5. FILTRE ADAPTIVE BAZATE PE MINIMIZAREA ERORII MEDII PATRATICE

Teoria filtrării optimale oferă soluția găsirii unui filtru optim, în sensul obținerii unei erori medii pătratice minime, în condițiile unui mediu staționar (semnalul de intrare și cel dorit sunt presupuse staționare cel puțin în sens larg). Dacă aceste condiții nu sunt îndeplinite, ar trebui, teoretic, să reeevaluăm aplicarea ecuației normale, la fiecare moment de timp. Dar nici această variantă nu este valabilă, pentru că ecuația normală necesită cunoașterea matricei de autocorelație **R** și a vectorului intercorelațiilor **p.** Acestea conțin însă funcții de corelație, deci medii statistice, a căror estimare este o problemă dificilă chiar și în cazul unui mediu staționar. În fine, în cazul unui filtru adaptiv capabil să opereze în timp real, prezintă importanță complexitatea aritmetică, iar aceasta este relativ ridicată în cazul unor filtre de lungime mare.

5.1 Metoda pantei descendente maxime (SD- Steepest Descent)

Vom renunța la ideea găsirii coeficienților optimi

$$\mathbf{w}(n) = [w_0(n), w_1(n), \cdots, w_{N-1}(n)]$$

într-un singur pas, folosind ecuația normală, și vom adopta o soluție iterativă, prin care să ne apropiem cu un mic pas de soluția optimă în fiecare iterație. Pentru aceasta vom porni de la următoarea

Teoremă

O funcție $f(\mathbf{z}, \mathbf{z}^*): \mathrm{C}^{2N} \to \mathrm{R}$ are direcția de variație (creștere) maximă în punctul $(\mathbf{z}, \mathbf{z}^*)$ dată de gradientul complex $\nabla_{\mathbf{z}^*}\{f(\mathbf{z}, \mathbf{z}^*)\}$.

Ca urmare, vom urmări în fiecare iterație ajustarea coeficienților cu un mic pas, în sensul reducerii maxime a funcției cost, urmând etapele de mai jos:

1. - Se porneşte de la o valoare iniţială a coeficienţilor $\mathbf{w}(0)$ (uzual $\mathbf{w}(0) = \mathbf{0}$).

2. - Se evaluează direcţia pantei maxime de creştere în jurul acestui punct pe suprafaţa $J(\mathbf{w})$. Aceasta este exprimată prin gradientul complex $\nabla_{\mathbf{w}^*}(J(\mathbf{w}))$.

3. - Se reactualizează coeficienţii printr-o deplasare pe direcţia pantei descendente maxime. Aceasta este echivalent cu o deplasare pe direcţia opusă gradientului cu un pas μ, $\mu \in \mathbf{R}_+$:

$$\mathbf{w}(n+1)=\mathbf{w}(n)-\mu\nabla_{w^*}(J(\mathbf{w}))$$

4. - Se reia procedeul din punctul 2.

După cum s-a văzut,

$$\nabla_{\mathbf{w}^*}(J(n))=-\mathbf{p}+\mathbf{R}\mathbf{w}(n)$$

deci ecuația de reactualizare a coeficienților este

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu(\mathbf{p}-\mathbf{R}\mathbf{w}(n))$$

Având în vedere relațiile

$$\mathbf{p}=\mathrm{E}\{\mathbf{x}(n)d^*(n)\},\quad \mathbf{R}\mathbf{w}(n)=\mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n)\},\quad e(n)=d(n)-\mathbf{w}^H(n)\mathbf{x}(n),$$

Relația de reactualizare mai poate fi scrisă

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu E(\mathbf{x}(n)e^*(n))$$

Analiza convergenței algoritmului.

Fiind un algoritm recursiv, se impune o analiză a convergenței, în sensul de a vedea în ce măsură, atunci când $n\to\infty$, souluția astfel obținută se apropie de aceea dată de fitrul optim. Vom aborda problema convergenței din două puncte de vedere

- *Analiza convergenței coeficienților* către coeficienții optimi, deci măsura în care $\mathbf{w}(n)\to\mathbf{w}_o$ când $n\to\infty$.
- *Analiza convergenței funcției cost*, deci măsura în care eroarea medie pătratică tinde către $J_{\min}$ când $n\to\infty$.

$$\mathbf{Q}^H\mathbf{c}(n+1)=(\mathbf{I}-\mu\mathbf{\Lambda})\mathbf{Q}^H\mathbf{c}(n)$$

a. Analiza convergenței coeficienților

Vom introduce ***vectorul eroare a coeficienţilor***, în raport cu valoarea optimă, dată de teoria filtrării optimale,

$$\mathbf{c}(n)=\mathbf{w}(n)-\mathbf{w}_o,\qquad \mathbf{R}\mathbf{w}_o=\mathbf{p}$$

Înlocuind în ecuația de reactualizare,

$$\mathbf{c}(n+1)=(\mathbf{I}-\mu\mathbf{R})\mathbf{c}(n)$$

$$\mathbf{c}(n+1)=\left(\mathbf{I}-\mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H\right)\mathbf{c}(n)$$

Vom introduce de asemenea ***vectorul eroare rotit***,

$$\mathbf{v}(n)=\mathbf{Q}^H\mathbf{c}(n)=\mathbf{Q}^H(\mathbf{w}(n)-\mathbf{w}_o)$$

Înmulțind la stânga cu $\mathbf{Q}^H$ ecuația recursivă obținută mai sus pentru vetorul eroare,

$$\mathbf{v}(n+1)=(\mathbf{I}-\mu\mathbf{\Lambda})\mathbf{v}(n)$$

Obținem o ecuație cu diferențe finite pentru acest vector, cu condițiile inițiale:

$$\mathbf{v}(0)=\mathbf{Q}^H(\mathbf{w}(0)-\mathbf{w}_o)$$

Ecuația cu diferențe finite poate fi exprimată în formă scalară, având în vedere caracterul de matrice diagonală al expresiei $(\mathbf{I}-\mu\mathbf{\Lambda})$, prin setul de relații

$$v_k(n+1)=(1-\mu\lambda_k)v_k(n),\qquad k=1,2,\cdots,N$$

cu soluția

$$v_k(n) = (1-\mu\lambda_k)^n v_k(0)$$

Convergenţa algoritmului impune ca $v_k(n)$ să tindă la 0 când $n \to \infty$. Pentru aceasta este necesar şi suficient ca rația progresiei geometrice respective să fie de modul subunitar:

$$|1-\mu\lambda_k| < 1 \quad \text{sau} \quad 0 < \mu < \frac{2}{\lambda_k}, \quad k = 1,2,\cdots N$$

evident îndeplinită dacă și numai dacă

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

După cum s-a văzut, o caracteristică importantă a unui filtru adaptiv este viteza de convergență, ceea ce în cazul de față reprezintă viteza cu care $v_k(n)$ tinde către zero, sau cu care $\mathbf{w}_k(n) \to \mathbf{w}_{o,k}$, $k = 1,\ldots,N$.

După cum s-a văzut, $\lambda_k \in \mathbf{R}_+$ așa încât $v_k(n)$ va descreşte către 0 fără oscilaţii. Distingem cazurile:

a) $0 < \mu\lambda_k < 1$ sau $0 < \mu < \frac{1}{\lambda_k}$, iar rația progresiei, $0 < 1-\mu\lambda_k < 1$ așa încât $v_k(n)$ descresc uniform (figura a).

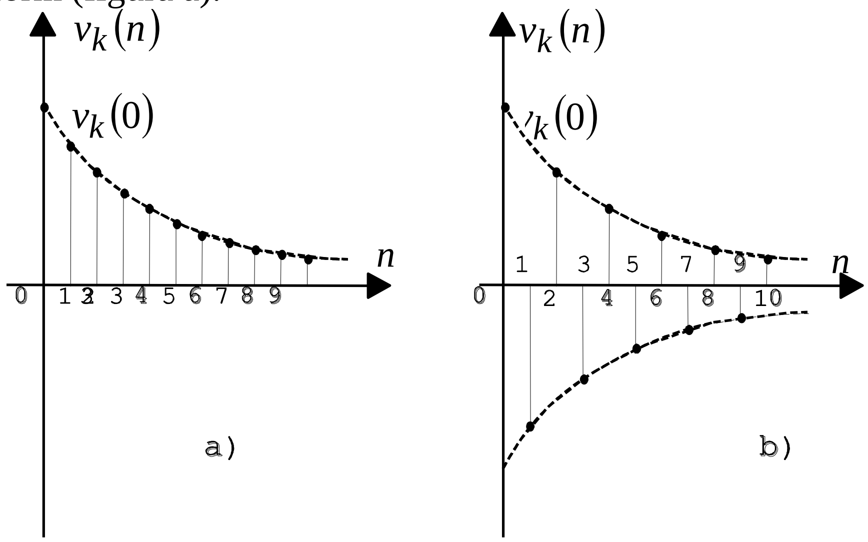

Figura 5.1. Evoluția erorii coeficinților, în cazurile a și b.

Se poate defini o constantă de timp, scriind

$$v_k(n) = (1-\mu\lambda_k)^n v_k(0) = \mathrm{e}^{-n/\tau_k}\, v_k(0),$$

De unde, logaritmând,

$$\tau_k = -\frac{1}{\ln(1-\mu\lambda_k)}.$$

Evident, cu cât constanta de timp este mai mică, cu atât convergența este mai rapidă. Acest lucru se întâmplă dacă μ se apropie de valoarea limită, $\mu \approx \frac{1}{\lambda_k}$. Din contra, o valoare foarte micăa a pasului la implica o convergență lentă.

b) $-1<1-\mu\lambda_k<0$ **sau** $\frac{1}{\lambda_k}<\mu<\frac{2}{\lambda_k}$ - $v_k(n)$ va fi reprezentat de un şir alternat, având în vedere că rația este negativă (fgura b),

$$v_k(n)=\mathrm{e}^{-n/\tau_k}(-1)^n v_k(0), \qquad \tau_k=-\frac{1}{\ln(\mu\lambda_k-1)}$$

Și în acest caz, viteza maximă se obține pentru $\mu\approx\frac{1}{\lambda_k}$, scăzând către limita superioară a intervalului.

c)$1-\mu\lambda_k=0 \;\Rightarrow\; v_k(n)=0, \quad n\geq 1$ - este o situație limită, ce ar putea fi îndeplinită doar pentru o valoare a lui *k*.
Dintre cele trei cazuri prezentate mai sus, importanță practică prezintă cazul a, care presupune alegerea pasului conform relației $0<\mu<\frac{1}{\lambda_{Max}}$. Cazul b nu conduce la avantaje privind viteza de convergență, dar prezintă un grad de risc, ca urmare a apropierii de pragul de convergență.

Relațiile obținute mai sus pot fi exprimate într-o formă unitară,

$$v_k(n)=\mathrm{e}^{-n/\tau_k}\, s_k^n v_k(0), \qquad \tau_k=-\frac{1}{\ln|1-\mu\lambda_k|}, \quad s_k=\mathrm{sgn}(1-\mu\lambda_k)$$

Coeficienţii $w_k(n)$ se pot exprima sub forma

$$\mathbf{w}(n)=\mathbf{w}_o+\mathbf{Q}\mathbf{v}(n)=\mathbf{w}_o+\sum_{k=1}^{N}\mathbf{q}_k v_k(n)$$

$$w_i(n)=w_{oi}+\sum_{k=1}^{N}\left(q_{ik}v_k(0)(1-\mu\lambda_k)^n\right)=w_{oi}+\sum_{k=1}^{N}\left(q_{ik}v_k(0)s_k^n\,\mathrm{e}^{-n/\tau_k}\right)$$

Aceste formule conduc la următoarele concluzii

- Convergenţa coeficienţilor are loc după o sumă ponderată de exponenţiale.
- Se obţine viteza de convergenţă maximă atunci când $q_{ik}v_k(0)$ sunt nuli, pentru toţi *k*, exceptând valoarea corespunzătoare lui λ_{Max}.
- Dacă nu sunt precizate condiţiile iniţiale, în cazul cel mai defavorabil, viteza de convergenţă este determinată de $\lambda_{\min}$, pentru care se obține constanta de timp maximă. Ca urmare a celor arătate mai sus, să presupunem o alegere a pasului de forma

$$\mu=r\frac{1}{\lambda_{\max}}, \qquad 0<r<1$$

$$w_i(n) = w_{oi} + \sum_{k=1}^{N}\left(q_{ik} v_k(0)\left(1 - r\frac{\lambda_k}{\lambda_{\max}}\right)^n \right)$$

Termenul cel mai lent descrescător este acela care conţine factorul

$$\left(1 - r\frac{\lambda_{\min}}{\lambda_{\max}}\right)^n$$

Acesta scade cu atât mai lent cu cât raportul $\lambda_{Max}/\lambda_{min}$ este mai mare, ceea ce se întâmplă dacă matricea **R** este rău condiţionată.

Analiza convergenței funcției cost

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

$$e(n) = d(n) - \mathbf{w}^H(n)\mathbf{x}(n) = d(n) - \mathbf{w}_o^H(n)\mathbf{x}(n) - \mathbf{c}^H(n)\mathbf{x}(n) = e_o(n) - \mathbf{c}^H(n)\mathbf{x}(n)$$

$$J(n) = \mathrm{E}\left\{\left(e_o(n) - \mathbf{c}^H(n)\mathbf{x}(n)\right)\left(e_o^*(n) - \mathbf{x}^H(n)\mathbf{c}(n)\right)\right\} = \mathrm{E}\left\{e_o(n)e_o^*(n)\right\} -$$
$$- \mathrm{E}\left\{e_o(n)\mathbf{x}^H(n)\mathbf{c}(n)\right\} - \mathrm{E}\left\{\mathbf{c}^H(n)\mathbf{x}(n)e_o^*(n)\right\} + \mathrm{E}\left\{\mathbf{c}^H(n)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{c}(n)\right\}$$

Având în vedere că $\mathbf{c}(n)$ este determinist, iar e_o satisface principiul ortogonalităţii,

$$J(n) = J_{\min} + \mathbf{c}^H(n)\mathbf{R}\mathbf{c}(n)$$

sau

$$J(n) = J_{\min} + \mathbf{c}^H(n)\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H\mathbf{c}(n) = J_{\min} + \mathbf{v}^H(n)\boldsymbol{\Lambda}\mathbf{v}(n)$$

sau scalar:

$$J(n) = J_{\min} + \sum_{k=1}^{N}\lambda_k |v_k(n)|^2 = J_{\min} + \sum_{k=1}^{N}\lambda_k (1 - \mu\lambda_k)^{2n} |v_k(0)|^2 =$$
$$= J_{\min} + \sum_{k=1}^{N}\lambda_k\, \mathrm{e}^{-2n/\tau_k} |v_k(0)|^2$$

Curba obţinută reprezentând $J(n)$ este numită *curba de învăţare* a algoritmului. Se constată că indiferent de condiţiile iniţiale $v_k(0)$, eroarea medie pătratică tinde către $J_{\min}$ dacă este îndeplinită condiţia de convergenţă.

De exemplu, pentru

$$N = 2 \quad \Rightarrow \quad J(n) - J_{\min} = \lambda_1 v_1^2(n) + \lambda_2 v_2^2(n)$$

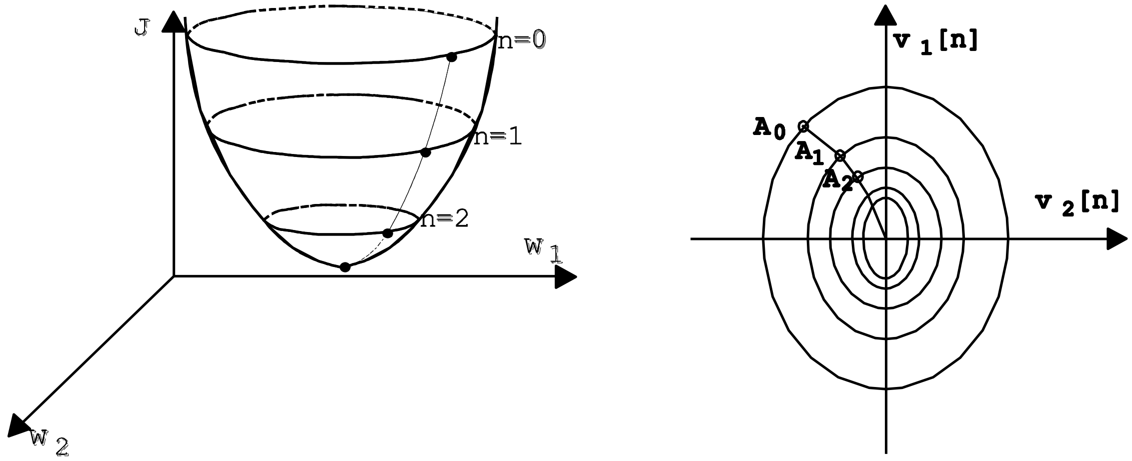

Figura 5.2 Aspectul funcției cost(stânga) și a unor secțiuni cu plane de *J*=const

Intersecţiile cu plane *J*=const sunt elipse cu semiaxele (Figura 5.2)

$$a=\left(\frac{J(n)-J_{\min}}{\lambda_1}\right)^{1/2}, \quad b=\left(\frac{J(n)-J_{\min}}{\lambda_2}\right)^{1/2}$$

5.2 Algoritmul gradientului stohastic

În cazul metodei SD ajustarea coeficienţilor se face pe baza gradientului erorii medii pătratice:

$$\nabla(J(n))=-\mathbf{p}+\mathbf{R}\mathbf{w}(n); \quad \mathbf{R}=\mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}; \quad \mathbf{p}=\mathrm{E}\{\mathbf{x}(n)d^*(n)\}$$

Mediile statistice în general nu sunt însă cunoscute. Se recurge la o estimare a gradientului utilizând nişte valori estimate pentru cele două matrice renunţând la operaţiile de mediere statistică:

$$\hat{\mathbf{R}}(n)=\mathbf{x}(n)\mathbf{x}^H(n); \quad \hat{\mathbf{p}}(n)=\mathbf{x}(n)d^*(n)$$

$$\hat{\nabla}(J(n))=-\mathbf{x}(n)d^*(n)+\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n)=-\mathbf{x}(n)e^*(n)$$

$$\boxed{\begin{gathered}\mathbf{w}(n+1)=\mathbf{w}(n)+\mu\mathbf{x}(n)\left(d^*(n)-\mathbf{x}^H(n)\mathbf{w}(n)\right)\\ \mathbf{w}(n+1)=\mathbf{w}(n)+\mu\mathbf{x}(n)e^*(n)\end{gathered}}$$

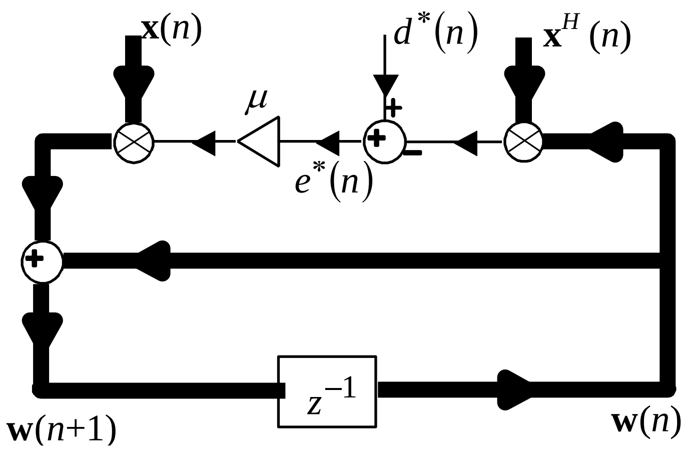

Figura 5.3. Reprezentare sub forma de graf a algoritmului LMS

$$\mathbf{w}(0)=\mathbf{0}$$
for $n=0,1,2,\cdots$
$$y(n)=\mathbf{w}^H(n)\mathbf{x}(n)$$
$$e(n)=d(n)-y(n)$$
$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu\mathbf{x}(n)e^*(n)$$
end

Observaţii

- Complexitate aritmetică: $2N+1$ înmulţiri şi $2N$ adunări pentru fiecare iteraţie
- Având în vedere criteriul de optimizare, se întâlneşte în limba engleză sub denumirea L*east* M*ean* S*quare* (LMS).
- Fiind calculat de fiecare dată utilizând setul de eşantioane **x**[n], ce au un caracter aleator, fără a efectua o mediere, gradientul estimat va avea de asemenea un caracter aleator.
- Înmulţirile cu **x**[n] din schema algoritmului dau acestuia un caracter neliniar.

Analiza convergenţei

În principiu, vom aborda problema convergenței algoritmului din aceleși două puncte de vedere, convergența coeficienților și convergența funcției de transfer, ca ți în cazul algoritmului SD. Intervine însă o diferență esențială. În cazul SD gradientul ce permitea reactualizarea coeficienților avea un caracter determinist, fiind media statistică a unor mărimi presupuse staționare. În cazul LMS, după cum s-a arătat, ca urmare a renunțării la mediere, gradientul va avea un caracter stohastic și la fel și coeficienții estimați la un moment dat și erorile acestora. Cu reformulările de rigoare, cele două abordări devin

- Convergența în medie a coeficienților. Va trebui testat în ce măsură tinde valoarea medie a vectorului **w**(n) către $\mathbf{w}_o$ atunci când $n \to \infty$. În caz afirmativ, se zice că algoritmul este *convergent în medie*.
- Convergența funcției cost. Va trebui testat în ce măsură tinde J(n) către o valoare finită atunci când $n \to \infty$? În caz afirmativ se spune că algoritmul este *convergent în medie pătratică*.

Caracterul aleatoriu a mărimilor ce intră în discuție, face ca analiza să fie foarte dificilă într-un caz foarte general, fără a face unele ipoteze asupra mărimilor de intrare. Un set uzual de ipoteze e cunoscut în literatură sub denumirea de

Ipoteze de independenţă:

- $\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(n)$ sunt statistic independenţi.
- $d(n)$ este statistic independent faţă de $d(1),...,d(\text{n-1})$ și faţă de $x(1), x(2), \cdots, x(n-1)$.
- vectorul de intrare **x**(n) şi d(n) formează împreună un set de variabile aleatoare gaussiene.

Desigur, aceste ipoteze sunt discutabile, dar ele ușurează analiza, iar rezultatele obținute sunt în general verificate în practică. În plus, pentru analiza convergenței vom presupune procesele aleatoare **x**(n) şi d(n) staționare și de medie nulă. Ultima dintre ipoteze, asociată cu aceea de medie nulă pentru semnalele de intrare implică echivalența dintre noțiunile de ortogonalitate, necorelare și independență statistică. Pornind de la aceste ipoteze și de la relațiile de reactualizare a acoeficienților,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n-1) e^*(n-1)$$

rezultă

$$E\{\mathbf{w}^H(n)\mathbf{x}(n)\} = 0 \text{ și } E\{\mathbf{c}^H(n)\mathbf{x}(n)\} = 0$$

deci independența vectorului eroare a coeficienților față de vectorul intrării la orice n.

Datorită lipsei medierii în calculul gradientului apare un *zgomot de gradient*, $\mathbf{N}(n)$,

$$\hat{\nabla}(J(n)) = \nabla(J(n)) + \mathbf{N}(n)$$

$$\mathbf{N}(n) = \hat{\nabla}(J(n)) - \mathrm{E}\{\hat{\nabla}(J(n))\} = -(\mathbf{x}(n)e^*(n) - \mathrm{E}\{\mathbf{x}(n)e^*(n)\})$$

Evident, este un vector de valoare medie nulă.

$$\mathrm{E}\{\mathbf{N}(n)\} = \mathbf{0}$$

Analiza convergenței în medie a coeficienților

Relaţia de reactualizare a coeficienţilor devine

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu(\nabla J(n) + \mathbf{N}(n)) = \mathbf{w}(n) + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}(n)) - \mu\mathbf{N}(n)$$

sau, introducând vectorul eroare,

$$\mathbf{w}_o + \mathbf{c}(n+1) = \mathbf{w}_o + \mathbf{c}(n) + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}_o - \mathbf{R}\mathbf{c}(n)) - \frac{1}{2}\mu\mathbf{N}(n)$$

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{R})\mathbf{c}(n) - \frac{1}{2}\mu\mathbf{N}(n)$$

Înmulțind la stânga cu $\mathbf{Q}^H$,

$$\mathbf{v}(n+1) = (\mathbf{I} - \mu\mathbf{\Lambda})\mathbf{v}(n) - \frac{1}{2}\mu\mathbf{Q}^H\mathbf{N}(n)$$

$$\mathrm{E}\{\mathbf{v}(n+1)\} = (\mathbf{I} - \mu\mathbf{\Lambda})\mathrm{E}\{\mathbf{v}(n)\}$$

Având în vedere analogia formală dintre această ecuaţie şi cea corespunzătoare în cazul algoritmului SD, concluziile trase acolo pentru vectorul eroare a coeficienţilor $\mathbf{v}(n)$se pot transpune aici pentru media acestui vector, $\mathrm{E}\{\mathbf{v}(n)\}$. Algoritmul LMS este deci ***convergent în medie*** dacă:

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

Observație. Trebuie însă reținută diferența esențială între convergența coeficienților în cazul SD și LMS. În cazul SD, era o scădere monotonă a erorii coeficienților, după o sumă de exponențiale, în timp ce în cazul LMS, media media acestei erori are această proprietate, în timp ce eroarea instantanee variază aleatoriu în jurul acestei medii (figura 5.4).

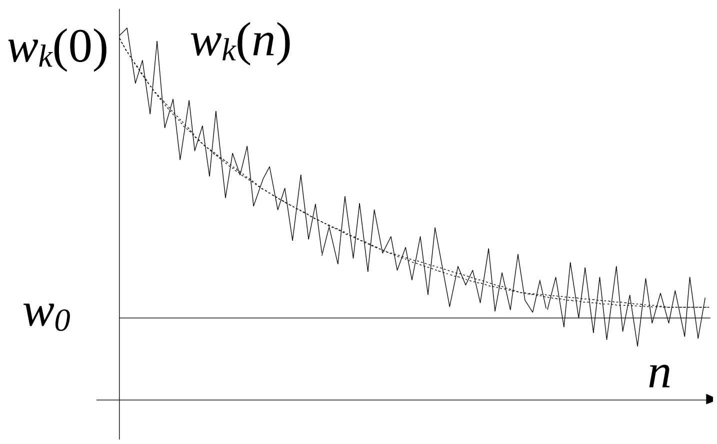

Figura 5.4 Comparație între convergența coeficienților în algoritmul SD (curba monotona) și în algoritmul LMS

Analiza convergenţei funcţiei cost (în medie pătratică)

După cum s-a văzut în cazul SD, funcția cost poate fi scrisă

$$J(n) = J_{\min} + \mathrm{E}\{\mathbf{c}^H(n)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{c}(n)\}$$

Dar,

$$\mathrm{E}\{\mathbf{c}^H(n)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{c}(n)\} = \mathrm{E}\{\mathrm{tr}\{\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{c}(n)\mathbf{c}^H(n)\}\} =$$
$$= \mathrm{tr}\{\mathrm{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}\mathrm{E}\{\mathbf{c}(n)\mathbf{c}^H(n)\}\} = \mathrm{tr}\{\mathbf{RC}(n)\}$$

unde

$$\mathbf{C}(n) = \mathrm{E}\{\mathbf{c}(n)\mathbf{c}^H(n)\}$$

În deducerea relației de mai sus s-au avut în vedere următoarele considerente
Fie doi vectori $\mathbf{u} = [u_1, \cdots, u_N]^T$, $\mathbf{v} = [v_1, \cdots, v_N]^T$, atunci,

$$\mathbf{v}^H\mathbf{u} = \mathrm{tr}\{\mathbf{u}\mathbf{v}^H\} = \sum_{i=1}^{N} u_i v_i^*$$

egalitate ce se aplică pentru $\mathbf{v}^H = \mathbf{c}^H(n)$, $\mathbf{u} = \mathbf{x}(n)\mathbf{x}^H(n)\mathbf{c}(n)$.

$$J(n) = J_{\min} + \mathrm{tr}\{\mathbf{RC}(n)\} = J_{\min} + J_{ex}(n)$$

unde $J_{ex}(n)$ reprezintă o ***eroare în exces*** în raport cu filtrul optimal.

$$J(n) = J_{\min} + \mathrm{tr}\{\mathbf{Q\Lambda Q}^H\mathbf{C}(n)\} = J_{\min} + \mathrm{tr}\{\mathbf{\Lambda Q}^H\mathbf{C}(n)\mathbf{Q}\} = J_{\min} + \mathrm{tr}\{\mathbf{\Lambda K}(n)\}$$

unde

$$\mathbf{K}(n) = \mathrm{E}\{\mathbf{v}(n)\mathbf{v}^H(n)\} = \mathbf{Q}^H\mathbf{C}(n)\mathbf{Q}$$

$$J_{ex}(n) = \sum_{i=1}^{N}\lambda_i k_{ii}(n)$$

În raport cu filtrul optimal, apare deci o eroare medie pătratică suplimentară, sau în exces, notată cu J_{ex}, ce poate fi pusă pe seama zgomotului de gradient. Pentru evaluarea sa sunt necesari termenii de pe diagonala principală a matricei $\mathbf{K}(n)$.

$$J_{ex}(n) = J_{ex}(\infty) + J_{tr}(n)$$

$$J_{ex}(\infty) = J_{\min}\frac{\sum_{i=1}^{N}\frac{\mu\lambda_i}{2-\mu\lambda_i}}{1-\sum_{i=1}^{N}\frac{\mu\lambda_i}{2-\mu\lambda_i}}$$

Se defineşte ***dezadaptarea*** (misadjustment) prin

$$M = \frac{J_{ex}(\infty)}{J_{\min}} = \frac{\sum_{i=1}^{N}\frac{\mu\lambda_i}{2-\mu\lambda_i}}{1-\sum_{i=1}^{N}\frac{\mu\lambda_i}{2-\mu\lambda_i}}$$

Pentru $\mu << \frac{2}{\lambda_{\max}}$

$$M \cong \frac{\mu}{2}\sum_{i=1}^{N}\lambda_i = \frac{\mu N}{2} r(0) = \frac{\mu N}{2}\lambda_{\text{med}}, \quad \text{unde} \quad \lambda_{\text{med}} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i$$

Are în general o tendinţă de scădere cu micşorarea pasului μ, de creştere cu ordinul filtrului *N*, şi e proporţional cu puterea medie a semnalului de intrare.

Componenta tranzitorie $J_{tr}(n) \to 0$ când $n \to \infty$ dacă

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad \cap \quad \sum_{i=1}^{N}\frac{\mu\lambda_i}{2-\mu\lambda_i} < 1$$

Dacă $0 < \mu << 1$ ambele condiţii sunt îndeplinite dacă

$$0 < \mu < \frac{2}{\sum_{j=1}^{N}\lambda_j}$$

$$\sum_{j=1}^{N}\lambda_j = Nr(0), \quad r(0) = \mathrm{E}\{x^*(n-k)x(n-k)\} = P_x, \quad k = 0,1,\cdots,N-1$$

P_x reprezintă ***puterea secvenţei*** $x(n)$, deci o formă simplificată a condiţiei de convergenţă este:

$$0 < \mu < \frac{2}{NP_x}; \qquad M = \frac{\mu N}{2}P_x$$

Se poate introduce o constantă de timp medie:

$$\tau_{\text{med}} = \frac{1}{2\mu\lambda_{\text{med}}}$$

care caracterizează viteza de scădere a părţii tranzitorii a erorii.

Se constată că dacă μ este mic, constanta de timp e mare, conducând la o adaptare lentă, dar dezadaptarea este mică.

Curba ce reprezintă funcția cost, deci eroarea pătratică medie, este denumită și *curba de învățare* (*learning curve*)

5.3. Metoda gradientului stohastic normalizat
(Metoda normalizată a minimizării erorii medii pătratice - NLMS)

Poate fi privită ca o problemă de optimizare cu constrângeri. Ne propunem să determinăm noile valori ***w***(n+1) ale coeficienţilor astfel încât să se minimizeze norma euclidiană a variaţiei:

$$\delta\mathbf{w}(n+1)=\mathbf{w}(n+1)-\mathbf{w}(n)$$

cu condiţia ca:

$$\mathbf{w}^H(n+1)\mathbf{x}(n)=d(n) \tag{1}$$

deci noii coeficienţi să aibă acele valori care, cu un tact mai înainte, ar fi anulat eroarea. Vom constitui ***funcţia cost*** reală:

$$J(n)=\|\delta\mathbf{w}(n+1)\|^2+\mathrm{Re}\{\lambda[\mathbf{w}^H(n+1)\mathbf{x}(n)-d(n)]\}$$

care îşi atinge minimul odată cu $\|\delta\mathbf{w}(n+1)\|$ dacă este îndeplinită condiţia (1)

$$\begin{aligned}J(n)&=\left(\mathbf{w}^H(n+1)-\mathbf{w}^H(n)\right)\left(\mathbf{w}(n+1)-\mathbf{w}(n)\right)+\\&+\frac{1}{2}\left[\lambda\left(\mathbf{w}^H(n+1)\mathbf{x}(n)-d(n)\right)+\lambda^*\left(\mathbf{x}^H(n)\mathbf{w}(n+1)-d^*(n)\right)\right]=\\&=\mathbf{w}^H(n+1)\mathbf{w}(n+1)-\mathbf{w}^H(n+1)\mathbf{w}(n)-\mathbf{w}^H(n)\mathbf{w}(n+1)+\mathbf{w}^H(n)\mathbf{w}(n)+\\&+\frac{1}{2}\left(\lambda\mathbf{w}^H(n+1)\mathbf{x}(n)+\lambda^*\mathbf{x}^H(n)\mathbf{w}(n+1)\right)-\frac{1}{2}\left(\lambda d(n)+\lambda^* d^*(n)\right)\end{aligned}$$

Pentru a găsi vectorul **w**(n+1) ce minimizează această expresie vom aplica metoda gradientului complex.

$$\nabla\left(\mathbf{w}^H(n+1)\mathbf{w}(n+1)\right)=2\mathbf{w}(n+1)$$
$$\nabla\left(\mathbf{w}^H(n+1)\mathbf{w}(n)+\mathbf{w}^H(n)\mathbf{w}(n+1)\right)=2\mathbf{w}(n)$$
$$\nabla\left(\lambda\mathbf{w}^H(n+1)\mathbf{x}(n)+\lambda^*\mathbf{x}^H(n)\mathbf{w}(n+1)\right)=2\lambda\mathbf{x}(n)$$

Rezultă:

$$\mathbf{w}(n+1)-\mathbf{w}(n)=-\frac{1}{2}\lambda\mathbf{x}(n)$$

λ se obţine punând condiţia (1). Pentru aceasta se înmulţeşte la stânga ultima relaţie cu $\mathbf{x}^H(n)$:

$$\mathbf{x}^H(n)\mathbf{w}(n+1)-\mathbf{x}^H(n)\mathbf{w}(n)=-\frac{1}{2}\lambda\mathbf{x}^H(n)\mathbf{x}(n)=-\frac{1}{2}\lambda\|\mathbf{x}(n)\|^2$$

$$\lambda=-\frac{2}{\|\mathbf{x}(n)\|^2}\left(d^*(n)-\mathbf{x}^H(n)\mathbf{w}(n)\right)=-\frac{2}{\|\mathbf{x}(n)\|^2}e^*(n)$$

Noii coeficienţi se vor calcula deci cu formula:

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\frac{1}{\|\mathbf{x}(n)\|^2}\mathbf{x}(n)e^*(n)$$

Se obişnuieşte să se introducă o scalare a pasului cu o constantă $\bar{\mu}$, deci:

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\frac{\bar{\mu}}{\|\mathbf{x}(n)\|^2}\mathbf{x}(n)e^*(n)$$

Poate fi echivalat cu algoritmul gradientului stohastic pentru:

$$\mu(n)=\frac{\bar{\mu}}{\|\mathbf{x}(n)\|^2}$$

în care pasul este variabil.
Deoarece putem aproxima, în cazul unor filtre de lungime mare

$$\|\mathbf{x}(n)\|^2 \cong N\sigma_x^2 = NP_x$$

Condiția de convergență dedusă pentru LMS conduce la:

$$0<\bar{\mu}<2$$

Evită prin normare *amplificarea zgomotului gradientului*, în expresia coeficienţilor $\mathbf{w}(n+1)$.
Apar în plus un număr de N înmulţiri şi N-1 adunări la calculul fiecărui coeficient, ca urmare a necesităţii evaluării lui $\|x(n)\|^2$. Eventual

$$\|x(n)\|^2 = \sum_{k=0}^{N-1}|x(n-k)|^2 = \|x(n-1)\|^2 + |x(n)|^2 - |x(n-N)|^2$$

Pentru a elimina riscul unei eventuale împărțiri prin zero (absență a semnalului pe o durată de cel puțin N tacte), formula de reactualizare utilizată este

$$\boxed{\mathbf{w}(n+1)=\mathbf{w}(n)+\frac{\bar{\mu}}{\delta+\|\mathbf{x}(n)\|^2}\mathbf{x}(n)e^*(n); \qquad \delta>0}$$

în care s-a introdus *factorul de regularizare* δ. Dacă acesta este privit doar ca o măsură de a elimina riscul împărțirii prin zero, el poate avea o valoare pozitivă mică, astfel încât să nu afecteze viteza de convergență.

Observații.

- Această variantă a algoritmului LMS este mai mult folosită decât cea originală, în primul rând pentru că nu necesită o estimare apriori a puterii semnalului de intrare pentru a putea alege valoarea pasului μ.
- Există în acest caz două grade de libertate în configurarea algoritmului, reprezentate prin parametrii μ si δ. Ei au în principiu acțiuni contrare, având în vedere ca pasul echivalent este $\mu(n)=\frac{\bar{\mu}}{\delta+\|\mathbf{x}(n)\|^2}$. Valori mari pentru $\bar{\mu}$, apropiate de valoarea 1, conduc la o convergență rapidă, dar și la o

dezadaptare redusă. Efectul lui δ este invers. O alegere corectă presupune un compromis rezonabil între cele două criterii.

- Apare atunci ideea unui algoritm cu pas $\bar{\mu}$ variabil. Într-adevăr, la pornirea algoritmului, ar fi util un $\bar{\mu}$ mare pentru a avea o convergență rapidă. O situație simnilară apare când se produce o schimbare bruscă a mediului. După intrarea în convergență, valoarea lui $\bar{\mu}$ poate fi redusă, urmărind reducerea erorii în exces.
- O abordare alternativă are în vedere optimizarea regularizării.

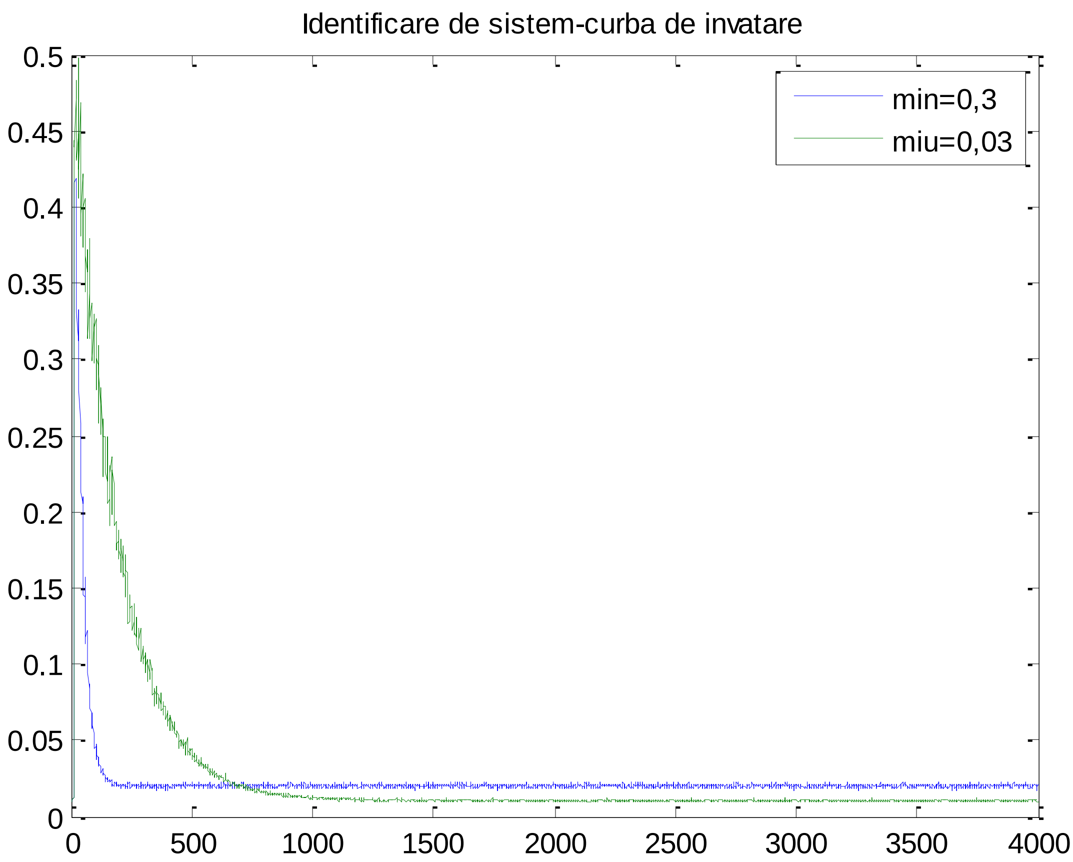

Figura 5.5 Evoluția funcției cost pentru două valori ale pasului

- O problemă caracteristică a algoritmilor te tip SD și a celor derivați din acesta, este convergența lentă în cazul unor semmnale rău condiționate, deci a unor semnale de intrare puternic corelate.

5.4 Metoda gradientului pentru structuri latice (GAL – Gradient adaptive lattice)

Algoritmii LMS/NLMS prezentați până aici presupuneau o realizare a fitrului adaptiv cu răspuns finit la impuls în formă transversală. În această secțiune se vor prezenta algoritmii de funcționare mai întâi pentru un predictor în forma latice, iar apoi, pentru un filtru adaptiv.

5.4.1 *Predictor adaptiv în structura latice*

Datorită structurii modulare, problema se rerzolvă separat, pentru fiecare celulă în parte. Vom considera deci o celulă, *m* din structura filtrului erorii de predicție (figura 5.6), pentru care singurul parametru ce trebuie reglat este coeficientul de reflexie k_m . Deoarece filtrul furnizeză ambele erori de predicție, în constituirea funcției cost se poate porni de la eroarea de predicţie directă, inversă, sau compusă (o formă care să le includă pe amândouă). În cele ce urmează ne vom propune să determinăm k_m aşa încât să fie minimizată eroarea medie pătratică de predicţie, în forma compusă,

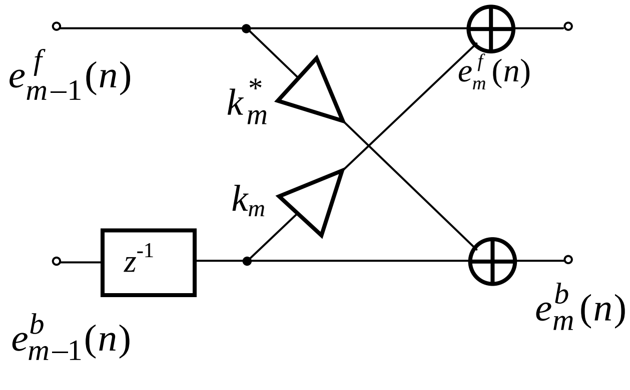

Figura 5.6 Celula latice

$$J_m(n)=\mathrm{E}\left\{\left|e^f_m(n)\right|^2+\left|e^b_m(n)\right|^2\right\}$$

Abordarea SD-GAL

Conform metodei gradientului, va trebui luat $k_m(n+1)$,

$$k_m(n+1)=k_m(n)-\mu_m\nabla_m(J_m(n))$$

unde s-a notat simplificat $\nabla_m(J_m(n))=\nabla_{k^*_m}(J_m(n))$ gradientul complex (derivata) în raport cu $k^*_m(n)$:

$$\nabla_m(J_m(n))=\mathrm{E}\left\{\nabla_m\left(e^{f*}_m(n)\right)e^f_m(n)+e^{f*}_m(n)\nabla_m\left(e^f_m(n)\right)\right\}+$$
$$+\mathrm{E}\left\{\nabla_m\left(e^{b*}_m(n)\right)e^b_m(n)+e^{b*}_m(n)\nabla_m\left(e^b_m(n)\right)\right\}$$

Folosind relaţiile de recurenţă:

$$\begin{aligned} e_m^f(n) &= e_{m-1}^f(n) + k_m e_{m-1}^b(n-1) \\ e_m^b(n) &= k_m^* e_{m-1}^f(n) + e_{m-1}^b(n-1) \end{aligned} \qquad (2)$$

$$k_m = k_m^r + \mathrm{j}k_m^i$$

$$\nabla_m\left(e_m^{f*}(n)\right) = \frac{1}{2}\left(\frac{\partial}{\partial k_m^r} + \mathrm{j}\frac{\partial}{\partial k_m^i}\right)\left(k_m^r - \mathrm{j}k_m^i\right)e_{m-1}^{b*}(n-1) = e_{m-1}^{b*}(n-1)$$

$$\nabla_m\left(e_m^f(n)\right) = \frac{1}{2}\left(\frac{\partial}{\partial k_m^r} + \mathrm{j}\frac{\partial}{\partial k_m^i}\right)\left(k_m^r + \mathrm{j}k_m^i\right)e_{m-1}^b(n-1) = 0$$

$$\nabla_m\left(e_m^b(n)\right) = \frac{1}{2}\left(\frac{\partial}{\partial k_m^r} + \mathrm{j}\frac{\partial}{\partial k_m^i}\right)\left(k_m^r - \mathrm{j}k_m^i\right)e_{m-1}^f(n) = e_{m-1}^f(n)$$

$$\nabla_m\left(e_m^{b*}(n)\right) = \frac{1}{2}\left(\frac{\partial}{\partial k_m^r} + \mathrm{j}\frac{\partial}{\partial k_m^i}\right)\left(k_m^r + \mathrm{j}k_m^i\right)e_{m-1}^{f*}(n) = 0$$

deci expresia gradientului este

$$\nabla_m(J(n)) = \mathrm{E}\left\{e_m^f(n)e_{m-1}^{b*}(n-1) + e_m^{b*}(n)e_{m-1}^f(n)\right\}$$

Rezultă:

$$k_m(n+1) = k_m(n) - \mu_m\,\mathrm{E}\left\{e_m^f(n)e_{m-1}^{b*}(n-1) + e_m^{b*}(n)e_{m-1}^f(n)\right\}$$

sau utilizând relațiile de recurență (2)

$$k_m(n+1) = k_m(n)\left[1 - \mu_m\,\mathrm{E}\left\{\left|e_{m-1}^b(n-1)\right|^2 + \left|e_{m-1}^f(n)\right|^2\right\}\right] - 2\mu_m\,\mathrm{E}\left\{e_{m-1}^f(n)e_{m-1}^{b*}(n-1)\right\}$$

Condiţia de stabilitate este:

$$\left|1 - \mu_m\,\mathrm{E}\left\{\left|e_{m-1}^b(n-1)\right|^2 + \left|e_{m-1}^f(n)\right|^2\right\}\right| < 1$$

Deci

$$0 < \mu_m < \frac{2}{\mathrm{E}\left\{\left|e_{m-1}^b(n-1)\right|^2 + \left|e_{m-1}^f(n)\right|^2\right\}}$$

Abordarea LMS-GAL

Cum însă mediile statistice nu sunt în general cunoscute, se preferă de obicei utilizarea *gradientului stohastic*, renunţând la operaţiile de mediere:

$$k_m(n+1) = k_m(n) - \mu_m\left(e_m^f(n)e_{m-1}^{b*}(n-1) + e_m^{b*}(n)e_{m-1}^f(n)\right)$$

Se poate utiliza un algoritm normalizat, normând incrementul la energia erorii de predicţie compusă, corespunzătoare intrărilor celulei m:

$$\mu_m = \mu_m(n) = \frac{1}{W_{m-1}(n)}$$

$$W_{m-1}(n) = \sum_{i=1}^{n}\left[\left|e_{m-1}^f(i)\right|^2 + \left|e_{m-1}^b(i-1)\right|^2\right] =$$

$$= W_{m-1}(n-1) + \left|e_{m-1}^f(n)\right|^2 + \left|e_{m-1}^b(n-1)\right|^2$$

Uneori se introduce un parametru 0<ß<1, care permite alocarea unor ponderi diferite pentru eşantioanele precedente şi cele actuale:

$$W_{m-1}(n) = \beta W_{m-1}(n-1) + (1-\beta)\left(\left|e_{m-1}^{f}(n)\right|^2 + \left|e_{m-1}^{b}(n-1)\right|^2\right)$$

In general se obţine o convergenţă mai rapidă utilizând structuri latice decât structuri transversale.

Suprafeţele $J(n)$ sunt pătratice în coeficienţii de reflexie, având un minim, dar secţiunile cu J=const. nu mai sunt în general elipse. Trebuie remarcat că fiecare coeficient k_m se evaluează independent urmărind minimizarea lui $J_m(n)$.

5.4.1 *Filtre adaptive în structura latice*

Algoritmul precedent permitea obţinerea unui predictor. Pentru a realiza un filtru adaptiv care să estimeze un semnal dorit, structura latice a predictorului se completează cu o structură în scară (figura 5.7).

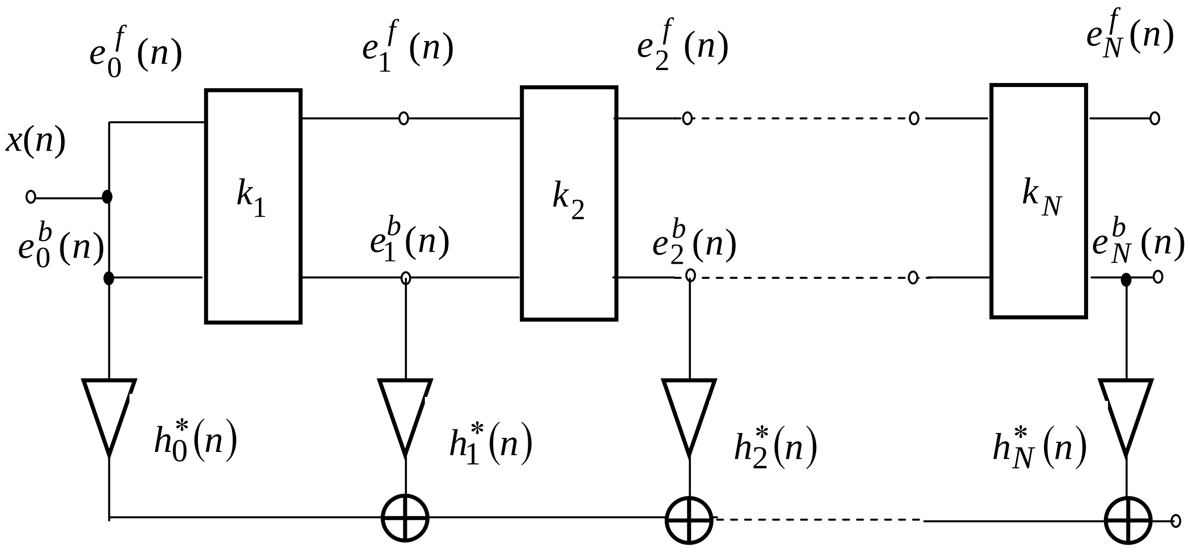

Figura 5.7 Filtru adaptiv în structura latice

În partea superioară a schemei se observă un predictor lattice, al cărui rol este de a furniza semnale decorelate $e_i^b(n),\ i = 0,\cdots,N$ pentru structura în scară. Structura în scară efectuează o însumare ponderată a acestora, cu ponderile $h_i^*(n),\ i = 0,\cdots,N$ astfel reglate încât ieşirea să estimeze cât mai corect un semnal dorit $d(n)$. Cu notațiile

$$\mathbf{h}(n) = (h_0(n), h_1(n), \cdots, h_N(n))^T$$
$$\mathbf{e}_N^b(n) = \left(e_o^b(n), e_1^b(n), \cdots, e_N^b(n)\right)^T$$

eroarea poate fi exprimată prin

$$e(n) = d(n) - y(n) = d(n) - \mathbf{h}^H(n)\mathbf{e}_N^b(n)$$

Coeficienţii k_i se calculează cu algoritmul GAL, iar h_i din condiţia minimizării funcţiei cost

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

Conform ecuaţiei Wiener-Hopf, coeficienţii optimi sunt daţi de

$$\mathbf{R}_e \mathbf{h}_o = \mathbf{r}_{ed}$$

unde

$$\mathbf{R}_e = \mathrm{E}\left\{\mathbf{e}_N^b(n)\mathbf{e}_N^{bH}(n)\right\}; \quad \mathbf{r}_{ed} = \mathrm{E}\left\{\mathbf{e}_N^b(n)d^*(n)\right\}$$

În cazul coeficienţilor k_i optimi, ieşirile laticei sunt ortogonale,

$$\mathrm{E}\left\{e_r^b(n)e_l^{b*}(n)\right\} = \begin{cases} 0, & r \neq l \\ \mathrm{E}\left(\left|e_r^b(n)\right|^2\right) = P_r, & r = l \end{cases}$$

Demonstrație

$$e_r^b(n) = \sum_{k=0}^{r} c_{r,k}\, x(n-k)$$

deci

$$\mathrm{E}\left\{e_r^b(n)e_l^{b*}(n)\right\} = \mathrm{E}\left\{\sum_{k=0}^{r} c_{r,k}\, x(n-k)e_l^{b*}(n)\right\} = \sum_{k=0}^{r} c_{r,k}\, \mathrm{E}\left\{x(n-k)e_l^{b*}(n)\right\}$$

Dar, conform ortogonalității

$$\mathrm{E}\left\{x(n-k)e_l^{b*}(n)\right\} = 0,\ k = 0,\cdots,l-1$$

Rezultă că

$$\mathrm{E}\left\{e_r^b(n)e_l^{b*}(n)\right\} = 0 \text{ pentru } r < l.$$

În mod asemănător, exprimând

$$e_l^b(n) = \sum_{k=0}^{l} c_{l,k}\, x(n-k)$$

se deduce egalitatea cu 0 pentru $l < r$. Rezultă că

$$\mathbf{R}_e = \mathrm{diag}\left\{P_0, P_1, \cdots, P_N\right\}$$

$$h_{o,k} = \frac{1}{P_k}\mathrm{E}\left\{e_k^b(n)d^*(n)\right\}$$

Pentru un algoritm recursiv, recurgem la metoda gradientului. În varianta SD:

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu\nabla_{\mathbf{h}^*} J(n)$$

$$\nabla_{\mathbf{h}^*}(J(n)) = -\mathbf{r}_{ed} + \mathbf{R}_e\mathbf{h}(n) =$$

$$= -\mathrm{E}\left\{\mathbf{e}_N^b(n)d^*(n)\right\} + \mathrm{E}\left\{\mathbf{e}_N^b(n)\mathbf{e}_N^{bH}(n)\right\}\mathbf{h}(n)$$

Cum valorile medii nu sunt în general cunoscute, aplicăm metoda gradientului stohastic (LMS),

$$\hat{\nabla}_{\mathbf{h}^*}(J(n)) = -\mathbf{e}_N^b(n)d^*(n) + \mathbf{e}_N^b(n)\mathbf{e}_N^{bH}(n)\mathbf{h}(n) = -\mathbf{e}_N^b(n)e^*(n)$$

rezultând ecuația de reactulizare a ponderilor

$$\mathbf{h}(n+1)=\mathbf{h}(n)+\mu e^{*}(n)\mathbf{e}_{N}^{b}(n)$$

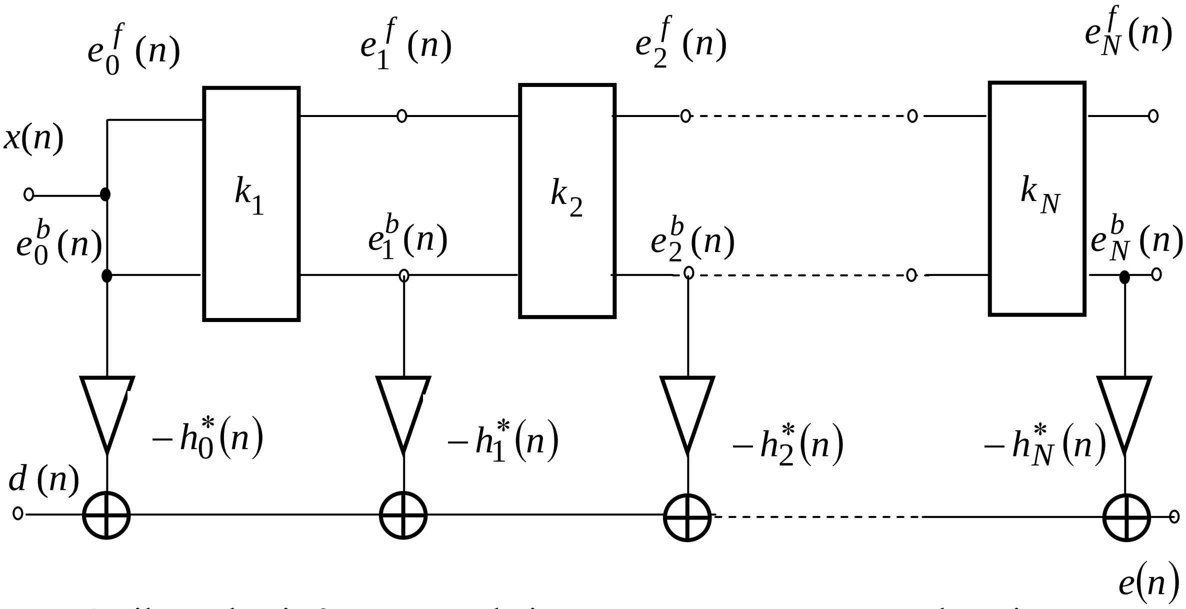

Figura 5.8 Filtru adaptiv în structura latice, care generează eroarea de estimare

Avantajul principal al acestor algoritmi este acela că datorită operației de decorelare efectuată de structura latice, convergență mai rapidă, în cazul unor date de intrare puternic corelate.
În figura 5.8 este dată o variantă de schemă care generează eroroarea de estimare.