SULLA PROPAGAZIONE DELL'ERRORE SISTEMATICO IN ALCUNI PROCEDIMENTI ITERATIVI (*)

di ALFREDO BELLEN e ALESSIO VOLČIČ (a Trieste) (**)

SOMMARIO. - *Si dà una maggiorazione della propagazione dell'errore sistematico nella ricerca dei punti fissi per le contrazioni generalizzate.*

SUMMARY. - *An upper bound is given for the roundoff in the computing the fixed points of a generalized contraction.*

I numerosi teoremi di punto unito, di cui la letteratura matematica abbonda, vengono principalmente usati per due scopi: quello di dare teoremi di esistenza e quello di fornire procedimenti iterativi atti al calcolo approssimato delle soluzioni ricercate.

In questa nota ci interessa mettere a fuoco un aspetto particolare di questo secondo modo di utilizzare i teoremi.

Supponiamo che l'equazione che vogliamo risolvere sia

$$Tu = u.$$

In generale la T non é valutabile con esattezza ed il metodo iterativo si applica all'equazione

$$Au = u$$

dove A é una trasformazione che approssima, in un certo senso, la T.

(*) Pervenuto in Redazione il 7 luglio 1973.

Lavoro eseguito col contributo del C. N. R. nell'ambito del Gruppo Nazionale per l'Analisi Funzionale e le sue Applicazioni.

(**) Indirizzo degli Autori: Istituto di Matematica dell'Università - Piazzale Europa 1, 34100 Trieste.

La trasformazione A, però, non gode in generale delle proprietà (continuità, contrattività, etc.) che garantiscono la convergenza delle successioni di Picard ed eventualmente l'unicità della soluzione per la T.

Lo scopo di questa nota è di indagare sull'attendibilità dei risultati che si ottengono applicando il procedimento iterativo alla trasformazione A invece che alla T, nell'ipotesi che sia possibile dare a priori una significativa maggiorazione dell'errore che si commette in ogni iterazione, errore che chiameremo *sistematico* inquanto ripetuto infinite volte.

Indichiamo con S uno spazio metrico completo e con T una trasformazione di S in sè (in seguito faremo delle ipotesi restrittive per la T). Sia inoltre A la trasformazione che approssima la T nel senso che esiste un $\eta > 0$ tale che

$$d(Au, Tu) < \eta \quad \forall u \in S.$$

Sotto convenienti ipotesi per la T, vogliamo studiare la successione di Picard $\{v_n\}_N$, dove $v_n = A^n u_0$ ($A^0 u_0 = u_0$, $A^n u_0 = A(A^{n-1} u_0)$), confrontandola con la analoga successione $\{u_n\}_N$ dove $u_n = T^n u_0$.

Supponiamo ora che la trasformazione T sia una *contrazione generalizzata*, che esistano cioè due costanti reali non negative α e β tali che $\alpha + 2\beta < 1$ ed inoltre

$$d(Tu, Tv) \leq \alpha\, d(u, v) + \beta\,[d(u, Tu) + d(v, Tv)] \qquad \forall u, v \in S.$$

Queste trasformazioni generalizzano le contrazioni di Banach-Caccioppoli [1] e [2] che si ottengono per $\beta = 0$, e quelle di Kannan [3] che si ottengono per $\alpha = 0$.

Ricordiamo che per le contrazioni generalizzate vale il seguente teorema dimostrato indipendentemente da S. Reich [4] e da I. A. Rus [5].

TEOREMA 1. *Se S è uno spazio metrico completo e T una contrazione generalizzata, allora esiste uno ed un solo punto unito per la T ed ogni successione di Picard $\{T^n u\}_N$ vi converge.*

Questo teorema ammette come corollario il teorema di Banach [1] e quello di Kannan [3].

Dimostriamo ora il risultato principale di questa nota.

TEOREMA 2. *Se S è uno spazio metrico completo, T una contrazione generalizzata, η una maggiorazione dell'errore sistematico ed $\tilde{u}$ l'unico punto unito della T, allora*

$$d(v_n, \tilde{u}) < \eta\left(\frac{1-\beta}{1-\alpha-2\beta}\right) + d(u_0, u_1)\,\beta\left[\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1} + \right.$$

$$\left. + \alpha\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-2} + \ldots + \alpha^{n-1}\right] + d(u_0, \tilde{u})\,\alpha^n \ (^1).$$

Risulta:

$$(1)\quad d(v_n, \tilde{u}) \leq d(v_n, Tv_{n-1}) + d(Tv_{n-1}, \tilde{u}) < \eta + d(Tv_{n-1}, \tilde{u}) \leq$$

$$\leq \eta + \alpha d(v_{n-1}, \tilde{u}) + \beta d(v_{n-1}, Tv_{n-1}).$$

Valutiamo ora

$$(2)\quad d(v_n, Tv_n) \leq d(v_n, Tv_{n-1}) + d(Tv_{n-1}, Tv_n) <$$

$$< \eta + d(Tv_{n-1}, Tv_n) \leq \eta + \alpha d(v_{n-1}, v_n) + \beta d(v_{n-1}, Tv_{n-1}) +$$

$$+ \beta d(v_n, Tv_n);$$

da cui

$$(1-\beta)\,d(v_n, Tv_n) < \eta + \alpha d(v_{n-1}, v_n) + \beta d(v_{n-1}, Tv_{n-1}) <$$

$$< \eta + \alpha d(v_{n-1}, Tv_{n-1}) + \alpha d(Tv_{n-1}, v_n) + \beta d(v_{n-1}, Tv_{n-1}) <$$

$$< \eta + \alpha\eta + (\alpha+\beta)\,d(v_{n-1}, Tv_{n-1}),$$

ovvero:

$$(3)\qquad d(v_n, Tv_n) < \eta\,\frac{1+\alpha}{1-\beta} + \frac{\alpha+\beta}{1-\beta}\,d(v_{n-1}, Tv_{n-1}).$$

(1) In accordo con le notazioni precedenti $v_n = A^n u_0$, $d(Au, Tu) < \eta$ $\forall u \in S$ ed $u_1 = Tu_0$.

Si noti che la quantità scritta in parentesi quadra è il termine n-esimo della serie prodotto alla Cauchy delle due serie assolutamente convergenti $\sum_0^\infty{}_n\ \alpha^n$ e $\sum_0^\infty{}_n \left(\frac{\alpha+\beta}{1-\beta}\right)^n$, perciò tende a zero al tendere di n all'infinito.

Dalla (3) si deduce, per ricorrenza, che

$$(4)\quad d(v_n, Tv_n) < \eta \frac{1+\alpha}{1-\beta}\left[1+\frac{\alpha+\beta}{1-\beta}+\left(\frac{\alpha+\beta}{1-\beta}\right)^2+\ldots+\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1}\right]+$$

$$+\left(\frac{\alpha+\beta}{1-\beta}\right)^n d(u_0, u_1).$$

Poiché risulta $0 \leq \frac{\alpha+\beta}{1-\beta} < 1$, possiamo maggiorare la quantità scritta in parentesi quadra con la somma della serie geometrica di ragione $\frac{\alpha+\beta}{1-\beta}$, ottenendo così

$$(5)\qquad d(v_n, Tv_n) < \eta \frac{1+\alpha}{1-\alpha-2\beta}+\left(\frac{\alpha+\beta}{1-\beta}\right)^n d(u_0, u_1).$$

Dalla (1) e dalla (5) si ha

$$(6)\quad d(v_n, \tilde{u}) < \eta + \alpha d(v_{n-1}, \tilde{u}) + \eta \frac{\beta+\alpha\beta}{1-\alpha-2\beta}+\beta\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1} d(u_0, u_1).$$

Dalla (6) si ottiene per ricorrenza che

$$(7)\qquad d(v_n, \tilde{u}) < \eta\left(1+\frac{\beta+\alpha\beta}{1-\alpha-2\beta}\right)(1+\alpha+\ldots+\alpha^{n-1})+$$

$$+\beta d(u_0, u_1)\left[\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1}+\alpha\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-2}+\ldots+\alpha^{n-1}\right]+$$

$$+\alpha^n d(u_0, \tilde{u}) < \eta\left(\frac{(1-\alpha)\cdot(1-\beta)}{1-\alpha-2\beta}\right)\frac{1}{1-\alpha}+$$

$$+\beta d(u_0, u_1)\left[\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1}+\alpha\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-2}+\ldots+\alpha^{n-1}\right]+\alpha^n d(u_0, \tilde{u}).$$

Notiamo che al tendere di n all'infinito la maggiorazione di $d(v_n \tilde{u})$ tende a

$$\eta \cdot \frac{1-\beta}{1-\alpha-2\beta}$$

che rappresenta quindi la *soglia* oltre la quale il procedimento iterativo approssimato non può spingersi. Nell'esempio conclusivo

vedremo che questa valutazione non può essere migliorata, almeno in generale.

Dal teorema 2 possiamo dedurre alcuni corollari.

COROLLARIO 1. *Se, nelle stesse ipotesi del teorema 2, risulta* $\beta = 0$, *si ha*

$$d(v_n, \tilde{u}) < \eta \frac{1}{1-\alpha} + \alpha^n d(u_0, \tilde{u})$$

Si noti che la maggiorazione dipende oltre che da n, α ed η, esclusivamente dalla distanza di u_0 dal punto unito.

Se l'errore sistematico si suppone nullo ($v_n = T^n u_0$), si ottiene la ben nota formula relativa alle contrazioni di Banach-Caccioppoli

$$d(T^n u_0, \tilde{u}) < \alpha^n d(u_0, \tilde{u})$$

COROLLARIO 2. *Se, nelle stesse ipotesi del teorema 2, risulta* $\alpha = 0$, si ha

$$d(v_n, \tilde{u}) < \eta \frac{1-\beta}{1-2\beta} + \beta\left(\frac{\beta}{1-\beta}\right)^{n-1} d(u_0, u_1).$$

Si noti che la maggiorazione dipende, oltre che da n, β ed η, esclusivamente dalla distanza tra u_0 e u_1 ($= Tu_0$).

COROLLARIO 3. *Se nel corollario 2 si suppone nullo l'errore sistematico, si ha*

$$d(T^n u_0, \tilde{u}) < \beta\left(\frac{\beta}{1-\beta}\right)^{n-1} d(u_0, u_1)$$

COROLLARIO 4. *Se nel teorema 2 si suppone nullo l'errore sistematico, si ha*

$$d(T^n u_0, \tilde{u}) < d(u_0, u_1)\,\beta\left[\left(\frac{\alpha+\beta}{1-\beta}\right)^{n-1} + \left(\frac{\alpha+\beta}{1-\beta}\right)^{n-2}\alpha + \ldots + \alpha^{n-1}\right] + \alpha^n d(u_0, \tilde{u}).$$

Diamo ora un esempio in cui si fa vedere che il valore della soglia di approssimazione da noi trovato non può essere migliorato.

Consideriamo la trasformazione

$$Tu = 0{,}5\,(u + 0{,}02 - 10^{-k}) \qquad \text{con } k \text{ intero e } k \geq 3.$$

T risulta una contrazione con $\alpha = 0{,}5$ e $\beta = 0$ il cui unico punto unito è $\tilde{u} = 0{,}02 - 10^{-k}$.

Supponiamo che la trasformazione approssimante Au sia quella che calcola Tu con $k+1$ cifre decimali esatte e poi tronchi il risultato alle prime due cifre decimali

$$Tu = i, a_1\, a_2 \ldots a_n \ldots$$

$$Au = i, a_1\, a_2$$

La minima maggiorazione che si può dare dell'errore sistematico è

$$\eta = 0{,}01$$

per cui il valore della soglia da noi trovato è

$$\eta \cdot \frac{1}{1-\alpha} = 0{,}02$$

Eseguendo l'iterazione $A^n u_0$ con $u_0 = -1$ si ha

$(k+1)$-esima cifra decimale

$$\downarrow$$

$$Tu_0 = -0{,}49000 \ldots 05 \qquad v_1 = Au_0 = -0{,}49$$

$$Tv_1 = -0{,}23500 \ldots 05 \qquad v_2 = Av_1 = -0{,}23$$

$$Tv_2 = -0{,}10500 \ldots 05 \qquad v_3 = Av_2 = -0{,}10$$

$$Tv_3 = -0{,}04000 \ldots 05 \qquad v_4 = Av_3 = -0{,}04$$

$$Tv_4 = -0{,}01000 \ldots 05 \qquad v_5 = Av_4 = -0{,}01$$

$$Tv_5 = +0{,}00499 \ldots 95 \qquad v_6 = Av_5 = 0$$

$$Tv_6 = +0{,}00999 \ldots 95 \qquad v_7 = Av_6 = 0$$

Dunque 0 è punto unito per la trasformazione A ed assumendolo come punto unito per la T si commette un errore pari a $0{,}02 - 10^{-k}$.

Al crescere di k si ottengono trasformazioni le cui approssimanti, nel senso sopra descritto, hanno sempre 0 come punto unito per cui l'errore che si commette è prossimo quanto si vuole alla soglia.

Si noti che A non è una contrazione generalizzata, infatti essa ammette più di un punto unito (è facile verificare che anche $0,01$ è unito per A), e non è neanche continua perché nell'intervallo $[-1,1]$ assume soltanto un numero finito di valori senza essere costante.

BIBLIOGRAFIA

[1] S. BANACH, *Théorie des opérations linéaires,* Waerszawa (1932).

[2] R. CACCIOPPOLI, *Un teorema generale sull'esistenza di punti uniti in una trasformazione funzionale,* Rend. Accad. Naz. Lincei, XI, (1930).

[3] R. KANNAN, *Some results on fixed point theorems,* Amer. Math. Monthly, 76, (1969), 405-408.

[4] S. REICH, *Some remarks concerning contraction mappings,* Can. Math. Bull,

[5] I. A. RUS, *Some fixed point theorems in metric spaces,* Rend. Ist. Mat. Univ. Trieste, vol. III, fasc. II, (1971), 169-172.
14-1, (1971), 121-126.