# L'EQUALIZZAZIONE AUTOMATICA

## 1 – Generalità.

La determinazione del valore dei pesi dell'equalizzatore presenta notevole difficoltà perché non è nota la forma dell'impulso di segnalazione $q(t)$ in arrivo al ricevitore Si tenga presente che, nella maggior parte dei casi (come ad esempio nella trasmissione wireless), le caratteristiche del canale variano con il tempo cosicché la risposta $q(t)$ non può ritenersi costante per tutto il periodo della trasmissione.

Il modo per determinare il valore dei pesi $c_i$ è basato su un algoritmo iterativo che consiste nell'aggiornare ad ogni istante di lettura $kT$ (al passo $k$) il valore dei pesi $\boldsymbol{c}$ ottenuto al passo presedente. Il criterio di aggiornamento dipende dal tipo di equalizzatoreedal criterio scelto per effettuare l'aggiornamento; nelle seguenti sezioni tali tecniche sono illustrate con riferimento ad equalizzatori di tipo ZF e di tipo MMSE.

In quel che segue si prendono in esame gli equalizzatori conformati utilizzando filtri a risposta impulsiva finita.

## 2 – Equalizzatori ZF.

Il problema della determinazione dei pesi del filtro trasversale è ricondotto (v. par. 8 del Cap. "L'equalizzazione di canale") alla soluzione del sistema $\boldsymbol{Q}\boldsymbol{Q}^H \boldsymbol{c}_o = \boldsymbol{Q}\boldsymbol{u}_D$.

Nel risolvere un tale sistema si presentano due difficoltà:

- non si conoscono gli elementi della matrice $\boldsymbol{Q}$ perché non è nota la forma dell'impulso di segnalazione;
- pur ammettendo di stimare gli elementi della matrice $\boldsymbol{Q}$, la determinazione dei pesi $\boldsymbol{c}_o$ comporta la determinazione della matrice $(\boldsymbol{Q}\boldsymbol{Q}^H)^{-1}\boldsymbol{Q}$ e tale operazione può richiedere, specie se $N$ è grande, un tempo di calcolo non compatibile con le alte velocità di trasmissione.

I pesi sono aggiornati in modo tale da annullare l'errore

(2.1) $$e_k = y_k - a_{k-D}$$

fra il valore corrente della variabile di decisione e il dato $a_{k-D}$ dove $DT$ denota il ritardo di trasmissione. Ricordando l'espressione della variabile di decisione $y_k = \boldsymbol{c}^H \cdot \boldsymbol{Q} \cdot \boldsymbol{a}_{k-D} + w_k$ si può scrivere:

(2.2) $$y_k = \tilde{\boldsymbol{q}}^H \cdot \boldsymbol{a}_{k-D} + w_k = \sum_n a_{k-D-n}\tilde{q}_n^* + w_k$$

dove si è posto $\tilde{\boldsymbol{q}} = \boldsymbol{Q}^H \boldsymbol{c}$ ed è $w_k = \boldsymbol{c}^H \boldsymbol{n}_k$. L'errore quindi diventa:

(2.3) $$e_k = \sum_n a_{k-D-n}\tilde{q}_n^* + w_k - a_{k-D}$$

che moltiplicata per $a_{k-i}^*$ e mediata fornisce:

(2.4) $$E\left\{e_k a_{k-i}^*\right\} = \sum_n E\left\{a_{k-D-n}a_{k-i}^*\right\}\tilde{q}_n^* + E\left\{w_k a_{k-i}^*\right\} - E\left\{a_{k-D}a_{k-i}^*\right\}$$

Se la sequenza dati è composta da elementi tali da aversi:

(2.5) $$E\left\{a_m a_n^*\right\} = \begin{cases} \sigma_a^2 & m = n \\ 0 & m \neq n \end{cases}$$

e la sequenza di rumore si suppone indipendente dalla sequenza dati, si ottiene:

(2.6) $$E\left\{e_k a_{k-i}^*\right\} = \sigma_a^2(\tilde{q}_{i-D}^* - \delta_{i-D}) = \begin{cases} \sigma_a^2(\tilde{q}_0^* - 1) & i = D \\ \sigma_a^2 \tilde{q}_{i-D}^* & i \neq D \end{cases}$$

Ciò comporta che, se è verificata la condizione di zero forcing data dalla $\tilde{q}_i = \delta_i$, si ha:

(2.7) $$E\left\{e_k a_{k-i}^*\right\} = 0$$

Appare pertanto naturale aggiornare i pesi secondo la seguente legge:

(2.8) $$c_i(k+1) = c_i(k) - \alpha E\left\{e_k a_{k-i}^*\right\}$$

dal momento che la condizione di equilibrio raggiunta dai pesi ($c_i(m) = \text{cost}$) comporta il verificarsi della condizione (2.7). La quantità $\alpha$ è scelta in modo da assicurare un'adeguata velocità di convergenza dell'algoritmo.

Se la sequenza dati si suppone ergodica si può scrivere, con $K$ sufficientemente elevato:

(2.9) $$\frac{1}{K}\sum_{k=1}^{K} e_k a_{k-i}^* \cong E\left\{e_k a_{k-i}^*\right\}$$

In tali ipotesi la legge di aggiornamento diviene:

(2.10) $$c_i(k+1) = c_i(k) - \frac{\alpha}{K}\sum_{k=1}^{K}(y_k - a_k)a_{k-i}^*$$

Per implementare tale algoritmo basta osservare che se la probabilità di errore di rivelazione risulta sufficientemente piccola, i simboli trasmessi $a_n$ possono essere considerati eguali ai simboli $\hat{a}_n$ in uscita dal decisore collegato a valle dell'equalizzatore. La precedente può pertanto essere approssimata dal cosiddetto *pseudo errore* per cui la legge di aggiornamento diventa:

(2.11) $$c_i(k+1) = c_i(k) - \frac{\alpha}{K}\sum_{k=1}^{K}(y_k - \hat{a}_k)\hat{a}_{k-i}^*$$

# 3 – Equalizzatori MMSE.

## 3.1 - Algoritmo del gradiente.

Ricordando il criterio adottato per gli equalizzatori di tipo MMSE il modo più naturale di aggiornare i pesi è quello che porta a minimizzare il valore quadratico medio dell'errore $e_k = y_k - a_{k-D}$. In questo caso una scelta naturale dell'algoritmo è quella del gradiente (o della discesa più ripida).

Ricordando la (9:4) del Cap "Equalizzazione di canale", l'errore quadratico medio è espresso dalla:

(3.1) $$J(\boldsymbol{c}) = \boldsymbol{c}^H \boldsymbol{R}\boldsymbol{c} - \boldsymbol{c}^H \boldsymbol{p} - \boldsymbol{p}^H \boldsymbol{c} + \sigma_a^2$$

dove

- $\boldsymbol{R}$ denota la matrice di autocorrelazione del vettore $\boldsymbol{x}_k$;
- $\boldsymbol{p}$ denota il vettore delle correlazioni mutue tra il vettore $\boldsymbol{x}_k$ e i dati $a_{k-D}$;
- $\sigma_a^2$ la varianza della sequenza dei dati.

Infatti, essendo la funzione obiettivo che si vuole minimizzare $J(\boldsymbol{c}) = E\left\{|e_k|^2\right\}$ una funzione quadratica semidefinita positiva delle variabili $c_i$, è l'aggiornamento dell'errore al passo $k$ avviene secondo una direzione opposta a quella del gradiente della funzione $J(\boldsymbol{c})$ calcolata in corrispondenza ai valori del vettore $\boldsymbol{c}$ al passo precedente. Ciò, in altri termini, equivale a porre la legge di aggiornamento dei pesi nella forma:

(3.2) $$\boldsymbol{c}(k+1) = \boldsymbol{c}(k) - \frac{\alpha}{2}\,\mathrm{grad}\left\{J\left[\boldsymbol{c}(k)\right]\right\}$$

dove $\alpha$ è un'opportuna costante. In questo modo l'aggiornamento dei pesi dell'equalizzatore avviene secondo le linee di maggior pendenza come è schematicamente indicato in Fig. 1.

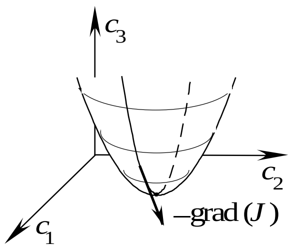

**Fig. 1** – $\mathrm{grad}(J)$

Ricordando l'espressione del gradiente della funzione $J(\boldsymbol{c})$, la legge di aggiornamento diventa:

(3.3) $$\boldsymbol{c}(k+1) = \boldsymbol{c}(k) - \alpha\left[\boldsymbol{R}\boldsymbol{c}(k) - \boldsymbol{p}\right]$$

Introducendo il vettore errore al passo $k$ definito dalla:

(3.4) $$\boldsymbol{w}(k) = \boldsymbol{c}(k) - \boldsymbol{c}_o$$

in cui $\boldsymbol{c}_o$ rappresenta il vettore dei pesi ottimo definito dalla $\boldsymbol{R}\boldsymbol{c}_o = \boldsymbol{p}$, la (3.3) diventa:

(3.5) $$\boldsymbol{w}(k+1) + \boldsymbol{c}_o = \boldsymbol{w}(k) + \boldsymbol{c}_o - \alpha\left[\boldsymbol{R}\boldsymbol{w}(k) + \boldsymbol{R}\boldsymbol{c}_o - \boldsymbol{p}\right]$$

che si semplifica nella:

(3.6) $$\boldsymbol{w}(k+1) = \left(\boldsymbol{I} - \alpha\boldsymbol{R}\right)\boldsymbol{w}(k)$$

Per determinare le condizioni che garantiscono la convergenza dell'algoritmo e cioè le condizioni per cui l'errore converge a zero, basta ricordare che la matrice $\boldsymbol{R}$ può essere posta nella seguente forma (*decomposizione diagonale*):

(3.7) $$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$$

dove

(3.8) $$\Lambda = \mathrm{diag}(\lambda_{-N},\ldots,\lambda_0,\ldots,\lambda_N) = \begin{bmatrix} \lambda_{-N} & 0 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \\ 0 & 0 & \lambda_0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \lambda_N \end{bmatrix}$$

è una matrice diagonale i cui elementi sono gli autovalori (supposti distinti) della matrice $\boldsymbol{R}$ e $\boldsymbol{U}$ è la matrice

(3.9) $$\boldsymbol{U} = \left[\boldsymbol{u}_{-N},\ldots,\boldsymbol{u}_0,\ldots,\boldsymbol{u}_N\right]$$

composta dagli autovettori normalizzati corrispondenti (v. Appendice A: Decomposizione diagonale della matrice $\boldsymbol{R}$). Premoltiplicando ambo i membri della (3.6) per $\boldsymbol{U}^{-1}$ e tenendo conto della decomposizione (3.7), si ottiene:

(3.10) $$\boldsymbol{U}^{-1}\boldsymbol{w}(k+1) = \boldsymbol{U}^{-1}\left(\boldsymbol{I} - \alpha\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\right)\boldsymbol{w}(k) = \boldsymbol{U}^{-1}\boldsymbol{w}(k) - \alpha\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\boldsymbol{w}(k) = \left(\boldsymbol{I} - \alpha\boldsymbol{\Lambda}\right)\boldsymbol{U}^{-1}\boldsymbol{w}(k)$$

che, introducendo la trasformazione di variabili:

(3.11) $$\overline{\boldsymbol{w}}(k) = \boldsymbol{U}^{-1}\boldsymbol{w}(k)$$

assume la forma:

(3.12) $$\overline{\boldsymbol{w}}(k+1) = \left(\boldsymbol{I} - \alpha\boldsymbol{\Lambda}\right)\overline{\boldsymbol{w}}(k)$$

che costituisce un insieme di $N+1$ equazioni disaccoppiate, la generica della quali è:

(3.13) $$\overline{w}_i(k+1) = \left(1-\alpha\lambda_i\right)\overline{w}_i(k) \quad i \in \{1,N\}$$

Iterando la precedente si ha:

(3.14) $$\overline{w}_i(k) = \left(1-\alpha\lambda_i\right)^k \overline{w}_i(0) \quad i \in \{1,N\}$$

dalla quale si evince che l'errore, al divergere nel numero $k$ di iterazioni tende a $0$ se tutti i coefficienti $1-\alpha\lambda_i$ sono contenuti nell'insieme $\{-1,1\}$ e cioè quando

(3.15) $$\phi(\alpha) = \max_i \left|1-\alpha\lambda_i\right| < 1$$

e quindi detto $\lambda_{\max}$ l'autovalore più grande, deve risultare:

(3.16) $$0 \le \alpha \le \frac{2}{\lambda_{\max}}$$

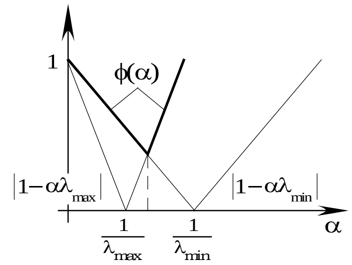

**Fig. 2** – Determinazione del valore ottimo del coefficiente $\alpha$

Infine per rendere più rapida la convergenza si deve scegliere il coefficiente $\alpha$ in modo da rendere minima la quantità $\phi(\alpha)$ perché in questo modo s'incrementa la velocità di convergenza del termine esponenziale della (3.14) a minor decadimento.

Se $\lambda_{\min}$ e $\lambda_{\max}$ denotano il più piccolo ed il più grande degli autovalori della matrice $\boldsymbol{R}$, il grafico della funzione $\phi(\alpha)$ è contenuto tra $\left|1-\alpha\lambda_{\max}\right|$ e $\left|1-\alpha\lambda_{\min}\right|$, come è riportato nella Fig. 2. Si deduce che il valore ottimo di $\alpha$ si ottiene per:

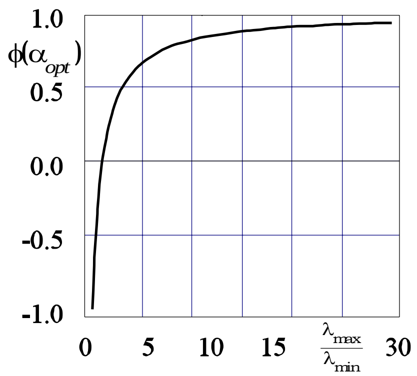

**Fig. 3** - Andamento del rapporto $\phi(\alpha_{opt})$ in funzione di $\lambda_{\max} / \lambda_{\min}$.

(3.17) $$1-\alpha_{\text{opt}}\lambda_{\min} = \alpha_{\text{opt}}\lambda_{\max} - 1$$

e cioè per

(3.18) $$\alpha_{\text{opt}} = \frac{2}{\lambda_{\max}+\lambda_{\min}}$$

L'errore pertanto è influenzato dal valore della quantità:

(3.19) $$\phi(\alpha_{\text{opt}}) = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} = \frac{\lambda_{\max}/\lambda_{\min}-1}{\lambda_{\max}/\lambda_{\min}+1}$$

che è rappresentato in Fig. 3 al variare di $\frac{\lambda_{\max}}{\lambda_{\min}}$.

La trasformazione (3.11) consente inoltre di controllare singolarmente i modi adeguando i valori del parametro $\alpha$. In altri termini scrivendo la (3.12) nella forma:

(3.20) $$\overline{\boldsymbol{w}}(k+1) = \left(\boldsymbol{I}-\boldsymbol{A\Lambda}\right)\overline{\boldsymbol{w}}(k)$$

dove:

(3.21) $$\boldsymbol{A} = \operatorname{diag}\left[\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_N\right]$$

e ponendo $\alpha_i = \alpha\lambda_i$, la (3.20) diventa:

(3.22) $$\overline{\boldsymbol{w}}(k+1) = \left(1-\alpha\right)\overline{\boldsymbol{w}}(k)$$

Tutti i modi hanno l stessa velocità do decadimento.

Nello studio dell'algoritmo del gradiente si è anche interessati alla convergenza dell'errore quadratico medio. Ricordando l'espressione $J(\boldsymbol{c}) = J_{\min} + (\boldsymbol{c}-\boldsymbol{c}_o)^H \boldsymbol{R}(\boldsymbol{c}-\boldsymbol{c}_o)$ ottenuta nel Cap.

"L'equalizzazione di canale", tenendo conto della posizione (3.4) e della trasformazione (3.11) si ha:

(3.23) $$J\left[\boldsymbol{c}(k)\right] = J_{\min} + \overline{\boldsymbol{w}}^H(k)\boldsymbol{U}^{-1}\boldsymbol{R}\boldsymbol{U}\overline{\boldsymbol{w}}(k)$$

dove si è tenuto conto che è $\boldsymbol{U}^H = \boldsymbol{U}^{-1}$. Ricordando infine la (3.7) la presedente si semplifica nella:

(3.24) $$J\left[\boldsymbol{c}(k)\right] = J_{\min} + \overline{\boldsymbol{w}}^H(k)\boldsymbol{\Lambda}\overline{\boldsymbol{w}}(k)$$

e cioè:

(3.25) $$J\left[\boldsymbol{c}(k)\right] = J_{\min} + \sum_{i=1}^{N}\lambda_i\overline{w}_i(k)\overline{w}_i^*(k)$$

che, tenendo conto della (3.14), diventa:

(3.26) $$J\left[\boldsymbol{c}(k)\right] = J_{\min} + \sum_{i=1}^{N}\lambda_i\left(1-\alpha\lambda_i\right)^{2k}\left|\overline{w}_i(0)\right|^2$$

dalla quale si evince che, se sono soddisfatte le condizioni di convergenza, l'errore quadratico medio converge a zero con una velocità maggiore di quella che compete ai pesi dell'equalizzatore.

### 3.2 – Algoritmo LMS.

È bene osservare che la legge di aggiornamento dei pesi (3.3) comporta la valutazione della matrice $\boldsymbol{R}$ e del vettore $\boldsymbol{p}$ che comportano conoscenze delle statistiche della sequenza $\boldsymbol{x}$ del segnale ricevuto, della sequenza $\boldsymbol{a}$ dei dati e di quella $\boldsymbol{n}$ del rumore. Poiché, in generale, queste informazioni non sono note al ricevitore si preferisce stimare la matrice $\boldsymbol{R}$ e il vettore $\boldsymbol{p}$ sulla base dei dati ricevuti. In particolare se si ipotizza che le sequenze coinvolte siano ergodiche i parametri in questione possono essere stimati mediante le medie temporali come segue:

(3.27) $$\hat{\boldsymbol{R}} \simeq \frac{1}{M}\sum_{n=k-(M-1)}^{k}\boldsymbol{x}_n\boldsymbol{x}_n^H \qquad \hat{\boldsymbol{p}} \simeq \frac{1}{M}\sum_{n=k-(M-1)}^{k}\boldsymbol{x}_n a_{k-D}^*$$

valide se $M$ è sufficientemente elevato.

Un'approssimazione più drastica basata sull'utilizzazione solo del dato corrente è stata introdotta da Widrow e Hoff. Essa consiste nell'usare le stime:

(3.28) $$\hat{\boldsymbol{R}} \simeq \boldsymbol{x}_n\boldsymbol{x}_n^H \qquad \hat{\boldsymbol{p}} \simeq \boldsymbol{x}_n a_{k-D}^*$$

che danno luogo alla seguente formulazione dell'algoritmo di aggiornamento dei pesi:

(3.29) $$\boldsymbol{c}(k+1) = \boldsymbol{c}(k) - \alpha\left[\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{c}(k) - \boldsymbol{x}_k a_{k-D}^*\right] = \boldsymbol{c}(k) - \alpha\boldsymbol{x}_k e^*(k)$$

essendo $e(k) = \boldsymbol{c}^H(k)\boldsymbol{x}_k - a_{k-D}$ l'errore al passo $k$.

È bene osservare che un tale algoritmo corrisponde ad utilizzare come funzione obiettivo il valore quadratico dell'errore di stima $\hat{J}(\boldsymbol{c}) = |e_k|^2 = \left|e_k e_k^*\right|$. Infatti, ricordando le posizioni introdotte nel Cap. "L'equalizzazione di canale", è:

(3.30) $$\begin{aligned}\operatorname{grad}\left\{\hat{J}(\boldsymbol{c})\right\} &= \operatorname{grad}\left\{e_k e_k^*\right\} = \operatorname{grad}\left[\left(y_k - a_{k-D}\right)\left(y_k - a_{k-D}\right)^*\right] = \\ &= \operatorname{grad}\left\{\left(\boldsymbol{c}^H\boldsymbol{x}_k - a_{k-D}\right)\left(\boldsymbol{c}^H\boldsymbol{x}_k - a_{k-D}\right)^H\right\} = \operatorname{grad}\left\{\left(\boldsymbol{c}^H\boldsymbol{x}_k - a_{k-D}\right)\left(\boldsymbol{x}_k^H\boldsymbol{c} - a_{k-D}^*\right)^H\right\} = \\ &= \operatorname{grad}\left\{\boldsymbol{c}^H\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{c} - \boldsymbol{c}^H\boldsymbol{x}_k a_{k-D}^* - \boldsymbol{x}_k^H\boldsymbol{c}a_{k-D} + \left|a_{k-D}\right|^2\right\} = 2\left(\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{c} - \boldsymbol{x}_k a_{k-D}^*\right) = 2\boldsymbol{x}_k e^*(k)\end{aligned}$$

Per questo motivo tale algoritmo prende il nome di algoritmo a ***minimo errore quadratico***

**LMS** (**L**east **M**inimum **S**quare *error*).

Come nel caso dell'algoritmo del gradiente anche in questo caso occorre indagare sulla convergenza dell'algoritmo; ma poiché l'algoritmo opera su variabili aleatorie la convergenza può essere accertata solo se si prende in esame la media statistica. In altre parole l'algoritmo del gradiente stocastico converge se risulta:

(3.31) $$\lim_{k\to\infty} E\{\boldsymbol{c}(k)\} = \boldsymbol{c}_o$$

dove $\boldsymbol{c}_o$ rappresenta la soluzione ottima e cioè la soluzione dell'equazione $\boldsymbol{R}\boldsymbol{c}_o = \boldsymbol{p}$.

La valutazione delle medie statistiche del vettore dei pesi non è matematicamente ottenibile. Per comodità di calcolo si è introdotta una particolare ipotesi nota come ***ipotesi di indipendenza***. Essa consiste nel supporre che:

i. il termine $\boldsymbol{x}_k$ della sequenza $\{\boldsymbol{x}_k\}$ è a media nulla e statisticamente indipendente da tutti gli altri termini precedenti $\boldsymbol{x}_n$ $(n<k)$;
ii. il termine $a_k$ della sequenza $\{a_k\}$ è a media nulla e statisticamente indipendente da tutti gli altri termini precedenti $a_n$ $(n<k)$;
iii. il termine $\boldsymbol{x}_k$ è statisticamente indipendente da tutti i termini $a_n$ $(n<k)$ precedenti della sequenza $\{a_k\}$.

È da tener presente che tale ipotesi in linea di principio non è accettabile. Ad esempio se si confronta il vettore $\boldsymbol{x}_k$

$$\boldsymbol{x}_k = \begin{bmatrix} x(k) & x(k-1) & \cdots & x(k-N+1)\end{bmatrix}^T$$

con il vettore $\boldsymbol{x}_{k-1}$

$$\boldsymbol{x}_{k-1} = \begin{bmatrix} x(k-1) & x(k-2) & \cdots & x(k-N)\end{bmatrix}^T$$

si deduce che essi contengono $N-1$ termini eguali, precisamente quelli compresi fra $x(k-N+1)$ e $x(k-1)$. Cionondimeno tale ipotesi è giustificata dalla circostanza che parecchie simulazioni al calcolatore hanno dimostrato che i risultati dedotti adoperando l'algoritmo LMS sono pressoché identici con quelli ottenuti introducendo l'ipotesi di indipendenza.

Prendendo le medie statistiche della (3.29), si ha:

(3.32) $$E\{\boldsymbol{c}(k+1)\} = E\{\boldsymbol{c}(k)\} - \alpha E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\boldsymbol{c}(k) - \boldsymbol{x}_k a_{k-D}^*\right\}$$

Il vettore dei pesi $\boldsymbol{c}(k)$ al passo $k$ dipende solo dai vettori $\boldsymbol{x}_n$ con $n<k$, per cui a causa dell'ipotesi di indipendenza, sopra enunciata, può ritenersi indipendente da $\boldsymbol{x}_k$. La precedente diviene:

(3.33) $$E\{\boldsymbol{c}(k+1)\} = E\{\boldsymbol{c}(k)\} - \alpha E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\right\}E\{\boldsymbol{c}(k)\} - \alpha E\left\{\boldsymbol{x}_k a_{k-D}^*\right\}$$

che, introducendo l'errore istantaneo $\boldsymbol{w}(k) = \boldsymbol{c}(k) - \boldsymbol{c}_o$, assume la forma:

(3.34) $$\begin{aligned} E\{\boldsymbol{w}(k+1)\} &= E\{\boldsymbol{w}(k)\} - \alpha E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\right\}\left(E\{\boldsymbol{w}(k)\} + \boldsymbol{c}_o\right) - \alpha E\left\{\boldsymbol{x}_k a_{k-D}^*\right\} = \\ &= \left[I - \alpha E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\right\}\right]E\{\boldsymbol{w}(k)\} - \alpha\left[E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\right\}\boldsymbol{c}_o - E\left\{\boldsymbol{x}_k a_{k-D}^*\right\}\right] \end{aligned}$$

che, tenendo conto che è $E\left\{\boldsymbol{x}_k\boldsymbol{x}_k^H\right\}\boldsymbol{c}_o - E\left\{\boldsymbol{x}_k a_{k-D}^*\right\} = \boldsymbol{R}\boldsymbol{c}_o - \boldsymbol{p} = \boldsymbol{0}$ si può porre nella forma:

(3.35) $$E\{\boldsymbol{w}(k+1)\} = \left(\boldsymbol{I} - \alpha\boldsymbol{R}\right)E\{\boldsymbol{w}(k)\}$$

Si è così pervenuto ad una equazione alle differenze identica alla (3.6) già incontrata a proposito dell'algoritmo del gradiente. Con considerazioni identiche a quelle esposte a proposito della convergenza dell'algoritmo del gradiente si può concludere che l'algoritmo LMS è convergente se il parametro $\alpha$ è contenuto nell'intervallo $\left(0, \frac{2}{\lambda_{\max}}\right)$.

La convergenza dell'algoritmo comporta che, in media, il vettore dei pesi tende al valore ottimo ma non assicura che tale convergenza avviene con piccole oscillazioni di $\boldsymbol{c}$ attorno a $\boldsymbol{c}_o$. Occorre pertanto completare l'indagine indagando sulla varianza del vettore $\boldsymbol{c}$. A tal proposito basta scrivere la (3.29) come segue:

(3.36) $$\boldsymbol{c}(k+1) = \boldsymbol{c}(k) - \frac{\alpha}{2}\,\mathrm{grad}\left\{\hat{J}\left[\boldsymbol{c}(k)\right]\right\}$$

in termini, cioè, del gradiente della stima $\hat{J}(k)$ della funzione obiettivo calcolata al passo $k$. La quantità $\mathrm{grad}\left\{\hat{J}(k)\right\}$ è una grandezza aleatoria; pertanto riesce conveniente esprimerla come somma del gradiente della funzione obiettivo

(3.37) $$\mathrm{grad}\left\{J\left[\boldsymbol{c}(k)\right]\right\} = 2\left(\boldsymbol{R}\boldsymbol{c}(k) - \boldsymbol{p}\right)$$

e di un termine aleatorio $2\boldsymbol{N}(k)$, supposto a media nulla. In altri termini, si pone:

(3.38) $$\mathrm{grad}\left\{\hat{J}\left[\boldsymbol{c}(k)\right]\right\} = \mathrm{grad}\left\{J\left[\boldsymbol{c}(k)\right]\right\} + 2\boldsymbol{N}(k)$$

L'algoritmo di aggiornamento (3.36) diventa:

(3.39) $$\boldsymbol{c}(k+1) = \boldsymbol{c}(k) - \alpha\left[\boldsymbol{R}\boldsymbol{c}(k) - \boldsymbol{p}\right] - \alpha\boldsymbol{N}(k)$$

In termini del vettore errore $\boldsymbol{w}(k)$ si ha:

(3.40) $$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \alpha\left\{\boldsymbol{R}\left[\boldsymbol{w}(k) + \boldsymbol{c}_o\right] - \boldsymbol{p}\right\} - \alpha\boldsymbol{N}(k) = \boldsymbol{w}(k) - \alpha\boldsymbol{R}\boldsymbol{w}(k) - \alpha\left[\boldsymbol{R}\boldsymbol{c}_o - \boldsymbol{p}\right] - \alpha\boldsymbol{N}(k)$$

ovvero:

(3.41) $$\boldsymbol{w}(k+1) = \left(\boldsymbol{I} - \alpha\boldsymbol{R}\right)\boldsymbol{w}(k) - \alpha\boldsymbol{N}(k)$$

La (3.41) consente di calcolare la matrice di correlazione $\boldsymbol{R}_{\boldsymbol{w}}(k)$ del vettore $\boldsymbol{w}(k)$ al passo $k$. Si ha:

(3.42) $$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{w}}(k+1) &= E\left\{\boldsymbol{w}(k+1)\boldsymbol{w}^H(k+1)\right\} = \\
&= E\left\{\left[\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)\boldsymbol{w}(k) - \alpha\boldsymbol{N}(k)\right]\left[\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)\boldsymbol{w}(k) - \alpha\boldsymbol{N}(k)\right]^H\right\} = \\
&= E\left\{\left[\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)\boldsymbol{w}(k) - \alpha\boldsymbol{N}(k)\right]\left[\boldsymbol{w}^H(k)\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)^H - \alpha\boldsymbol{N}^H(k)\right]\right\} = \\
&= E\left\{\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)\boldsymbol{w}(k)\boldsymbol{w}^H(k)\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)^H\right\} + \\
&\quad - \alpha E\left\{\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)\boldsymbol{w}(k)\boldsymbol{N}^H(k) + \boldsymbol{N}(k)\boldsymbol{w}^H(k)\left(\boldsymbol{I}-\alpha\boldsymbol{R}\right)^H\right\} + \\
&\quad + \alpha^2 E\left\{\boldsymbol{N}(k)\boldsymbol{N}^H(k)\right\}
\end{aligned}$$

Una notevole semplificazione si ottiene quando l'algoritmo si trova in prossimità della fase di regime e cioè quando $\boldsymbol{c}(k)$ si può identificare con il valore ottimo $\boldsymbol{c}_o$. In tali condizioni, dalla (3.38), si deduce:

(3.43) $$\mathrm{grad}\left\{\hat{J}\left[\boldsymbol{c}(k)\right]\right\} \cong \mathrm{grad}\left\{J(\boldsymbol{c}_o)\right\} + 2\boldsymbol{N}(k) = 2\boldsymbol{N}(k)$$

essendo $\mathrm{grad}\left\{J(\boldsymbol{c}_o)\right\} = \boldsymbol{R}\boldsymbol{c}_o - \boldsymbol{p} = \boldsymbol{0}$. Tenendo infine conto delle (3.30) e (3.38) è:

(3.44) $$\boldsymbol{N}(k) \simeq \boldsymbol{x}_k e_k^*$$

Dalla (3.44) si desume che $\boldsymbol{N}(k)$ dipende da $\boldsymbol{x}_k$ e $a_{k-D}$, mentre dalla (3.41) si evince che $\boldsymbol{w}(k)$ dipende da $\boldsymbol{x}_{k-1}$ e $\boldsymbol{N}(k-1)$. Sulla base dell'ipotesi di indipendenza si può pertanto concludere che $\boldsymbol{w}(k)$ e $\boldsymbol{N}(k)$ sono statisticamente indipendenti, per cui, essendo $\boldsymbol{N}(k)$ a media nulla si può scrivere:

(3.45) $$E\left\{\boldsymbol{w}(k)\boldsymbol{N}^H(k)\right\} = E\left\{\boldsymbol{N}^H(k)\boldsymbol{w}(k)\right\} = 0$$

Inoltre, ricordando che, sempre in virtù dell'ipotesi di indipendenza, $\boldsymbol{w}(k)$ è indipendente da $\boldsymbol{x}_k(k)$ si ha:

(3.46) $$\boldsymbol{R}_w(k+1) = E\left\{(\boldsymbol{I} - \alpha\boldsymbol{R})\right\}\boldsymbol{R}_w(k)E\left\{(\boldsymbol{I} - \alpha\boldsymbol{R})\right\} + \alpha^2 E\left\{\boldsymbol{N}(k)\boldsymbol{N}^H(k)\right\}$$

essendo manifestamente $(\boldsymbol{I} - \alpha\boldsymbol{R})^H = \boldsymbol{I} - \alpha\boldsymbol{R}$ poiché $\boldsymbol{R}$ è una matrice hermitiana.

Per valutare la matrice $E\left\{\boldsymbol{N}(k)\boldsymbol{N}^H(k)\right\}$ occorre introdurre un'ulteriore ipotesi che consiste nel supporre che $\boldsymbol{x}_k$ sia statisticamente indipendente dall'errore $e_k$. Tenendo presente la (3.44) si ottiene così:

(3.47) $$E\left\{\boldsymbol{N}(k)\boldsymbol{N}^H(k)\right\} = E\left\{|e_k|^2\right\}E\left\{\boldsymbol{x}_k\boldsymbol{x}_k\right\} = \boldsymbol{R}J_{\min}$$

dove $J_{\min} = E\left\{|e_k|^2\right\} = E\left\{\left|\boldsymbol{c}_o^H\boldsymbol{x}_k - a_{k-D}\right|^2\right\} = E\left\{\left|a_k{}^2\right|\right\} - \boldsymbol{c}_o^H\boldsymbol{R}\boldsymbol{c}_o$.

La (3.46) allora diventa:

(3.48) $$\boldsymbol{R}_w(k+1) = (\boldsymbol{I} - \alpha\boldsymbol{R})\boldsymbol{R}_w(k)(\boldsymbol{I} - \alpha\boldsymbol{R}) + \alpha^2 J_{\min}\boldsymbol{R}$$

che introducendo la trasformazione (3.11), assume la forma:

(3.49) $$\boldsymbol{U}^{-1}\boldsymbol{R}_{\overline{w}}(k+1)\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{I} - \alpha\Lambda)\boldsymbol{R}_{\overline{w}}(k)(\boldsymbol{I} - \alpha\Lambda)\boldsymbol{U}^{-1} + \alpha^2 J_{\min}\boldsymbol{U}\Lambda\boldsymbol{U}^{-1}$$

dove $\boldsymbol{R}_{\overline{w}}(k)$ denota la matrice di correlazione del vettore $\overline{\boldsymbol{w}}(k)$. Si ha:

(3.50) $$\boldsymbol{R}_{\overline{w}}(k+1) = (\boldsymbol{I} - \alpha\Lambda)\boldsymbol{R}_{\overline{w}}(k)(\boldsymbol{I} - \alpha\Lambda) + \alpha^2 J_{\min}\Lambda$$

Poiché si è interessati soltanto alle varianze delle componenti dei vettori $\overline{\boldsymbol{w}}(k)$ occorre prendere in considerazione solo gli elementi della diagonale principale della matrice di auto-correlazione $\boldsymbol{R}_{\overline{w}}(k)$. Denotando con $\gamma_i(k)$ il generico termine della diagonale principale della matrice $\boldsymbol{R}_{\overline{w}}(k)$ si ha:

(3.51) $$\gamma_i(k+1) = (1 - \alpha\lambda_i)^2\gamma_i(k) + \alpha^2 J_{\min}\lambda_i$$

che iterata dà luogo alla:

(3.52) $$\gamma_i(k) = (1 - \alpha\lambda_i)^{2k}\gamma_i(0) + \alpha^2 J_{\min}\lambda_i\sum_{n=0}^{k/2}(1 - \alpha\lambda_i)^{2n}$$

Poiché, per la convergenza in media è richiesta la condizione $|1 - \alpha\lambda_i| < 1$si deduce che risulta:

(3.53) $$\lim_{k\to\infty}\gamma_i(k) = \alpha^2 J_{\min}\lambda_i\frac{1}{1 - (1 - \alpha\lambda_i)^2}$$

È da osservare che poiché la varianza del generico componente del vettore $\overline{\boldsymbol{w}}(k)$ converge ad un valore non nullo, il vettore $\overline{\boldsymbol{w}}(k)$ continua a fluttuare attorno al suo valore limite $\overline{\boldsymbol{w}}_o$.

# APPENDICE

## PROPIRETÀ DI UNA MATRICE HERMITIANA

**1 - Se $R$ denota una matrice hermitiana, gli autovalori sono reali**.

Infatti se $\boldsymbol{u}$ è un autovettore corrispondente all'autovalore $\lambda$ è

$$\boldsymbol{Ru} = \lambda \boldsymbol{u}$$

che premoltiplicata per $\boldsymbol{u}^H$ fornisce:

(1) $$\boldsymbol{u}^H \boldsymbol{Ru} = \lambda \boldsymbol{u}^H \boldsymbol{u}$$

Prendendo ancora la trasposta coniugata, si ottiene:

$$(\boldsymbol{u}^H \boldsymbol{Ru})^H = \lambda^* (\boldsymbol{u}^H \boldsymbol{u})^H$$

e cioè

(2) $$\boldsymbol{u}^H \boldsymbol{Ru} = \lambda^* \boldsymbol{u}^H \boldsymbol{u}$$

dove si è tenuto conto che è $\boldsymbol{R}^H = \boldsymbol{R}$. Dalle (1) e (2) è facile dedurre:

(3) $$\lambda = \lambda^*$$

e cioè l'autovalore $\lambda$ è reale.

**2 – Gli autovettori di una matrice hermitiana sono ortogonali**.

Infatti siano $\boldsymbol{u}_i$ e $\boldsymbol{u}_j$ due autovettori corrispondenti agli autovalori $\lambda_i$ e $\lambda_j$, supposti distinti. Si ha:

(4) $$\boldsymbol{Ru}_i = \lambda_i \boldsymbol{u}_i \ \text{ e } \ \boldsymbol{Ru}_j = \lambda_j \boldsymbol{u}_j$$

Premoltiplicando la prima delle (a) per $\boldsymbol{u}_j^H$ e la seconda per $\boldsymbol{u}_i^H$, si ha:

(5) $$\boldsymbol{u}_j^H \boldsymbol{Ru}_i = \lambda_i \boldsymbol{u}_j^H \boldsymbol{u}_i \ \text{ e } \ \boldsymbol{u}_i^H \boldsymbol{Ru}_j = \lambda_j \boldsymbol{u}_i^H \boldsymbol{u}_j$$

Prendendo la coniugata trasposta della prima della (b) è:

(6) $$\boldsymbol{u}_i^H \boldsymbol{Ru}_j = \lambda_i^* \boldsymbol{u}_i^H \boldsymbol{u}_j$$

od anche, essendo $\lambda_i = \lambda_i^*$:

(7) $$\boldsymbol{u}_i^H \boldsymbol{Ru}_j = \lambda_i \boldsymbol{u}_i^H \boldsymbol{u}_j$$

Dal confronto della (7) con la seconda della (5) si deduce:

(8) $$(\lambda_i - \lambda_j)\boldsymbol{u}_i^H \boldsymbol{u}_j = 0$$

che, essendo gli autovalori distinti comporta:

(9) $$\boldsymbol{u}_i^H \boldsymbol{u}_j = 0 \qquad i \neq j$$

Imponendo inoltre la normalizzazione si ha:

(10) $$\boldsymbol{u}_i^H \boldsymbol{u}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

**3 – Diagonalizzazione di una matrice**.

Se $\{\boldsymbol{u}_i\}_{i=-N}^{N}$ denotano gli autovettori corrispondenti agli autovalori $\{\lambda_i\}_{i=-N}^{N}$, supposti distinti, si può scrivere:

(11)
$$\begin{aligned} \boldsymbol{R}\boldsymbol{u}_{-N} &= \lambda_{-N}\boldsymbol{u}_{-N} \\ &\cdots\cdots\cdots\cdots \\ \boldsymbol{R}\boldsymbol{u}_{N} &= \lambda_{N}\boldsymbol{u}_{N} \end{aligned}$$

che si può porre anche nella seguente forma matriciale:

(12)
$$\boldsymbol{R}\left[\boldsymbol{u}_{-N} \quad \dots \quad \boldsymbol{u}_0 \quad \dots \quad \boldsymbol{u}_N\right] = \left[\boldsymbol{u}_{-N} \quad \dots \quad \boldsymbol{u}_0 \quad \dots \quad \boldsymbol{u}_N\right] \begin{bmatrix} \lambda_{-N} & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & \lambda_0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \lambda_N \end{bmatrix}$$

Denotando con $\boldsymbol{U}$ e $\Lambda$ le matrici:

(13)
$$\boldsymbol{U} = \left[\boldsymbol{u}_{-N} \quad \dots \quad \boldsymbol{u}_0 \quad \dots \quad \boldsymbol{u}_N\right]$$
$$\Lambda = \mathrm{diag}(\lambda_{-N}, \dots, \lambda_0, \dots, \lambda_N) = \begin{bmatrix} \lambda_{-N} & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & \lambda_0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \lambda_N \end{bmatrix}$$

si scrive:

(14)
$$\boldsymbol{R}\boldsymbol{U} = \boldsymbol{U}\Lambda$$

Dalla (14) si deduce:

(15)
$$\boldsymbol{R} = \boldsymbol{U}\Lambda\boldsymbol{U}^{-1}$$

D'altra parte sulla base della (9) si ha:

(16)
$$\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}$$

che comporta $\boldsymbol{U}^{-1} = \boldsymbol{U}^H$. Si ha quindi:

(17)
$$\boldsymbol{R} = \boldsymbol{U}\Lambda\boldsymbol{U}^{-1} = \boldsymbol{U}\Lambda\boldsymbol{U}^H$$

e cioè

(18)
$$\boldsymbol{R} = \sum_{i=-N}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^H$$