8.6 Convergence Behaviour of RLS

- To date, we have examined the principles and organization of RLS. Now it's time to check its convergence.

- Test bed for convergence study:

$$d(n) = e_o(n) + \underline{w}_o^t \underline{u}(n) \qquad \text{and} \quad \lambda = 1$$

where
- measurement error $e_o(n)$ is zero mean, white, variance σ_e^2, independent of $\underline{u}(i)$ for all i
- parameter vector $\underline{w}_o$ is constant
- where needed, $\underline{u}(i)$ is independent, identically distributed sequence (iid), with zero mean and cov matrix R

- Quantities of interest:
  - mean value of estimate $\hat{\underline{w}}(n)$
  - covariance matrix of estimate $\hat{\underline{w}}(n)$
  - mean square value of $e(n)$, $J(n)$
  - mean square value of $\xi(n)$, $J'(n)$, to compare with LMS.

- With startup by Solution 2 (block solution at time $n \geq M$) or approximately by Solution 3 (soft constrained, $\Phi^{-1}(0) = \delta^{-1} I$), the RLS solution is the LS solution, just computed efficiently. Therefore these properties also apply to LS solutions generally, when applied to test bed.

- 8.6.1 The Solution Has the Right Mean
  - We know from Section 7.4 of notes, Section 11.8 of Haykin, that LS estimates are unbiased if $e_o(i)$ process has zero mean. Therefore

$$E_e\left[\hat{w}(n)\right] = \underline{w}_o$$

  - In order to obtain a useful relation en route, Haykin provides re-derivation

$$\hat{w}(n) = \Phi(n)^{-1} \underline{z}(n) = \Phi(n)^{-1} \sum_{i=1}^{n} \underline{u}(i) d^*(i)$$

$$\left[d(i) = e_o(i) + \underline{w}_o^+ \underline{u}(n) \right]$$

$$= \Phi(n)^{-1} \sum_{i=1}^{n} \underline{u}(i) \underline{u}^+(i) \underline{w}_o + \Phi(n)^{-1} \sum_{i=1}^{n} \underline{u}(i) e_o^*(i)$$

$$= \underline{w}_o + \Phi(n)^{-1} \sum_{i=1}^{n} \underline{u}(i) e_o^*(i)$$

Since measurement error is assumed independent of $\underline{u}(i)$ process, and zero mean,

$$E\left[\hat{\underline{w}}(n)\right] = \underline{w}_o \quad \text{for } n \geq M \text{ (i.e., after block startup)}$$

  - RLS (LS) is convergent in the mean for $n \geq M$, unlike LMS, where $E\left[\hat{w}(n)\right] \to \underline{w}_o$ for $n \to \infty$.

- 8.6.2 Covariance Matrix of the Solution $\underline{w}$
  - We know from Section 7.4 of notes, Section 11.8 of Haykin, that covariance matrix of LS estimates is $\sigma^2 \Phi^{-1}(n)$.
  - Haykin re-derives it as the start of a more useful conclusion. Define weight error vector

$$\underline{\epsilon}(n) = \underline{\hat{w}}(n) - \underline{w}_o = \Phi^{-1}(n) \sum_{i=1}^{n} \underline{u}(i)\, e_o^*(i) \qquad \text{(from 8.6.2)}$$

  - So cov matrix of the solution is

$$K(n) = E_{\underline{u}, e_o}\left[(\underline{\hat{w}}(n) - \underline{w}_o)(\underline{\hat{w}}(n) - \underline{w}_o)^+\right] = E_{\underline{u}, e_o}\left[\underline{\epsilon}(n)\,\underline{\epsilon}^+(n)\right]$$

$$= E_{\underline{u}}\left[\Phi^{-1}(n) \sum_{i=1}^{n}\sum_{j=1}^{n} \underline{u}(i)\, \underbrace{e_o^*(i)\, e_o(j)}_{\sigma^2 \delta_{ij}}\, \underline{u}^+(j)\, \Phi^{-1}(n)\right]$$

$$= \sigma^2 E_{\underline{u}}\left[\Phi^{-1}(n) \underbrace{\sum_{i=1}^{n} \underline{u}(i)\,\underline{u}^+(i)}_{\Phi(n)} \Phi^{-1}(n)\right] = \sigma^2 E_{\underline{u}}\left[\Phi^{-1}(n)\right]$$

  - Use of $\underline{u}(i)$ stats gives

$$K(n) = \frac{\sigma^2}{n-M-1} R^{-1} \qquad n > M+1$$

see Haykin 13.6

assuming $\underline{u}(i)$ iid, zero mean, cov R

  - This makes the m.s. misadjustment

$$E\left[\underline{\epsilon}^+(n)\,\underline{\epsilon}(n)\right] = \mathrm{tr}\left[K(n)\right] = \frac{\sigma^2}{n-M-1} \sum_{i=1}^{M} \frac{1}{\lambda_i}$$

(since eigenvals of R^{-1} are reciprocals of eigenvals λ_i of R)

Hence - decreases asymptotically as $1/n$

- small eigenvals λ_i of of R appear to cause problems
- converges in mean square, no jitter in $\underline{\hat{w}}(n)$
- excess mse ?

- A clearer way to view this result is to recall that the small eigenvalue directions are less coupled to the output, so the error in estimating $d(n)$ may not be excessive.
- As in LMS analysis, expand using eigenvector basis:

$$\underline{\varepsilon}(n) = Q\,\underline{v}(n) \qquad \underline{v}(n) = Q^{+}\underline{\varepsilon}(n)$$

$$E\left[\underline{v}(n)\,\underline{v}^{\dagger}(n)\right] = \frac{\sigma^2}{n-M-1} Q^{+} R^{-1} Q = \frac{\sigma^2}{n-M-1} \begin{bmatrix} \lambda_1^{-1} & & 0 \\ & \lambda_2^{-1} & \\ 0 & & \ddots \\ & & & \lambda_M^{-1} \end{bmatrix}$$

The standard deviation along each axis is ppl to $\frac{1}{\sqrt{\lambda_i}}$

This is exactly what we achieved in LMS <u>after</u> orthogonalization, so we expect RLS to be similarly insensitive to eigenvalue spreads

• 8.6.3 Mean Squared Estimation Error (Learning Curve)

- For many purposes, the real test of convergence is the estimation error. Use $\xi(n)$, as in LMS analysis. Since we initialize with $\hat{\underline{w}}(0) = \underline{0}$, the ms error $E[|\xi(1)|^2] = |d(1)|^2$ but should decrease from that value.

- The a priori estimation error at time n is

$$\xi(n) = d(n) - \hat{w}^{\dagger}(n-1)\,\underline{u}(n)$$

$$= e_o(n) + \underline{w}_o^{\dagger}\,\underline{u}(n) - \hat{\underline{w}}^{\dagger}(n-1)\,\underline{u}(n)$$

$$= e_o(n) - \underline{\epsilon}^{\dagger}(n-1)\,\underline{u}(n)$$

- The ms error is defined

$$J'(n) = E\left[|\xi(n)|^2\right]$$

although remember the cost function is $\sum_{i=1}^{n} |e(i)|^2$

$$= \overline{|e_o(n)|^2} - \overline{\underline{\epsilon}^{\dagger}(n-1)\,\underline{u}(n)\,e_o^*(n)} - \overline{e_o(n)\,\underline{u}^{\dagger}(n)\,\underline{\epsilon}(n-1)}$$

$$+ \overline{\underline{u}^{\dagger}(n)\,\underline{\epsilon}(n-1)\,\underline{\epsilon}^{\dagger}(n-1)\,\underline{u}(n)}$$

Terms 2 and 3 are zero because $e_o(n)$ is indep of $\underline{u}(n)$ and all past values (test bed definition) and zero mean.

So

$$J'(n) = \sigma^2 + \overline{\underline{u}^{\dagger}(n)\,\underline{\epsilon}(n-1)\,\underline{\epsilon}^{\dagger}(n-1)\,\underline{u}(n)}$$

Time to invoke the independence assumption: $\underline{\epsilon}(n-1)$ depends on $\underline{u}(i), d(i)$, $i<n$ and $\underline{u}(n)$ is assumed iid (nonsense, of course, since $\underline{u}(n)_k = \underline{u}(n-1)_{k-1}$, $k=2..M$, but we'll run with it). Therefore $\underline{\epsilon}(n-1)$ is indep of $\underline{u}(n)$.

$$J'(n) = \sigma^2 + E_{\underline{u}}\left[\underline{u}^{\dagger}(n)\,E_{\underline{\epsilon}}\left[\underline{\epsilon}(n-1)\,\underline{\epsilon}^{\dagger}(n-1)\right]\underline{u}(n)\right]$$

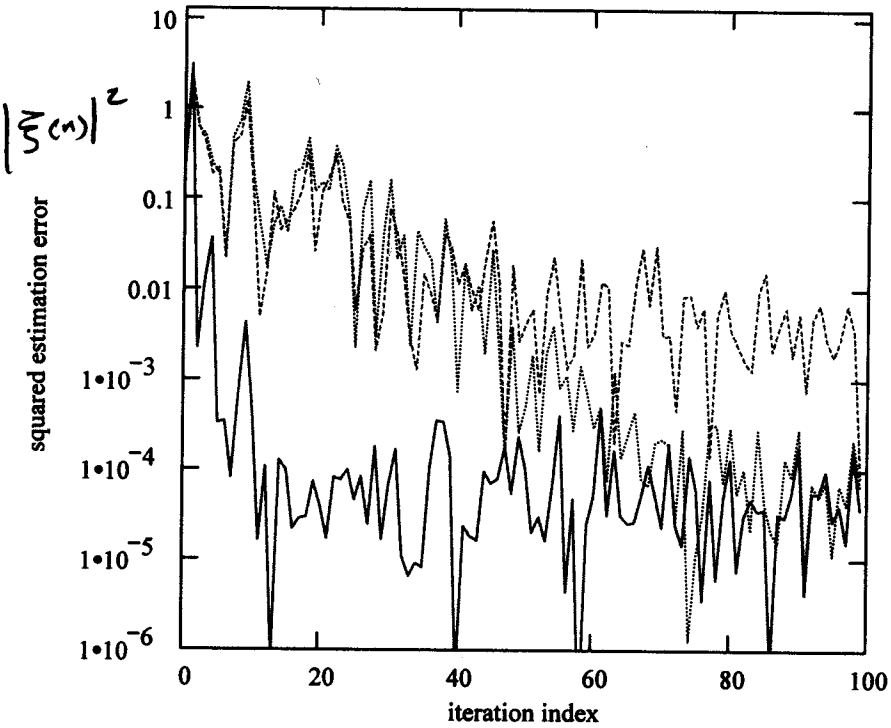
$$= \sigma^2 + \frac{\sigma^2}{n-M-1}\underbrace{E_{\underline{u}}\left[\underline{u}^{\dagger}(n)\,R^{-1}\,\underline{u}(n)\right]}_{= \mathrm{tr}\left[R^{-1}R\right]}$$

$$= \sigma^2 + \frac{\sigma^2 M}{n-M-1} = \sigma^2\left(\frac{n-1}{n-1-M}\right) \qquad n > M+1$$

No excess mse — $J'(n) \to \sigma^2 = J'_{min}$

No dependence on eigenvalue spread!

- Summary:
  - The ensemble mean of the solution vector is correct $E[\hat{w}(n)] = \underline{w}_o$ for all $n \geq M$
  - The jitter in weights $K(n) = E[\underline{\epsilon}(n)\underline{\epsilon}^t(n)]$ is shaped, so that less jitter in directions strongly coupled to the estimation error. Unlike standard LMS, but like orthogonalized LMS
  - The jitter in weights converges in mean square inversely with time to zero as $n \to \infty$. However, if $\lambda < 1$, then this isn't quite true (but we want some tracking ability).
  - The a priori mse $J'(n)$ convergence is insensitive to eigenvalue spread
  - By $n = 2M+1$, the ensemble averaged $J'(n)$ is only twice the minimum value $J'(2M+1) = 2\sigma^2 = 2J_{min}$ from a start value of $J'(1) = E[|d(1)|^2]$. $J'(\infty) = J_{min}$ for $\lambda = 1$.
  - Comparison with LMS?
    - if small eigenval ratio, then about the same speed
    - if large spread, RLS unchanged, LMS much slower
    - RLS achieves lower error than LMS, since no excess mse
    - LMS can track for nonzero μ; RLS needs $\lambda < 1$ to track, so they should be compared on basis of same tracking ability



RLS
decorr. LMS, alpha for 10% excess mse
LMS, mu for 10% excess mse

Estim'n Error: RLS, Decorr. LMS, LMS

$w = 0.45$

$M = 3$

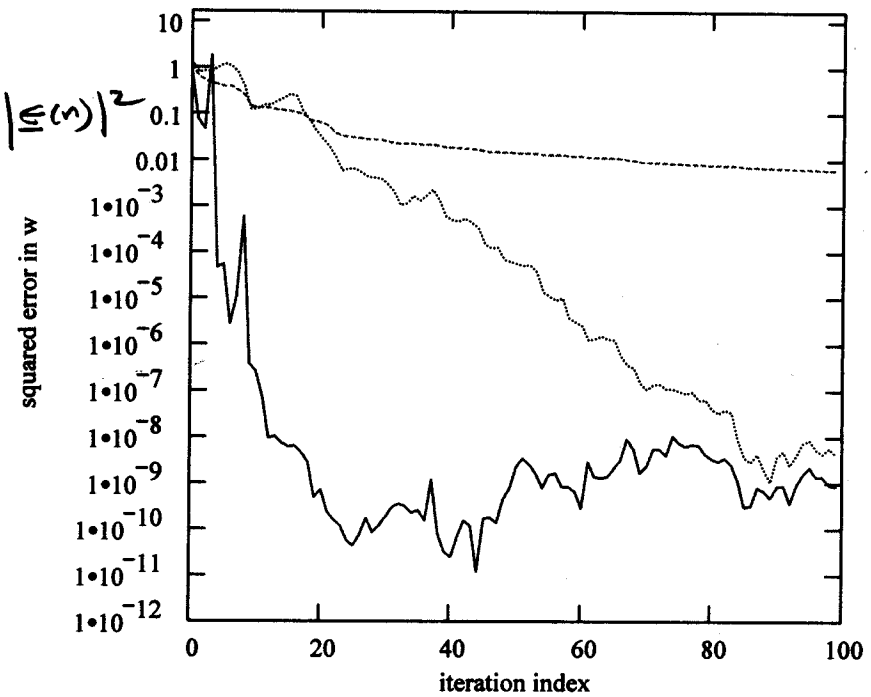
$\frac{\max(\lambda)}{\min(\lambda)} = 31.4$

$\delta = 0.063$

$J_{min} = 6.607 \cdot 10^{-5}$

$\lambda = 0.95$

$\gamma = 1 \cdot 10^{-3}$



RLS
decorr. LMS, alpha for 10% excess mse
LMS, mu for 10% excess mse

Weight Error: RLS, Decorr. LMS, LMS

- Here's a summary of our set of methods for adaptation

order \ time	block	recursive
block	• estimate $r_u(n)$, $r_{ud}(n)$; solve $R\underline{w} = \underline{p}$; mmse • LS; use $\underline{U}$, $\underline{d}$; solve $\Phi \underline{w} = \underline{z}$; minimize $\|\underline{e}\|^2$	• LMS; converges to mmse, but not a solution until convergence • RLS; a solution at every step; $\min \|\underline{e}\|^2$ converges to mmse
recursive	• block adaptive lattice (Burg criterion); solution for all orders	• gradient lattice; converges to mmse; not a solution en route • RLS lattice; a solution at every step; converges to mmse; solution all orders