❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

- Modelele din clasa ARMAX pot fi identificate şi cu ajutorul altor metode decît MCMMP şi MVI.
- Cu cît modelul este mai complex, cu atît metoda utilizată este mai laborioasă.
- O parte dintre metodele utilizate în acest scop se bazează atît pe revizuirea/adaptarea MCMMP, cît şi pe algoritmi care folosesc Tehnici de Optimizare.

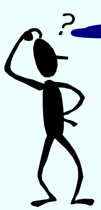
Proceduri recursive cu ajutorul cărora un punct de optim (maxim sau minim) al unei funcţii criteriu este aproximat printr-un proces iterativ exprimat de o ecuaţie cu reactualizare aditivă.

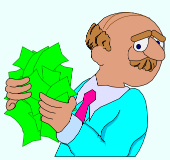
$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{\Delta}_k \quad \forall k \in \mathbb{N}$$

valori succesive ale aproximaţiei (vector sau scalar)

corecţie (aditivă)

☝ Se pleacă de la o iniţializare. $\boldsymbol{\theta}_0$





Cum poate fi cuantificată precizia aproximaţiei?

Folosind norma Euclidiană.

Aproximaţia următoare este mai precisă decît aproximaţia curentă dacă:

$$\left\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}^*\right\| < \left\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\right\|$$

punct de optim

Necunoscut !

☝ Pentru a stopa procesul iterativ, trebuie utilizată o inegalitate în care punctul de optim să nu apară explicit.

Test practic de stop

$$\left\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\right\| = \left\|\boldsymbol{\Delta}_k\right\| < \varepsilon$$

Test ideal de stop

$$\left\|\boldsymbol{\theta}_k - \boldsymbol{\theta}^*\right\| < \varepsilon$$

☹ Neimplementabil.

prag minimal de precizie

Exemplu

$\varepsilon = 10^{-4}$

θ_k & θ_{k+1} (scalari) coincid pînă la a patra zecimală inclusiv

❹.❹ Metode bazate pe optimizarea parametrilor

Obiectiv

Determinarea expresiei generale a corecției în funcție de criteriul de optimizare. $\mathcal{V}$

Metoda Newton-Raphson (MNR)

Metodă aplicabilă în cazul funcțiilor criteriu derivabile de cel puțin 2 ori.

$\mathcal{V}$

Gradientul

funcție care trebuie să fie derivabilă

$$\nabla\mathcal{V}(\boldsymbol{\theta}) \overset{def}{=} \left[\frac{\partial\mathcal{V}}{\partial\theta_1}(\boldsymbol{\theta}) \quad \frac{\partial\mathcal{V}}{\partial\theta_2}(\boldsymbol{\theta}) \quad \cdots \quad \frac{\partial\mathcal{V}}{\partial\theta_{n\theta}}(\boldsymbol{\theta})\right]^T \in \mathbb{R}^{n\theta}$$

☝ Vector coloană.

Matricea Hessiană

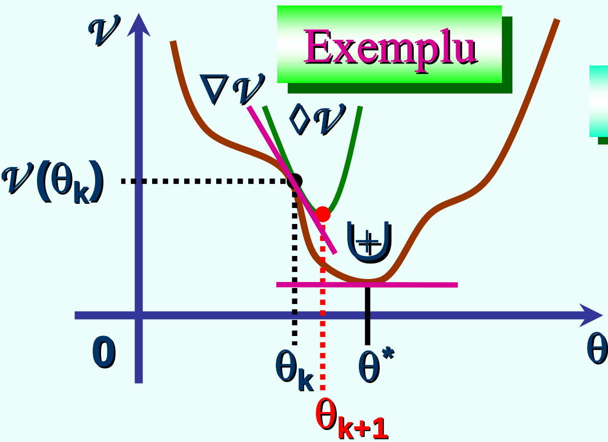
funcție care trebuie să fie continuă

$$\Diamond\mathcal{V}(\boldsymbol{\theta}) \overset{def}{=} \boxed{\nabla(\nabla\mathcal{V})(\boldsymbol{\theta})} = \left[\frac{\partial^2\mathcal{V}}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta})\right]_{i,j\in\overline{1,n\theta}} \in \mathbb{R}^{n\theta\times n\theta}$$

Jacobianul gradientului

☝ Matrice simetrică.

Exemplu



Principiul metodei

- În jurul punctului de optim, prima derivată (sau norma sa) este aproximativ nulă, iar derivata a doua este strict pozitiv/negativ definită.
- Se construieşte parabola (paraboloidul) care trece prin punctul curent $(\theta_k, \mathcal{V}(\theta_k))$ şi are aceeaşi tangentă şi derivată secundă ca funcția criteriu.
- Punctul de minim/maxim al parabolei este θ_{k+1}.

Metoda Newton-Raphson (continuare)

Aşadar Trebuie construit paraboloidul care verifică 3 proprietăţi de coincidenţă cu funcţia criteriu în punctul curent.

$\mathcal{V}_k$

Cum poate fi construit?

Ar trebui exploatată proprietatea funcţiei criteriu de a fi de 2 ori derivabilă.

$\mathcal{V}_k(\theta_k) = \mathcal{V}(\theta_k)$ (derivata 0)

$\nabla\mathcal{V}_k(\theta_k) = \nabla\mathcal{V}(\theta_k)$ (derivata 1)

$\Diamond\mathcal{V}_k(\theta_k) = \Diamond\mathcal{V}(\theta_k)$ (derivata 2)

Dezvoltare în serie Taylor pînă la ordinul 2

Paraboloidul Taylor

$$\mathcal{V}_k(\theta) = \mathcal{V}(\theta_k) + \left[\nabla\mathcal{V}(\theta_k)\right]^T(\theta - \theta_k) + \frac{1}{2}(\theta - \theta_k)^T\left[\Diamond\mathcal{V}(\theta_k)\right](\theta - \theta_k) \quad \forall\, \theta \in \mathbb{R}^{n\theta}$$

Exerciţiu

- Arătaţi că paraboloidul Taylor verifică cele 3 proprietăţi de coincidenţă.

Pentru a determina punctul de optim al paraboloidului Taylor, se va anula gradientul acestuia.

Reguli de derivare

$$\nabla(\mathbf{r}^T\theta) = \mathbf{r}$$

$$\nabla(\theta^T\mathbf{R}\theta) = (\mathbf{R} + \mathbf{R}^T)\theta$$

☝ Matricea Hessiană este simetrică şi inversabilă în vecinătatea optimului.

$$\nabla\mathcal{V}_k(\theta) = \nabla\mathcal{V}(\theta_k) + \left[\Diamond\mathcal{V}(\theta_k)\right](\theta - \theta_k) = \mathbf{0}$$

$$\theta_{k+1} = \theta_k \underbrace{-\left[\Diamond\mathcal{V}(\theta_k)\right]^{-1}\nabla\mathcal{V}(\theta_k)}_{+\Delta_k} \quad \forall\, k \in \mathbb{N}$$

❹.❹ Metode bazate pe optimizarea parametrilor

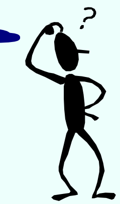
Metoda Newton-Raphson (continuare)

$$\Delta_k \overset{def}{=} -\left[\lozenge \mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1} \nabla \mathcal{V}(\boldsymbol{\theta}_k) \quad \forall k \in \mathbb{N}$$

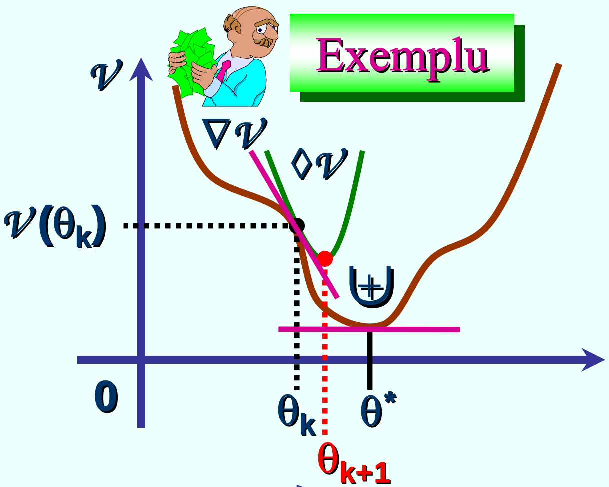
☝ minus

Termenul corector are semn opus gradientului, doarece matricea Hessiană este strict pozitiv/negativ definită.

Ce semnificaţie are acest rezultat?



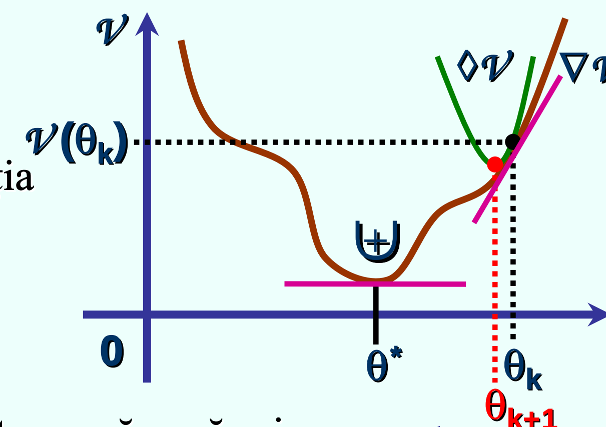
Exemplu



➔ Semnul primei derivate obligă aproximaţia curentă să se deplaseze către optimul criteriului.

➔ Aproximaţia curentă fiind inferioară optimului, iar derivata fiind negativă, corecţia se adaugă aproximaţiei pentru a se apropia mai mult de optim.

➔ Dacă este adăugată o cantitate prea mare, se ajunge în zona de derivată pozitivă şi următorul factor de corecţie se va scădea din aproximaţia curentă, superioară optimului.



☝ Convergenţa la punctul de optim poate fi realizată fie printr-un şir monoton de aproximaţii, fie prin intermediul unuia oscilant.

Testul de stop

$$\left\| \Delta_k \right\| = \left\| \left[\lozenge \mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1} \nabla \mathcal{V}(\boldsymbol{\theta}_k) \right\| < \varepsilon$$

- Eficienţa se măsoară prin numărul de iteraţii necesar verificării testului de stop.

☝ În vecinătatea optimului, norma gradientului este aproximativ nulă.

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Metoda Newton-Raphson (continuare)

Algoritmul Newton-Raphson cu pas constant

Date de intrare
- $\mathcal{V} : \mathbb{R}^{n\theta} \to \mathbb{R}$ (funcţie criteriu, de 2 ori derivabilă)
- $\varepsilon > 0$ (prag de precizie)

Iniţializare
- $\boldsymbol{\theta}_0 \in \mathbb{R}^{n\theta}$ ➔ ales fie arbitrar, fie cu ajutorul unui algoritm specific
- $\nabla\mathcal{V}(\boldsymbol{\theta}_0)$ ➔ gradientul iniţial
- $\Diamond\mathcal{V}(\boldsymbol{\theta}_k)$ ➔ matricea Hessiană iniţială
- $k = 0$ ➔ indicele iterativ iniţial

Concept definit în continuare.

Buclă iterativă

Cît timp $\left\| \left[\Diamond\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1} \nabla\mathcal{V}(\boldsymbol{\theta}_k) \right\| \geq \varepsilon$

① Se reactualizează aproximaţia optimului $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \left[\Diamond\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1} \nabla\mathcal{V}(\boldsymbol{\theta}_k)$

② Se reactualizează gradientul şi matricea Hessiană $\nabla\mathcal{V}(\boldsymbol{\theta}_k) \leftarrow \nabla\mathcal{V}(\boldsymbol{\theta}_{k+1})$, $\Diamond\mathcal{V}(\boldsymbol{\theta}_k) \leftarrow \Diamond\mathcal{V}(\boldsymbol{\theta}_{k+1})$

③ Se trece la pasul următor $k \leftarrow k + 1$

Date de ieşire
- $\boldsymbol{\theta}_k$ ➔ aproximaţia de precizie ε
- k ➔ numărul de iteraţii pentru atingerea preciziei dorite

☝ Algoritmul este sensibil la iniţializare, în sensul că, în general, va converge către optimul cel mai apropiat de aceasta.

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

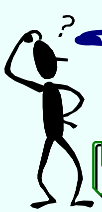
Metoda Newton-Raphson (continuare)

☹ Utilizarea MNR (cu pas constant) este deficitară în următoarele cazuri principale:

- ➔ Funcţia criteriu are o regularitate slabă în vecinătatea optimului vizat (adică prezintă oscilaţii mari în acea vecinătate).
- ➔ Funcţia criteriu este afectată de zgomote importante în vecinătatea optimului vizat.
- ➔ Optimul vizat al funcţiei criteriu este situat pe un platou.

- De regulă, în primele 2 cazuri, se renunţă la MNR şi se adoptă o strategie evoluţionistă.
- În ultimul caz, MNR conduce la un număr imens de iteraţii, din cauza faptului că, pe platou, atît norma gradientului cît şi spectrul matricii Hessiene sunt aproximativ nule.

☝ Corecţia este imprecis determinată, în special din cauza problemelor de natură numerică induse de inversarea matricii Hessiene.



Ce se poate face?

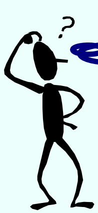
Termenul corector poate fi ajustat adaptiv cu un scalar, astfel încît viteza de convergenţă să crească prin efectuarea unui salt peste platou.

$$\theta_{k+1} = \theta_k - \underbrace{}_{1}\left[\lozenge\mathcal{V}(\theta_k)\right]^{-1}\nabla\mathcal{V}(\theta_k) \quad \forall k\in\mathbb{N}$$

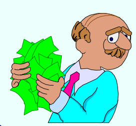
1 pas constant

$$\theta_{k+1} = \theta_k - \alpha_k\left[\lozenge\mathcal{V}(\theta_k)\right]^{-1}\nabla\mathcal{V}(\theta_k) \quad \forall k\in\mathbb{N}$$

pas variabil



Cum poate fi reactualizat pasul variabil?



Prin utilizarea MNR cu pas constant pentru optimizarea unei funcţii criteriu scalare adaptive.

$$\mathcal{F}_k(\alpha) \overset{def}{=} \mathcal{V}\left(\theta_k - \alpha\left[\lozenge\mathcal{V}(\theta_k)\right]^{-1}\nabla\mathcal{V}(\theta_k)\right) \quad \forall \alpha\in\mathbb{R}$$

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Metoda Newton-Raphson (continuare)

Algoritmul Newton-Raphson cu pas variabil

Date de intrare
- $\mathcal{V}:\mathbb{R}^{n\theta}\to\mathbb{R}$ (funcţie criteriu, de 2 ori derivabilă)
- $\varepsilon>0$ (prag de precizie)

Iniţializare
- $\boldsymbol{\theta}_0\in\mathbb{R}^{n\theta}$, $\alpha_0\in\mathbb{R}^*$ ➔ alese fie arbitrar, fie cu ajutorul unui algoritm specific
- $\nabla\mathcal{V}(\boldsymbol{\theta}_0)$ ➔ gradientul iniţial
- $\lozenge\mathcal{V}(\boldsymbol{\theta}_k)$ ➔ matricea Hessiană iniţială
- $k=0$ ➔ indicele iterativ iniţial

Exerciţiu

- Justificaţi acest algoritm în manieră riguroasă.

Buclă iterativă

Cît timp $|\alpha_k|\cdot\left\|\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1}\nabla\mathcal{V}(\boldsymbol{\theta}_k)\right\|\geq\varepsilon$

☝ Se vor utiliza regulile cunoscute de derivare pentru reactualizarea pasului variabil.

① Se reactualizează aproximaţia optimului

$$\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_k-\alpha_k\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1}\nabla\mathcal{V}(\boldsymbol{\theta}_k)$$

② Se reactualizează pasul variabil

$$\alpha_{k+1}=\alpha_k+\frac{\left[\nabla\mathcal{V}(\boldsymbol{\theta}_{k+1})\right]^T\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1}\nabla\mathcal{V}(\boldsymbol{\theta}_k)}{\left[\nabla\mathcal{V}(\boldsymbol{\theta}_k)\right]^T\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1}\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_{k+1})\right]\left[\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\right]^{-1}\nabla\mathcal{V}(\boldsymbol{\theta}_k)}$$

③ Se reactualizează gradientul şi matricea Hessiană

$$\nabla\mathcal{V}(\boldsymbol{\theta}_k)\leftarrow\nabla\mathcal{V}(\boldsymbol{\theta}_{k+1})$$
$$\lozenge\mathcal{V}(\boldsymbol{\theta}_k)\leftarrow\lozenge\mathcal{V}(\boldsymbol{\theta}_{k+1})$$

④ Se trece la pasul următor $k\leftarrow k+1$

Date de ieşire
- $\boldsymbol{\theta}_k$ ➔ aproximaţia de precizie ε
- k ➔ numărul de iteraţii pentru atingerea preciziei dorite

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Metoda Newton-Raphson (continuare)

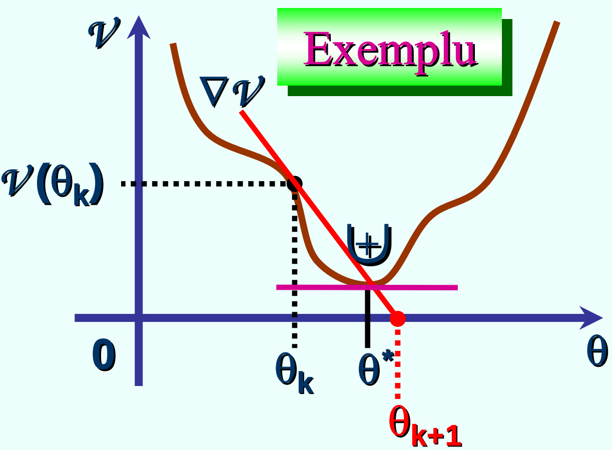
- **MNR** poate fi utilizată şi pentru a **rezolva unele ecuaţii transcendente** sau în **aproximarea unor numere transcendente** cum ar fi π, e, $\ln 2$, etc.

Exemplu $\sin(x) = 1$ ⇒ $x = \frac{\pi}{2}$

- Aproximarea numărului π/2 se poate realiza plecînd de la funcția criteriu a cărei primă derivată este: $f'(x) = \sin(x) - 1$
- Este evident că anularea derivatei conduce la optimizarea funcţiei: $f(x) = -\cos(x) - x + C$

Exerciţiu
- Folosind exemplul anterior, determinați numărul π cu 7 zecimale exacte.

- Pentru a diminua complexitatea metodei, **se poate renunța la a doua derivată**.



Metoda gradientului (M∇)

- De această dată, aproximația următoare se determină prin **intersectarea tangentei (hiperplanului tangent)** care trece prin punctul curent $(\theta_k, \mathcal{V}(\theta_k))$ **cu axa orizontală (hiperplanul principal)**.

Exerciţiu

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \nabla\mathcal{V}(\boldsymbol{\theta}_k) \quad \forall k \in \mathbb{N}$$

M∇ cu pas constant

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_k \nabla\mathcal{V}(\boldsymbol{\theta}_k)$$

$$\alpha_{k+1} = \alpha_k + \left[\nabla\mathcal{V}(\boldsymbol{\theta}_{k+1})\right]^T \nabla\mathcal{V}(\boldsymbol{\theta}_k) \quad \forall k \in \mathbb{N}$$

M∇ cu pas variabil

☺ **M∇** este utilizată atunci cînd **derivata a doua fie nu există, fie nu se poate evalua**.

☹ Viteza de convergență a **M∇** este însă **modestă chiar şi în cazul pasului variabil**.

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Metoda Gauss-Newton (MGN)

Metodă aplicabilă în cazul funcțiilor criteriu **de tip pătratic** (cum sunt şi cele din IS).

$$\mathcal{V}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \varepsilon^2[n,\boldsymbol{\theta}]$$

reziduu

- Dacă funcția criteriu este **de 2 ori derivabilă**, MNR poate fi adaptată astfel încît **să se evite evaluarea matricii Hessiene**.

Există 2 abordări

Aproximarea matricii Hessiene | **Liniarizarea reziduurilor**

Exercițiu

$$\mathcal{V}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \varepsilon^2[n,\boldsymbol{\theta}]$$

$$\nabla\mathcal{V}(\boldsymbol{\theta}) = 2\sum_{n=1}^{N} \varepsilon[n,\boldsymbol{\theta}]\nabla\varepsilon[n,\boldsymbol{\theta}] \quad \forall\,\boldsymbol{\theta}\in\mathbb{R}^{n\theta}$$

$$\Diamond\mathcal{V}(\boldsymbol{\theta}) = \underbrace{2\sum_{n=1}^{N} \nabla\varepsilon[n,\boldsymbol{\theta}]\big[\nabla\varepsilon[n,\boldsymbol{\theta}]\big]^T}_{\text{termen principal}} + \underbrace{2\sum_{n=1}^{N} \boxed{\varepsilon[n,\boldsymbol{\theta}]}\Diamond\varepsilon[n,\boldsymbol{\theta}]}_{\text{termen parazit}} \quad \forall\,\boldsymbol{\theta}\in\mathbb{R}^{n\theta}$$

- Dacă se efectuează un număr suficient de mare de iterații, **reziduurile depondereză puternic** matricea Hessiană a acestora în termenul parazit.

$$\Diamond\mathcal{V}(\boldsymbol{\theta}) \cong 2\sum_{n=1}^{N} \nabla\varepsilon[n,\boldsymbol{\theta}]\big[\nabla\varepsilon[n,\boldsymbol{\theta}]\big]^T \quad \forall\,\boldsymbol{\theta}\in\mathbb{R}^{n\theta}$$

☝ În vecinătatea optimului, reziduul are valori neglijabile.

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Metoda Gauss-Newton (continuare)

MNR

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_k \left[\Diamond \mathcal{V}(\boldsymbol{\theta}_k) \right]^{-1} \nabla \mathcal{V}(\boldsymbol{\theta}_k) \qquad \forall k \in \mathbb{N}$$

$$\mathcal{V}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \varepsilon^2[n,\boldsymbol{\theta}]$$

$$\nabla \mathcal{V}(\boldsymbol{\theta}) = 2\sum_{n=1}^{N} \varepsilon[n,\boldsymbol{\theta}] \nabla \varepsilon[n,\boldsymbol{\theta}] \qquad \forall \boldsymbol{\theta} \in \mathbb{R}^{n\theta}$$

$$\Diamond \mathcal{V}(\boldsymbol{\theta}) \cong 2\sum_{n=1}^{N} \nabla \varepsilon[n,\boldsymbol{\theta}] \left[\nabla \varepsilon[n,\boldsymbol{\theta}] \right]^T \qquad \forall \boldsymbol{\theta} \in \mathbb{R}^{n\theta}$$

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_k \left[\sum_{n=1}^{N} \nabla \varepsilon[n,\boldsymbol{\theta}_k] \left[\nabla \varepsilon[n,\boldsymbol{\theta}_k] \right]^T \right]^{-1} \left[\sum_{n=1}^{N} \varepsilon[n,\boldsymbol{\theta}_k] \nabla \varepsilon[n,\boldsymbol{\theta}_k] \right] \qquad \forall k \in \mathbb{N}$$

☝ Deşi matricea Hessiană a criteriului pătratic nu a fost înlăturată din expresia iterativă principală, în calculul acesteia nu intervine decît gradientul reziduurilor.

Exerciţiu

- A doua abordare se bazează pe liniarizarea reziduurilor, adică pe aproximarea reziduurilor prin dezvoltarea acestora în serie Taylor de ordin I.
- Să se arate că ecuaţia iterativă a MGN obţinută din MNR prin liniarizarea reziduurilor, coincide cu cea din abordarea anterioară.

$$\varepsilon[n,\boldsymbol{\theta}] \cong \varepsilon[n,\boldsymbol{\theta}_k] + \left[\nabla \varepsilon[n,\boldsymbol{\theta}_k] \right]^T (\boldsymbol{\theta} - \boldsymbol{\theta}_k) \qquad \forall n \in \overline{1,N} \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{n\theta}$$

❹ Metode de identificare şi validare

❹.❹ Metode bazate pe optimizarea parametrilor

Caracteristici ale tehnicilor de optimizare în IS

- Majoritatea covîrşitoare a metodelor de optimizare nu pot garanta că optimul aproximat este cel global, ci, eventual cel mai apropiat de inițializarea procesului iterativ.
- Metodele de optimizare funcționează numai pentru funcțiile criteriu cărora li se poate evalua gradientul (prima derivată) şi, eventual, matricea Hessiană (derivata a doua).

☝ În multe aplicații, această exigență este imposibil de satisfăcut.

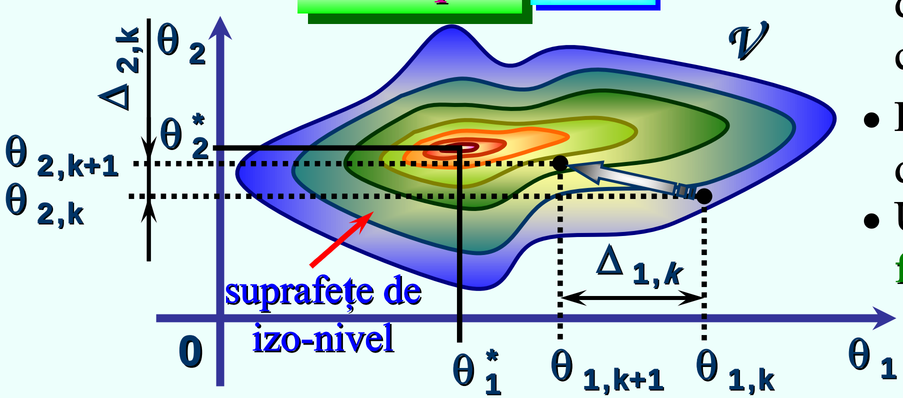
- Pentru a asigura eficiența algoritmilor bazați pe TO, este de dorit ca funcțiile criteriu să fie cît mai netede, cu oscilații puține în jurul optimului vizat şi fără paliere.

☝ Din cauza perturbațiilor care corup datele măsurate, funcțiile criteriu din cadrul IS sunt în general extrem de neregulate.

Tehnicile de optimizare ca MNR sau MGN sunt utilizate în IS mai mult ca instrumente auxiliare integrate în alte metode mai precise.

- Parametrii care pot fi identificați cu ajutorul tehnicilor de optimizare au, de regulă, precizii diferite.

Exemplu $n\theta = 2$



- Aproximația de ordin $k+1$ se obține din aproximația de ordin k prin contribuția a două componente de corecție diferite: una mai mare şi alta mai mică.
- Parametrul θ_1 este mai puțin sensibil (identificabil) decît parametrul θ_2.
- Un nou test de stop poate asigura precizia minimă a fiecărei componente (cu scăderea vitezei de convergență):

$$\max_{i\in 1,n\theta} \left|\theta_{i,k+1} - \theta_{i,k}\right| < \frac{\varepsilon}{n\theta}$$