
- 第二章 LMS自适应滤波

2.3 LMS算法稳定性分析

(Stability of LMS filter)

- 一 复习自相关阵及相关特性
- 二 LMS算法稳定性分析

LMS filter 的特点:

- 非线性;
- 反馈(feedback)

理论分
析困难

近似
分析

2.3 LMS算法稳定性分析

一 复习自相关阵及相关特性 ($x(n)$: 平稳信号)

1 自相关阵

$$\mathbf{R} = \mathbf{R}_{xx} = E[\mathbf{X}(n)\mathbf{X}^T(n)]$$
$$= \begin{bmatrix} r(0), & r(1), & \dots, & r(N-1) \\ r(1), & r(0), & \dots, & r(N-2) \\ \dots & \dots & \circ & \dots & \dots \\ r(N-1), & \dots, & \dots & r(0) \end{bmatrix}$$

$$r(p) = E[x(n)x(n-p)]$$

自相关阵特点:

1) \mathbf{R} 是对称阵, 即 $\mathbf{R}^T = \mathbf{R}$ (实数情况) 或 (复数情况) $\mathbf{R}^H = \mathbf{R}$ 且具有Toplitz性质

2) \mathbf{R} 是非负的

$$\tilde{y} = \mathbf{H}^T \mathbf{X} = \mathbf{X}^T \mathbf{H} \Rightarrow$$

$$E[|\tilde{y}|^2] = E[\mathbf{H}^T \mathbf{X} \mathbf{X}^T \mathbf{H}] = \mathbf{H}^T \mathbf{R}_{xx} \mathbf{H} \geq 0$$

3) 含噪信号的 \mathbf{R} 正定,
非奇异 \rightarrow \mathbf{R} 非奇异,
可逆

2.3 LMS算法稳定性分析

2 特征值和特征向量

定义：一个 $N \times N$ 阶方阵 \mathbf{R} ，它的特征矢量 \mathbf{v}_i （列矢量）定义为

$$\mathbf{R}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad 0 \leq i \leq N-1$$

称 λ_i 为 \mathbf{R} 的特征值（或特征根） λ_i 所对应的特征矢量

；
特征值 λ_i 是特征方程的解

$$\det[\mathbf{R} - \lambda \mathbf{I}_N] = 0$$

White noise

$\mathbf{R} = \text{diag}(\sigma^2, \sigma^2, \dots, \sigma^2)$, σ^2 - 方差

\mathbf{R} 有 N 个相同的特征值, 任意 $N \times 1$ 的列矢量均是其特征矢量(随机性)

Complex Sinusoidal

$$e^{j(\omega n + \varphi)} \Rightarrow r(p) = e^{j\omega p}$$

$$\mathbf{R} = \begin{bmatrix} 1 & e^{j\omega} & \dots & e^{j(N-1)\omega} \\ e^{-j\omega} & 1 & \dots & e^{j(N-2)\omega} \\ \dots & \dots & \dots & \dots \\ e^{-j(N-1)\omega} & e^{-j(N-2)\omega} & \dots & 1 \end{bmatrix}$$

\mathbf{R} 的特征矢量: $\mathbf{q} = [1, e^{-j\omega}, \dots, e^{-j(N-1)\omega}]^T$
(特征值为 N)

由信号 $e^{j\omega n}$ 组成的矢量相差一个共扼

2.3 LMS算法稳定性分析

3 矩阵的对角化

对于一个实对称阵（**R**阵），其不同特征值对应的特征矢量是正交的（欧氏空间内积为零）。 **$\mathbf{R}\mathbf{q}_j = \lambda_j \mathbf{q}_j$**

$$\|\mathbf{q}_i\| = 1, i = 0, 1, 2, \dots, N-1$$

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$$\mathbf{Q} = [\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{N-1}]$$

$$\mathbf{Q}^T \mathbf{R} \mathbf{Q} = \mathbf{\Lambda}' \quad \text{or} \quad \mathbf{R} = \mathbf{Q} \mathbf{\Lambda}' \mathbf{Q}^T$$

$$\mathbf{\Lambda}' = \text{diag}(\lambda_i) = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$$

$$\mathbf{Q} \mathbf{Q}^T = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}$$

$$\mathbf{q}_i^T \mathbf{q}_j = \frac{1}{\lambda_j} \mathbf{q}_i^T \mathbf{R} \mathbf{q}_j$$
$$\mathbf{R} \mathbf{q}_i = \lambda_i \mathbf{q}_i$$
$$\mathbf{q}_i^T \mathbf{q}_j = \frac{\lambda_i}{\lambda_j} \mathbf{q}_i^T \mathbf{q}_j$$

2.3 LMS算法稳定性分析

4 平稳随机过程自相关阵 \mathbf{R} 的特征值和特征向量的性质

1) $\det(\mathbf{R}) = \prod_{i=0}^{N-1} \lambda_i \longrightarrow \lambda_i \text{ 不等于 } 0 \text{ 是 } \mathbf{R} \text{ 可逆的充要条件}$

$$0 \leq i \leq N-1$$

2) $Tr(\mathbf{R}) = \sum_{i=0}^{N-1} \lambda_i = Nr(0) = N\sigma_x^2$

$$\lambda_i = \frac{\mathbf{q}_i^T \mathbf{R} \mathbf{q}_i}{\mathbf{q}_i^T \mathbf{q}_i}$$

3) 因为矩阵 \mathbf{R} 是非负的, 故有 $\lambda_i \geq 0, 0 \leq i \leq N-1$

4) 定义矩阵范数
矩阵的条件数 $\|A\| = (\lambda_{\max} \text{ of } A^T A)^{1/2}$
 $X(A) = \|A\| \|A^{-1}\|$

则有 $\|\mathbf{R}\| = \lambda_{\max}$ $\|\mathbf{R}^{-1}\| = 1/\lambda_{\min}$
 $X(\mathbf{R}) = \lambda_{\max}/\lambda_{\min} \longrightarrow \text{特征值的分散程度(大:病态)}$

5) KL变换 (Karhunen-Loeve展开)

当N个正交矢量被获得后, N维信号矢量可以表示成这些基矢量的线性组合, 这就是KL展开.

$$\mathbf{X}(n) = \sum_{i=0}^{N-1} c_i(n) \mathbf{q}_i \quad \mathbf{q}_i, i = 0, 1, \dots, N-1 \quad \text{归一化正交}$$

$$c_i(n) = \mathbf{q}_i^T \mathbf{X}(n), \quad i = 0, 1, \dots, N-1$$

$$\mathbf{R} = E[\mathbf{X}(n) \mathbf{X}^T(n)] = E\left[\sum_{i=0}^{N-1} c_i(n) \mathbf{q}_i \cdot \sum_{j=0}^{N-1} c_j(n) \mathbf{q}_j^T\right]$$

$$= E\left[\sum_{i,j=0}^{N-1} c_i(n) c_j(n) \mathbf{q}_i \mathbf{q}_j^T\right] = \sum_{i=0}^{N-1} E[c_i^2(n)] \mathbf{q}_i \mathbf{q}_i^T$$

$$\mathbf{R} = \mathbf{Q} \operatorname{diag}(\lambda_i) \mathbf{Q}^T = \sum_{i=0}^{N-1} \lambda_i \mathbf{q}_i \mathbf{q}_i^T$$

$$E[c_i^2(n)] = \lambda_i$$

特征值可以看成是信号矢量在特征矢量上投影的功率

$$J(\mathbf{H}) = E[e^2(n)] \Rightarrow J(n) = e^2(n)$$

二 LMS算法稳定性分析 ($\mathbf{x}(n)$ 平稳的情况下)

自适应步长 δ ,自相关矩阵 \mathbf{R} 起着决定性作用。

δ 的选择满足下面两种收敛要求:

- 1) 均值收敛: 指系数 $\mathbf{H}(n)$ 的均值收敛到维纳最优解 \mathbf{H}_{opt}
- 2) 均方收敛: 指均方误差 $J(n)$ 收敛到一个最小值

定义: 权系数误差矢量 $\mathbf{c}(n) = \mathbf{H}(n) - \mathbf{H}_{\text{opt}}$

$E[\mathbf{c}(n)]$ (一阶矩) 帮助我们分析LMS算法在均值意义下的收敛条件

$E[\mathbf{c}(n)\mathbf{c}^T(n)]$ (二阶矩) 帮助我们分析LMS算法在均方意义下收敛条件

由 $\mathbf{H}(n+1) = \mathbf{H}(n) + \delta e(n+1)\mathbf{X}(n+1)$
可得 $\mathbf{c}(n+1) = [\mathbf{I}_N - \delta \mathbf{X}(n+1)\mathbf{X}^T(n+1)]\mathbf{c}(n) + \delta \mathbf{X}(n+1)e_0(n+1)$
其中 $e_0(n+1) = y(n+1) - \mathbf{X}^T(n+1)\mathbf{H}_{\text{opt}}$



维纳滤波误差 (最优线性滤波误差)

$$e(n+1) = y(n+1) - \mathbf{H}^T(n)\mathbf{X}(n+1)$$

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \delta e(n+1)\mathbf{X}(n+1), n = 0, 1, 2, \dots$$

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \delta e(n+1)\mathbf{X}(n+1)$$

$$= \mathbf{H}(n) + \delta \mathbf{X}(n+1)[y(n+1) - \mathbf{X}^T(n+1)\mathbf{H}(n)]$$

$$= [\mathbf{I}_N - \delta \mathbf{X}(n+1)\mathbf{X}^T(n+1)]\mathbf{H}(n) + \delta y(n+1)\mathbf{X}(n+1)$$

$$= [\mathbf{I}_N - \delta \mathbf{X}(n+1)\mathbf{X}^T(n+1)]\mathbf{c}(n) + \mathbf{H}_{opt}$$

$$- \delta \mathbf{X}(n+1)\mathbf{X}^T(n+1)\mathbf{H}_{opt} + \delta y(n+1)\mathbf{X}(n+1)$$

$$\mathbf{C}(n) = \mathbf{H}(n) - \mathbf{H}_{opt}$$

$$\mathbf{H}(n) = \mathbf{C}(n) + \mathbf{H}_{opt}$$



$$e_0(n+1) = y(n+1) - \mathbf{X}^T(n+1)\mathbf{H}_{opt}$$

$$\mathbf{c}(n+1) = [\mathbf{I}_N - \delta \mathbf{X}(n+1)\mathbf{X}^T(n+1)]\mathbf{c}(n) + \delta \mathbf{X}(n+1)e_0(n+1)$$

$$\mathbf{c}(n+1) = [\mathbf{I}_N - \delta \mathbf{X}(n+1) \mathbf{X}^T(n+1)] \mathbf{c}(n) + \delta \mathbf{X}(n+1) e_0(n+1)$$

$$\mathbf{C}(n) = \mathbf{H}(n) - \mathbf{H}_{\text{opt}}$$

(一) 均值收敛分析

假设 $H(n)$ 统计独立于 $X(n+1)$, $y(n+1)$, 而只与 $n+1$ 时刻以前的量有关 (注意: 在很多实际应用中, 该假设不一定成立; 但实验表明, 基于此假设的分析结果基本正确)

$$\begin{aligned} E[\mathbf{c}(n+1)] &= E[(\mathbf{I}_N - \delta \mathbf{X}(n+1) \mathbf{X}^T(n+1)) \mathbf{c}(n)] + \delta E[\mathbf{X}(n+1) e_0(n+1)] \\ &= (\mathbf{I}_N - \delta E[\mathbf{X}(n+1) \mathbf{X}^T(n+1)]) E[\mathbf{c}(n)] \\ &= [\mathbf{I}_N - \delta \mathbf{R}] E[\mathbf{c}(n)] = [\mathbf{I}_N - \delta \mathbf{Q} \mathbf{\Lambda}' \mathbf{Q}^T] E[\mathbf{c}(n)] \end{aligned}$$

= 0 (正交原理)

$$\mathbf{Q}^T E[\mathbf{c}(n+1)] = [\mathbf{I}_N - \delta \mathbf{\Lambda}'] \mathbf{Q}^T E[\mathbf{c}(n)]$$

$$\boldsymbol{\alpha}(n) = \mathbf{Q}^T \mathbf{c}(n) = \mathbf{Q}^T [\mathbf{H}(n) - \mathbf{H}_{\text{opt}}]$$

$$E[\boldsymbol{\alpha}(n+1)] = [\mathbf{I}_N - \delta \text{diag}(\lambda_i)] E[\boldsymbol{\alpha}(n)]$$

初始值, 设 $\mathbf{H}(0)=0$, $E[\boldsymbol{\alpha}(0)] = \mathbf{Q}^T E[\mathbf{H}(0) - \mathbf{H}_{\text{opt}}] = -\mathbf{Q}^T \mathbf{H}_{\text{opt}}$

$$E[\mathbf{a}(n+1)] = [\mathbf{I} - \delta \text{diag}(\lambda_i)] E[\mathbf{a}(n)]$$

$$\begin{bmatrix} E[\alpha_0(n+1)] \\ E[\alpha_1(n+1)] \\ \dots \\ E[\alpha_{N-1}(n+1)] \end{bmatrix} = \begin{bmatrix} 1 - \delta\lambda_0 & 0 & \dots & 0 \\ 0 & 1 - \delta\lambda_1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 - \delta\lambda_{N-1} \end{bmatrix} \begin{bmatrix} E[\alpha_0(n)] \\ E[\alpha_1(n)] \\ \dots \\ E[\alpha_{N-1}(n)] \end{bmatrix}$$

$$E[\alpha_k(n+1)] = (1 - \delta\lambda_k) E[\alpha_k(n)], k = 0, 1, \dots, N-1$$

$$E[\alpha_k(n)] = (1 - \delta\lambda_k)^n E[\alpha_k(0)], k = 0, 1, \dots, N-1$$

均值收敛条件

$$|1 - \delta\lambda_k| < 1, \quad \text{for all } k$$

$$0 < \delta < 2 / \lambda_{\max}$$

均值收敛: $E[\mathbf{H}(n)] \rightarrow \mathbf{H}_{\text{opt}}$



$$E[\mathbf{a}(n)] = E[\mathbf{Q}^T \mathbf{c}(n)]$$

$$= \mathbf{Q}^T E[\mathbf{H}(n) - \mathbf{H}_{\text{opt}}] \Rightarrow 0$$

注意: 尽管收敛条件由最大特征值决定, 但均值收敛的快慢, 受最小特征值 λ_{\min} 决定.

均方收敛：指均方误差J(n)收敛到一个最小值

(二) 均方收敛分析

$$\mathbf{a}(n+1) = \mathbf{Q}^T [\mathbf{H}(n+1) - \mathbf{H}_{opt}] = \mathbf{Q}^T [\mathbf{H}(n) + \delta e(n+1)\mathbf{X}(n+1) - \mathbf{H}_{opt}]$$

$$= \mathbf{a}(n) + \delta \mathbf{Q}^T e(n+1)\mathbf{X}(n+1)$$

$$e(n+1) = y(n+1) - \mathbf{H}^T(n)\mathbf{X}(n+1)$$

$$= y(n+1) - \mathbf{H}_{opt}^T \mathbf{X}(n+1) - \mathbf{H}^T(n)\mathbf{X}(n+1) + \mathbf{H}_{opt}^T \mathbf{X}(n+1)$$

$$= e_0(n+1) - [\mathbf{H}(n) - \mathbf{H}_{opt}]^T \mathbf{X}(n+1)$$

$$= e_0(n+1) - [\mathbf{a}(n)]^T \mathbf{Q}^T \mathbf{X}(n+1)$$

$$\mathbf{a}(n) = \mathbf{Q}^T [\mathbf{H}(n) - \mathbf{H}_{opt}]$$

$e_0(n+1)$ 是和 $\mathbf{X}(n+1)$ 不相关的，进一步假设假设 $\mathbf{H}(n)$ 统计独立于 $\mathbf{X}(n+1)$, $y(n+1)$ ，而只与 $n+1$ 时刻以前的量有关，则将 $e(n+1)$ 代入 $\mathbf{a}(n+1)$ 可以得到：

$$E[[\mathbf{a}(n+1)][\mathbf{a}(n+1)]^T] =$$

$$\{E[\mathbf{a}(n)][\mathbf{a}(n)]^T - \delta \text{diag}(\lambda_i) E[\mathbf{a}(n)][\mathbf{a}(n)]^T -$$

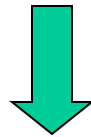
$$\delta E[\mathbf{a}(n)][\mathbf{a}(n)]^T \text{diag}(\lambda_i)\} + \delta^2 E[e^2(n+1)] \cdot \text{diag}(\lambda_i)$$

$e_0(n+1)$ 是和 $X(n+1)$ 不相关的，进一步假设假设 $H(n)$ 统计独立于 $X(n+1)$, $y(n+1)$, 而只与 $n+1$ 时刻以前的量有关

$$+ \delta^2 E[e^2(n+1)] \cdot \text{diag}(\lambda_i)$$

$$\mathbf{a}(n+1) = \mathbf{a}(n) + \delta \mathbf{Q}^T e(n+1) \mathbf{X}(n+1)$$

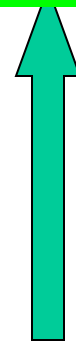
$$[\mathbf{a}(n+1)]^T = [\mathbf{a}(n)]^T + \delta \mathbf{X}^T(n+1) e(n+1) \mathbf{Q}$$



$$\begin{aligned} [\mathbf{a}(n+1)][\mathbf{a}(n+1)]^T &= [\mathbf{a}(n)][\mathbf{a}(n)]^T + [\mathbf{a}(n)]\delta e(n+1)\mathbf{X}^T(n+1)\mathbf{Q} \\ &\quad + \delta \mathbf{Q}^T \mathbf{X}(n+1)e(n+1)[\mathbf{a}(n)]^T \\ &\quad + \delta^2 \mathbf{Q}^T e(n+1)\mathbf{X}(n+1)\mathbf{X}^T(n+1)e(n+1)\mathbf{Q} \end{aligned}$$



$$\begin{aligned} e(n+1) &= e_0(n+1) - [\mathbf{a}(n)]^T \mathbf{Q}^T \mathbf{X}(n+1) \\ &= e_0(n+1) - \mathbf{X}^T(n+1)\mathbf{Q}[\mathbf{a}(n)] \end{aligned}$$



$$\begin{aligned}
 e(n+1) &= e_0(n+1) - [\mathbf{a}(n)]^T \mathbf{Q}^T \mathbf{X}(n+1) \\
 &= e_0(n+1) - \mathbf{X}^T(n+1) \mathbf{Q} [\mathbf{a}(n)]
 \end{aligned}$$

由于

$$J(n) = E[e^2(n+1)] = E[e_0^2(n+1)] + E[\underbrace{[\mathbf{a}(n)]^T \mathbf{Q}^T [\mathbf{X}(n+1) \mathbf{X}^T(n+1)] \mathbf{Q} [\mathbf{a}(n)]}_{\text{标量}}]$$

$$E[e_0^2(n+1)] = J_{\min}$$

标量

$$E[[\mathbf{a}(n)]^T \mathbf{Q}^T \mathbf{X}(n+1) \mathbf{X}^T(n+1) \mathbf{Q} [\mathbf{a}(n)]]$$

$$= E[\text{Tr}\{[\mathbf{a}(n)]^T \mathbf{Q}^T \mathbf{X}(n+1) \mathbf{X}^T(n+1) \mathbf{Q} [\mathbf{a}(n)]\}]$$

$$= E[\text{Tr}\{\mathbf{Q}^T \mathbf{X}(n+1) \mathbf{X}^T(n+1) \mathbf{Q} [\mathbf{a}(n)] [\mathbf{a}(n)]^T\}]$$

$$= \text{Tr}\{\mathbf{Q}^T E[\mathbf{X}(n+1) \mathbf{X}^T(n+1)] \mathbf{Q} E[[\mathbf{a}(n)] [\mathbf{a}(n)]^T]\}$$

$$= \text{Tr}\{\text{diag}(\lambda_i) E[[\mathbf{a}(n)] [\mathbf{a}(n)]^T]\}$$

$$= \sum_{i=0}^{N-1} \lambda_i E[a_i^2(n)]$$

$$J(n) = J_{\min} + \sum_{i=0}^{N-1} \lambda_i E[a_i^2(n)]$$

$$J(n) = J_{\min} + \sum_{i=0}^{N-1} \lambda_i E[\alpha_i^2(n)]$$

$$\mathbf{a}(n) = \mathbf{Q}^T \mathbf{c}(n) = \mathbf{Q}^T [\mathbf{H}(n) - \mathbf{H}_{opt}]$$

令

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \dots \\ \lambda_{N-1} \end{bmatrix} \quad [\mathbf{a}^2(n)] = \begin{bmatrix} \alpha_0^2(n) \\ \alpha_1^2(n) \\ \dots \\ \alpha_{N-1}^2(n) \end{bmatrix}$$

$$E[[\mathbf{a}(n+1)][\mathbf{a}(n+1)]^T] = \{ E[\mathbf{a}(n)][\mathbf{a}(n)]^T - \delta E[\mathbf{a}(n)][\mathbf{a}(n)]^T \text{diag}(\lambda_i) - \delta \text{diag}(\lambda_i) E[\mathbf{a}(n)][\mathbf{a}(n)]^T \} + \delta^2 \text{diag}(\lambda_i) E[e^2(n+1)]$$

$$E[e^2(n+1)] = J(n)$$

令对角线上的元素相等

$$\begin{cases} J(n) = J_{\min} + \mathbf{\Lambda}^T E[\mathbf{a}^2(n)] \\ E[\mathbf{a}^2(n+1)] = [\mathbf{I}_N - 2\delta \text{diag}(\lambda_i) + \delta^2 \mathbf{\Lambda} \mathbf{\Lambda}^T] E[\mathbf{a}^2(n)] + \delta^2 J_{\min} \mathbf{\Lambda} \end{cases}$$

为简化表达式, 令 $\mathbf{\beta}(n) = E[\mathbf{a}^2(n)]$

则上式写成 $\mathbf{\beta}(n+1) = \mathbf{B} \mathbf{\beta}(n) + \delta^2 J_{\min} \mathbf{\Lambda}$

其中 \mathbf{B} 是个方阵, 全部元素为正的实数的对称阵

$$b_{ij} = \begin{cases} (1 - \delta \lambda_i)^2 & i = j \\ \delta^2 \lambda_i \lambda_j & i \neq j \end{cases}$$

$$J(n) = J_{\min} + \Lambda^T E[\alpha^2(n)]$$

分析:

$$(1) \quad J(n) = J_{\min} + \Lambda^T E[\alpha^2(n)]$$

LMS算法均方收敛特性由 $\beta(n) = E[\alpha^2(n)]$ 决定，而LMS算法的均方收敛的过渡特性是由 $\beta(n) = E[\alpha^2(n)]$ 的进化来分析。定义超量误差 $J_{ex}(n)$

$$J(n) = J_{\min} + J_{ex}(n)$$

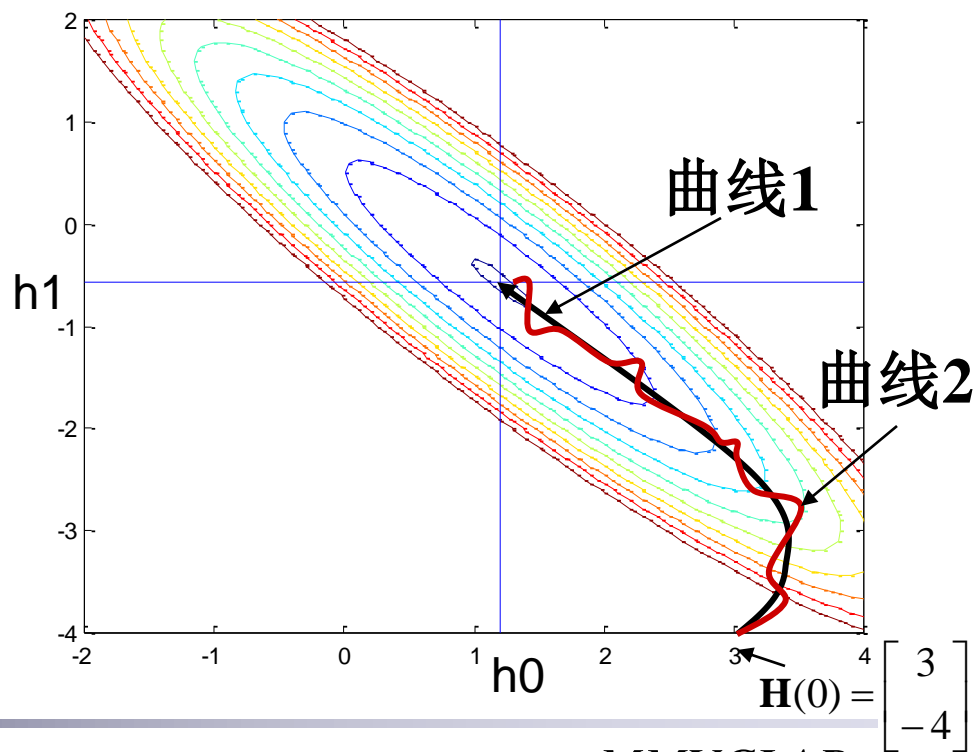
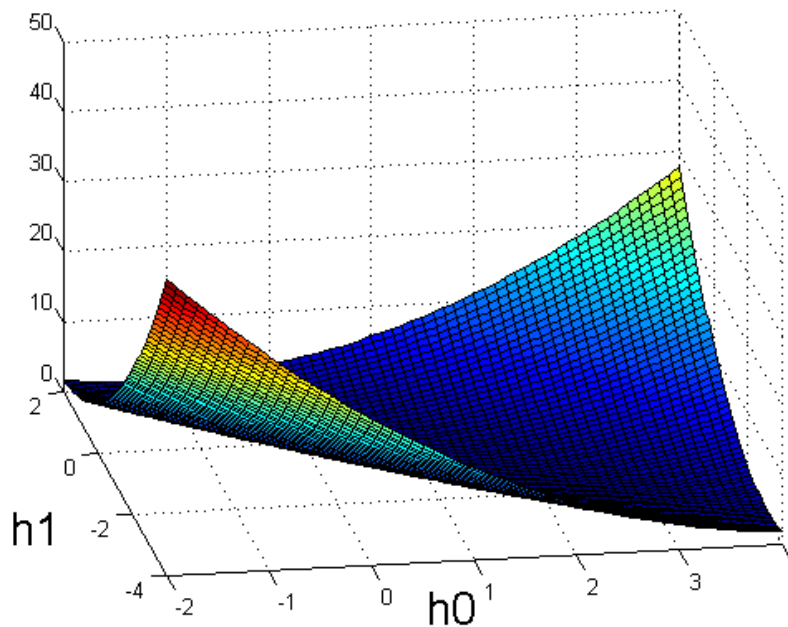
$$\text{则} \quad J_{ex}(n) = \Lambda^T \beta(n) = \sum_{i=0}^{N-1} \lambda_i E[\alpha_i^2(n)]$$

算法收敛时， $J_{ex}(n)$ 达到一个稳定值，除非 $E[\alpha^2(n)]$ 为零矢量，否则 $J(n) \neq J_{\min}$

例：从 $\mathbf{H}(n)$ 调整过程理解

$$J(n) = E[y^2(n)] - 2\mathbf{r}_{yx}^T \mathbf{H} + \mathbf{H}^T \mathbf{R}_{xx} \mathbf{H}$$

$$= 0.55 + h_0^2 + h_1^2 + 2h_0h_1 \cos \frac{\pi}{8} - \sqrt{2}h_0 \cos \frac{\pi}{10} - \sqrt{2}h_1 \cos \frac{9\pi}{40}$$



2) 均方收敛条件

从公式 $\beta(n+1) = \mathbf{B}\beta(n) + \delta^2 J_{\min} \Lambda$

得到收敛条件是 B 中行元素的绝对值之和必须小于1。

其中 \mathbf{B} 是个方阵，全部元素为正的实数的对称阵

$$b_{ij} = \begin{cases} (1 - \delta\lambda_i)^2 & i = j \\ \delta^2 \lambda_i \lambda_j & i \neq j \end{cases}$$

故有: $0 < (1 - \delta\lambda_i)^2 + \delta^2 \lambda_i \sum_{\substack{j=0 \\ j \neq i}}^{N-1} \lambda_j < 1$

始终成立

或

$$0 < 1 - 2\delta\lambda_i + \delta^2 \lambda_i \sum_{j=0}^{N-1} \lambda_j < 1$$

可得

$$0 < \delta < \frac{2}{\sum_{i=0}^{N-1} \lambda_j}$$

$$0 < 1 + \delta\lambda_i [\delta \sum_{j=0}^{N-1} \lambda_j - 2] < 1$$

$$0 < \delta < \frac{2}{\sum_{i=0}^{N-1} \lambda_i}$$

这时我们称LMS算法在均方意义下收敛;平稳输入下有

$$Tr(\mathbf{R}) = \sum_{i=0}^{N-1} \lambda_i = Nr(0) = N\sigma_x^2$$

所以均方收敛条件: $0 < \delta < \frac{2}{N\sigma_x^2}$

(3) 收敛时, $n \rightarrow \infty$, 由

$$\begin{aligned} E[\mathbf{a}(n+1)\mathbf{a}(n+1)^T] &= \{ E[\mathbf{a}(n)\mathbf{a}(n)^T] - \delta E[\mathbf{a}(n)\mathbf{a}(n)^T] \text{diag}(\lambda_i) \\ &\quad - \delta \text{diag}(\lambda_i) E[\mathbf{a}(n)\mathbf{a}(n)^T] \} \\ &\quad + \delta^2 \text{diag}(\lambda_i) E[e^2(n+1)] \end{aligned}$$

可得:

$$E[\mathbf{a}(\infty)\mathbf{a}(\infty)^T] \text{diag}(\lambda_i) + \text{diag}(\lambda_i) E[\mathbf{a}(\infty)\mathbf{a}(\infty)^T] = \delta \text{diag}(\lambda_i) J(\infty) \mathbf{I}_N$$

$$E\left[\left[\mathbf{a}(\infty)\right]\left[\mathbf{a}(\infty)\right]^T\right]diag(\lambda_i)+diag(\lambda_i)E\left[\left[\mathbf{a}(\infty)\right]\left[\mathbf{a}(\infty)\right]^T\right]=\delta diag(\lambda_i)J(\infty)\mathbf{I}_N$$

$$E\left[\alpha_i^2(\infty)\right]=\beta_i(\infty)=\frac{\delta}{2}J(\infty)$$

$$J(n)=J_{\min}+J_{ex}(n)=J_{\min}+\mathbf{\Lambda}^T\boldsymbol{\beta}(n)$$

$$J(\infty)=J_{\min}+\underbrace{\frac{\delta}{2}J(\infty)\sum_{i=0}^{N-1}\lambda_i}_{\text{可得}}$$

$$J(\infty)=J_{\min}/(1-\frac{\delta}{2}\sum_{i=0}^{N-1}\lambda_i)=J_{\min}/(1-\frac{\delta}{2}N\sigma_x^2)$$

$$J_{ex}(\infty)=\frac{\delta}{2}[J_{\min}/(1-\frac{\delta}{2}N\sigma_x^2)][N\sigma_x^2]$$

$$=J_{\min}\bullet\frac{\delta}{2}N\sigma_x^2/(1-\frac{\delta}{2}N\sigma_x^2)$$

$$\boldsymbol{\beta}(n)=E\left[\mathbf{a}^2(n)\right]=E\begin{bmatrix}\alpha_0^2(n)\\\alpha_1^2(n)\\\dots\\\alpha_{N-1}^2(n)\end{bmatrix}$$

$$\mathbf{\Lambda}=\begin{bmatrix}\lambda_0\\\lambda_1\\\dots\\\lambda_{N-1}\end{bmatrix}$$

2.4 LMS算法性能分析

(Performance Analysis of LMS filter)

2.4 LMS算法性能分析

$$E[\alpha_k(n)] = (1 - \delta\lambda_k)^n E[\alpha_k(0)], \\ k = 0, 1, \dots, N-1$$

若假设初始条件为 $\beta(0)$ ，则可由递推式生成

$$\beta(n) = \mathbf{B}^n \beta(0) + \delta^2 J_{\min} \sum_{i=0}^{n-1} \mathbf{B}^i \Lambda \quad \beta(n+1) = \mathbf{B}\beta(n) + \delta^2 J_{\min} \Lambda$$

可见是 \mathbf{B}^n 控制着收敛，或者说是受输入信号的相关阵 \mathbf{R} 的特征值 λ_i 的控制。若忽略 \mathbf{B} 中与 δ^2 有关的量，则

$$b_{ij} \approx \begin{cases} 1 - 2\delta\lambda_i & i = j \\ 0 & i \neq j \end{cases} \quad b_{ij} = \begin{cases} (1 - \delta\lambda_i)^2 & i = j \\ \delta^2 \lambda_i \lambda_j & i \neq j \end{cases}$$

$$\mathbf{B}^n \approx \text{diag}(1 - 2\delta\lambda_i)^n, \quad \delta\lambda_i \ll 1$$

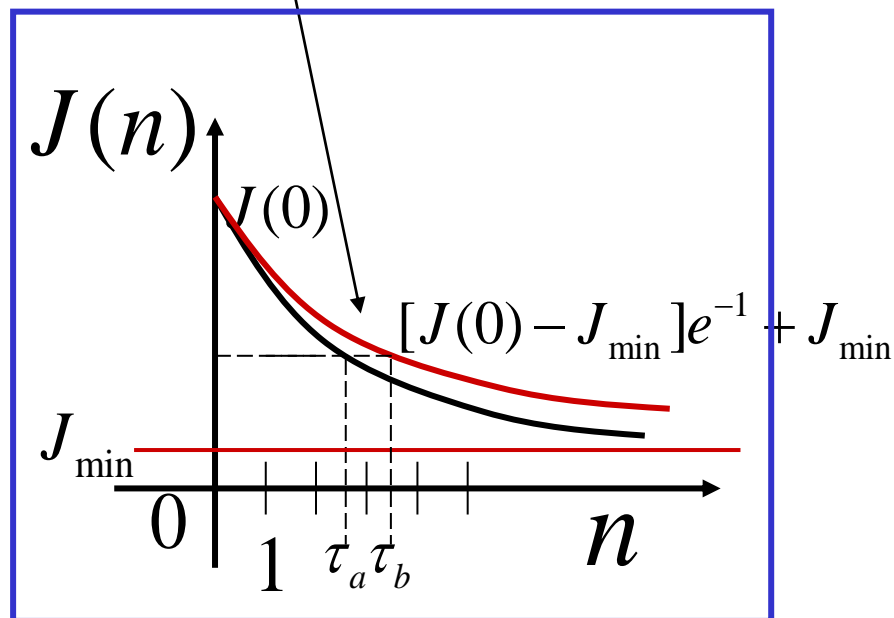
随着 n 的增大， λ_i 中最小的一项 λ_{\min} 控制着收敛过程，可见，若条件数 $X(\mathbf{R}) = \lambda_{\max}/\lambda_{\min}$ 大，即 λ_i 分布范围广，则收敛慢。收敛速度的快慢，一般用时间常数的概念来表示：

$$J(n) = E[e^2(n)] \propto n$$

学习曲线

$$J(n) = J_{\min} + \Lambda^T E[\mathbf{a}^2(n)] \quad \text{Learning curve}$$

$$J(n) - J_{\min} = e^{-\frac{n}{\tau}} [J(0) - J_{\min}]$$



2.4 LMS算法性能分析

一 时间常数

在均值收敛中 $E[\alpha_k(n)] = (1 - \delta\lambda_k)^n E[\alpha_k(0)]$

可以用指数曲线来拟合 $E[\alpha_k(n)] = e^{-\frac{n}{\tau_k}} E[\alpha_k(0)]$

可以求得 $\tau_k = \frac{-1}{\ln(1 - \delta\lambda_k)}$

时间常数是指衰减到初始值的 e^{-1} 倍时所需的时间
当 δ 足够小, $\delta \ll 1$, $\delta \lambda_k \ll 1$, 则

$$\ln(1 - \delta\lambda_k) \approx -\delta\lambda_k \quad \tau_k \approx \frac{1}{\delta\lambda_k}$$

$E[H(n)]$ 整体收敛时间常数 τ_a , 可看成介于

$$\frac{-1}{\ln(1 - \delta\lambda_{\max})} \leq \tau_a \leq \frac{-1}{\ln(1 - \delta\lambda_{\min})}$$

对于均方收敛，若忽略2次项

$$\mathbf{B}^n \approx \text{diag}(1 - 2\delta\lambda_k)^n$$

要求

$$(1 - 2\delta\lambda_k)^n = e^{-n/\tau_k}$$

可得

$$\tau_k = \frac{-1}{\ln(1 - 2\delta\lambda_k)} \approx \frac{1}{2\delta\lambda_k}$$

如果R的特征值不太分散的话， λ_i 都用 $\bar{\lambda} = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i$ 代替，可得

$$\tau_e \approx \frac{1}{2\delta\bar{\lambda}} = \frac{1}{2\delta\sigma_x^2}$$

$$J(\infty) = J_{\min} / (1 - \frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i) = J_{\min} / (1 - \frac{\delta}{2} N \sigma_x^2)$$

二、失调 Mis-adjustment

定义 $M_{adj} = \frac{J(\infty)}{J_{\min}}$

由前面讨论可得, $M_{adj} = \frac{1}{1 - \frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i}$

若 $\frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i = \frac{\delta}{2} N \sigma_x^2 \ll 1$, $M_{adj} \approx 1 + \frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i$
 $= 1 + \frac{\delta}{2} N \sigma_x^2$

三、讨论

1. 归纳LMS算法

参数

N 权系数个数，**FIR**滤波器的阶数

δ ,自适应步距

初始条件

$H(0)=0$

数据

$X(n)$, $N \times 1$, 输入数据矢量 (**n**时刻)

$y(n)$,参考信号, 期望的响应 (**n**时刻)

计算

for $n=0,1,2,\dots$

$$e(n+1)=y(n+1)-H^T(n)X(n+1)$$

$$H(n+1)=H(n)+\delta e(n+1)X(n+1)$$

2. 三个基本因素影响LMS算法的性能

$$M_{adj} \approx 1 + \frac{\delta}{2} N \sigma_x^2$$
$$\tau_e \approx \frac{1}{2\delta\sigma_x^2}$$

1) δ

采用小的 δ 值，自适应较慢，时间常数较大，相应收敛后的均方误差要小，需要较大量的数据来完成自适应过程；

当 δ 值较大时，自适应相对较快，代价是增加了收敛后的平均超量误差,需要较少量的数据来完成自适应过程；

因此 δ 的倒数可以被看成是LMS算法的Memory长度。

2) N

由于算法均方收敛的条件是

$$0 < \delta < \frac{2}{\sum_{i=0}^{N-1} \lambda_i} \quad \text{或} \quad 0 < \delta < \frac{2}{N\sigma_x^2}$$

所以均方误差 $J(n)$ 的收敛特性和阶次 N 有关。

$$\mathbf{a}(n) = \mathbf{Q}^T \mathbf{c}(n) = \mathbf{Q}^T [\mathbf{H}(n) - \mathbf{H}_{opt}]$$

而 $\mathbf{H}(n)$ 收敛到 \mathbf{H}_{opt} 的条件是（均值收敛条件）：

$$[\mathbf{a}^2(n)] = \begin{bmatrix} \alpha_0^2(n) \\ \alpha_1^2(n) \\ \dots \\ \alpha_{N-1}^2(n) \end{bmatrix}$$

$$0 < \delta < \frac{2}{\lambda_{\max}}$$

$$\boldsymbol{\beta}(n) = E[\mathbf{a}^2(n)]$$

$$J(n) = J_{\min} + J_{ex}(n) = J_{\min} + \boldsymbol{\Lambda}^T \boldsymbol{\beta}(n)$$

当 $\mathbf{H}(n)$ 均值收敛到 \mathbf{H}_{opt} 还不够。希望 $\mathbf{H}(n)$ 的起伏即 $\mathbf{H}(n) - \mathbf{H}_{opt}$ 的方差要小，即 $\mathbf{J}_{ex}(n)$ 要收敛，且要小。

由于

$$0 < \delta < \frac{2}{\sum_{i=0}^{N-1} \lambda_j} \quad \longrightarrow \quad 0 < \delta < \frac{2}{\lambda_{\max}}$$

所以考虑到实际应用中不但要求均值收敛而且一般会要求均方收敛，所以关于叠代步长的选取一般按均方收敛要求去取：

$$0 < \delta < \frac{2}{\sum_{i=0}^{N-1} \lambda_j} = \frac{2}{N\sigma_x^2}$$

$$\tau_k = \frac{-1}{\ln(1-2\delta\lambda_k)} \approx \frac{1}{2\delta\lambda_k}$$

$$J(\infty) = J_{\min} / (1 - \frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i) \approx J_{\min} + J_{\min} \frac{\delta}{2} \sum_{i=0}^{N-1} \lambda_i$$

3) λ_i

当输入的相关阵**R**的特征值比较分散时，LMS算法的超量均方误差主要由最大特征值决定。而权系数矢量均值收敛到**H_{opt}**所需的时间受最小特征值限制。在特征值很分散（输入相关阵是病态的）时，LMS算法的收敛较慢。

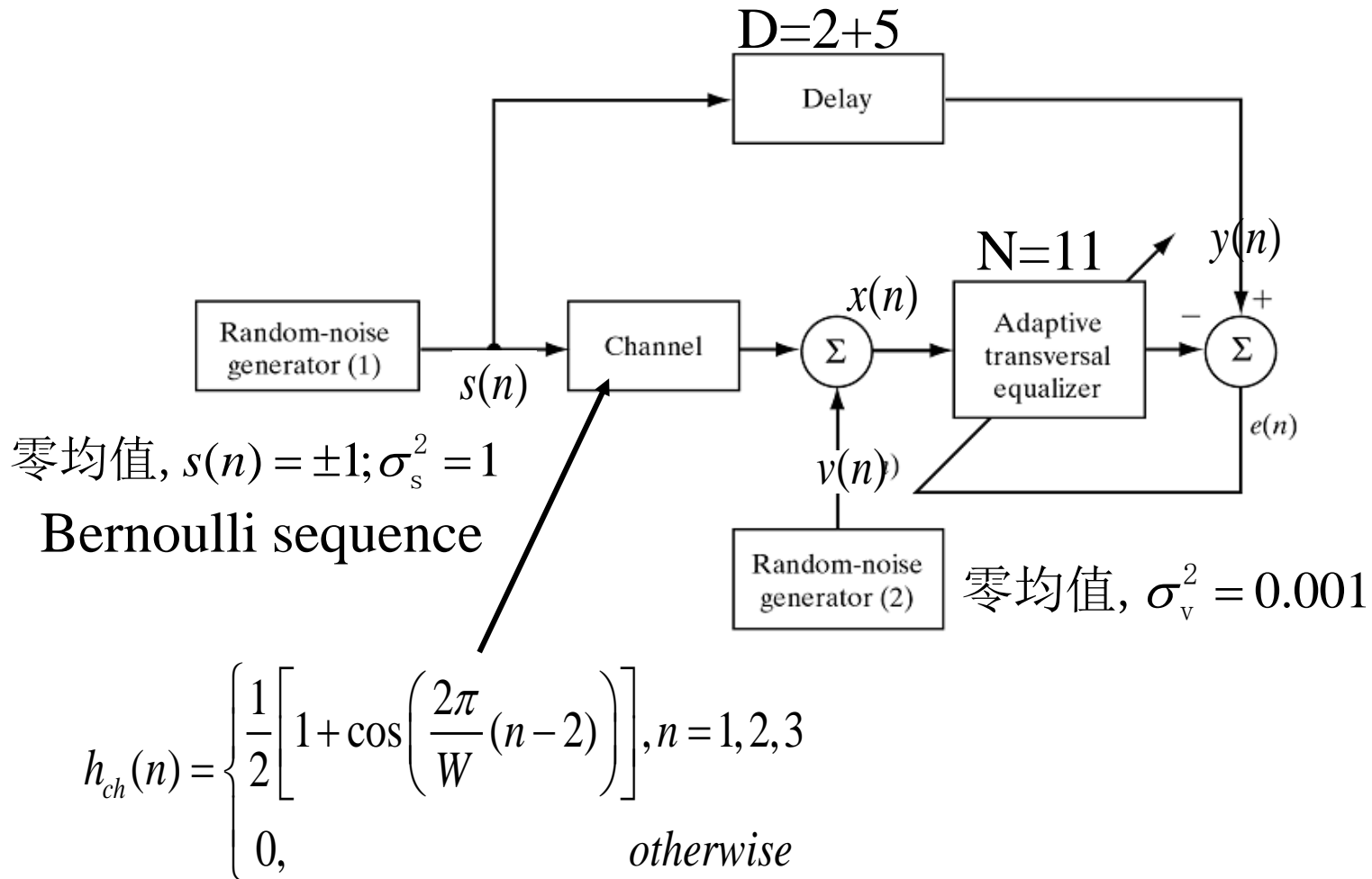
3. 独立理论

上面的结论都是基于一个有关量的独立性假设，且是平稳的。

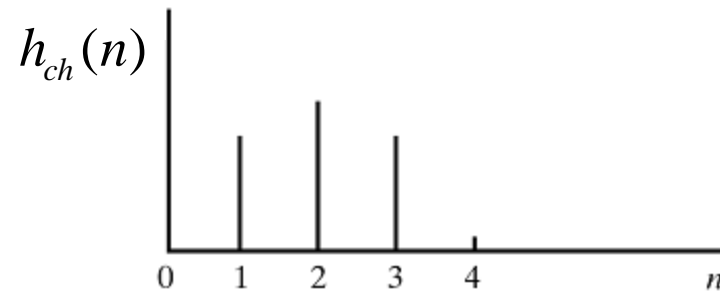
有关的独立假设等价于假定输入矢量序列是统计独立的，导致**H(n)**和**y(n+1)**、**X(n+1)**是独立的，事实上这一假设是有问题的，它忽略了在算法收敛过程中，从一次迭代到下一次迭代时，梯度方向之间是统计相关的。但是，由独立假设预测的结果和实验及计算机模拟结果多数符合得相当好。

2.4 LMS算法性能分析

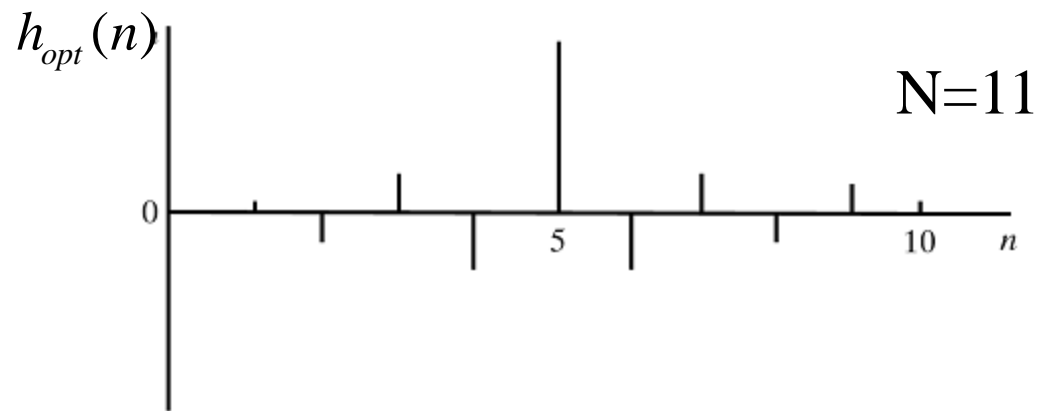
四、Computer experiment on adaptive equalization



2.4 LMS算法性能分析



(a)



(b)

(a) Impulse response of channel; (b) impulse response of optimum transversal equalizer.

Correlation Matrix of the Equalizer Input

$$x(n) = \sum_{k=1}^3 h_{ch}(k)s(n-k) + v(n)$$

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & r(2) & 0 & \dots & 0 \\ r(1) & r(0) & r(1) & r(2) & \dots & 0 \\ r(2) & r(1) & r(0) & r(1) & \dots & 0 \\ 0 & r(2) & r(1) & r(0) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & r(0) \end{bmatrix}_{N \times N}$$

$$r(0) = h_{ch}^2(1) + h_{ch}^2(2) + h_{ch}^2(3) + \sigma_v^2,$$

$$r(1) = h_{ch}(1)h_{ch}(2) + h_{ch}(2)h_{ch}(3),$$

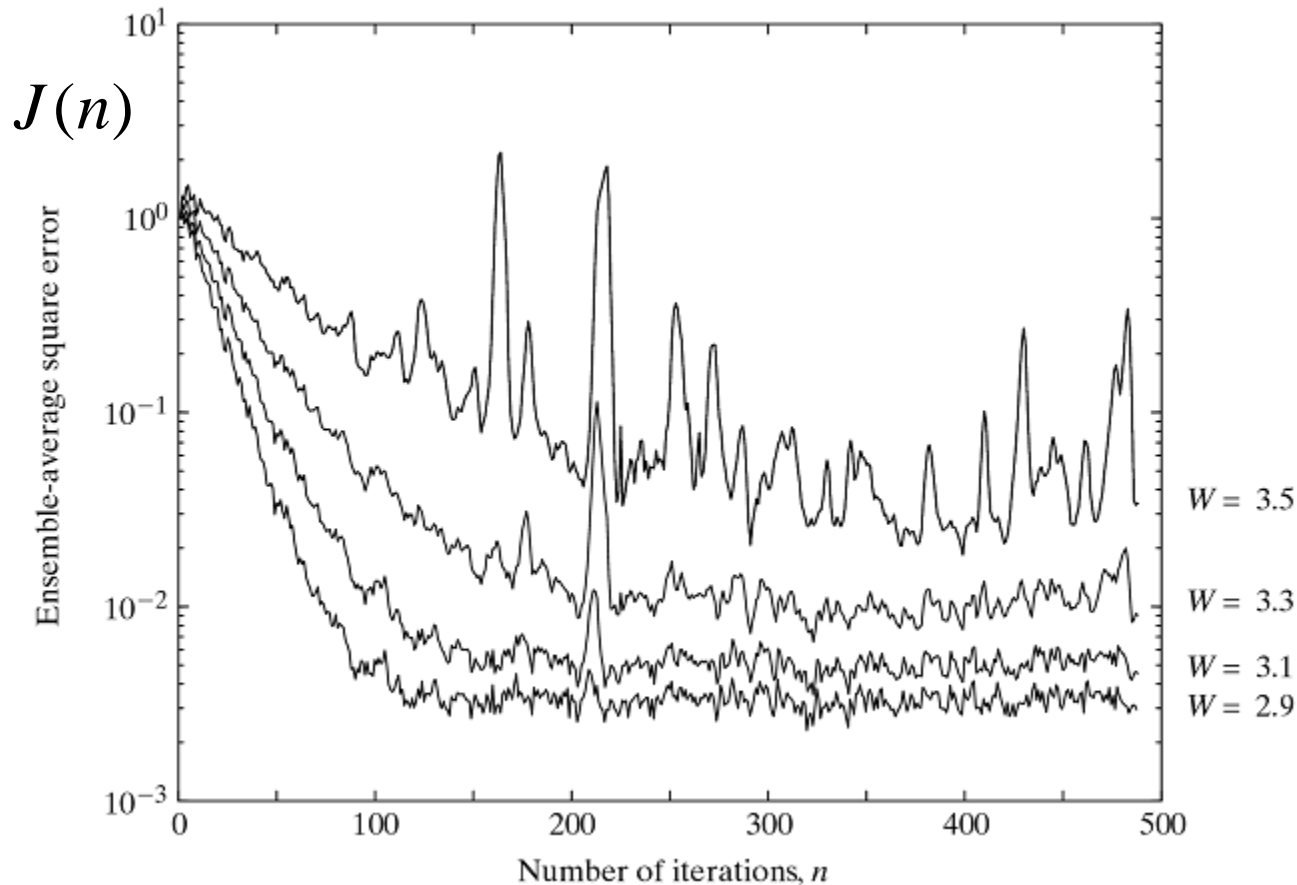
$$r(2) = h_{ch}(1)h_{ch}(3)$$

Summary of parameters for the experiment on
adaptive equalization

W	2.9	3.1	3.3	3.5
$r(0)$	1.0963	1.1568	1.2264	1.3022
$r(1)$	0.4388	0.5596	0.6729	0.7774
$r(2)$	0.0481	0.0783	0.1132	0.1511
λ_{\min}	0.3339	0.2136	0.1256	0.0656
λ_{\max}	2.0295	2.3761	2.7263	3.0707
$\chi(\mathbf{R}) = \lambda_{\max} / \lambda_{\min}$	6.0782	11.1238	21.7132	46.8216

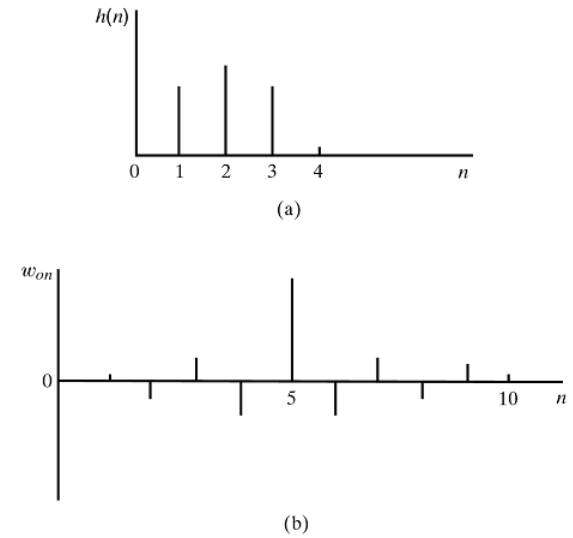
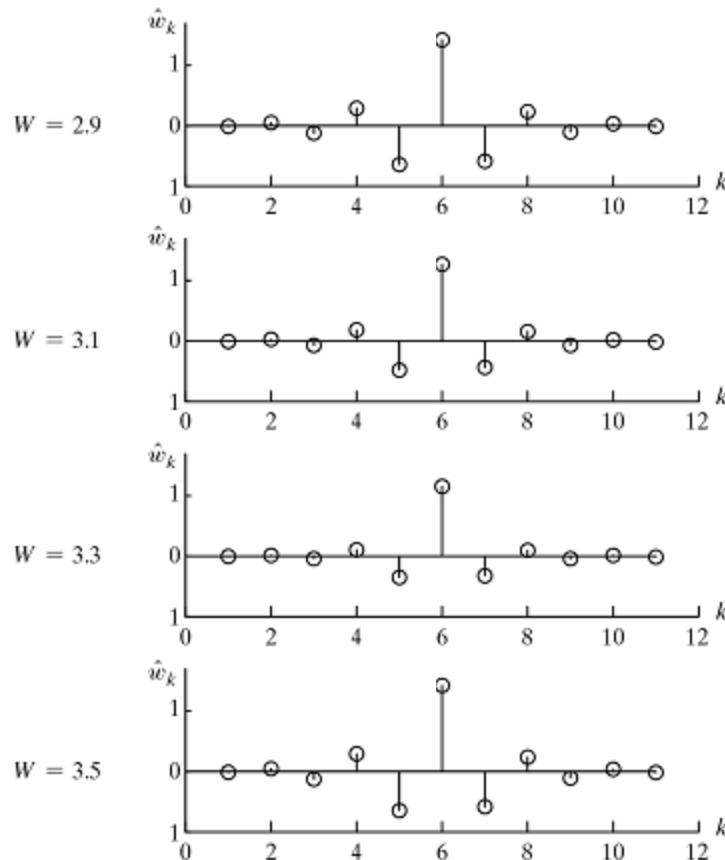
Experiment 1: Effect of Eigenvalue Spread

2.4 LMS算法性能分析



Learning curves of the LMS algorithm for an adaptive equalizer with number of taps $N=11$, step-size $\delta = 0.075$, and varying eigenvalue spread $\chi(\mathbf{R})$

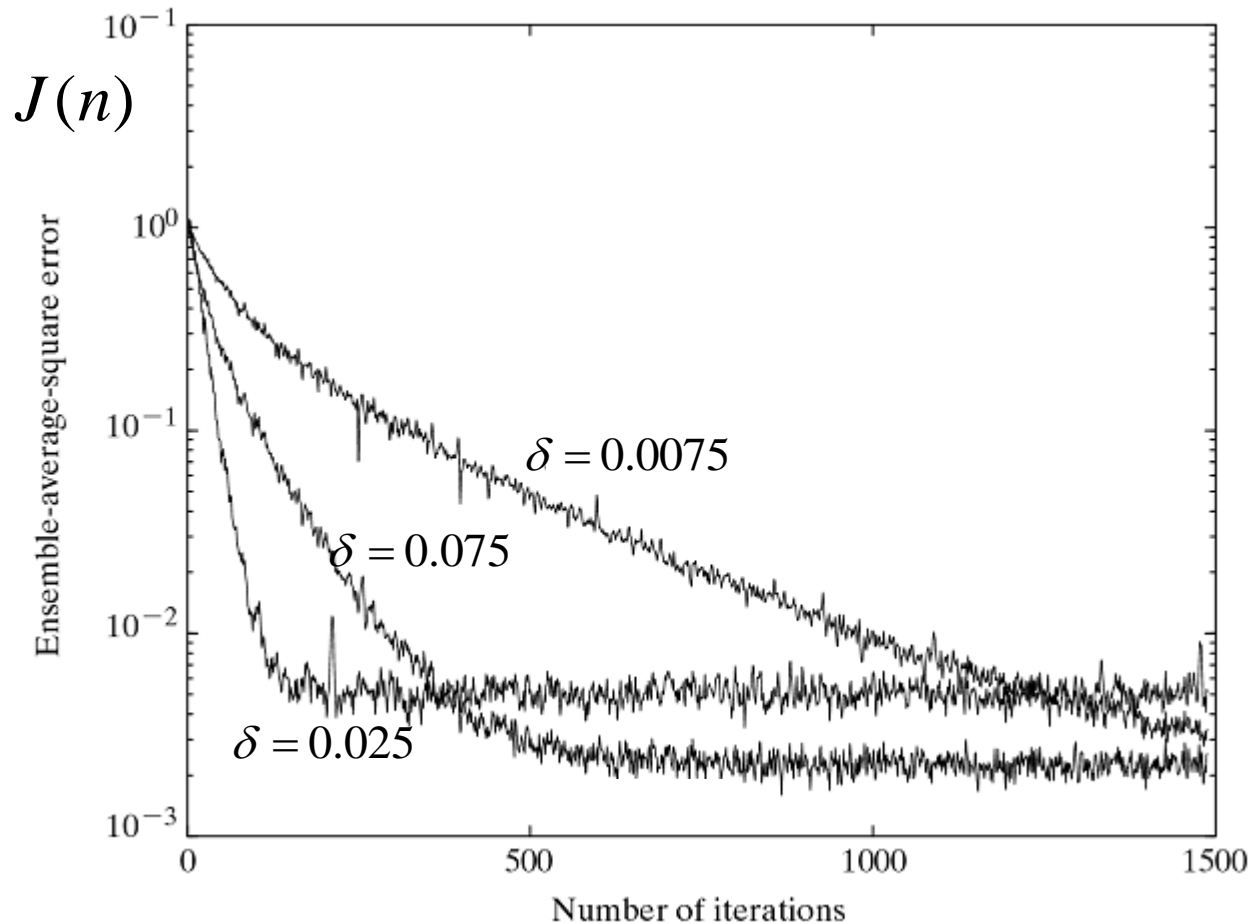
2.4 LMS算法性能分析



Ensemble-average impulse response of the adaptive equalizer (after 1000 iterations) for each of four different eigenvalue spreads

Experiment 2: Effect of Step-size Parameter

2.4 LMS算法性能分析



Learning curves of the LMS algorithm for an adaptive equalizer with number of taps $N=11$, fixed eigenvalue spread, and varying step-size δ