(F) Calculul soluțiilor

(1) Introducerea

In paragrafele precedente, am încercat să răspund unor întrebări legate de ecuațiile (92) obținute pentru circuitul de problemă de probă (fig 22)

In felul acesta am justificat promisa corelare între A.C.F și metoda (α) (v pag 11). In ceea ce privește metoda (c), a rezolvării pe calculator, acest paragraf își propune să arate modul cum această rezolvare poate fi la rândul ei intovărășită de metodele A.C.F.

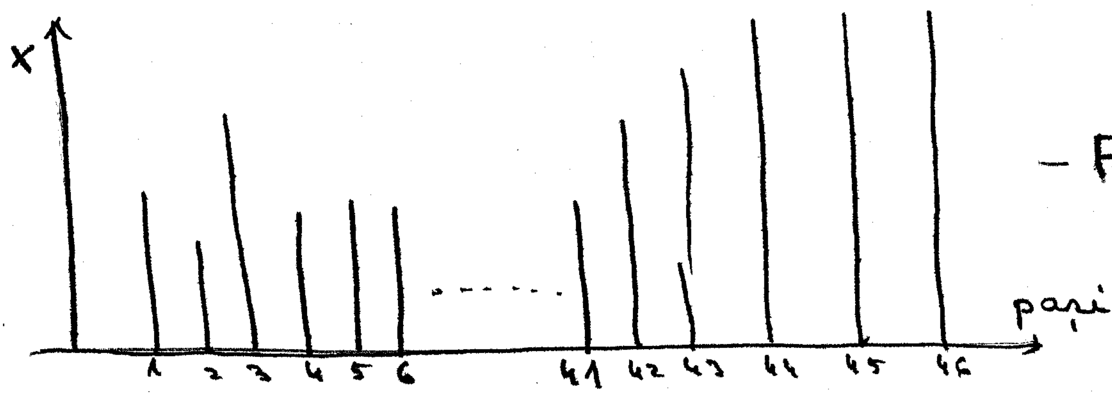
Prima observație care trebuie făcută este că, pentru ca un algoritm de calcul să fie definit, trebuie ca ecuația la care el se referă să fie bine definită (să existe o soluție, unică)

Putem deci să evităm cheltuiala inutilă de timp și de bani, pe care o reprezintă încărcarea calculatorului cu o problemă de acest tip, căreia să-i spunem: „problemă interzisă."

Din păcate, nu aceasta este singurul tip de problemă la care calculatorul, s-ar putea să reacționeze nedorit.

Mai există o categorie de cazuri, în care, deşi ecuaţia $F(x) = a$, introdusă în calculator are o soluţie, unică, programul pe care se bazează calculatorul nu este în mod cert conceput încât să o rezolve convergent. Astfel de situaţii, să le spunem „probleme nedecise" pot avea ca urmare imposibilitatea continuării calculului la un moment dat, sau o continuare la infinit (algoritmul nu e convergent).

Ca idee, să remarcăm posibilitatea unor răspunsuri eronate, după cum se sugerează în figură:



— Fig 39 —

Drumul imaginar al unui algoritm de iteraţie „cu capcană"

Ne putem foarte bine pune întrebarea dacă nu cumva, iteraţia a căzut într-o „capcană" între paşii 6 - 41, astfel încât, dacă testul de oprire a calculatorului este: $\|x_k - x_{k-1}\| < \varepsilon$, el să ne-o dea pe x_{41} ca rezultat final, deşi dacă ar continea, procesul ar converge poate în altă parte. (desigur că dacă apropierea de soluție e demonstrată monotonă, se elimină astfel de posibilități.

Se poate ajunge în situațiile mai sus menționate, pentru că, din păcate, algoritmii de rezolvare a sistemelor de ecuații algebrice neliniare nu sînt universali, și au la bază anumite teoreme de convergență, care cer satisfacerea anumitor condiții de ipoteză, Si mai neplăcut este faptul că aceste aspecte nu sînt în general scoase în relief, pentru ca utilizatorul să le poată sesiza.

De aceea, dacă rezultatele din paragrafele precedente ar putea sta la baza eliminării, din start a circuitelor „nepermise", pentru cele „nedecise", avem nevoie de noi teoreme de calitate, de data aceasta privind algoritmii de calcul ai ecuației: $F(x) = a$.

Orice metodă de calcul numeric are două faze: 1) separarea rădăcinilor (să știm că într-o anumită regiune V, ecuația are o soluție unică)

2) aplicarea formulei de recurență convergente.

Ori, se remarcă (v [7]) că nu există algoritmi generali de separare a soluțiilor unei ecuații neliniare. Ajungem, din nou, la importanța pe care ar putea-o avea o demonstrare prealabilă a existenței și unicității soluției – în spațiul R^n, sau un rezultat care să ne ofere <u>zona</u> în care soluțiile se pot afla.

Există mai multe metode:

- metode bazate pe liniarizarea pe porțiuni (v. lucrările respective) <u>Nu ne ocupăm de ele, aici</u>
- metode bazate pe o teoremă de punct fix.
- metode de tip Newton
- bazate pe coborârea gradientului etc.

(2) Metode de punct fix

(a) Definiții

Vom folosi mai jos noțiunile introduse la C4 (pag 84)

Astfel să amintim că, după introducerea unei metrici d, cu proprietățile

$$(137)\quad \begin{cases} d(x,y) \geq 0\ , & d(x,y)=0 \Rightarrow x=y \\ d(x,y) = d(y,x) \\ d(x,y)+d(y,z) \geq d(x,z) \end{cases}$$

se arată că spațiul R^n este complet (deci orice șir ~~convergent~~ Cauchy: $\left(\forall\varepsilon\ \exists N(\varepsilon)\ \text{a.î.}\ d(x_n - x_m) < \varepsilon \text{ pt un } m, n \geq N(\varepsilon)\right)$

este un șir convergent: (138)

$$\left(\exists a \text{ aî } \forall \varepsilon,\ \exists N(\varepsilon) \text{ pt care } |x_n - a| < \varepsilon \text{ pt } n \geq N(\varepsilon)\right)$$

Să mai introducem și noțiunea de contracție, (operator contractant T): dacă $\exists\ \alpha \in (0,1)$ aî:

$$d(Tx, Ty) \leq \alpha\, d(x,y) \quad \forall x, y \in R^n \qquad (139)$$

iar un punct fix al unui operator este acela care verifică:

$$(140)\quad \begin{array}{ll} Tx = x & \text{(notat ca operator)} \\ T(x) = x & \text{(notat ca funcție)} \end{array}$$

ⓑ Teorema lui Banach T1

Enunț: Dacă E este un spațiu metric complet (de pildă R^n) atunci operatorul contractant T are un punct fix, unic.

Demonstrație (v [4])

Fie $x_0 \in E$ fixat și fie șirul

$$x_1 = Tx_0,\ x_2 = Tx_1, \ldots\ldots\ x_n = Tx_{n-1}$$

Avem:

$d(x_{k+1}, x_k) = d(Tx_k, Tx_{k-1}) \leq \alpha\, d(x_k, x_{k-1})$, care aplicată succesiv duce la:

$$d(x_{k+1}, x_k) \leq \alpha^k d(x_1, x_0) \quad k = 0, 1, 2 \ldots \qquad (141)$$

Pentru $m > n$, două numere naturale:

$$d(x_m, x_n) \overset{(137)}{\leq} d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \dots + d(x_{n+1}, x_n)$$
$$\overset{(141)}{\leq} \left(\alpha^n + \alpha^{n+1} + \dots + \alpha^{m-1}\right) d(x_1, x_0)$$

Dar $\alpha^n(1 + \alpha + \dots + \alpha^{m-n-1}) \leq \alpha^n(1 + \alpha + \alpha^2 + \dots) = \dfrac{\alpha^n}{1-\alpha}$

(α subunitar)

Deci:

$$d(x_m, x_n) \leq \frac{\alpha^n}{1-\alpha} d(x_1, x_0)$$

De aici, deducem că x_n e un şir Cauchy (138) şi automat, într-un spaţiu complet, un şir convergent la o limită x. În plus:

$$d(x, Tx) \leq d(x, x_n) + d(x_n, Tx_n) + d(Tx_n, Tx) \leq$$
$$\leq 2d(x_n, x) + d(x_n, x_{n+1}) \qquad (142)$$

Dar în relaţia (142), ambii membri pot fi făcuţi oricât de mici, ~~...~~ primul pentru că x_n e convergent, al 2lea pentru că e şir Cauchy (v (138))

În concluzie $d(x, Tx) = 0 \Rightarrow \boxed{x = Tx}$ (v 137)

şi x era chiar punct fix. Să arătăm că e unic:

Dacă ar mai exista un punct fix x^0, : $Tx^0 = x^0$, am avea $d(x, x^0) = d(Tx, Tx^0) \leq \alpha d(x, x^0)$

sau $d(x, x^0)(\alpha - 1) \geq 0 \Rightarrow d(x, x^0) = 0$
$(d(x, x^0) \geq 0)(\alpha < 1)$

$\Rightarrow x = x^0$ q.e.d.

© Ideea folosirii teoremei lui Banach

Pornind de la $F(x) = 0$, se caută un operator φ, a.î. $\varphi(x) = x \Longleftrightarrow F(x) = 0$. Apoi, dacă reușim să arătăm că φ e o contracție, va rezulta din teorema lui Banach că șirul:

(143) $\boxed{\varphi(x_n) = x_{n+1}}$

converge la x, punctul fix al contracției și soluție a ecuației.

Să remarcăm că dispunem de o mare libertate în privința φ-ului ales și a metricii față de care facem verificarea condiției de contracție.

Pe această libertate au profitat în ultimul timp unii analiști care au elaborat metode adaptate ecuațiilor care se obțin în circuitele cu tranzistori:

(d) $\boxed{\text{Teorema 2}}$ (Sandberg) - de punct fix.
(144)

Dacă: 1) $\boxed{F \in \mathfrak{F}^n}$ ($f_1, f_2 \dots f_n$ $\boxed{\in \mathfrak{F}}$, adică sunt funcții strict crescătoare, ducînd R pe R)

(145) 2) $\dfrac{f_i(\alpha) - f_i(\beta)}{\alpha - \beta} \geq \varepsilon \quad \forall\, \alpha, \beta \in R$

3) Matricea A e slab dominantă pe linii:

(146) $$a_{ii} \geq \sum_{j \neq i} |a_{ij}| \quad \text{pentru } i = 1 \dots n$$

Atunci, pentru ecuația $\boxed{F(x) + Ax = B}$ (147)

Algoritmul : $\cancel{X}_{n+1} = ((F + B) + D)^{-1} (-\Delta) x_n$ (148)

$$\text{unde} \begin{cases} D = \text{diag } A \\ \Delta = A - \text{diag } A \end{cases} \qquad (149)$$

este convergent la soluție.

———.

<u>Observație</u> : Notația 148 vrea să spună că se va rezolva ecuația

$$(F(\cancel{X}_{n+1}) - B) + D\cancel{X}_{n+1} = -\Delta x_n \qquad (148')$$

<u>Justificare</u>

Se scrie ecuația 147 în forma:

$$F(\cancel{X}) + Ax = B \Rightarrow F(x) + Dx + \Delta x = B$$

$$\Rightarrow \quad F(x) - B + Dx = -\Delta x \qquad (147')$$

În această formă, pentru a putea aplica T lui Banach de punct fix, ar trebui să folosim un operator T a.î. $\cancel{X} = T\cancel{X}$ să reprezinte o soluție pentru (147') deci

$F(Tx) - B + D(Tx) = -\Delta x$ sau notat operațional : $Tx = ((F(\ ) - B) + D)^{-1} (-\Delta) x$

(v rel. 148)

Verificare : $Tx = [(F(\ ) - B) + D(\ )]^{-1} (-\Delta) x \Leftrightarrow$

$$\begin{cases} F(Tx) - B + D(Tx) = -\Delta x & \rightarrow \text{prin construcție} \\ F(x) - B + D x = -\Delta x & \rightarrow \text{ecuația dată} \end{cases}$$

Din ultimele relații se vede tocmai faptul că $x = Tx$ va fi și o soluție a ecuației.

Justificarea se termină (v. art [10-11]) arătînd că T este o contracție.

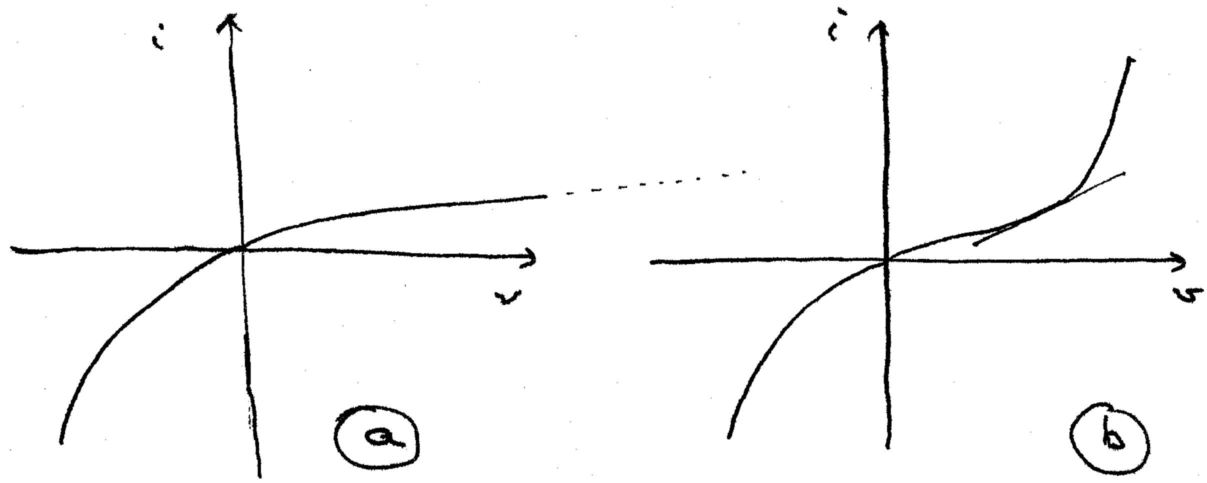
e) Aplicație pentru circuitul de probă

În primul rînd să urmărim verificarea condițiilor din ipoteză.

Condițiile asupra funcțiilor sînt cele care trebuie discutate mai întîi. Ele vor genera o discuție valabilă și pentru paragrafele care urmează, de aceea o vom face aici mai în detaliu.

Mulți algoritmi (cum este acesta) reclamă ca funcțiile f care intervin în modelele diodelor să fie surjective, adică să aplice R^n pe R^n (clasa $\tilde{F}$ - era clasa funcțiilor strict crescătoare și surjective). Desigur că diodele standard nu au această proprietate.

Dacă se ține cont de anumite efecte fizice (ca efect de rezistența de ex) modelatorul poate lua în considerație o pantă mică, dar pozitivă ca în fig 40a sau să țină cont și de efectul de avalanșă (în limita reversibilă (fig 40b)



— fig 40 — : „Diode F" (surjective, strict crescătoare)

Dincolo de aspectele fizice mai sus menționate, putem considera fig 40 și ca o prelungire a caracteristicii standard cu domeniu limitat. In acest sens vor trebui făcute întotdeauna verificările finale de coincidență (v pag 136)

Cea de a doua condiție impusă funcțiilor f_1, f_2, era ca ele să aibă panta (panta unei coarde oarecare) mărginită inferior de un număr pozitiv c. Ambele figuri din 40 satisfac această pretenție. (chiar în b, deși funcția devine din concavă - convexă, panta nu ajunge la 0) (Derivata I are o discontinuitate)

In sfîrșit să trecem la condițiile impuse matricii A (la noi $T^{-1}G$, calculată la anterior, pentru a vedea ce condiții obținem în cazul concret:

$$(150)\quad A = \frac{1}{1-\alpha_r+\alpha_t}\begin{bmatrix}1 & \alpha_r\\ \alpha_t & 1\end{bmatrix} G = k\begin{bmatrix} G_3(1-\alpha_r) & G_3(\alpha_r-1)+\alpha_r(G_1+G_2)\\ G_3(\alpha_t-1) & G_1+G_2+G_3(1-\alpha_t)\end{bmatrix}$$

şi condiţiile (146) devin

$$(151)\quad \begin{cases} a_{11} \geq |a_{12}| \\ a_{22} \geq |a_{21}| \end{cases} \quad \text{sau} \quad \begin{cases} G_3(1-\alpha_r) \geq |G_3(\alpha_r-1)+\alpha_r(G_1+G_2)| \\ G_1+G_2+G_3(1-\alpha_t) \geq |G_3(\alpha_t-1)| \end{cases}$$

A doua relaţie, dacă, $\boxed{G_1, G_2, G_3 > 0,\ \alpha_r, \alpha_t \in (0,1)}$ (152)

devine:

$G_1+G_2+G_3(1-\alpha_t) \geq G_3(1-\alpha_t)$ şi este evident

îndeplinită pentru $G_1+G_2>0$.

Pentru prima relaţie avem cazurile:

1) $-G_3(1-\alpha_r)+\alpha_r(G_1+G_2) \leq 0 \iff$

$G_3(1-\alpha_r) > \alpha_r(G_1+G_2) \iff G_3 \geq \frac{\alpha_r}{1-\alpha_r}(G_1+G_2)$ (153')

în care caz avem:

$G_3\cancel{(1-\alpha_r)} \geq \cancel{(1-\alpha_r)}G_3 - \alpha_r(G_1+G_2)$ evident îndeplinită

2) $-G_3(1-\alpha_r)+\alpha_r(G_1+G_2) \geq 0$, deci

$G_3 \leq \frac{\alpha_r}{1-\alpha_r}(G_1+G_2)$ relaţia fiind:

$G_3(1-\alpha_r) \geq -G_3(1-\alpha_r)+\alpha_r(G_1+G_2)$ sau

$2G_3(1-\alpha_r) \geq \alpha_r(G_1+G_2)$ sau:

$$G_3 \geq \frac{\alpha_r}{2(1-\alpha_r)}(G_1+G_2) \quad (153'')$$

Concluzia va fi deci că (146) sunt îndeplinite

dacă $\boxed{G_3 \geq \frac{\alpha_r}{2(1-\alpha_r)}(G_1+G_2)}$ (153)

Am obținut așadar concluzia (pt ex. de probă):

$\boxed{\text{Concluzia 1}}$

Dacă ex. de probă (fig 22) satisface:

$$\begin{cases} 1)\ f_1, f_2 \in \tilde{\mathcal{F}} \\ 2)\ \dfrac{f(\alpha)-f(\beta)}{\alpha-\beta} \geq \varepsilon \quad \forall\, \alpha, \beta \\ 3)\ G_1, G_2, G_3 > 0 \qquad \alpha_r, \alpha_t \in (0,1) \\ 4)\ G_3 \geq \dfrac{\alpha_r}{2(1-\alpha_r)}\,(G_1+G_2) \end{cases} \qquad (154)$$

atunci se poate aplica algoritmul (sigur convergent) descris mai jos

. ——— .

__Algoritm__

$v^{(n+1)}_{n+1} = \begin{pmatrix} v^{(n+1)}_1 \\ v^{(n+1)}_2 \end{pmatrix}$ este obținut din ecuația:

$$f\left(v_n^{(n+1)}\right) - b + \operatorname{diag} A \cdot v^{(n+1)} = -(A - \operatorname{diag} A)\, v^{(n)}$$

Mai clar:

$$\begin{cases} f_1\left(v_1^{(n+1)}\right) - b_1 + a_{11} v_1^{(n+1)} = -a_{12} v_2^{(n)} \\ f_2\left(v_2^{(n+1)}\right) - b_2 + a_{22} v_2^{(n+1)} = -a_{21} v_1^{(n)} \end{cases} \qquad (154')$$

$$\left(\text{cu } \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -G_3 \\ G_2+G_3 \end{pmatrix} E_c. \right)$$

$a_{11}, a_{12}, a_{21}, a_{22}$ date în relația (150).

. ——— .

__Observație__

1) Faptul că relațiile (154) sunt bine definite se vede imediat dacă observăm că funcțiile:

$t_1(x) + a_{11}x - b_1 = \hat{\mathcal{F}}_1$

$t_2(x) + a_{22}x - b_2 = \hat{\mathcal{F}}_2$ <u>sînt bijecții</u>, deoarece $t_1, t_2 \in \hat{\mathcal{F}}$, iar $a_{11}, a_{22} > 0$ (aceasta rezultă întotdeauna din 146), deci $a_{11}x, a_{22}x$ sînt strict crescătoare și surjective. Deci și suma e strict crescătoare și surjectivă, deci bijectivă. Aceasta înseamnă printre altele că ecuațiile 154 <u>au întotdeauna o soluție unică</u>.

2) Să mai remarcăm faptul că, s-a reușit aducerea ecuației inițiale la două ecuații neliniare <u>de o singură variabilă</u>, pentru care dispunem de numeroase proceduri de calcul.

3) Ultima remarcă este <u>dependența concluziilor de îndeplinirea anumitor condiții</u>, fapt semnalat anterior (și universal). Totuși acestea sînt condiții suficiente, nu și necesare de convergență și nu este exclus ca, deși nu sînt îndeplinite condițiile amintite un anumit proces să fie convergent.

4) În sfîrșit să nu uităm că $t_1, t_2 \in \hat{\mathcal{F}}$, așa că dacă am obținut aceasta printr-o prelungire, putem avea surpriza ca soluția calculată să fie în afara domeniului

recal de definiție. Se la baza deci în final verificarea de coincidență, în astfel de cazuri.

③ Metode de tip Newton

Nu vom intra în detalii generale privind această metodă. Ideea este o generalizare liniară a cunoscutei metode Newton pentru cazul ecuațiilor cu o singură variabilă (înlocuirea funcției cu tangenta într-un punct etc)

Astfel că se ajunge la algoritmi de forma:

(155) $x^{k+1} = x^{k} - [J_k]^{-1} f(x^{k})$ $\quad k = 0, 1, 2, \dots$

(J_k matricea Jacobiană)

De la început să remarcăm că formula este bine definită dacă $\exists J_k^{-1}$, adică $\det J_k \neq 0$, condiție care se aseamănă amintind și de aceea din teoremele lui Palais privind existența și unicitatea.

Aceasta însă nu este singura condiție necesară procesului iterativ (155). Un exemplu de condiții ni-l oferă teorema lui Kantorovici, pe care o dăm mai jos, doar cu caracter ilustrativ:

Teorema 3 (Kantorovici) – detalii in [?]

<u>Dacă</u> în sfera deschisă

$S: \{x \in X,\ \|x - x_0\| \leq r\}$ sunt satisfăcute condițiile

a) J_f pentru $x = x_0$ are o inversă cu proprietatea că $\|J_f^{-1}\| \leq A_0$

b) $\|J_f^{-1} \cdot f(x_0)\| = \|x^{(1)} - x^{(0)}\| \leq B_0$

c) $\sum_{p=1}^{n} \frac{\partial^2 f_i(x)}{\partial x_j \partial x_p} \leq C \quad i, j = 1 \dots n,\ x \in S$

d) constantele A, B, C, satisfac

$$h_0 = 2n A_0 B_0 C \leq 1$$

<u>Atunci</u> pentru aproximația inițială x^0 șirul (15) este convergent la <u>soluția ecuației</u>

Doar câteva comentarii (observații):

1) dependența de valoarea inițială x_0

2) Condiția a) înseamnă printre altele:

$\left\| \frac{1}{\det J_f} J_f^{*} \right\| \leq A_0$, de unde se vede clar că nu va mai fi suficient ca $\det J_f \neq 0$ și se impune ca $\det J_f \geq \varepsilon$ $(\varepsilon > 0)$

3) Dependența tuturor condițiilor de diverși parametri din scrierea ecuației.

Verificările sus puse ar trebui făcute de fiecare

care dată, fapt inoperant. De aceea

4) Dacă am putea dispune de anumite teoreme a căror valabilitate să fie verificabilă într-o manieră care să depindă mai puțin critic de valorile parametrilor, avantajul ar fi evident. În acest sens a fost dirijat efortul unor cercetători (art 7-10-11) care au stabilit deja unele rezultate promițătoare. Unele din acestea sînt urmărite în continuare.

$\boxed{\text{Teorema 4}}$ (Wilson)

Fie ecuația $F(x)+Ax=B$. <u>Dacă</u>:

(156)

1) $F \in \mathfrak{F}$ (F e o aplicație diagonală cu f_k strict crescătoare și surjectivă)
2) Ori toate funcțiile f_i sînt convexe, ori concave
3) $a_{ii} \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|$ pt $i=1,\dots n$
4) $a_{ij} \leq 0 \quad i \neq j$

<u>Atunci</u>: algoritmul de mai jos (157) este convergent.

<u>Algoritm</u>: (157) $x^{k+1} = \left[F'(x^k)+A\right]^{-1}\left[B - F(x^k)+F'(x^k)\,x^k\right]$

unde $F'(x_k) = \begin{pmatrix} f'_1(x_1) & & \\ & \ddots & \\ & & f'_n(x_n) \end{pmatrix}$

(157')

Observații

1) Și acest algoritm este bine definit, căci este bine de arătat că $f \in \mathcal{F} \Rightarrow f$ continuă și mai mult că există cel puțin derivata la dreapta în orice punct $f'_{kd}(x)$, care intervine în (157')

2) ideea de pornire (tip Newton) este de a înlocui pe $F(x)$ din ecuația de pornire.

$$F(x) \simeq F(x_k) + F'(x_k)(x - x_k)$$ (se folosește pentru toate coordonatele aproximarea dreptei curbei prin tangenta în x_k). Deci, ecuația aproximată este:

$$F(x_{k+1}) + A x_{k+1} = B \text{ sau}$$

$$F(x_k) + F'(x_k)(x_{k+1} - x_k) + A x_{k+1} = B \text{ sau}$$

$$[F'(x_k) + A] x_{k+1} = B - F(x_k) + F'(x_k) \cdot x_k,$$ de unde imediat, relația 157. $F'(x_k) + A$ va avea în mod cert inversă, căci $f'_k \geq 0$ ($f_k \in \mathcal{F}$) și A satisface proprietatea 3) (151), deci $F'(x_k) + A$ este tare dominantă pe linii (inegalitatea 3 este ușurată de apariția lui F'). Ori (T7, IV B) arată înseamnă că $F' + A$ este nesingulară.

Să trecem acum la exemplul de practică

A) Condițiile 1 – 4) devin:

1) $F \in \tilde{\mathcal{F}}$, aceeași din cazul aplicației precedente.

2) Condiția de concavitate sau convexitate este o restricție mai puternică. Sînt respinse variantele din fig 40b, dar ceea ce este mai grav e că <u>că toate joncțiunile sînt obligate să fie de aceeași tip</u> (pn, sau np)

3) 4) Condiția impusă matricii A, a fost discutată la pag 163, mai exact prima variantă ($a_{12} \leq 0$). Prelucrăm rezultatul:

$$(158)' \quad \boxed{G_3 \geq \frac{\alpha_r}{1-\alpha_r}(G_1+G_2)}$$

Am obținut așadar concluzia

$\boxed{\text{Concluzia 2}}$

<u>Dacă</u>

1) $k_1, k_2 \in \tilde{\mathcal{F}}$

2) $G_1, G_2, G_3 > 0 \qquad \alpha_r, \alpha_f \in (0,1)$ $\qquad (158)'$

3) $G_3 \geq \dfrac{\alpha_r}{1-\alpha_r}(G_1+G_2)$

4) Toate funcțiile f_i sînt convexe ori concave.

<u>Atunci</u>: se poate folosi algoritmul (sigur conv.)

$$(159)\quad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_{k+1} = \left\{ \begin{bmatrix} k_1'\big|_{v_1^{(k)}} & 0 \\ 0 & k_2'\big|_{v_2^{(k)}} \end{bmatrix} + A \right\}^{-1} \left\{ B - \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}_k + \begin{pmatrix} k_1'\big|_{v_1^{(k)}} & 0 \\ 0 & k_2'\big|_{v_2^{(k)}} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_k \right\}$$

Comentariu (comparație între C.1 și C 2)

In primul rînd vom remarca faptul că relațiile (158) definesc condiții mai grele decît 154. Ele impun restricții suplimentare atît pentru elementele matricei (o condiție de două ori mai grea pentru c_1, c_2, c_3) dar mai ales pentru funcții (toate diodele să fie pn, sau toate np). Aceasta înseamnă că în acest caz, metoda de punct fix e mai bună decît cea de tip Newton, ~~d. p. d.~~ din punctul de vedere al lejerității condițiilor care asigură convergența.

Totuși, dacă ambele procedee sînt valabile pentru o problemă dată (sînt satisfăcute cond. 154 și 158) nu putem întrebare care procedeu trebuie preferat.

Pot exista mai mulți factori de decizie. (vizînd în final economia de timp pentru calculator). De exemplu să observăm că (157) cere inversarea unei matrici, la fiecare pas, iar (154'), rezolvarea a două ecuații implicite. (diferența de rigoare o poate face un specialist)

Un alt factor care ne-ar putea interesa este viteza de convergență a procesului. O comparație între cele două tipuri de metode se va face la par 7.

(4) Metoda coborârii gradientului

Oricare dintre teoremele precedente a impus verificarea anumitor condiţii, de exemplu aceea de slabă dominanţă pe linii. (v 154, 158).

Ori, noi am demonstrat, pentru problema noastră, existenţa şi unicitatea soluţiei. Deci necesitatea asigurării convergenţei algoritmilor 159, 154' s-a impus restricţii care nu ne convin.

Rezultatul din teorema care urmează îşi are tocmai aici punctul tare (forţa): el certifică convergenţa unui anumit algoritm, exact în acele aşi condiţii care de obicei sînt folosite pentru a demonstra existenţa şi unicitatea soluţiilor.

Astfel, din discuţiile anterioare, s-a văzut că $A \in P_0$ a fost o condiţie urmărită. Teoremele 3 şi 14 din cap IV (C şi D) ne asigură că, dacă $F \in \hat{F}^n$, atunci

$$\cancel{A}F(x) + Ax = B \qquad (160)$$

$$AF(x) + Bx = C \qquad (161)$$

(formele standard)

au o soluție unică dacă

$$\boxed{\begin{array}{c} A \in P_0 \\ \text{respectiv} \\ (A, B) \in \mathcal{W}_0 \end{array}} \begin{array}{c} (162) \\ \\ (163) \end{array}$$

Direcția care urmează enunțul teoremei își propune:

– să expună mecanismul demonstrației (măcar parțial), adică al principiului coborîrii gradientului

– să justifice rolul clasei $\mathcal{W}_0$, anticipînd astfel introducerea ei la cap IV.

– să sublinieze legătura dintre $\mathcal{W}_0$ și P_0.

– să redea <u>efectiv</u>, mersul de calcul, pe exemple de probă.

Teorema 5

<u>Dacă</u> pentru ecuația $A\,F(\alpha) + B\,\alpha = c$ (160):

1) $(A, B) \in \mathcal{W}_0$ (164)

2) $F \in \tilde{\mathcal{F}}^m \cap C^1$ (165)

<u>Atunci</u> este întotdeauna convergent la soluție algoritmul de mai jos: . —— .

<u>Algoritm</u>

– Se ia o matrice oarecare U $(n \times n)$ pozitiv definită simetrică, cu elem. reale.

– Se construiește funcția:

(166) $\varphi(v) = [A F(v) + Bv - c]^T M [A F(v) + Bv - c]$

– Se construiește gradientul și $\nabla\varphi = \begin{bmatrix} \frac{\partial\varphi}{\partial v_1} \\ \vdots \\ \frac{\partial\varphi}{\partial v_m} \end{bmatrix}$ (167)

– Se construiește, pentru r oarecare funcția de control:

$$g(v,r) = \begin{cases} \dfrac{\varphi(v) - \varphi[v - r\nabla\varphi(v)]}{r\,\|\nabla\varphi(v)\|^2} & r>0 \\ 1 & r=0 \end{cases} \qquad (168)$$

– ~~Atunci~~ Se alege un σ oarecare: $0 \le \sigma \le \frac{1}{2}$, și un x_0 arbitrar în E^n și se definește șirul:

$$\boxed{x^{k+1} = x^k - r^k \nabla\varphi(x^k)} \qquad (169)$$

– Unde numărul r^k (nu e putere!), trebuie să verifice condiția:

$$(170) \quad \begin{cases} \sigma \le g(x^k, r^k) \le 1-\sigma & \text{pentru } g(x^k,1) \le \sigma \\ r^k = 1 & \text{pentru } g(x^k,1) > \sigma \end{cases}$$

———

Ori, se arată în cap IV (și de altfel am arătat și mai în par anterioare), că dacă $T^{-1} G \in P_0$, atunci $(T, G) \in W_0$, și ca urmare teorema este aplicabilă cazului nostru. În general, dacă observăm că $(I, A) \in W_0$, dacă $A \in P_0$ (I matricea unitate) obținem corolarul:

Teorema 5 bis

<u>Dacă</u> pentru ecuația $F(x) + Ax = B \quad (161)$:

1) $A \in P_0$
2) $F \in \mathcal{F}^n \cap C^1$

<u>atunci</u> este întotdeauna convergent algoritmul:

<u>Algoritm</u> exact ca cel precedent, cu singura deosebire că

$$\varphi(v) = [F(v) + Av - B]^T M [F(v) + Av - B] \quad (167\,bis)$$

<u>Comentariile</u> - se vor face (pentru simplitate), pe cazul particular al problemei exemplu, în paralel cu exemplificarea procedurii.

Să pornim de la ecuația:

$$(171) \quad T F(v) + G v = B \qquad , \text{ sau pe larg:}$$

(171')

$$\begin{pmatrix} 1 & -\alpha_5 \\ -\alpha_4 & 1 \end{pmatrix} \begin{pmatrix} f_1(v_1) \\ f_2(v_2) \end{pmatrix} + \begin{pmatrix} G_3 & -G_3 \\ -G_3 & G_1+G_2+G_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -G_3 \\ G_2+G_3 \end{pmatrix} E_c.$$

și să aplicăm procedura descrisă:

1) Se formează mai întâi

$$\varphi(v) = [T F(v) + G v - B]^T M [T F(v) + G v - B]$$

unde M este liberă, cu singura condiție <u>de a fi pozitiv definită și simetrică</u>. Dar aceasta va însemna că forma $\varphi(v)$ este și ea pozitiv definită $(>0, \forall v \neq 0)$

—176—

Vectorul ~~Matricea~~ $TF(v) + Gv - B$ este:

$$\begin{bmatrix} k_1(v_1) - \alpha_r k_2(v_2) + G_3 v_1 - G_3 v_2 + E_c G_3 \\ -\alpha_f k_1(v_1) + k_2(v_2) - G_3 v_1 + (G_1+G_2+G_3) v_2 - (G_2+G_3) E_c \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = u \qquad (172)$$

Aşa încât forma patratică este:

$$Q(v) = m_{11} u_1^2 + \overbrace{(m_{12}+m_{21})}^{2m_{12}} u_1 u_2 + m_{22} u_2^2 > 0 \qquad (173')$$

Să ne hotărâm de exemplu pentru

$$(174) \quad M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{deci}$$

$$Q(v) = u^t I u = u^t u = \|u\|^2 \qquad \text{şi regăsim}$$

$$(173) \quad Q(v) = u_1^2 + u_2^2 \qquad \left(\text{din 173 cu } \begin{matrix} m_{12} = m_{21} = 0 \\ m_{11} = m_{22} = 1 \end{matrix}\right)$$

2) Să calculăm gradientul lui Q (v. 167)

$$\nabla Q(v) \overset{(174)}{=} \begin{bmatrix} 2u_1 \frac{\partial u_1}{\partial v_1} + 2u_2 \frac{\partial u_2}{\partial v_1} \\ 2u_1 \frac{\partial u_1}{\partial v_2} + 2u_2 \frac{\partial u_2}{\partial v_2} \end{bmatrix} = 2 \begin{bmatrix} u_1(k_1' + G_3) + u_2(-\alpha_f k_1' - G_3) \\ u_1(-\alpha_r k_2' - G_3) + u_2(k_2' + G_1 + G_2 + G_3) \end{bmatrix}$$

$$(175) = \begin{bmatrix} k_1' + G_3 & -\alpha_f k_1' - G_3 \\ -\alpha_r k_2' - G_3 & k_2' + G_1 + G_2 + G_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \left\{ \begin{pmatrix} 1 & -\alpha_f \\ -\alpha_r & 1 \end{pmatrix} \begin{pmatrix} k_1' & 0 \\ 0 & k_2' \end{pmatrix} + G \right\} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad (176) = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$

Discuţie

Am ajuns acum la o observaţie evidentă. Dacă am putea fi siguri că $\det\left\{ T \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} + G \right\} \neq 0 \ \forall d_1, d_2 > 0$ (căci $k_1, k_2 \in \tilde{F} \cap C^1 \Rightarrow k_1', k_2' > 0$) atunci relaţia (176) ne arată că $\nabla Q(v)$ se anulează numai pentru $u = \Theta$

Ori $u = 0 \Rightarrow \begin{matrix} u_1 = 0 \\ u_2 = 0 \end{matrix}$ este (172) tocmai ecuația noastră

Așadar $\nabla Q(v)$ se anulează numai în soluțiile ecuației (171), dacă

$$\det\left\{ T \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} + G \right\} \neq 0 \quad \forall\, d_1, d_2 > 0. \qquad (177)$$

Condiția (177) nu este întâmplătoare: am obținut-o ca suficientă pentru unicitatea soluțiilor ecuației $T F(v) + G v = B$. Se arată că dacă $F \in \tilde{\mathcal{F}}^1$ se asigură și existența. Vom arăta deci că algoritmul e convergent oricând am demonstrat existența și unicitatea pe această cale.

Mai mult, condiția (177) este exact condiția de definiție a clasei matricilor $\mathcal{K}_0$: $(T, G) \in \mathcal{K}_0$.

În sfârșit, profitând de faptul că $\exists\, T^{-1}$, mai am obținut forma $\det\left\{\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} + T^{-1} G\right\} \neq 0$, sau cu $A = T^{-1} G$, $A \in P_0$. Aceasta sugerează legătura dintre clasa de matrici P_0 și clasa de perechi $\mathcal{K}_0$.

Continuarea demonstrației (v. art [13]) se prevalează de faptul că $\nabla Q(v) = 0$ numai pentru sol. ecuației, că $Q(v)$ este pozitiv definită, o tratează de coborârea gradientului într-o anumită sferă, analizează și ajunge în câțiva pași la concluzia finală din T6.

3) Să calculăm acum și funcția de control (168)

$$\varphi(v) - \varphi(v - \gamma\nabla\varphi) = u_1^2 + u_2^2 - \varphi[v - \gamma\nabla\varphi] \quad \text{dar}$$

$$v - \gamma\nabla\varphi = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \gamma\left\{T\begin{pmatrix} k_1' & 0 \\ 0 & k_2' \end{pmatrix} + G\right\}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$ și de aici:

(178)
$$v - \gamma\nabla\varphi = \begin{bmatrix} v_1 - \gamma[(k_1' + G_3)u_1 - (k_1' + G_3)u_2] \\ v_2 - \gamma[-(k_1 + k_2' + G_3)u_1 + (k_2' + G_1 + G_2 + G_3)u_2] \end{bmatrix} = \begin{bmatrix} v_1 - \gamma w_1 \\ v_2 - \gamma w_2 \end{bmatrix}$$

Deci $\varphi(v) - \varphi(v - \gamma\nabla\varphi) = u_1^2 + u_2^2 - (v_1 - \gamma w_1)^2 - (v_2 - \gamma w_2)^2$

$\|\nabla\varphi(v)\| = w_1^2 + w_2^2$

și

(179)
$$g(v,\gamma) = \begin{cases} \dfrac{u_1^2 + u_2^2 - (v_1 - \gamma w_1)^2 - (v_2 - \gamma w_2)^2}{\gamma(w_1^2 + w_2^2)} & \text{pt } \gamma > 0 \\ 1 & \text{pt } \gamma = 0 \end{cases}$$

cu u_1, u_2 definiți la (172), w_1, w_2 la (178).

Observație

Demonstrarea completă a teoremei trebuie să ~~demonstr~~ arate ~~toate acum~~ că algoritmul e bine definit, deci că ecuația se poate folosi (170) pentru a obține un (~~...~~) γ^k care să fie introdus în (169). Pentru aceasta se demonstrează că funcția $g(v,\gamma)$ definită mai sus e continuă în γ, pe $[0,\infty)$.

4) De aceea examinând procesul de aflare a lui γ^k vedem că, după ce se calculează un x^k (la început x^0) se începe cu iterarea:

$$i:\ g(x^k,1) \le \sigma \ \text{sau}\ g(x^k,1) \ge \sigma\ ?$$
$$\downarrow i_1 \qquad\qquad \downarrow i_2$$

Dacă răspunsul este i_2, se alege pur şi simplu $r^k = 1$.
Dacă răspunsul este i_1, trebuie găsit un r^k a.î.

$$\sigma \le g(x^k, r^k) \le 1-\sigma. \qquad (170')$$

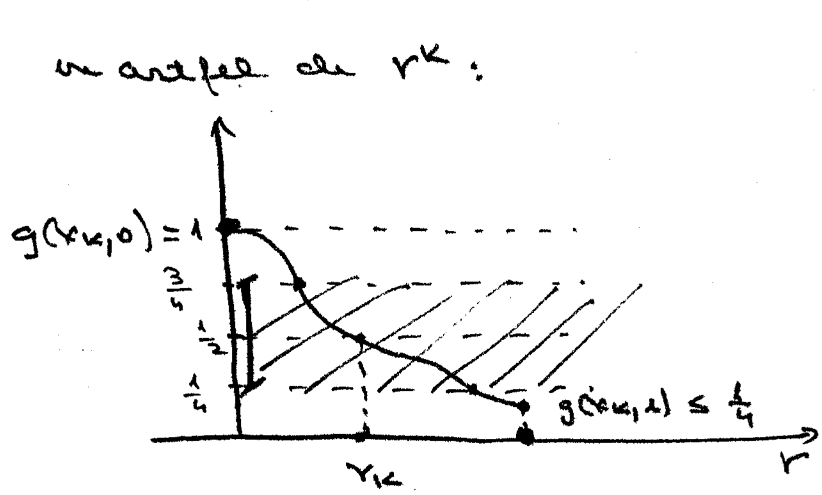
Să presupunem că ne hotărâm pentru $\boxed{\sigma = \frac{1}{4}}$ $\left(\le \frac{1}{2}\right)$

atunci testul devine:

$$\begin{cases} g(x^k,1) \ge \frac{1}{4} \Longrightarrow r^k = 1 \\ g(x^k,1) \le \frac{1}{4} \Longrightarrow r^k \text{ rezultă din relaţia:} \end{cases} \qquad (180)$$

$$\frac{1}{4} \le g(x^k, r^k) \le \frac{3}{4}.$$

În figură se arată că întotdeauna se va găsi un astfel de r^k:



— fig 41 —

Este evident că $\exists\, r_k$ a.î. $g(x_k, r_k)$ să ia o anumită valoare între $\frac{1}{4}$ şi $\frac{3}{4}$. (în general $(\sigma, 1-\sigma)$)

Dacă dorim de exemplu să definim exact pe r^k, am putea căuta pe r^k a.î. $g(x^k, r^k) = \frac{\sigma + 1 - \sigma}{2} = \frac{1}{2}$) ceea ce, o dată obţinut (179) reprezintă o ecuaţie. E posibil însă ca mai practic să fie proba directă a câtorva valori r^k introduse în (179), urmată de (170')

Rezumatul operațiilor de efectuat ($\boxed{\text{Concluzia 3}}$)

(1) Se pregătește calculul lui $\Phi(v)$ (173), u (172), $\nabla\varphi$ (176), $g(v,v)$ (179) (ψ (176) → poate fi introdus în $\nabla\varphi$ și $g(v,v)$)

Se alege δ $\left(0 < \delta \leq \frac{1}{4} \right)$

(2) Se face $k=0$

(3) Se pornește de la x^k (x^0 este dat, oarecare) și se face verificarea $g(x^k, 1) \leq \delta$?

(4) Dacă nu, $\gamma^k = 1$

(5) Dacă da, se caută γ^k a.î. $\delta \leq g(x^k, \gamma^k) \leq 1-\delta$

Se poate proceda prin încercări, sau se rezolvă ecuația $g(x^k, \gamma^k) = \frac{1}{2}$ (în γ^k - care are riguros măcar o soluție, între 0 și 1).

(6) Cu γ^k dat de pașii precedenți se calculează:

$$x^{k+1} = x^k - \gamma^k \nabla\varphi(v)$$

(7) Se face $k \to k+1$ și se revine la (3)

Se va introduce eventual și o condiție de stopare, cum ar fi aceea ca

$$|x^{k+1} - x^k| \leq \varepsilon \quad , \quad \varepsilon \text{ dat.}$$

In această manieră, algoritmul se pretează imediat la scrierea schemei logice și implementarea pe calculator.

Se poate face o subrutină pentru punctul (5).

(Menționez că s-a demonstrat că procedeul duce riguros la soluția unică, deci Concluzia 3)