用酉矩阵 $\boldsymbol{S}$ 对输入数据向量 $\boldsymbol{u}(n)$ 进行酉变换，得到

$$\boldsymbol{x}(n) = \boldsymbol{S}\boldsymbol{u}(n) \tag{4.5.24}$$

即变换后的输入数据变为 $\boldsymbol{x}(n)$。对应地，酉变换后的权向量 $\boldsymbol{w}(n-1)$ 变为

$$\hat{\boldsymbol{w}}(n-1) = \frac{1}{\beta}\boldsymbol{S}\boldsymbol{w}(n-1) \tag{4.5.25}$$

它就是我们需要更新估计的变换域自适应滤波器的权向量。

因此，原预测误差 $e(n) = d(n) - \boldsymbol{w}^{\mathrm{H}}(n-1)\boldsymbol{u}(n)$ 可改用变换后的输入数据向量 $\boldsymbol{x}(n)$ 和滤波器权向量 $\hat{w}(n-1)$ 表示，即

$$e(n) = d(n) - \hat{\boldsymbol{w}}^{\mathrm{H}}(n-1)\boldsymbol{x}(n) \tag{4.5.26}$$

将变换前后的输入数据向量 $\boldsymbol{u}(n)$ 和 $\boldsymbol{x}(n)$ 比较知，原数据向量的元素是 $u(n-i+1)$ 的移位形式，它们相关性强，而 $\boldsymbol{x}(n) = [x_1(n), x_2(n), \cdots, x_M(n)]^{\mathrm{T}}$ 的元素则相当于 M 信道的信号，可以期望，它们具有比原信号 $u(n)$ 更弱的相关性。换言之，通过酉变换，在变换域实现了某种程度的解相关。从滤波器的角度讲，原来的单信道 M 阶 FIR 横向滤波器变成了等价的多信道滤波器，而原输入信号 $\boldsymbol{u}(n)$ 则等价通过一含有 M 个滤波器的滤波器组。

总结以上分析，很容易得到变换域 LMS 算法如下。

算法 4.5.4 (变换域 LMS 算法)

步骤 1 初始化： $\hat{w}(0) = \boldsymbol{0}$;

步骤 2 给定一酉变换矩阵 $\boldsymbol{S}$

更新： $n = 1, 2, \cdots$

$$\boldsymbol{x}(n) = \boldsymbol{S}\boldsymbol{u}(n)$$
$$e(n) = d(n) - \hat{\boldsymbol{w}}^{\mathrm{H}}(n-1)\boldsymbol{x}(n)$$
$$\hat{\boldsymbol{w}}(n) = \hat{\boldsymbol{w}}(n-1) + \mu(n)\boldsymbol{x}(n)e(n)$$

特别地，若酉变换采用 DFT，则 $\boldsymbol{u}$ 变成输入数据向量 $\boldsymbol{u}(n)$ 的滑动窗 Fourier 变换。这表明，被估计的权向量 $\hat{w}(k)$ 是时域滤波器 $\boldsymbol{w}(n)$ 的频率响应。因此，我们可以说，自适应发生在频域，即此时的滤波为频域自适应滤波。

4.5.3 学习速率参数的选择

LMS 算法中的步长参数 μ 决定抽头权向量在每步迭代中的更新量，是影响算法收敛速率的关键参数。由于 LMS 算法的目的是在更新过程中使抽头权向量逼近 Wiener 滤波器，所以权向量的更新过程可以视为一种学习过程，而 μ 决定 LMS 算

法学习过程的快慢。从这个意义上讲，步长参数 μ 也称学习速率参数。下面从 LMS 算法收敛的角度，讨论学习速率参数的选择问题。

基本 LMS 算法的收敛可分为均值收敛与均方收敛两种 [86]：

(1) 由式 (4.5.14)，可以建立基本 LMS 算法收敛必须满足的条件，即

$$\mathrm{E}\{e(n)\} \to 0 \qquad \text{若 } n \to \infty$$

或等价为 $\hat{w}(n)$ 的值收敛为最优 Wiener 滤波器，即

$$\lim_{n\to\infty} \mathrm{E}\{\hat{w}(n)\} = w_{\mathrm{opt}} \tag{4.5.27}$$

这一收敛称为均值收敛。

(2) LMS 算法称为均方收敛，若当迭代次数 n 趋于无穷大时，误差信号 $\epsilon(n) = d(n) - \boldsymbol{w}^{\mathrm{H}}(n)\boldsymbol{u}(n)$ 的均方值收敛为一常数，即

$$\lim_{n\to\infty} \mathrm{E}\left\{|\epsilon(n)|^2\right\} = c \tag{4.5.28}$$

式中 c 为一正的常数。

1. 均值收敛的条件

对权向量更新公式 (4.5.14) 两边取数学期望，得

$$\begin{aligned}
\mathrm{E}\{\boldsymbol{w}(n)\} &= \mathrm{E}\{\boldsymbol{w}(n-1)\} + \mu E\{e^*(n)\boldsymbol{u}(n)\} \\
&= \mathrm{E}\{\boldsymbol{w}(n-1)\} + \mu \mathrm{E}\{\boldsymbol{u}(n)[d(n) - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{w}^*(n-1)]^*\} \\
&= (\boldsymbol{I} - \mu\boldsymbol{R})\mathrm{E}\{\boldsymbol{w}(n-1)\} + \mu\boldsymbol{r}
\end{aligned} \tag{4.5.29}$$

式中 $\boldsymbol{R} = \mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\}$ 和 $\boldsymbol{r} = \mathrm{E}\{\boldsymbol{u}(n)d^*(n)\}$。

当 $k = 1$ 时，式 (4.5.29) 可写作

$$\mathrm{E}\{\boldsymbol{w}(1)\} = (\boldsymbol{I} - \mu\boldsymbol{R})\mathrm{E}\{\boldsymbol{w}(0)\} + \mu\boldsymbol{r} \tag{4.5.30}$$

式中，令 $\mathrm{E}\{\boldsymbol{w}(0)\} = \boldsymbol{w}(0)$。

当 $k = 2$ 时，利用式 (4.5.29) 及式 (4.5.30)，则有

$$\begin{aligned}
\mathrm{E}\{\boldsymbol{w}(2)\} &= (\boldsymbol{I} - \mu\boldsymbol{R})\mathrm{E}\{\boldsymbol{w}(1)\} + \mu\boldsymbol{r} \\
&= (\boldsymbol{I} - \mu\boldsymbol{R})^2\boldsymbol{w}(0) + \mu\sum_{i=0}^{1}(\boldsymbol{I} - \mu\boldsymbol{R})^i\boldsymbol{r}
\end{aligned}$$

仿此，又可得到

$$\mathrm{E}\{\boldsymbol{w}(n)\} = (\boldsymbol{I} - \mu\boldsymbol{R})^n\boldsymbol{w}(0) + \mu\sum_{i=0}^{n-1}(\boldsymbol{I} - \mu\boldsymbol{R})^i\boldsymbol{r} \tag{4.5.31}$$

令共轭对称的相关矩阵 $\boldsymbol{R}$ 的特征值分解为

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}} \tag{4.5.32}$$

式中 $\boldsymbol{U}$ 为酉矩阵，$\Sigma = \mathrm{diag}(\lambda_1, \cdots, \lambda_M)$ 是对角阵，其对角元素 λ_i 为矩阵 $\boldsymbol{R}$ 的特征值。

利用式 (4.5.32)，可以将式 (4.5.31) 重写为

$$\mathrm{E}\{\boldsymbol{w}(n)\} = (\boldsymbol{I} - \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}})^n\boldsymbol{w}(0) + \mu\sum_{i=0}^{n-1}(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}})^i\boldsymbol{r} \tag{4.5.33}$$

容易证明

$$\begin{aligned}(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}})^i &= (\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} - \mu\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}})^i \\ &= \left[\boldsymbol{U}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})\boldsymbol{U}^{\mathrm{H}}\right]^i \\ &= \boldsymbol{U}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})\boldsymbol{U}^{\mathrm{H}}\cdots\boldsymbol{U}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{U}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})^i\boldsymbol{U}^{\mathrm{H}}\end{aligned} \tag{4.5.34}$$

若对所有特征值 λ_i 有 $|1 - \mu\lambda_i| < 1$，则 $\sum_{i=0}^{\infty}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})^i = (\mu\boldsymbol{\Sigma})^{-1}$。在这一条件下，式 (4.5.34) 可以改写作

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=0}^{n-1}(\boldsymbol{I} - \mu\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}})^i &= \sum_{i=0}^{\infty}\boldsymbol{U}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})^i\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{U}\left[(\mu\boldsymbol{\Sigma})^{-1}\right]\boldsymbol{U}^{\mathrm{H}}\end{aligned} \tag{4.5.35a}$$

由于对角矩阵 $(\boldsymbol{I} - \mu\boldsymbol{\Sigma})$ 所有对角元素均小于 1，故

$$\lim_{n\to\infty}(\boldsymbol{I} - \mu\boldsymbol{\Sigma})^n = 0 \tag{4.5.35b}$$

将式 (4.5.34) 和 (4.5.35) 分别代入式 (4.5.33)，则有

$$\lim_{n\to\infty}\mathrm{E}\{\boldsymbol{w}(n)\} = 0 + \boldsymbol{U}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{r} = \boldsymbol{R}^{-1}\boldsymbol{r} = \boldsymbol{w}_{\mathrm{opt}}$$

这表明，权向量是均值收敛的，其条件是得对角矩阵 $(\boldsymbol{I}-\mu\Sigma)$ 所有对角元素均小于 1，即

$$|1-\mu\lambda_{\max}|<1$$

其解为

$$0<\mu<\frac{2}{\lambda_{\max}} \tag{4.5.36}$$

这就是权向量均值收敛时，学习速率参数 $\mu(n)$ 必须满足的条件。

2. 均方收敛的条件

在实际应用中，相关矩阵 $\boldsymbol{R}$ 的最大特征值 $\lambda_{\max}$ 通常是未知的，因此，式 (4.5.36) 的应用并不方便。为了克服这一困难，我们来分析 LMS 算法的均方收敛条件。

LMS 算法均方收敛的详细分析比其均值收敛分析复杂得多。可以证明 (对证明感兴趣的读者可参考文献 [206])，LMS 算法均方收敛的条件是：学习速率参数 μ 满足不等式

$$0<\mu<\frac{2}{\mathrm{tr}[\boldsymbol{R}]} \tag{4.5.37}$$

式中，$\mathrm{tr}[\boldsymbol{R}]$ 是相关矩阵 $\boldsymbol{R}$ 的迹。根据矩阵代数知

$$\mathrm{tr}[\boldsymbol{R}]=\sum_{k=1}^{M}\lambda_k\geqslant\lambda_{\max} \tag{4.5.38}$$

故由式 (4.5.36) ~ 式 (4.5.38)，有

$$0<\mu<\frac{2}{\mathrm{tr}[\boldsymbol{R}]}\leqslant\frac{2}{\lambda_{\max}} \tag{4.5.39}$$

这表明，若学习速率参数满足 LMS 算法均方收敛条件 (4.5.37)，则它必然满足 LMS 算法均值收敛的条件 (4.5.36)。换句话说，若 LMS 算法是均方收敛的，则它必定也是均值收敛的。

根据矩阵代数，我们还知道，一个正方矩阵的迹等于它的对角元素之和。当横向滤波器是空域滤波器时，其 M 个输入端的输入信号分别是 M 个传感器的观测数据。此时，空域滤波器的输入信号向量为 $\boldsymbol{u}(n)=[u_1,\cdots,u_M(n)]^{\mathrm{T}}$。由于相关矩阵 $\boldsymbol{R}=\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\}$ 的第 i 个对角元素为 $R_i(0)=\mathrm{E}\{u_i(n)u_i^*(n)\}$ 是输入信号 $u_i(n)$ 的均方值或能量，因此，相关矩阵 $\boldsymbol{R}$ 的迹等于在滤波器 M 个输入端上测得的总的输入能量，即

$$\mathrm{tr}[\boldsymbol{R}]=\sum_{i=1}^{M}R_i(0)=\sum_{i=1}^{M}\mathrm{E}\{|u_i|^2\}=\text{总的输入能量}$$

这表明，LMS 算法均方收敛的条件 (4.5.37) 也可以写成

$$0 < \mu < \frac{2}{\text{总的输入能量}} \tag{4.5.40}$$

注意，对于图 4.5.1 的横向滤波器，式 (4.5.40) 的分母等于输入能量 $\mathrm{E}\{|u(n)|^2\}$ 的 M 倍。

3. 自适应学习速率参数

上面从 LMS 算法的均值收敛和均方收敛的角度分别得到了学习速率参数应该满足的条件。在 LMS 算法中最简单的学习速率参数选择是取 $\mu(n)$ 为一常数，即

$$\mu(n) = \mu \tag{4.5.41}$$

式中 μ 由式 (4.5.36)、式 (4.5.39) 或式 (4.5.40) 确定。然而，这种方法会引起收敛与稳态性能的矛盾：大的学习速率能够提高滤波器的收敛速率，但稳态性能就会降低；反之，为了提高稳态性能而采用小的学习速率时，收敛就会慢。因此，学习速率的选择应该兼顾稳态性能与收敛速率。简单而有效的方法就是在不同的迭代时间使用不同的学习速率参数，即采用时变的学习速率。

时变学习速率的思想最早可追溯到 Robbins 与 Monro 于 1951 年提出的随机逼近法 [158]。最简单的时变学习速率为

$$\mu(n) = \frac{c}{n} \tag{4.5.42}$$

式中，c 为一常数。这种选择常称模拟退火法则。需要注意的是，当参数 c 比较大时，LMS 算法有可能在经过若干迭代后即陷于发散。

更好的方法是在暂态即过渡阶段使用大的学习速率，而在稳态使用小的学习速率。学习速率参数的这种选择称为“换档变速方法” (gear-shifting approach)[206]。例如，“固定 + 时变”的学习速率就是典型的换档变速方法。下面试举两个典型例子。

第一个例子是使用所谓的“先搜索、后收敛”的法则 [55]

$$\mu(n) = \frac{\mu_0}{1 + (n/\tau)} \tag{4.5.43}$$

式中 μ_0 为一固定的学习速率参数，而 τ 表示一“搜索时间常数”。由式 (4.5.43) 可以看出，这种法则在 $n \ll \tau$ 的迭代时间内使用近似固定的学习速率 μ_0；而当迭代时间 n 比搜索时间参数 τ 大时，学习速率则随时间衰减，并且衰减速度越来越快。

第二个例子是“先固定、后指数衰减”的法则 [209]

$$\mu(n) = \begin{cases} \mu_0, & n \leqslant N_0 \\ \mu_0 e^{-N_d(n-N_0)}, & n > N_0 \end{cases} \tag{4.5.41}$$

式中 μ_0 和 N_d 分别为正的常数；N_0 为正整数。

上述时变的学习速率是预先确定的，与 LMS 算法的实际运行状态并没有直接的关系。如果时变的学习速率是由 LMS 算法的实际运行状态控制的，则这类时变的学习速率称为自适应学习速率，也称“学习规则的学习”(learning of learning rules)，是 Amari 于 1967 年提出的 [3]。已提出许多方法选择自适应学习速率，这里介绍三个例子。

- Harries 等人 [84] 通过检验 LMS 算法估计误差相邻样本值的极性控制学习速率。若估计误差存在 m_0 个相邻的符号变化，则适当减小学习速率；而如果存在 m_1 个相邻的同符号，则适当增加学习速率。
- Kwong 和 Johston[107] 提出根据预测误差的平方来调节学习速率。
- 以上方法需要用户选择一些附加的常数和初始学习速率，它们都是根据诸如“初始阶段使用大的学习速率，稳态阶段使用小的学习速率”的语言规则，并把这些语言规则转换成数学模型进行学习速率参数的调节。自然地，学习速率的调节也可以直接使用模糊系统理论和语言模型来实现，构成所谓的模糊步长调节。对这种方法感兴趣的读者可参考文献 [70]。

4.5.4 LMS 算法的统计性能分析

前面介绍了基本的 LMS 自适应算法及其学习速率的选择。下面应用独立性理论对 LMS 算法的统计性能进行分析。

LMS 算法的独立性理论最早是 Widrow 等人 [205] 和 Mazo[128] 提出的。它的核心是下面的独立性假设：

(1) 输入向量 $\boldsymbol{u}(1), \boldsymbol{u}(2), \cdots, \boldsymbol{u}(n)$ 相互统计独立；

(2) 在时刻 n，输入向量 $\boldsymbol{u}(n)$ 与所有过去时刻的期望响应 $d(1), d(2), \cdots, d(n-1)$ 统计独立；

(3) 时刻 n 的期望响应 $d(n)$ 与 n 时刻的输入向量 $\boldsymbol{u}(n)$ 相关，但与过去时刻的输入向量统计独立；

(4) 输入向量 $\boldsymbol{u}(n)$ 和期望响应 $d(n)$ 对所有 n 组成联合高斯分布的随机变量。

令 $\boldsymbol{w}_{\rm opt}$ 表示最优 Wiener 滤波器，则权误差向量定义为

$$\boldsymbol{\epsilon}(n) \stackrel{\rm def}{=} \boldsymbol{w}(n) - \boldsymbol{w}_{\rm opt} \tag{4.5.45}$$

于是，由 LMS 算法产生的估计误差可以定义为

$$\begin{aligned} e(n) &\stackrel{\rm def}{=} d(n) - \boldsymbol{w}^{\rm H}(n)\boldsymbol{u}(n) \\ &= d(n) - \boldsymbol{w}_{\rm opt}\boldsymbol{u}(n) - \boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n) \\ &= e_{\rm opt}(n) - \boldsymbol{\epsilon}^{\rm H}\boldsymbol{u}(n) \end{aligned} \tag{4.5.46}$$

式中 $e_{\rm opt}(n)$ 是最优 Wiener 滤波器的估计误差。考虑抽头权向量 $\boldsymbol{w}(n)$ 的估计误差的均方值，简称均方误差，记作

$$\xi(n) = {\rm MSE}(\boldsymbol{w}(n)) = {\rm E}\left\{|e(n)|^2\right\} \tag{4.5.47}$$

利用独立性假设，易得

$$\begin{aligned} \xi(n) &= {\rm E}\left\{\left[e_{\rm opt}(n) - \boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n)\right]\left[e_{\rm opt}^*(n) - \boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\right]\right\} \\ &= \xi_{\rm min} + {\rm E}\left\{\boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\right\} \end{aligned} \tag{4.5.48}$$

式中

$$\xi_{\rm min}(n) = {\rm E}\left\{|e_{\rm opt}(n)|^2\right\} = {\rm E}\{e_{\rm opt}(n)e_{\rm opt}^*(n)\} \tag{4.5.49}$$

是最优 Wiener 滤波器产生的最小均方误差.

计算式 (4.5.48) 右边的第二项，有

$$\begin{aligned} {\rm E}\left\{\boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\right\} &= {\rm E}\left\{{\rm tr}\left[\boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\right]\right\} \\ &= {\rm E}\left\{{\rm tr}\left[\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\boldsymbol{\epsilon}^{\rm H}(n)\right]\right\} \\ &= {\rm tr}\left[{\rm E}\left\{\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\boldsymbol{\epsilon}^{\rm H}(n)\right\}\right] \end{aligned} \tag{4.5.50}$$

式中使用了矩阵迹的性质 ${\rm tr}[\boldsymbol{AB}] = {\rm tr}[\boldsymbol{BA}]$，并假定 $\boldsymbol{\epsilon}(n)$ 与 $\boldsymbol{u}(n)$ 统计独立。

利用独立性假设，式 (4.5.50) 又可写作

$$\begin{aligned} {\rm E}\left\{\boldsymbol{\epsilon}^{\rm H}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\boldsymbol{\epsilon}(n)\right\} &= {\rm tr}\left[{\rm E}\left\{\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\right\}{\rm E}\left\{\boldsymbol{\epsilon}(n)\boldsymbol{\epsilon}^{\rm H}(n)\right\}\right] \\ &= {\rm tr}\left[\boldsymbol{RK}(n)\right] \end{aligned} \tag{4.5.51}$$

式中 $\boldsymbol{R} = {\rm E}\left\{\boldsymbol{u}(n)\boldsymbol{u}^{\rm H}(n)\right\}$ 是输入向量的相关矩阵，而 $\boldsymbol{K}(n) = {\rm E}\left\{\boldsymbol{\epsilon}(n)\boldsymbol{\epsilon}^{\rm H}(n)\right\}$ 表示 n 时刻的滤波器权误差向量的相关矩阵，简称权误差相关矩阵。

将式 (4.5.51) 代入式 (4.5.48)，即可将 LMS 算法中的均方误差表示为

$$\xi(n) = \xi_{\min} + \mathrm{tr}[\boldsymbol{RK}(n)] \tag{4.5.52}$$

由 n 时刻的自适应算法产生的均方误差 $\xi(n)$ 与由最优 Wiener 滤波器产生的最小均方误差 $\xi_{\min}$ 之差称为 n 时刻的自适应算法的剩余均方误差，记作 $\xi_{\mathrm{ex}}(n)$. 即有

$$\xi_{\mathrm{ex}}(n) = \xi(n) - \xi_{\min} = \mathrm{tr}[\boldsymbol{RK}(n)] \tag{4.5.53}$$

当 n 趋于无穷大时，剩余均方误差的极限值称为稳态剩余均方误差 (或称渐近剩余均方误差)，记作

$$\xi_{\mathrm{ex}} = \xi_{\mathrm{ex}}(\infty) = \lim_{n\to\infty} \mathrm{tr}[\boldsymbol{RK}(n)] \tag{4.5.54}$$

本节最后考虑期望响应 $d(n)$ 的一种特殊选择 $d(n) \equiv 0$ 。此时，最小均方误差 (MMSE) 准则的代价函数式 (4.5.4) 变作

$$J(n) = \mathrm{E}\left\{\left|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{u}(n)\right|^2\right\} \tag{4.5.55}$$

由于上该式右边边代表滤波器输出的能量，所以上式的最小化称为最小输出能量 (minimum output energy, MOE) 准则。

仿照均方误差的定义式 (4.5.29)，我们可以定义 n 时刻滤波器抽头权向量 $\boldsymbol{w}(n)$ 的平均输出能量 (mean output energy):

$$\eta(n) \stackrel{\mathrm{def}}{=} \mathrm{MOE}\,(\boldsymbol{w}(n)) = \mathrm{E}\left\{\left|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{u}(n)\right|^2\right\} \tag{4.5.56}$$

由于权误差向量 $\boldsymbol{\epsilon}(n) = \boldsymbol{w}(n) - \boldsymbol{w}_{\mathrm{opt}}$，并且 $\boldsymbol{w}_{\mathrm{opt}}$ 与滤波器输入向量 $\boldsymbol{u}(n)$ 统计独立，故由式 (4.5.56) 得

$$\begin{aligned}\eta(n) &= \mathrm{E}\left\{\left|[\boldsymbol{w}_{\mathrm{opt}} + \boldsymbol{\epsilon}(n)]^{\mathrm{H}}\boldsymbol{u}(n)\right|^2\right\} \\ &= \mathrm{E}\left\{|\boldsymbol{w}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{u}(n)|^2\right\} + \mathrm{E}\left\{\boldsymbol{\epsilon}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\epsilon}(n)\right\} \\ &= \eta_{\min} + \mathrm{E}\left\{\boldsymbol{\epsilon}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\epsilon}(n)\right\}\end{aligned} \tag{4.5.57}$$

式中 $\eta_{\min} = \mathrm{E}\left\{|\boldsymbol{w}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{u}(n)|^2\right\}$ 表示最优滤波器的输出能量，它就是自适应滤波器所能够达到的最小输出能量。定义剩余输出能量

$$\eta_{\mathrm{ex}}(n) \stackrel{\mathrm{def}}{=} \eta(n) - \eta_{\min} \tag{4.5.58}$$

并利用式 (4.5.51)，则有

$$\eta_{\mathrm{ex}}(n) = \mathrm{E}\left\{\boldsymbol{\epsilon}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\epsilon}(n)\right\} = \mathrm{tr}[\boldsymbol{RK}(n)] \tag{4.5.59}$$

比较式 (4.5.59) 与式 (4.5.53) 知

$$\eta_{\text{ex}}(\infty) = \xi_{\text{ex}}(\infty) \tag{4.5.60}$$

即是说，滤波器的稳态剩余输出能量与稳态剩余输出均方误差等价。

以上分析表明，尽管根据 MMSE 准则与 MOE 准则设计的滤波器抽头权向量可能不同，但它们的稳态剩余均方误差和稳态剩余输出能量等价。特别地，实际测量的剩余均方误差 $\xi_{\text{ex}}(n)$ 相对于迭代时间 n 的变化曲线称为 LMS 算法的学习曲线，它是一条随时间衰减的曲线，从中可以看出 LMS 算法的收敛性能 (收敛的快慢与稳态剩余均方误差的大小)。

4.5.5 LMS 算法的跟踪性能

上面对 LMS 算法统计性能的分析是在 Wiener 滤波器固定不变的基本假设下进行的。因此，这些统计性能是标准 LMS 算法所具有的 “平均性能”，它们适合于平稳环境。

在非平稳的环境下，系统的参数是时变的，因而 Wiener 滤波器的参数也应该是时变的，以跟踪系统的动态变化。评价 LMS 算法对非平稳环境的适应能力的指标是 LMS 算法的跟踪性能。根据参数随时间变化的快慢，时变系统有快时变和慢时变之分。这里只研究慢时变环境。

一个未知的动态系统可以用一横向滤波器建模，该滤波器的抽头权向量即冲激响应向量 $\boldsymbol{w}_{\text{opt}}(n)$ 服从一阶 Markov 过程：

$$\boldsymbol{w}_{\text{opt}}(n+1) = a\boldsymbol{w}_{\text{opt}}(n) + \boldsymbol{\omega}(n) \tag{4.5.61}$$

式中 a 为一固定的参数，对于慢时变系统，a 是一个非常接近于 1 的正数；$\boldsymbol{\omega}(n)$ 为过程噪声，其均值为零，相关矩阵为 $\boldsymbol{Q}$。

横向滤波器的输出 $\boldsymbol{w}_{\text{opt}}^{\text{H}}(n)\boldsymbol{u}(n)$ 逼近期望响应，其逼近误差 $v(n)$ 称为测量噪声。因此，横向滤波器的期望响应可以表示为

$$d(n) = \boldsymbol{w}_{\text{opt}}^{\text{H}}(n)\boldsymbol{u}(n) + v(n) \tag{4.5.62}$$

对滤波器的输入、过程噪声和测量噪声作如下假设：

(1) 过程噪声向量 $\boldsymbol{\omega}(n)$ 与输入向量 $\boldsymbol{u}(n)$、测量噪声向量 $\boldsymbol{v}(n)$ 独立；

(2) 输入向量 $\boldsymbol{u}(n)$ 和测量噪声 $v(n)$ 相互独立；

(3) 测量噪声 $v(n)$ 为白噪声，它具有零均值和有限方差 $\sigma_v^2 < \infty$。

为描述模型的"快"和"慢"变化，Macchi [121],[122] 将时变系统的"非平稳度"(degree of nonstationarity) 定义为由过程噪声向量 $\omega(n)$ 引起的平均噪声功率与测量噪声引起的平均噪声功率之比，即

$$\alpha \stackrel{\text{def}}{=} \left(\frac{\mathrm{E}\{|\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)|^2\}}{\mathrm{E}\{|v(n)|^2\}}\right)^{1/2} \tag{4.5.63}$$

注意，非平稳度 α 只是时变系统的一个特征描述，它并不对自适应滤波器作任何描述。

利用过程噪声向量 $\omega(n)$ 与输入向量 $\boldsymbol{u}(n)$ 之间的统计独立性，并注意到对于标量 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$ 而言，$\mathrm{E}\{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}\} = \mathrm{tr}[\mathrm{E}\{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}\}]$，容易得到式 (4.5.63) 的分子为 [86]

$$\begin{aligned}
\mathrm{E}\{|\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)|^2\} &= \mathrm{E}\{\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\omega}(n)\} \\
&= \mathrm{tr}[\mathrm{E}\{\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\omega}(n)\}] \\
&= \mathrm{E}\{\mathrm{tr}[\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\boldsymbol{\omega}(n)]\} \\
&= \mathrm{E}\{\mathrm{tr}[\boldsymbol{\omega}(n)\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)]\} \\
&= \mathrm{tr}[\mathrm{E}\{\boldsymbol{\omega}(n)\boldsymbol{\omega}^{\mathrm{H}}(n)\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\}] \\
&= \mathrm{tr}[\mathrm{E}\{\boldsymbol{\omega}(n)\boldsymbol{\omega}^{\mathrm{H}}(n)\}\mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\}] \\
&= \mathrm{tr}[\boldsymbol{Q}\boldsymbol{R}]
\end{aligned} \tag{4.5.64}$$

式中 $\boldsymbol{Q} = \mathrm{E}\{\boldsymbol{\omega}(n)\boldsymbol{\omega}^{\mathrm{H}}(n)\}$ 是过程噪声向量 $\boldsymbol{\omega}(n)$ 的相关矩阵；$\boldsymbol{R} = \mathrm{E}\{\boldsymbol{u}(n)\boldsymbol{u}^{\mathrm{H}}(n)\}$ 表示输入向量 $\boldsymbol{u}(n)$ 的相关矩阵。

另一方面，式 (4.5.63) 的分母是零均值的测量噪声 $v(n)$ 的方差 σ_v^2。将这一结果和式 (4.5.64) 代入式 (4.5.63)，即可将非平稳度简写作

$$\alpha = \frac{1}{\sigma_v}(\mathrm{tr}[\boldsymbol{Q}\boldsymbol{R}])^{1/2} = \frac{1}{\sigma_v}(\mathrm{tr}[\boldsymbol{R}\boldsymbol{Q}])^{1/2} \tag{4.5.65}$$

式中 $\mathrm{tr}[\boldsymbol{Q}\boldsymbol{R}] = \mathrm{tr}[\boldsymbol{R}\boldsymbol{Q}]$ 是因为矩阵乘积 $\boldsymbol{Q}\boldsymbol{R}$ 和 $\boldsymbol{R}\boldsymbol{Q}$ 具有相同的对角线元素。

除了前面介绍过的收敛速率外，失调 (misadjustment) 是衡量自适应滤波器性能的另一个重要测度。自适应滤波器的失调定义为滤波器稳态剩余均方误差 J_{ex} 与滤波器最小均方误差 $J_{\min}$ 之比，即

$$\mathcal{M} \stackrel{\text{def}}{=} \frac{J_{\mathrm{ex}}}{J_{\min}} \tag{4.5.66}$$

式中，稳态剩余均方误差定义为滤波器输出的实际均方误差与最小均方误差之差，即 $J_{\mathrm{out}} - J_{\min}$。显然，当 $J_{\mathrm{ex}} = 0$ 时，滤波器的输出刚好达到最小输出均方误差，因