5. FILTRE ADAPTIVE BAZATE PE MINIMIZAREA ERORII MEDII PATRATICE

5. FILTRE ADAPTIVE BAZATE PE MINIMIZAREA ERORII MEDII PATRATICE

- funcţia cost este eroarea medie pătratică, J(n)
- ecuaţiile Wiener-Hopf nu oferă o soluţie practică
  - complexitate aritmetică ridicată
  - necesită funcţiile de autocorelaţie
  - se doreşte un algoritm recursiv

Minimizarea unei funcții reale de variabilă complexă

Teoremă

O funcție $f(\boldsymbol{z},\boldsymbol{z}^*): \mathbb{C}^N \to \mathbb{R}$ are direcția de variație maximă dată de gradientul complex $\nabla_{z^*}\{f(\boldsymbol{z},\boldsymbol{z}^*)\}$.

$$\nabla_z J = \left[\frac{\partial J}{\partial z_0}, \frac{\partial J}{\partial z_1}, \ldots, \frac{\partial J}{\partial z_{N-1}}\right]^T$$

$$\frac{\partial J}{\partial z_i} = \frac{1}{2}\left(\frac{\partial J}{\partial a_i} - \mathrm{j}\frac{\partial J}{\partial b_i}\right), \quad \frac{\partial J}{\partial z^*_i} = \frac{1}{2}\left(\frac{\partial J}{\partial a_i} + \mathrm{j}\frac{\partial J}{\partial b_i}\right), \quad z_i = a_i + b_i$$

$$\nabla_z\{\mathbf{a}^H\mathbf{z}\} = \mathbf{a}^* \qquad \nabla_{z^*}\{\mathbf{a}^H\mathbf{z}\} = \mathbf{0}$$

$$\nabla_z\{\mathbf{z}^H\mathbf{a}\} = \mathbf{0} \qquad \nabla_{z^*}\{\mathbf{z}^H\mathbf{a}\} = \mathbf{a}$$

$$\nabla_z\{\mathbf{z}^H\mathbf{A}\mathbf{z}\} = (\mathbf{A}\mathbf{z})^* \qquad \nabla_{z^*}\{\mathbf{z}^H\mathbf{A}\mathbf{z}\} = (\mathbf{A}\mathbf{z})$$

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

1. - Se porneşte de la o valoare iniţială a coeficienţilor $\mathbf{w}(0)$ (eventual $\mathbf{w}(0)=\mathbf{0}$).
2. - Se evaluează direcţia pantei maxime de creştere în jurul acestui punct pe suprafaţa $J(\mathbf{w})$. Aceasta este exprimată prin gradientul complex $\nabla_{\mathbf{w}^*}\left(J(\mathbf{w})\right)$.
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, $\mu \in \mathbf{R}_+$:

$$\mathbf{w}(n+1)=\mathbf{w}(n)-\mu\nabla\left(J(n)\right)$$

4. - Se reia procedeul din punctul 2.

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

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

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

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\nabla\left(J(n)\right)$$

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza stabilităţii algoritmului

Vom introduce ***vectorul eroare a coeficienţilor***,

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

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

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza stabilității algoritmului

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

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

Vom introduce ***vectorul eroare rotit***, $\mathbf{v}(n)$,

$$\mathbf{v}(n)=\mathbf{Q}^{H}\mathbf{c}(n)=\mathbf{Q}^{H}\left(\mathbf{w}(n)-\mathbf{w}_{o}\right)$$

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza stabilităţii algoritmului

Condiţii iniţiale:

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

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

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

Stabilitatea schemei, deci convergenţa algoritmului impune ca $v_k(n)$ să tindă la 0 când $n \to \infty$. Pentru aceasta este necesar şi suficient ca:

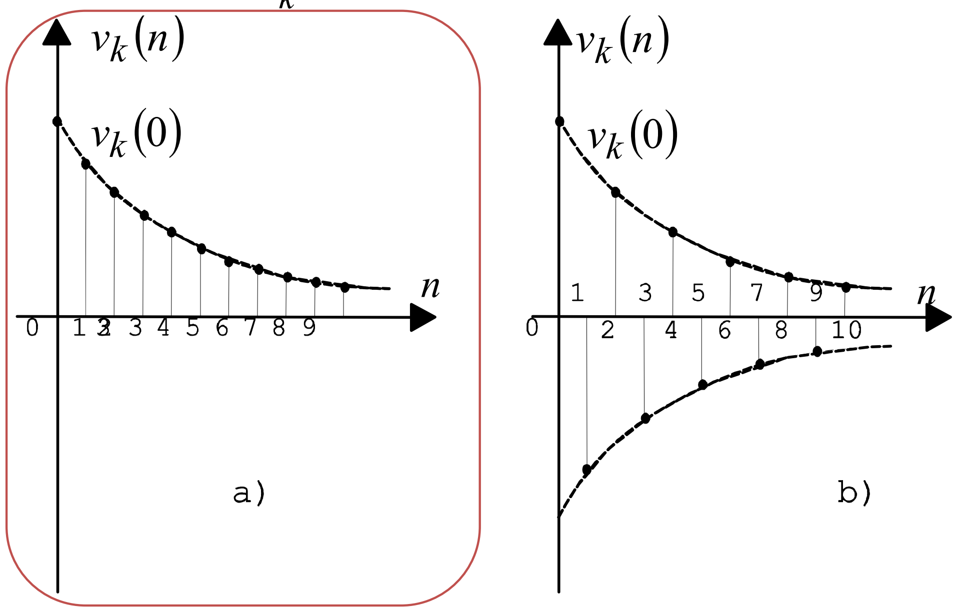
$$\left|1 - \mu\lambda_k\right| < 1 \quad \text{sau} \quad 0 < \mu < \frac{2}{\lambda_k}, \quad k = 1,2,\cdots N$$

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)
Analiza convergenţei

$\lambda_k \in \mathbf{R}_+ \quad \Rightarrow \quad v_k(n)$ va descreşte către 0 fără oscilaţii. Distingem cazurile:

a) $0 < \mu\lambda_k < 1$ sau $0 < \mu < \dfrac{1}{\lambda_k} \quad \Rightarrow \quad v_k(n)$ descresc uniform

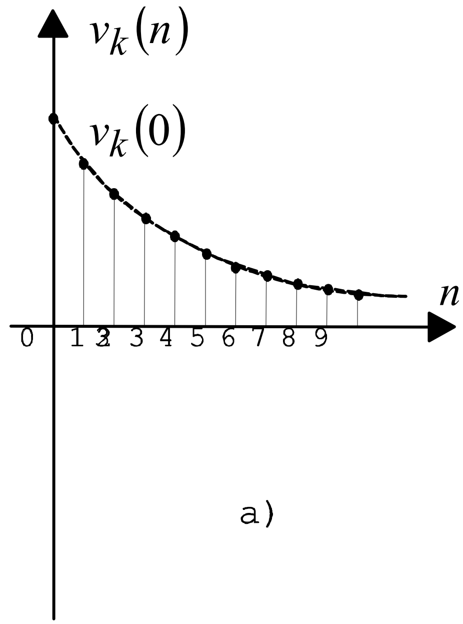


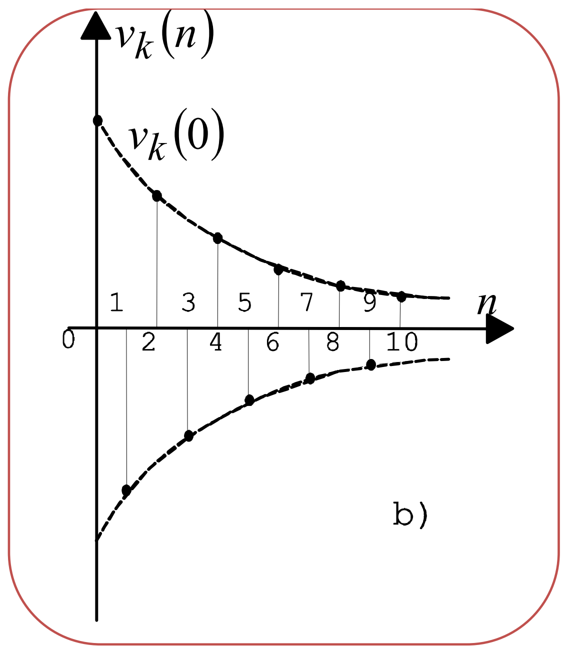
$$v_k(n) = \mathrm{e}^{-n/\tau_k}\, v_k(0), \qquad \tau_k = -\frac{1}{\ln(1-\mu\lambda_k)}$$

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)
Analiza convergenţei

b) $-1 < 1-\mu\lambda_k < 0$ **sau** $\dfrac{1}{\lambda_k} < \mu < \dfrac{2}{\lambda_k}$ - şir alternat

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





5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza convergenţei

c) $1-\mu\lambda_k = 0 \quad \Rightarrow \quad v_k(n) = 0, \quad n \geq 1$

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

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

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

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza convergenţei

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

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

- 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 $\lambda_{\min}$.

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Analiza convergenţei

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

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

Termenul cel mai lent descrescător este acela care conţine factorul

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

R rău condiţionată (raport $\lambda_{Max}/\lambda_{min}$ mare) $\Rightarrow$ convergenţă lentă.

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Curba de învăţare

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

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

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

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

c[n] este determinist, e_o satisface principiul ortogonalităţii $\Rightarrow$

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

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Curba de învăţare

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

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:

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

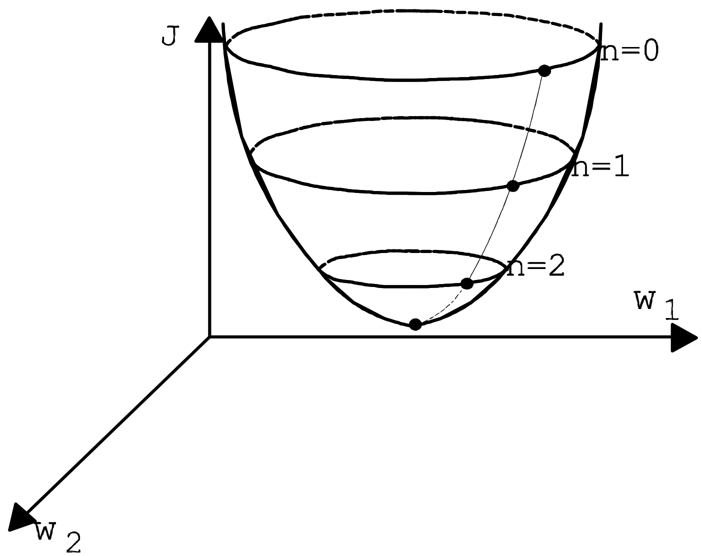
$$= J_{\min} + \sum_{k=1}^{N} \lambda_k \, \mathrm{e}^{-2n/\tau_k} |v_k(0)|^2$$

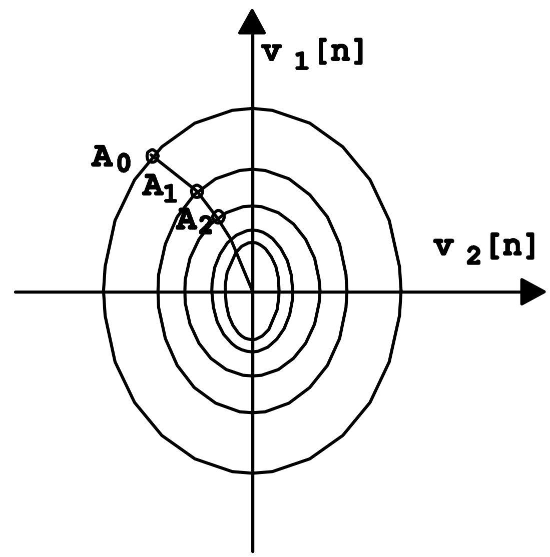
Curba obţinută reprezentând $J(n)$ este *curba de învăţare* a algoritmului. Indiferent de condiţiile iniţiale $v_k(0)$, eroarea medie pătratică tinde către $J_{\min}$ dacă este îndeplinită condiţia de stabilitate.

5.1 Metoda pantei descendente maxime (**S**teepest **D**escent - SD)

Curba de învăţare

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





Intersecţiile cu plane J=const sunt elipse cu semiaxele

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

5.2 Algoritmul gradientului stohastic (LMS)

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

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

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:

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

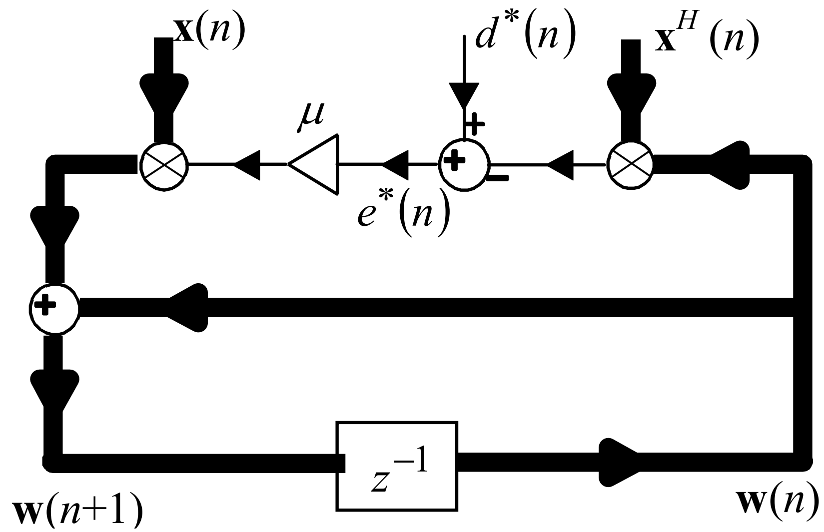
$$\hat{\nabla}\left(J(n)\right) = -\mathbf{x}(n)d^*(n) + \mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n) = -\mathbf{x}(n)e^*(n)$$

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

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

5.2 Algoritmul gradientului stohastic (LMS)

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



5.2 Algoritmul gradientului stohastic (LMS)

$$\mathbf{w}(0) = \mathbf{0}$$

$$\text{for} \quad n = 0,1,2,\cdots$$

$$y(n) = \mathbf{w}^H(n)\mathbf{x}(n)$$

$$e(n) = d(n) - y(n)$$

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

$$\text{end}$$

5.2 Algoritmul gradientului stohastic (LMS)

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

5.2 Algoritmul gradientului stohastic (LMS)

Analiza convergenţei

Probleme de convergenţă

- Tinde valoarea medie a vectorului $\mathbf{w}(n)$ către $\mathbf{w}_o$ atunci când $n \to \infty$? În caz afirmativ, se zice că algoritmul este *convergent în medie*.
- Tinde $J(n)$ către o valoare finită atunci când $n \to \infty$? In caz afirmativ se spune că algoritmul este *convergent în medie pătratică*.

5.2 Algoritmul gradientului stohastic (LMS)

Ipoteze de independenţă:

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

Datorită lipsei medierii în calculul gradientului apare un ***zgomot de gradient***

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

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

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

5.2 Algoritmul gradientului stohastic (LMS)

Pe de altă parte, relaţia de reactualizare a coeficienţilor devine

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

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

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

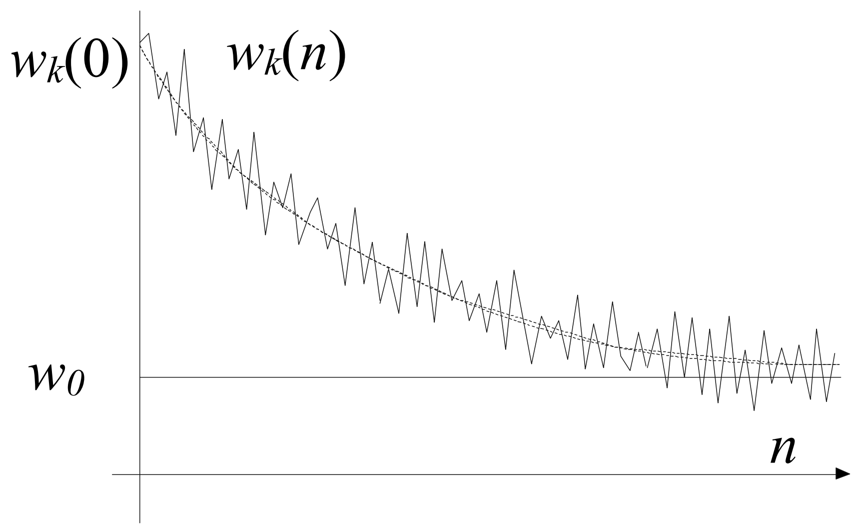
$$\mathbf{v}(n+1)=\left(\mathbf{I}-\mu\mathbf{\Lambda}\right)\mathbf{v}(n)-\mu\mathbf{Q}^H\mathbf{N}(n)$$

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

5.2 Algoritmul gradientului stohastic (LMS)

Având în vedere analogia formală dintre această ecuaţie şi cea corespunzătoare în cazul algoritmului SD, concluziile trase acolo pentru vectorul eroare a coeficienţilor se pot transpune aici pentru media acestui vector. Algoritmul LMS este ***convergent în medie*** dacă:

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



5.2 Algoritmul gradientului stohastic (LMS)

Analiza convergenţei în medie pătratică

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

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

$$= \mathrm{tr}\left\{\mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\}\mathrm{E}\left\{\mathbf{c}(n)\mathbf{c}^H(n)\right\}\right\} = \mathrm{tr}\left\{\mathbf{R}\mathbf{C}(n)\right\}$$

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

$$J(n) = J_{\min} + \mathrm{tr}\left\{\mathbf{R}\mathbf{C}(n)\right\} = J_{\min} + J_{ex}(n)$$

5.2 Algoritmul gradientului stohastic (LMS)

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

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

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

unde

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

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

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 $\mathbf{K}(n)$.

5.2 Algoritmul gradientului stohastic (LMS)

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

Comp. staţionară

Comp.tranzitorie

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

5.2 Algoritmul gradientului stohastic (LMS)

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

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

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

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

5.2 Algoritmul gradientului stohastic (LMS)

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

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

5.2 Algoritmul gradientului stohastic (LMS)

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

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

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

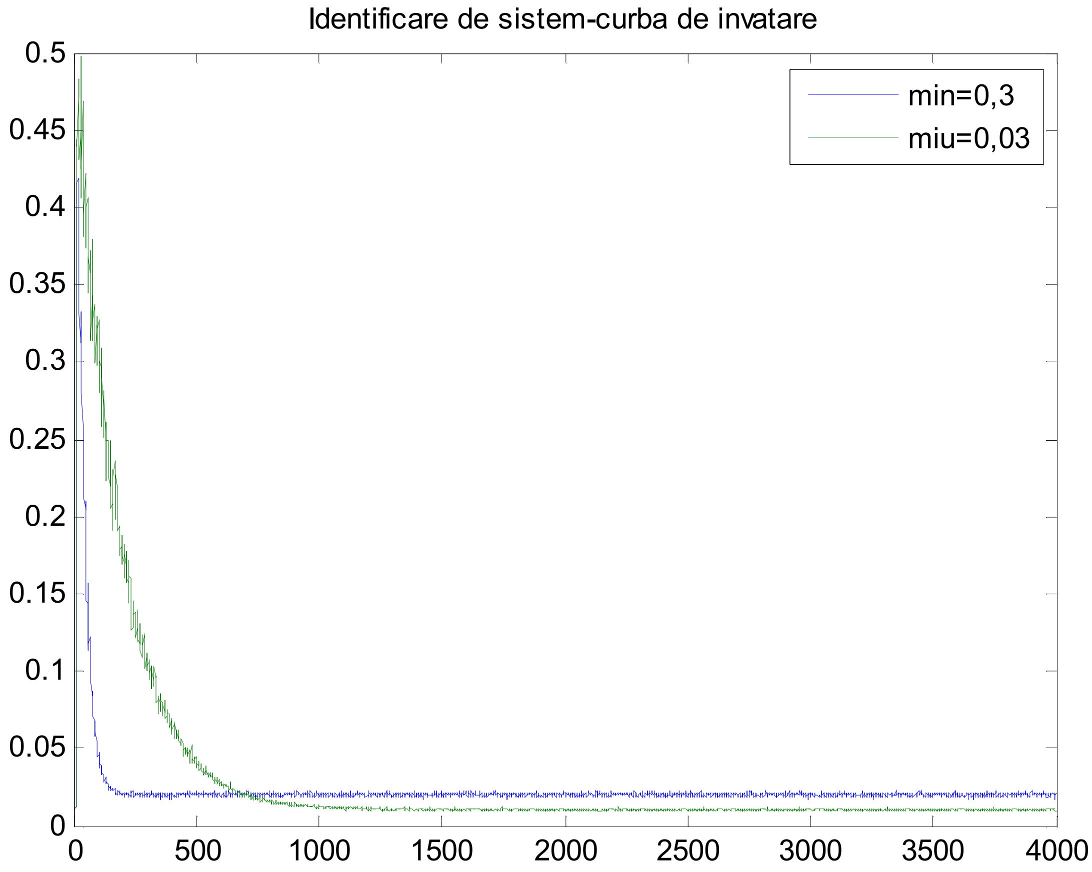
Se poate introduce o constantă de timp medie:

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

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.

Exemplu – efectul valorii pasului μ



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:

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

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

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

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

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

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

$$\nabla_{\mathbf{w}^*(n+1)}\left(\mathbf{w}^H(n+1)\mathbf{w}(n+1)\right) = \mathbf{w}(n+1)$$

$$\nabla_{\mathbf{w}^*(n+1)}\left(\mathbf{w}^H(n+1)\mathbf{w}(n) + \mathbf{w}^H(n)\mathbf{w}(n+1)\right) = \mathbf{w}(n)$$

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

Rezultă:

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

$$\nabla_{\mathbf{w}^*(n+1)} = \mathbf{0} \quad \Rightarrow \quad \mathbf{w}(n+1) - \mathbf{w}(n) = -\frac{1}{2}\lambda\mathbf{x}(n)$$

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

λ se obţine punând condiţia

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

Pentru aceasta se înmulţeşte la stânga cu $\mathbf{x}^H(n)$ relaţia:

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

$$\Downarrow$$

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

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

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

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 $\overline{\mu}$, deci:

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

Poate fi echivalat cu algoritmul gradientului stohastic pentru:

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

în care ***pasul este variabil***.

Convergenţa în medie pătratică este asigurată dacă:

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

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

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

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

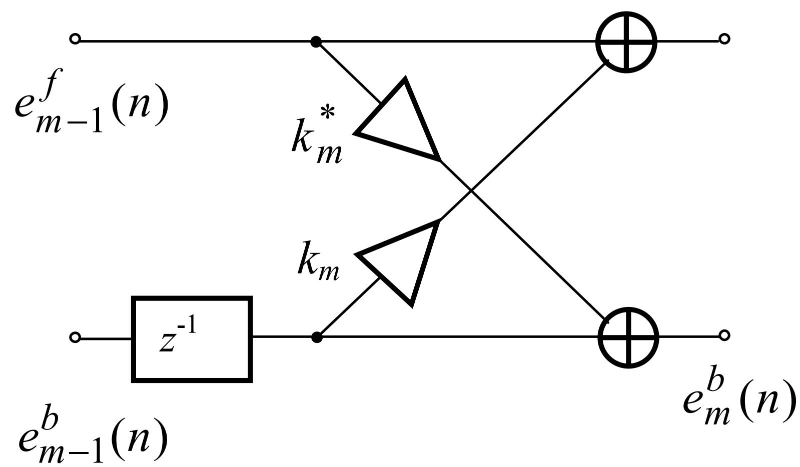
Frecvent

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

unde a este o constantă mică, adăugată pentru a evita împărţirea prin zero.

5.4 Metoda gradientului pentru structuri latice (GAL)

Ne punem, pentru început, problema proiectării unui ***predictor latice adaptiv***.
Vor trebui deci optimizaţi coeficienţii k_m iar funcţia cost poate fi definită pornind de la eroarea de predicţie directă, inversă, sau compusă.
Să considerăm o celulă m a filtrului şi să determinăm k_m aşa încât să fie minimizată eroarea medie pătratică de predicţie, în forma compusă,



$$J_m(n) = \mathrm{E}\left\{\left|e_m^f(n)\right|^2 + \left|e_m^b(n)\right|^2\right\}$$

5.4 Metoda gradientului pentru structuri latice (GAL)

Abordarea SD.

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

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

Va trebui deci evaluat gradientul complex în raport cu k^*_m:

$$\nabla_m(J_m(n)) = \mathrm{E}\left\{\nabla_m\left(e_m^{f*}(n)\right)e_m^f(n) + e_m^{f*}(n)\nabla_m\left(e_m^f(n)\right)\right\} + \\ + \mathrm{E}\left\{\nabla_m\left(e_m^{b*}(n)\right)e_m^b(n) + e_m^{b*}(n)\nabla_m\left(e_m^b(n)\right)\right\}$$

Vom folosind relaţiile de recurenţă:

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

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

5.4 Metoda gradientului pentru structuri latice (GAL)

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

$$\nabla_m\left(e_m^{f}(n)\right)=\nabla_{k_m^*}\left(e_{m-1}^{f}(n)+k_m e_{m-1}^{b}(n-1)\right)=0$$

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

$$\nabla_m\left(e_m^{b*}(n)\right)=\nabla_{k_m^*}\left(k_m e_{m-1}^{f*}(n)+e_{m-1}^{b*}(n-1)\right)=0$$

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

5.4 Metoda gradientului pentru structuri latice (GAL)

Rezultă deci:

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

sau aplicând (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:

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

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

5.4 Metoda gradientului pentru structuri latice (GAL)

Abordarea LMS

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

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

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

5.4 Metoda gradientului pentru structuri latice (GAL)

Abordarea NLMS

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

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

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

5.4 Metoda gradientului pentru structuri latice (GAL)

Aproximativ

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

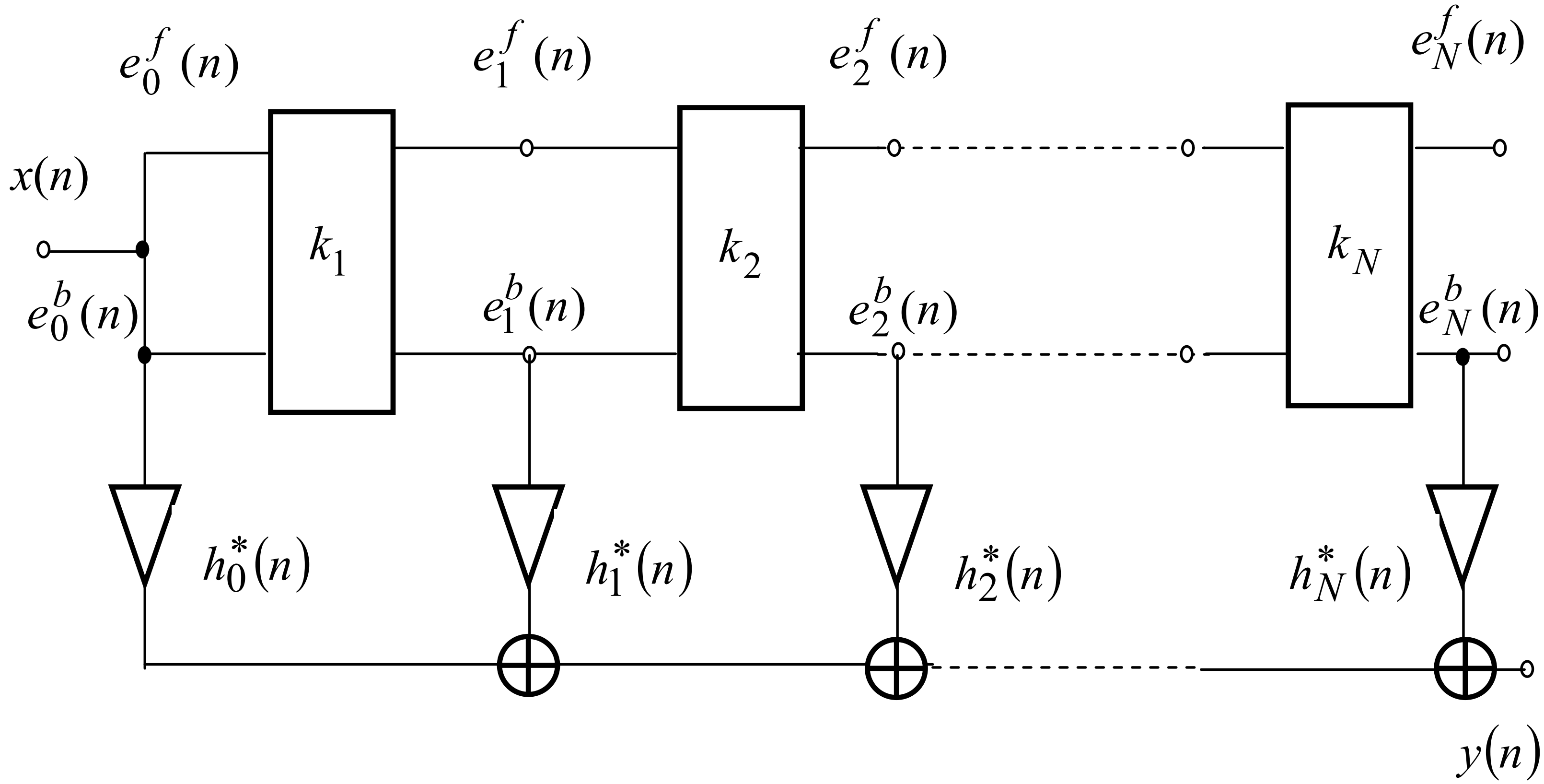
$$=\frac{1}{n}\left((n-1)W_{m-1}(n-1)+\left|e_{m-1}^{f}(n)\right|^{2}+\left|e_{m-1}^{b}(n-1)\right|^{2}\right)$$

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

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

5.5 Algoritmul LMS pentru structuri latice-scară

Algoritmul GAL permitea obţinerea unui predictor. Pentru a obţine un Filtru adaptiv se adaugă o structură în scară.



5.5 Algoritmul LMS pentru structuri latice-scară

Abordarea SD

Punem condiţia ca ieşirea să estimeze cât mai corect un semnal dorit $d(n)$.

$$\mathbf{h}(n) = \left(h_0(n), h_1(n), \cdots, h_N(n)\right)^T$$

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

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

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

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

5.5 Algoritmul LMS pentru structuri latice-scară

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,

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

$$\mathbf{R}_e = \mathrm{diag}\left\{J_0^b, J_1^b, \cdots, J_N^b\right\}$$

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

5.5 Algoritmul LMS pentru structuri latice-scară

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

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \mu \nabla J_N(n)$$

$$\nabla\left(J_N(n)\right) = -\mathbf{r}_{ed} + \mathbf{R}_e \mathbf{h}(n) =$$

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

Abordarea LMS

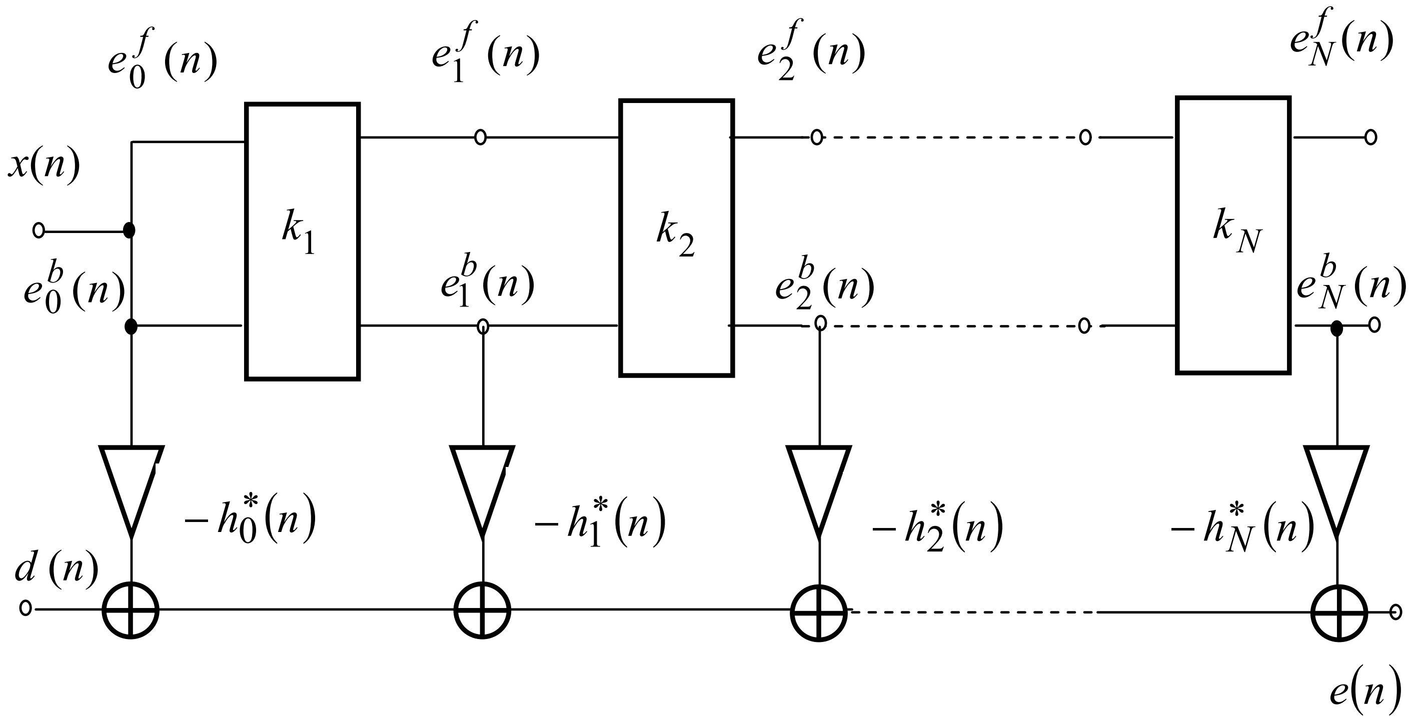
Cum valorile medii nu sunt în general cunoscute, aplicăm metoda gradientului stohastic (LMS),

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

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

5.5 Algoritmul LMS pentru structuri latice-scară

Variantă pentru generarea erorii de estimare.



Avantaj: convergenţă mai rapidă, datorită decorelării datelor, de către structura latice.
Dezavantaj: complexitate aritmetică mai ridicată