$= tr\left(K(n)\right)$$

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

MSE at time n

$$J(n) = J_{min} + tr\left(R_Y K(n)\right)$$

$K(n) = Q P(n) Q^*$
$$= J_{min} + tr\left(R_Y Q P(n) Q^*\right)$$

$$= J_{min} + tr\left(\underbrace{Q^* R_Y Q}_{\Lambda} P(n)\right)$$

$$= J_{min} + tr\left(\Lambda P(n)\right)$$

$$= J_{min} + \sum_{k=1}^{M} \lambda_k \, p_k(n)$$

where $p_k(n)$ is the k^{th} diagonal entry of $P(n)$

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

Note that $p_k(n) = E|\tilde{u}_k(n)|^2$

Let us see how $\underline{\tilde{u}}(n)$ evolves with time n.

$$\underline{\tilde{u}}(n+1) = (I - \mu\Lambda)\,\underline{\tilde{u}}(n) + \underline{\phi}(n)$$

→ Looking at k^{th} coordinate

$$\tilde{u}_k(n+1) = (1-\mu\lambda_k)\,\tilde{u}_k(n) + \phi_k(n)$$

$\phi_k(n)$ is an i.i.d sequence with

Variance $\mu^2 J_{min}\lambda_k$

(Recall that $Cov(\underline{\phi}(n)) = \mu^2 J_{min}\Lambda$)

Solving the difference equation

$$\tilde{u}_k(n) = \underbrace{(1-\mu\lambda_k)^n\,\tilde{u}_k(0)}_{\text{Natural mode}} + \underbrace{\sum_{i=0}^{n-1}(1-\mu\lambda_k)^{n-1-i}\,\phi_k(i)}_{\text{forced mode}}$$

Using this solution, we have

Mean : $E\{\tilde{u}_k(n)\} = \tilde{u}_k(0).\,(1-\mu\lambda_k)^n$

Mean square · $E\{|\tilde{u}_k(n)|^2\} = \frac{\mu J_{min}}{2-\mu\lambda_k} + (1-\mu\lambda_k)^{2n}|\tilde{u}_k(0)|^2 - (1-\mu\lambda_k)^{2n}\frac{\mu J_{min}}{2-\mu\lambda_k}$

using the fact that $\phi_k(n)$ is zero mean i.i.d.

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

$$= \sum_{k=1}^{M} E|\tilde{u}_k(n)|^2$$

$$= \sum_{k=1}^{M} (\;\cdot\;)$$

And MSE $J(n) = J_{min} + tr(\Lambda P(n))$

$$= J_{min} + \sum_{k=1}^{n} \lambda_k E|\tilde{u}_k(n)|^2$$

$$J(n) = J_{min} + \frac{\mu J_{min}}{2}\sum_{k=1}^{M}\frac{\lambda_k}{2-\mu\lambda_k} + \sum_{k=1}^{M}\lambda_k\left(|\tilde{u}_k(0)|^2 - \frac{\mu J_{min}}{2-\mu\lambda_k}\right)(1-\mu\lambda_k)^{2n}$$

for small μ $\quad 2-\mu\lambda_k \approx 2$

$$J(n) = J_{min} + \frac{\mu J_{min}}{2} \sum_{k=1}^{M} \lambda_k + \sum_{k=1}^{M} \lambda_k \left(\|\tilde{u}_k(0)\|^2 - \frac{\mu J_{min}}{2} \right) (1-\mu\lambda_k)^{2n}$$

Assuming $|1-\mu\lambda_k| < 1$ for all k,

$$J(\infty) = J_{min} + \frac{\mu J_{min}}{2} \sum_{k=1}^{M} \lambda_k$$

$$\begin{aligned} EMSE &= \frac{\mu J_{min}}{2} \operatorname{tr}(\Lambda) \\ &= \frac{\mu J_{min}}{2} \operatorname{tr}(\Lambda Q^* Q) \\ &= \frac{\mu J_{min}}{2} \operatorname{tr}(Q \Lambda Q^*) \\ &= \frac{\mu J_{min}}{2} \operatorname{tr}(R_y) \end{aligned}$$

Same as what we got from steady state analysis.

x ———— x

24/10 Practical choice of Step Size μ

Recall we need $0 < \mu < \frac{2}{\lambda_{max}}$

for mean convergence of LMS.

$\lambda_{max} \rightarrow$ largest eigen value of R_Y

Note that the statistics R_Y is not known while running LMS.

We can do some conservative estimates of λ_{max} as follows

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

$$\lambda_{max} \leq \sum_{i=1}^{M} \lambda_i = tr(\Lambda)$$

$$= tr(R_Y)$$

$$tr(Q \Lambda Q^*)$$

$$= tr(\Lambda \underbrace{Q^* Q}_{I})$$

We can easily estimate $tr(R_Y)$

$\underline{Y}(n) \rightarrow$ Assume that entries in $\underline{Y}(n)$ have same variance σ_Y^2

$$E\{|Y_i(n)|^2\} = \sigma_Y^2 \quad \text{for } i = 1 \text{ to } M$$

$$tr(R_Y) = M\sigma_Y^2$$

$M \rightarrow$ is the number of taps in adaptive filter.

$\sigma_Y^2 \rightarrow$ estimated by averaging the mean squared value of input $\underline{Y}(n)$

We can choose step size as

$$0 < \mu < \frac{2}{M\sigma_Y^2}$$

Robustness of LMS

Consider system identification model

$$X(n) = \underline{w}_o^{*} \underline{Y}(n) + V(n)$$ → measurement noise

$\underline{w}_0$ → unknown true system

We want to adaptively identify the system

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

We update the weights by looking at the sequences $\{X(n)\}$, $\{\underline{Y}(n)\}$

Let weight vector $\underline{w}(n+1)$ be obtained from $\{X(n), \ldots, X(1)\}$, $\{\underline{Y}(n), \ldots, \underline{Y}(1)\}$

Let $f\left((X(n), \ldots, X(1)), (Y(n), \ldots, Y(1))\right)$ be a function which gives the weight vector $\underline{w}(n+1)$

Let μ be real positive number such that

$$0 < \mu \leq \frac{1}{\|\underline{Y}(n)\|^2}$$

Consider the apriori estimation error

$$e_a(n) = \underline{w}_o^* \underline{Y}(n) - \underline{w}^*(n)\underline{Y}(n)$$

$$\min_{f} \max_{\underline{w}(0),\, v(n)} \frac{|e_a(n)|^2}{\mu^{-1}\|\underline{w}(0) - \underline{w}_o\|^2 + \sum_{i=1}^{n} |v(i)|^2}$$

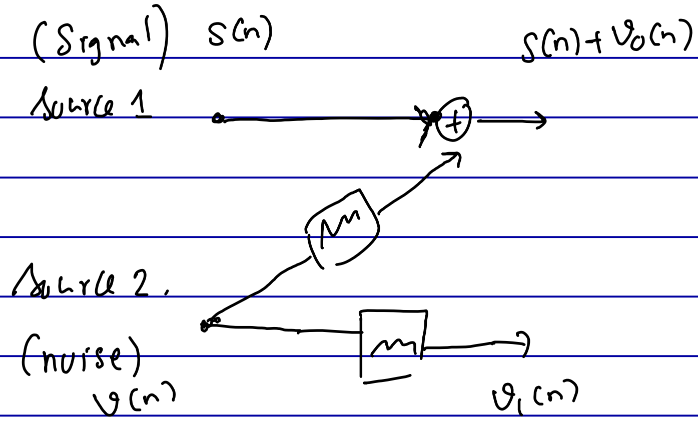
$\Downarrow$ comes from Robust Control.

$\Downarrow$ called $H(\infty)$ criterion.

$\rightarrow$ LMS is a solution for this minimax criterion

x ——— x

Adaptive noise cancellation (using LMS)



Signal s(n) and noise v(n) are uncorrelated.

$V_0(n)$ and $V_1(n)$ are some distorted version of $V(n)$

Goal is to remove the noise from

$s(n) + v_0(n)$

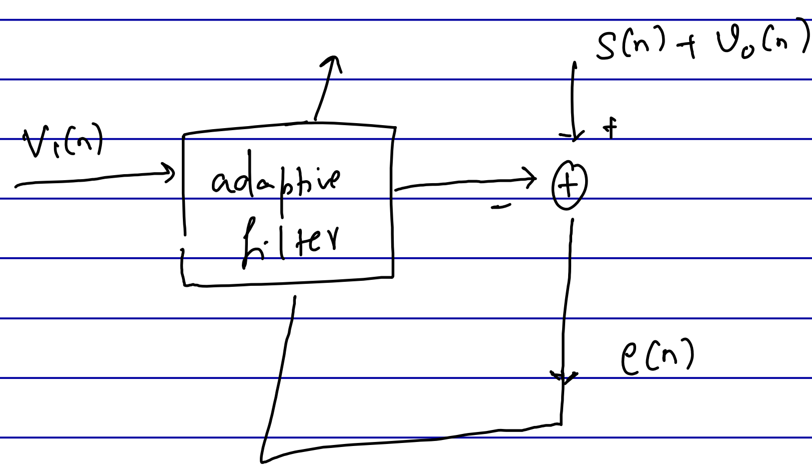
Take desired signal $x(n) = s(n) + v_0(n)$

input $y(n) = v_1(n)$

minimize $E\{|x(n) - \underline{w}^* y(n)|^2\}$

$$= E\{|s(n) + v_0(n) - \underline{w}^* v_1(n)|^2\}$$

$$= E\{|s(n)|^2\} + E\{|v_0(n) - w^* v_1(n)|^2\}$$



Once the noise $v_0(n)$ is cancelled,

the error $e(n)$ will have only the desired signal $s(n)$