

È VERO CHE ANCHE I COMPUTER SBAGLIANO? OVVERO, I PERICOLI DOVUTI AGLI ERRORI DI ARROTONDAMENTO

1. INTRODUZIONE

La crescente capillare diffusione degli strumenti di calcolo automatico sia nelle famiglie che nelle aziende può condurre persone di formazione non specialistica a ritenere sostanzialmente superate aritmetica, algebra ed analisi matematica. In questo articolo presentiamo una serie di esempi, alcuni dei quali davvero elementari, di calcoli aritmetici *erronei* effettuati su un «personal computer». Lo scopo è quello di illustrare quanto sia importante la conoscenza dei concetti essenziali dell'analisi matematica per poter bene utilizzare un elaboratore nell'esecuzione di calcoli aritmetici anche fra i più semplici.

È noto che ogni elaboratore digitale lavora con un numero limitato di cifre significative (*). Esso quindi sostituisce il sistema dei numeri reali con un sistema numerico semplificato. Consideriamo, ad esempio, un calcolatore che utilizzi - nella rappresentazione decimale dei numeri - esattamente sette cifre significative; allora con tale calcolatore il numero 111.1313144 viene rappresentato mediante il numero «equivalente» $0.1111313 \cdot 10^3$. Supponiamo ora di eseguire con questo calcolatore l'operazione $2/3$: se il calcolatore utilizza la tecnica di *troncamento numerico* («chopping») otteniamo il risultato approssimato 0.6666666; se invece il calcolatore utilizza la tecnica (più appropriata) di *arrotondamento* («rounding») otteniamo 0.6666667. In generale gli errori dovuti appunto al fatto che un calcolatore non può fare altro che approssimare quei numeri la cui rappresentazione decimale è di lunghezza superiore alla portata del calcolatore stesso sono denominati complessivamente errori di *arrotondamento* («roundoff»).

(*) Qui facciamo riferimento alla rappresentazione a virgola mobile («floating point»), che è quella più frequentemente utilizzata. Si tenga presente che vi sono altri tipi di rappresentazione quale ad esempio quello a virgola fissa. Per brevità non faremo riferimento alle restrizioni riguardanti l'esponente del numero 10 nella rappresentazione a virgola mobile ed ai relativi problemi di «overflow» ed «underflow». Infine seguendo l'uso anglosassone, utilizziamo il punto, anziché la virgola, nella rappresentazione decimale.

Vediamo un semplice esempio. Proviamo a cercare le (eventuali) radici reali dell'equazione di secondo grado

$$ax^2 + bx + c = 0,$$

quando $a = 6$, $b = -31.6$, $c = 41.61$. Le radici si ottengono mediante la ben nota regola di calcolo

$$x_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

In questo caso il calcolo esatto del discriminante dà $\Delta := b^2 - 4ac = 998.56 - 998.64 = -0.08 < 0$, provando così che non esistono radici reali. Se d'altra parte svolgiamo i calcoli lavorando solo con quattro cifre ed arrotondando ad ogni passo, otteniamo erroneamente $\Delta = 0$, e quindi due radici reali coincidenti di valore 2.633. Una delle proprietà più insidiose degli errori di arrotondamento si riscontra quando si effettuano calcoli in sequenza (l'esito del 1° calcolo viene utilizzato nel 2° calcolo, e così via). In situazioni di questo tipo, può verificarsi un'accumulazione di errore, quando gli errori di arrotondamento vengono a propagarsi da un passo al successivo con amplificazioni successive. Si tratta di un fenomeno noto come «instabilità numerica».

In questo articolo vedremo vari esempi nei quali gli errori di arrotondamento si presentano in maniera vistosa. Per una trattazione più esauriente, rimandiamo comunque ai testi di analisi numerica citati in bibliografia [1] — [8]. Un interessante articolo che illustra gli inconvenienti connessi con l'uso della rappresentazione a virgola mobile è dovuto a Forsythe [9].

Prima di passare agli esempi, è bene comunque ricordare per completezza che nel calcolo numerico bisogna tenere conto — oltreché degli errori di arrotondamento — anche di altri tipi di errori. In aggiunta agli errori dovuti alla eventuale incertezza nei dati che si utilizzano per i calcoli, vi sono gli errori *inerenti* (analoghi a quelli di arrotondamento) che si manifestano quando un calcolatore converte i dati dalla base 10 ad altra base, e poi, dopo aver eseguito i calcoli, riconverte i risultati alla base 10. Vi sono poi gli errori detti di *discretizzazione* (o di *troncamento*, da non confondere con il troncamento numerico citato sopra); gli errori di discretizzazione si verificano quando un procedimento matematico, il quale sia per sua natura infinito, viene rimpiazzato da un procedimento finito. Ad esempio, la nota formula di Taylor

$$f(x+h) = f(x) + h \frac{df}{dx}(x) + \frac{h^2}{2} \frac{d^2f}{dx^2}(x) + \dots,$$

se troncata dopo il 2° termine, dà l'espressione approssimata $\frac{f(x+h) - f(x)}{h}$ per la derivata $\frac{df}{dx}(x)$. Ancora dalla formula di Taylor possiamo trarre l'approssimazione

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}$$



per $\sin(x)$, con x espresso in radianti. Ad esempio, questa formula approssimata dà una stima per $\sin(\pi/3)$ che è esatta a meno di un errore di circa $4 \cdot 10^{-6}$. Viceversa per $x = \pi/3 + 2\pi$, la medesima formula dà -107.6 , mentre per $x = \pi/3 + 10\pi$ avremmo addirittura $-7.244 \cdot 10^6$. Espressive rappresentazioni grafiche per errori di questo tipo si trovano in [10, pag. 413-416] e [11, pag. 44].

Negli esempi che seguono i calcoli sono stati effettuati su un personal computer Olivetti M24. Non c'è bisogno di sottolineare che le difficoltà che intendiamo indicare non sono certamente eliminabili mediante l'utilizzazione di un elaboratore che lavori con un numero maggiore di cifre significative.

2. ESEMPI

Gli esempi che riportiamo servono ad illustrare le difficoltà dovute agli errori di «arrotondamento» (roundoff). Molti calcolatori consentono di eseguire i calcoli sia in singola precisione (cioè con un determinato numero k di cifre significative) che in doppia precisione ($2k$ cifre significative). Se si sospetta che una determinata procedura di calcolo sia inficiata da errori di arrotondamento è conveniente ripetere i calcoli sia in singola che in doppia precisione (o su due calcolatori aventi numeri diversi di cifre significative). Una forte discrepanza fra i risultati è sufficiente per concludere che gli errori di arrotondamento nel calcolo in singola precisione sono rilevanti (ma non per concludere che i calcoli in doppia precisione sono affidabili!).

Esempio 1

La tavola 1 illustra il calcolo di $c_n = (1 - 10^n) + 10^n$ per $n = 1, 2, 3, \dots$. Chiaramente il risultato esatto è $c_n = 1$. Lavorando in singola precisione (cioè nella configurazione usuale di calcolo) otteniamo $c_n = 0$ già per $n = 8$. Il fenomeno si manifesta anche in doppia precisione, ma solo per $n \geq 17$.

TAVOLA 1. - Calcolo di $c_n = (1 - 10^n) + 10^n$. Con $(c_n)_S$ denotiamo l'esito del calcolo in singola precisione, con $(c_n)_D$ quello del calcolo in doppia precisione. Si noti inoltre che la formula $d_n = 1 + (-10^n + 10^n)$ dà invece il risultato esatto $d_n = 1$ sia in singola che in doppia precisione. Il fenomeno dell'«overflow» si manifesta per valori di n sufficientemente grandi, ma non intendiamo parlarne in questo articolo.

n	1	2	3	4	5	6	7	8	9	10	12	14	16	18
$(c_n)_S$	1	1	1	1	1	1	1	0	0	0	0	0	0	0
$(c_n)_D$	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Esempio 2

Poniamo $f_1(x, \delta) = 2x + \delta$ e $f_2(x, \delta) = \frac{(x + \delta)^2 - x^2}{\delta}$. È chiaro che, per $x, \delta \in \mathbb{R}$ e $\delta \neq 0$, le due formule danno il valore del rapporto incrementale della funzione $x \rightarrow x^2$ nel punto x . Risulta che $f_1(x, \delta) = f_2(x, \delta) \rightarrow 2x$ per $\delta \rightarrow 0$. La tavola 2 riporta i valori calcolati di f_1 e f_2 per $x = 1$. Si noti come per valori di $|\delta|$ inferiori a 10^{-5} , il calcolo di f_2 fornisce valori che si allontanano da quello esatto: essi si avvicinano a zero anziché a $2x$ per $\delta \rightarrow 0$.

TAVOLA 2. - Confronto fra i valori (calcolati in singola precisione) di f_1 e di f_2 per $x = 1$.

Incremento δ	Valore di f_1	Valore di f_2
10^{-1}	2.1	2.100001
10^{-2}	2.01	2.01
10^{-3}	2.001	2.001047
10^{-4}	2.0001	2.000332
10^{-5}	2.00001	2.002716
10^{-6}	2.000001	1.907394
10^{-7}	2	2.384186
10^{-8}	2	0
10^{-9}	2	0

Esempio 3

La base dei logaritmi naturali è il numero di Nepero $e = 2.7182818284590452...$. Esso può essere definito come limite per $n \rightarrow +\infty$ della successione crescente

$$x_n = \left(1 + \frac{1}{n}\right)^n$$

La convergenza è molto lenta. Per accelerarla consideriamo la sottosuccessione definita da

$$y_n = \left(1 + \frac{1}{10^n}\right)^{10^n}$$

La tavola 3 riporta i valori di y_n per $n = 1, 2, 3, \dots$. Si noti la perdita di precisione dopo $n = 4$ in singola e dopo $n = 9$ in doppia precisione.

TAVOLA 3. - Calcolo di y_n in singola precisione, $(y_n)_S$, ed in doppia precisione, $(y_n)_D$.

n	$(y_n)_S$	$(y_n)_D$
1	2.593743	2.5937424601
2	2.704813	2.704813829421528
3	2.717042	2.716923932235903
4	2.718023	2.718145926824935
5	2.695423	2.718268237177208
6	2.481655	2.718280469322095
7	2.28484	2.718281692623132
8	1	2.718281813436853
9	1	2.718281825669121
10	1	2.718282053234787
12	1	2.71829714290784
14	1	2.716110034087023
16	1	3.035035206549264
18	1	1
20	1	1

Esempio 4

Per l'equazione di secondo grado $ax^2 + x - 1 = 0$ ($a > 0$) possiamo scrivere l'algoritmo di soluzione nelle due forme

$$x_{\pm} = \frac{-1 \pm \sqrt{1+4a}}{2a}$$

$$z_{\pm} = \frac{2}{1 \pm \sqrt{1+4a}}$$

Ovviamente $x_{\pm} = z_{\pm}$ per ogni $a > 0$. Risulta poi $x_{+} = z_{+} \rightarrow 1$ per $a \rightarrow 0^{+}$. Nella tavola 4 abbiamo riportato i valori calcolati di x_{+} e z_{+} per valori decrescenti di a . Si noti l'erroneo comportamento del valore calcolato di x_{+} . Per un'analisi dettagliata del problema vedi [12].

TAVOLA 4. - Confronto tra i valori di $z_+ = 2/(1 + \sqrt{1 + 4a})$ ed i valori di $x_+ = (-1 + \sqrt{1 + 4a})/2a$, calcolati sia in singola che in doppia precisione.

Coefficiente a	Valore di z_+ calcolato in s.p.	Valore di x_+ calcolato in s.p.
10^{-1}	0.9160797	0.9160799
10^{-2}	0.9901951	0.9901941
10^{-3}	0.9990021	0.9989739
10^{-4}	0.9999001	0.99957
10^{-5}	0.99999	1.001358
10^{-6}	0.9999991	1.013279
10^{-7}	1	0.5960465
10^{-8}	1	0
10^{-9}	1	0

Coefficiente a	Valore di z_+ calcolato in d.p.	Valore di x_+ calcolato in d.p.
10^{-1}	0.9160797830996161	0.9160797830996161
10^{-3}	0.9990019950139581	0.9990019950139595
10^{-5}	0.999990000199995	0.9999900002000129
10^{-7}	0.99999900000002	0.999999001086834
10^{-9}	0.999999999	0.9999999994736442
10^{-11}	0.99999999999	1.000000082740371
10^{-13}	0.9999999999999	1.00003338943111
10^{-15}	0.99999999999999	0.9992007221626409
10^{-17}	1	0
10^{-19}	1	0

Esempio 5

Consideriamo il problema di determinare x_n , definito dalla formula ricorrente

$$x_{n+1} = \left(9 + \frac{1}{9}\right)x_n - x_{n-1}, \quad n = 1, 2, 3, \dots$$

supponendo di avere assegnato x_0 e x_1 . La soluzione esatta è

$$x_n = \alpha 9^n + \beta 9^{-n}$$

dove $\alpha = (9x_1 - x_0)/80$ e $\beta = (-9x_1 + 81x_0)/80$. Allora per $x_0 = 1$ e $x_1 = 1/9$, abbiamo $\alpha = 0$, $\beta = 1$; in questo caso quindi la soluzione esatta

è

$$x_n = 9^{-n}.$$

Nella tavola 5 abbiamo riportato i valori calcolati di x_n . Si noti come il rapido propagarsi dell'errore generi dei risultati che si allontanano indefinitamente da quelli esatti.

TAVOLA 5. - $(x_n)_S$ e $(x_n)_D$ sono i valori calcolati in singola e doppia precisione, rispettivamente, utilizzando la formula di ricorrenza $x_{n+1} = (9 + 1/9)x_n - x_{n-1}$ a partire da $x_0 = 1$ e $x_1 = 1/9$. La soluzione esatta è data da $x_n = 9^{-n}$.

n	9^{-n}	$(x_n)_S$	$(x_n)_D$
1	0.1111111111	0.1111111	0.111111111111111
2	1.23456790123E-02	1.234567E-02	1.234567901234568D-02
3	1.37174211248E-03	1.371667E-03	1.371742112482846D-03
4	1.52415790275E-04	1.517367E-04	1.524157902758111D-04
5	1.69350878084E-05	1.082267E-05	1.693508780787681D-05
6	1.88167642315E-06	-5.313015E-05	1.881676418177666D-06
7	2.09075158128E-07	-4.948974E-04	2.090751132974755D-07
8	2.32305731254E-08	-4.455935E-03	2.323016964377758D-08
9	2.58117479171E-09	-4.010362E-02	2.577543456942437D-09
10	2.86797199079E-10	-0.3609326	2.541151861424004D-10
11	3.18663554532E-11	-3.248393	-2.622717609783439D-10
12	3.54070616147E-12	-29.23554	-2.643702341722867D-09
13	3.93411795719E-13	-263.1198	-2.382479401916334D-08
14	4.37124217465E-14	-2368.078	-2.144266431662098D-07
15	4.85693574962E-15	-21312.7	-1.929840177050748D-06
16	5.39659527735E-16	-191814.3	-1.736856163662949D-05
17	5.99621697484E-17	-1726329	-1.563170547344624D-04
18	6.66246330538E-18	-1.553696E+07	-1.406853492610695D-03
19	7.40273700597E-19	-1.398326E+08	-1.266168143349631D-02
20	8.22526333997E-20	-1.258493E+09	-0.1139551329014668
21	9.13918148886E-21	-1.132644E+10	-1.025596196113201
22	1.01546460987E-21	-1.019379E+11	-9.230365765018812
23	1.12829401097E-22	-9.174413E+11	-83.07329188516931
24	1.25366001219E-23	-8.256972E+12	-747.6596269665238
25	1.39295556910E-24	-7.431273E+13	-6728.936642698714

Esempio 6

Consideriamo la somma parziale $S_n = \sum_{k=0}^n 10^{-k}$; risulta che $S_0 = 1$, $S_1 = 1.1$, $S_2 = 1.11$, ecc. Poniamo ora

$$Q_n = \sum_{k=0}^n (10^k + 10^{-k}) - \sum_{k=0}^n 10^k, \quad R_n = \sum_{k=0}^n ((10^k + 10^{-k}) - 10^k).$$

È chiaro che per ogni $n \in \{0, 1, 2, \dots\}$ si ha $S_n = Q_n = R_n$. Nella tavola 6 abbiamo riportato i valori di Q_n e R_n , calcolati sia in singola che in doppia precisione.

TAVOLA 6. - Valori calcolati per S_n , Q_n , R_n dell'esempio 6.

n	S_n calcolato in s.p.	Q_n calcolato in s.p.	R_n calcolato in s.p.
1	1.1	1.1	1.1
2	1.11	1.110001	1.110003
3	1.111	1.110962	1.110979
4	1.1111	1.111328	1.110979
5	1.11111	1.109375	1.110979
6	1.111111	1.125	1.110979
7	1.111111	1	1.110979
8	1.111111	0	1.110979
9	1.111111	0	1.110979

n	Q_n calcolato in d.p.	R_n calcolato in d.p.
1	1.1	1.1
2	1.1099999999999999	1.11
3	1.1109999999999999	1.1110000000000005
4	1.1111000000000079	1.1110999999999998
5	1.1111099999999826	1.1111099999999727
6	1.111110000005621	1.11111099999279
7	1.11111109983176	1.111111099877136
8	1.111111108213663	1.111111109190361
9	1.111111104488373	1.111111109190361
10	1.111111164093018	1.111111109190361
12	1.111114501953125	1.111111109190361
14	1.111328125	1.111111109190361
15	1.109375	1.111111109190361
16	1	1.111111109190361
17	0	1.111111109190361
18	0	1.111111109190361

Esempio 7

Com'è noto, sviluppando in serie di Taylor con centro nell'origine la funzione seno otteniamo

$$\operatorname{sen} x = \lim_{n \rightarrow \infty} S_n(x), \quad x \in C,$$

dove

$$S_n(x) = \sum_{k=0}^n (-1)^k x^{2k+1} / (2k+1)!$$

Nella tavola 7 abbiamo riportato i valori calcolati della somma parziale S_n nel caso $x = (8 + 1/6)\pi$, per valori crescenti di n . Invitiamo il lettore a condurre esperimenti di «convergenza numerica» di questo tipo — ad esempio, ponendo $x = (2r + 1/6)\pi$ ed assegnando ad r valori positivi interi prefissati.

TAVOLA 7. — Valori calcolati per $S_n = \sum_{k=0}^n (-1)^k x^{2k+1} / (2k+1)!$ con $x = (8 + 1/6)\pi$. Per aggirare la difficoltà di «overflow» si è proceduto ponendo: $S_n = S_{n-1} + Z_n$, $Z_n = -\frac{1}{2n(2n+1)} \cdot x^2 \cdot Z_{n-1}$, $n \geq 1$, a partire da $S_0 = x$ e $Z_0 = x$. I valori numerici delle somme parziali $(S_n)_1$ e $(S_n)_2$ sono stati ottenuti mediante calcoli in singola precisione utilizzando, rispettivamente, le seguenti rappresentazioni di π per difetto e per eccesso: $\pi_1 = 3.141592$ e $\pi_2 = 3.141593$. Un calcolo esatto del limite nel punto $x = (8 + 1/6)\pi$ dovrebbe dare $\lim_{n \rightarrow \infty} S_n = 0.5$. Analoghi risultati che qui omettiamo di riportare si ottengono anche lavorando in doppia precisione.

Indice della Somma parziale n	Somma parziale calcolata con π approssimato	
	per difetto $(S_n)_1$	per eccesso $(S_n)_2$
1	-2789.047	-2789.049
4	1.191156E+07	1.191159E+07
7	-7.829177E+08	-7.829213E+08
10	4.535224E+09	4.535252E+09
13	-4.759204E+09	-4.759243E+09
16	1.342456E+09	1.342469E+09
19	-1.302176E+08	-1.302192E+08
20	4.865791E+07	4.865837E+07
21	-1.65383E+07	-1.653862E+07
22	5136064	5136018
23	-1462964	-1463097
24	383887.5	383780
25	-92851.75	-92966.12
26	21013.44	20900.77
27	-4222.752	-4335.815
28	981.4024	868.4229
29	-19.65601	-132.6521
30	160.3834	47.39038
32	134.8436	21.85018
36	134.2177	21.22417
40	134.2175	21.22399
50	134.2175	21.22399

Esempio 8

Il problema di ricorrenza non-lineare

$$x_n = r x_{n-1} (1 - x_{n-1}), \quad n \in \{1, 2, 3, \dots\},$$

con $x_0 \in [0, 1]$ e $r \in [0, 4]$ assegnati, è stato proposto in letteratura quale modello per lo studio dell'evoluzione nel tempo di certe popolazioni di insetti. Si noti che se $r \in [0, 4]$, allora $x_0 \in [0, 1]$ implica $x_n \in [0, 1]$ per $n = 1, 2, 3, \dots$. L'equazione è nota come equazione «logistica» a tempo discreto. Il problema è apparentemente molto semplice, eppure per valori di r sufficientemente elevati l'andamento delle soluzioni diviene estremamente complicato. Problemi di questo tipo sono stati estesamente studiati nell'ultimo decennio nell'ambito della «Teoria dei Sistemi Dinamici». È noto (vedi ad esempio [10], p. 533-539) che quando $4 > r > r^* = 3.8284\dots$ le soluzioni manifestano andamento «caotico». La tavola 8 e la figura 1 illustrano un esempio di calcolo della soluzione x_n

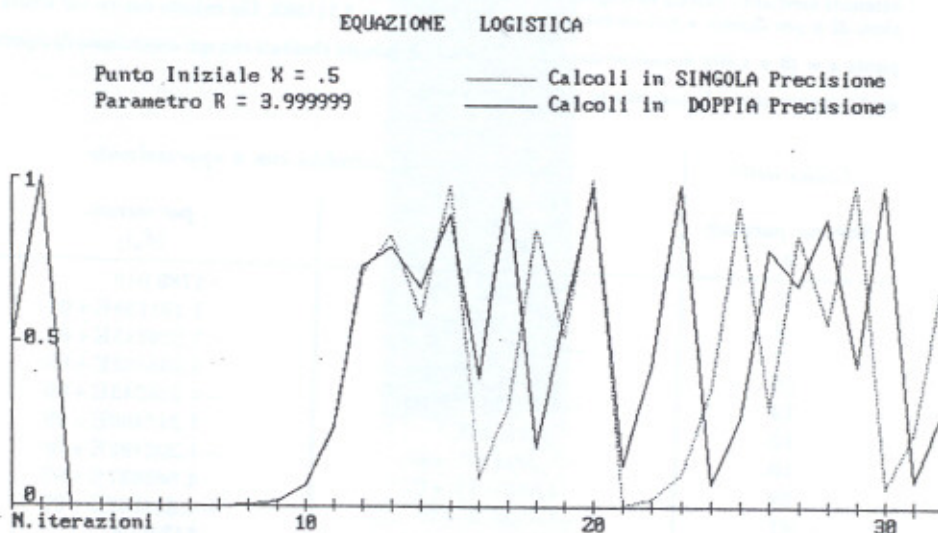


Figura 1.

dell'equazione logistica a partire da $x_0 = 0.5$ quando $r = 3.999999$ (cioè in regime «caotico»). Si noti dalla tavola 8 il progressivo distacco, al crescere di n , fra i valori di x_n calcolati in singola precisione e quelli calcolati in doppia precisione.

TAVOLA 8. - Calcolo in singola e in doppia precisione della soluzione dell'equazione logistica $x_n = r x_{n-1} (1 - x_{n-1})$, con $x_0 = 0.5$ e $r = 3.999999$.

n	x_n calcolato in s.p.	x_n calcolato in d.p.
1	0.9999998	0.99999975
2	9.536739E-07	9.999994999733071D-07
3	3.814691E-06	3.999992999898728D-06
4	1.52587E-05	1.599990399984192D-05
5	6.103386E-05	6.399857601200765D-05
6	2.441205E-04	2.559778567826241D-04
7	9.762433E-04	1.023649072565512D-03
8	3.90116E-03	4.090403837965777D-03
9	1.554376E-02	1.629468566396010D-02
10	6.120859E-02	6.411665950312243D-02
11	0.2298483	0.2400227939034189
12	0.7080721	0.7296472268290109
13	0.8268238	0.7890484075768282
14	0.5727448	0.6658039058581813
15	0.9788326	0.8900360366996206
16	8.287734E-02	0.3914874624307187
17	0.3040347	0.952899878536072
18	0.8463901	0.1795267552063448
19	0.5200554	0.5891874501888044
20	0.9983909	0.9681821528696794
22	2.553878E-02	0.432152813174636
24	0.3585469	7.229663069491792D-02
26	0.2945216	0.7852217695218489
28	0.5614529	0.8780675209787751
30	5.951131E-02	0.979413067553601
32	0.6950281	0.296590362738905

3. COMMENTI E CONCLUSIONI

Con i semplici esempi riportati sopra abbiamo voluto dare al lettore qualche indicazione sull'importanza di un uso «intelligente» dell'elaboratore digitale nel calcolo numerico. Il passaggio alla «doppia precisione» o ad elaboratori sempre più potenti e dispendiosi non può rimpiazzare un attento studio teorico degli algoritmi utilizzati.

Il lettore che voglia approfondire le questioni trattate in questo articolo potrà fare riferimento alla letteratura, [1] - [8]; qui, per concludere presentiamo qualche ulteriore commento ed osservazione.

La chiave per l'interpretazione dell'esempio 3 è l'osservazione che il calcolatore approssima, ad esempio, il numero reale $1 + a$ con 1 quando $|a| < 10^{-n}$ dove n è il numero di cifre significative utilizzate. Gli esempi 2

e 4 illustrano invece l'effetto di amplificazione dell'errore relativo di arrotondamento che si ha quando si effettua la sottrazione fra due numeri reali pressochè uguali. Siano ad esempio $a = 0.1237$ e $b = 0.1234$; allora $c = a - b = 0.0003$. Lavorando ad esempio con 3 cifre significative, otteniamo per arrotondamento le approssimazioni $\tilde{a} = 0.124$ e $\tilde{b} = 0.123$. Se ora effettuiamo la sottrazione di $\tilde{b}$ da $\tilde{a}$ otteniamo $\tilde{c} = \tilde{a} - \tilde{b} = 0.001$. In questo caso gli errori relativi sono: $|\tilde{a} - a|/|a| \simeq 0,24\%$; $|\tilde{b} - b|/|b| \simeq 0,32\%$, mentre $|\tilde{c} - c|/|c| \simeq 233\%$. Abbiamo così avuto un effetto di amplificazione dell'ordine di 10^5 .

Una conseguenza immediata del manifestarsi di errori di questo tipo è che — nell'implementazione numerica dell'operazione di somma in virgola mobile — viene a cadere la proprietà associativa. Questo fatto è illustrato dagli esempi 1 e 6. Con riferimento all'esempio 1, si noti che l'esito del calcolo di c_n con l'espressione $1 - 10^n + 10^n$ risulta ben presto diverso da quello ottenuto con l'espressione riordinata $10^n - 10^n + 1$. Allora — anche se la proprietà commutativa per la somma di due addendi continua a valere — non risulterà in generale equivalente dal punto di vista numerico calcolare (per $n \geq 3$) $x_1 + x_2 + \dots + x_n$ in quest'ordine, oppure in un ordine diverso.

L'esempio 7 esibisce il fenomeno della «cancellazione catastrofica». Se si esaminano i valori $(S_n)_1$ e $(S_n)_2$ calcolati in singola precisione, vediamo come un piccolo errore di arrotondamento (dell'ordine di $\pm 10^{-6}$) nel valore assegnato a π , in congiunzione con gli errori di arrotondamento che si accumulano passo per passo porti ad errori drammatici nella stima di $\sin((8 + 1/6)\pi)$. Un confronto con i calcoli in doppia precisione mostra, ad esempio, come per $10 \leq n \leq 20$ l'errore relativo nelle stime $(S_n)_1$ e $(S_n)_2$ di S_n si mantenga molto piccolo (dell'ordine di $10^{-30}\%$). Tuttavia, poiché per quei medesimi valori di n risulta che $|S_n|$ è dell'ordine di 10^9 i minuscoli errori relativi corrispondono ad errori effettivi dell'ordine di 10^4 . Essi sono quindi molto più grandi del valore finale teorico 0.5. Nei passi successivi errori così alti vengono riassorbiti solo in parte. In generale fenomeni di instabilità numerica di questo tipo si manifestano nelle serie a segno alterno quando le somme parziali raggiungono valori molto elevati (in modulo) rispetto al valore della somma (finale) della serie.

Ritorniamo ora ad esaminare gli esempi 1, 2, 4 e 6. Abbiamo visto che in ciascuno di questi casi è possibile sostituire un algoritmo (procedura di calcolo) che conduce — se implementato su calcolatore — a risultati non accettabili con un algoritmo la cui implementazione conduce a risultati adeguati. I due algoritmi sono equivalenti se il calcolo viene eseguito in modo esatto (cioè, senza l'introduzione di errori di arrotondamento). Nella letteratura si parla in riferimento ad un medesimo proble-

ma di algoritmi *numericamente instabili* e di algoritmi *numericamente stabili*. D'altra parte vi sono problemi che sono intrinsecamente *mal condizionati (mal posti)*. Sono problemi caratterizzati dal fatto che un «piccolo» errore nei dati si ripercuote in un «grande» errore nei risultati. Tipicamente, nella risoluzione numerica di questi problemi l'effetto del «rumore numerico» (cioè degli arrotondamenti) è analogo a quello di una perturbazione dei dati. La conseguenza è il manifestarsi di errori significativi. Questo è quanto avviene nei calcoli relativi agli esempi 5 ed 8.

Esaminiamo il caso dell'esempio 5. Abbiamo visto che la soluzione esatta di $x_{n+1} = (9 + 1/9)x_n - x_{n-1}$ è data da $\alpha 9^n + \beta 9^{-n}$ dove $\alpha = (9x_1 - x_0)/80$ e $\beta = (-9x_1 + 81x_0)/80$. Nel nostro caso $x_0 = 1$ e $x_1 = 1/9$, cosicché $\alpha = 0$, $\beta = 1$ e $x_n = 9^{-n}$ (esattamente). D'altra parte il calcolo numerico di x_1 richiede l'effettuazione della divisione $1/9$. Denotiamo con $\tilde{x}_1$ l'esito di tale implementazione numerica e con δ_1 l'errore di arrotondamento introdotto: quindi $\tilde{x}_1 = x_1 + \delta_1 = 1/9 + \delta_1$. Supponiamo ora di implementare in modo esatto i calcoli successivi che danno x_2, x_3 , ecc., a partire da $x_0 = 1$ e da $\tilde{x}_1$. Otterremo $\tilde{x}_n = \tilde{\alpha} 9^n + \tilde{\beta} 9^{-n}$ con $\tilde{\alpha} = (9\tilde{x}_1 - x_0)/80 = 9\delta_1/80$ e $\tilde{\beta} = 1 - 9\delta_1/80$. Allora per n sufficientemente grande $\tilde{x}_n \approx (81/80)\delta_1 9^{n-1}$, da confrontare con il risultato esatto $x_n = 9^{-n}$. L'errore introdotto nel calcolo di x_1 si è «propagato» amplificandosi da un passo all'altro («instabilità numerica»). Si noti che in effetti ad ogni passo ($n = 1, 2, \dots$) il calcolo di $(9 + 1/9)x_n - x_{n-1}$ introduce ulteriori errori di arrotondamento nella sequenza dei calcoli. Schemi numerici adatti a problemi ricorrenti del tipo dell'esempio 5 sono stati presentati ad esempio in [13].

L'esempio 8 riguarda un problema ricorsivo non-lineare in regime caotico. Si noti dalla tavola 8 che per $n \geq 15$ la soluzione numerica in singola precisione non concorda con quella in doppia precisione neppure per quel che riguarda la prima cifra decimale. Anche qui si tratta di un problema intrinsecamente «mal posto». Si osservi che se il parametro r e la condizione iniziale x_0 sono razionali, è possibile — anche se estremamente dispendioso — procedere ad un calcolo esatto di x_n mediante calcolatore digitale, per valori crescenti di n , utilizzando una rappresentazione del tipo $x_n = p_n/q_n$, dove p_n e q_n sono numeri interi che vengono calcolati ricorsivamente e senza errore di arrotondamento (vedi ad esempio [14]).

Ritorniamo ora al nostro discorso originale sull'implementazione su calcolatore di calcoli aritmetici mediante rappresentazione a virgola mobile («floating point»). In Analisi Numerica sono stati sviluppati vari metodi di esame della natura dell'algoritmo utilizzato e di stima dell'errore. Citiamo — senza entrare in dettaglio — le tecniche di analisi dell'errore (*a priori*, e *a posteriori*), di aritmetica degli intervalli, e di

stima statistica dell'errore. Vi è poi l'accorgimento — da noi ripetutamente utilizzato — di confrontare l'esito del calcolo in singola precisione con quello del calcolo in doppia precisione. Non va dimenticato, infine, l'accorgimento di confrontare l'esito dei calcoli numerici con risultati ottenuti analiticamente (o per approssimazione asintotica) su problemi «prototipo». Rimandiamo comunque alla letteratura per tutte queste questioni.

Chiudiamo sottolineando che se da una parte è vero che la facile accessibilità ai calcolatori e la vasta diffusione dei «personal» e degli «home computer» hanno incoraggiato calcoli inutili, mal congegnati o semplicemente errati, dall'altra ci sembra evidente che i calcolatori hanno favorito nuovi filoni scientifici (ad esempio l'analisi numerica) e consentito ricerche altrimenti impossibili. Ebbene, questo nostro articolo nel suo piccolo ha lo scopo di scoraggiare un uso passivo e pigro del calcolatore.

4. RINGRAZIAMENTI

Gli autori sono grati a F. Fontanella, L. Gori, L. Lo Cascio, L. Pasquini e al relatore di Archimede per aver letto e commentato una versione preliminare di questo articolo.

5. BIBLIOGRAFIA COMMENTATA

Nei settori toccati da questo articolo, la letteratura è molto vasta, e rapidamente crescente. Come indicato nel testo, quali introduzioni al soggetto suggeriamo di consultare ad esempio:

- [1] FONTANELLA F., PASQUALI A., *Calcolo Numerico: Metodi e Algoritmi*, vol. 1. e 2., Bologna, Pitagora, 1978 e 1982.
- [2] HILDEBRAND F. B., *Analisi Numerica*, Milano, Ambrosiana, 1967.
- [3] STOER J., *Introduzione all'Analisi Numerica 1*, Bologna, Zanichelli, 1974.
- [4] STOER J., BULIRSCH R., *Introduzione all'Analisi Numerica 2*, Bologna, Zanichelli, 1975.
- [5] CONTE S. D., DE BOOR C., *Elementary Numerical Analysis*, 3° ed., New York, Mc Graw-Hill, 1980.
- [6] DAHLQUIST G., BJÖRCK Å., *Numerical Methods*, Englewood Cliffs, Prentice Hall, 1974.
- [7] FORSYTHE G. E., MALCOLM M. A., MOLER C. B., *Computer Methods for Mathematical Computations*, Englewood Cliffs, Prentice-Hall, 1977.
- [8] KING J. T., *Introduction to Numerical Computations*, New York, Mc Graw Hill, 1984.

e anche l'articolo introduttivo:

- [9] FORSYTHE G. E., *Pitfalls in computation, or why a Math Book isn't enough*, «American Mathematical Monthly», 77, 931-956 (1970).

Nel testo si è fatto inoltre riferimento a quanto segue:

- [10] GEYMONAT G., *Lezioni di Matematica*, vol. I, Torino, Levrotto & Bella, 1981.
- [11] POSTON T., STEWART I., *Catastrophe Theory and its Applications*, London, Pitman, 1978.
- [12] FORSYTHE G. E., *Come risolvere un'equazione quadratica su un calcolatore*, da «Le Scienze Matematiche», pag. 159-176, Bologna, Ed. UMI, Zanichelli, 1973.
- [13] WIMP J., *Computation with Recurrence Relations*, London, Pitman, 1984.

Software caratterizzato da aritmetica razionale a infinita precisione è stato sviluppato, ad esempio, da Rich e Stoutmeyer nel 1971. È incluso nel sistema muMATH. Vedi:

- [14] SQUIRE W., *muMATH System Effective Tool for Algebra*, «SIAM News», 17, Novembre 1984.

Concludiamo dando qualche breve indicazione riguardante l'aritmetica a virgola mobile («floating point»). Facciamo riferimento in particolare all'autorevole ed esauriente:

- [15] KNUTH D. E., *The art of Computer Programming*, vol. 2, Seminumerical Algorithms, II ed., Reading, Addison Wesley, 1981.

oltreché a:

- [16] STERBENZ P., *Floating Point Computation*, Englewood Cliffs, Prentice Hall, 1974.

Fra le due tecniche di arrotondamento, quello di troncamento numerico («chopping») e quello di arrotondamento propriamente detto («rounding»), vi è un consenso generale fra gli analisti numerici a favore del «rounding». D'altra parte alcune case costruttrici tendono a privilegiare il troncamento numerico a livello di hardware ed anche di software. Rimandiamo a Knuth [15, § 4.2.2.] per una discussione. Supponiamo ora di utilizzare un'aritmetica a virgola mobile tipo «rounding». Per fissare le idee supponiamo inoltre da qui in avanti che x sia un numero compreso fra 0.5 ed 1. Una rappresentazione a virgola mobile di x in base decimale con n «cifre significative esatte» comporta un errore massimo pari a $\frac{1}{2} \cdot 10^{-n}$. D'altra parte la maggioranza dei calcolatori lavora con base $b \neq 10$ (in genere viene utilizzata la base binaria, $b = 2$, oppure la base ottale, $b = 8$, od anche la base esadecimale, $b = 16$). La rappresentazione del numero x in base b con m «cifre significative esatte» comporta un errore massimo di $\frac{1}{2} \cdot b^{-m}$. Se ora riportassimo il numero x alla base decimale (idealmente senza ulteriori errori) avremmo una rappresentazione di x con un numero di «cifre significative esatte» pari a $m \cdot \log_{10} b$. Si noti che, paradossalmente, tale numero è in generale non intero. Le questioni relative all'aritmetica a virgola mobile sono oggetto di studio e ricerca. Si veda, ad esempio, [15] e [16], oppure:

- [17] KULISH U. W., MIRANKER W. L., *The arithmetic of the digital computer: a new approach*, «SIAM Review», 28, 1-40 (1986).

GIOVANNI FROSALI

Istituto Matematica Applicata «G. Sansone»
Facoltà di Ingegneria, Università- Firenze

STEFANO L. PAVERI-FONTANA

Dip.to Metodi e Modelli Matematici
Facoltà di Ingegneria, Università «La Sapienza»-Roma