Riv. Mat. Univ. Parma (4) 2 (1976), 221-228

GIOVANNI MONEGATO (*)

Nota su due metodi recenti per il calcolo numerico degli integrali iterati della funzione degli errori. (**)

1. – Il problema del calcolo numerico degli integrali iterati della funzione degli errori è stato ripetutamente trattato da W. Gautschi $[3]_2$, $[3]_3$, $[3]_4$, $[3]_5$. Facendo uso della relazione di ricorrenza a tre termini che tali integrali soddisfano, Gautschi risolve il problema dell'instabilità numerica cui tale relazione dà luogo, usando, se necessario, la relazione stessa a ritroso.

A. Ascari [1] propone un metodo alternativo che conduce ad un sistema triangolare inferiore. Tuttavia, la soluzione numerica di tale sistema mostra, almeno nei casi esaminati, una precisione inferiore a quella ottenuta applicando direttamente la nota relazione ricorsiva.

In questa Nota, partendo dai risultati di Ascari, deriviamo un'espressione molto semplice per la soluzione del suddetto sistema. Risulta poi immediato osservare che il metodo, scritto in quest'ultima forma, coincide essenzialmente con l'applicazione diretta della già menzionata relazione ricorsiva a tre termini.

Successivamente, esaminiamo un secondo metodo, recentemente proposto da V. Dose e C. Semini [2]. Anche in questo caso, dopo una breve analisi del metodo, empiricamente osserviamo che l'accuratezza dei risultati non è mai superiore a quella ottenuta con la più volte accennata relazione.

2. – Seguendo le notazioni usate in [1], poniamo

$$\tag{2.1} \begin{cases} F_{-1}(x) = \dfrac{2}{\sqrt{\pi}} \exp[-x^2] \\[2ex] F_n(x) \;= \dfrac{2}{\sqrt{\pi}} \displaystyle\int_x^{\infty} \dfrac{(t-x)^n}{n!} \exp[-t^2]\, dt \qquad (n = 0, 1, 2, \dots). \end{cases}$$

(*) Indirizzo: Istituto di Calcoli Numerici, Università, Via Carlo Alberto 10, Torino, Italia.

(**) Lavoro eseguito presso il Dipartimento di Computer Sciences, Purdue University, West Lafayette, Indiana - U.S.A., quando l'Autore usufruiva di una borsa di studio del C.N.R. per l'estero. – Ricevuto: 17-II-1976.

È allora noto che $F_n(x)$ soddisfa la seguente relazione ricorsiva

$$F_{n+1}(x) + \frac{x}{n+1} F_n(x) - \frac{1}{2(n+1)} F_{n-1}(x) = 0 \qquad (n = 0, 1, 2, \dots). \tag{2.2}$$

Partendo dalla considerazione che gli integrali iterati $F_n \equiv F_n(x)$ e i polinomi di Hermite $H_n \equiv H_n(x)$ sono due casi particolari della funzione del cilindro parabolico, $D_\nu(z)$ con ν intero, Ascari deduce la seguente relazione bilineare tra F_n e H_n:

$$\sum_{k=0}^{n} \frac{1}{2^k k!} H_k F_{n-k} = \frac{1}{2^n n!} H_{n-1} F_{-1} \qquad (n = 1, 2, \dots). \tag{2.3}$$

È a questo punto utile osservare che la (2.3) vale per ogni x complesso.

Per il calcolo delle F_n $(n = 1, 2, \dots, N)$ è allora sufficiente risolvere il seguente sistema triangolare inferiore

$$\sum_{k=0}^{n-1} \frac{1}{2^k k!} H_k F_{n-k} = \frac{1}{2^n n!} (H_{n-1} F_{-1} - H_n F_0) \qquad (n = 1, 2, \dots, N), \tag{2.4}$$

dove, fissato x, i valori di $H_0, H_1, \dots, H_n, F_{-1}, F_0$ sono facilmente calcolabili.

Introducendo le nuove quantità $a_k = \big(1/(2^k k!)\big) H_k$, il sistema (2.4) assume la forma

$$\sum_{k=0}^{n-1} a_k F_{n-k} = \frac{1}{2n} a_{n-1} F_{-1} - a_n F_0 \qquad (n = 1, 2, \dots, N), \tag{2.5}$$

dove le $a_k \equiv a_k(x)$ possono essere calcolate mediante la relazione ricorsiva

$$a_{n+1} - \frac{x}{n+1} a_n + \frac{1}{2(n+1)} a_{n-1} = 0 \qquad (n = 0, 1, \dots, N-1),$$

con $a_{-1} = 0$ e $a_0 = 1$.

Allo scopo di scrivere la (2.5) in forma più compatta, poniamo

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_1 & 1 & 0 & \dots & 0 \\ a_2 & a_1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{N-1} & a_{N-2} & a_{N-3} & \dots & 1 \end{bmatrix}, \qquad c = [c_1, c_2, c_3, \dots, c_N]^T,$$

con

$$c_n = \frac{a_{n-1}}{2n} F_{-1} - a_n F_0 \,, \qquad (n = 1, 2, \ldots, N) . \tag{2.6}$$

Allora $f = [F_1, F_2, F_3, \ldots, F_N]^T$ è la soluzione del seguente sistema lineare, triangolare inferiore,

$$Af = c . \tag{2.7}$$

Poichè A è nonsingolare, abbiamo

$$f = A^{-1} c \,, \qquad A^{-1} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ b_1 & 1 & 0 & \ldots & 0 \\ b_2 & b_1 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{N-1} & b_{N-2} & b_{N-3} & \ldots & 1 \end{bmatrix} . \tag{2.8}$$

Le quantità b_n $(n = 1, 2, \ldots, N-1)$ sono date dalla soluzione del sistema

$$b_n + a_1 b_{n-1} + \ldots + a_{n-1} b_1 + a_n = 0 \qquad (n = 1, 2, \ldots, N-1) \,, \tag{2.9}$$

anch'esso triangolare inferiore.

3. – Vogliamo ora dimostrare che la funzione $b_n = b_n(x)$, soluzione del sistema (2.9), soddisfa in realtà la relazione (2.2). A tale scopo, riscriviamo la (2.5),

$$\sum_{k=0}^{n} a_k(x) F_{n-k}(x) = \frac{a_{n-1}(x)}{2n} F_{-1}(x) \qquad (n = 1, 2, \ldots, N) . \tag{3.1}$$

Sostituendo x con $-x$, e osservando che

$$a_k(-x) = (-1)^k a_k(x) \qquad \text{e} \qquad F_{-1}(-x) = F_{-1}(x) \,,$$

la (3.1) diventa

$$\sum_{k=0}^{n} (-1)^k a_k(x) F_{n-k}(-x) = (-1)^{n-1} \frac{a_{n-1}(x)}{2n} F_{-1}(x) .$$

Moltiplicando quest'ultima per $(-1)^n$ e sommando il risultato alla (3.1), otteniamo

$$\sum_{k=0}^{n} a_k(x) [F_{n-k}(x) + (-1)^{n-k} F_{n-k}(-x)] = 0 . \tag{3.2}$$

È però noto che ([**3**]$_1$, eq. (7.2.11))

$$(-1)^n F_n(x) + F_n(-x) = \frac{i^{-n}}{2^{n-1} n!} H_n(ix) ,$$

o, scritta in altra forma,

$$\frac{1}{2}[F_n(x) + (-1)^n F_n(-x)] = \frac{i^n}{2^n n!} H_n(ix) \equiv G_n(x) .$$

La (3.2) allora diventa

$$\sum_{k=0}^{n} a_k(x) G_{n-k}(x) = 0 \qquad (n = 1, 2, \ldots, N) . \tag{3.3}$$

Osservando che $G_0(x) = 1$, il confronto della (3.3) con la (2.9) ci dà $b_k(x) \equiv \equiv G_k(x)$, ed inoltre, poichè $G_k(x)$ soddisfa la relazione (2.2),

$$b_{n+1}(x) + \frac{x}{n+1} b_n(x) - \frac{1}{2(n+1)} b_{n-1}(x) = 0 \qquad (n = 0, 1, 2, \ldots), \tag{3.4}$$

con $b_{-1}(x) \equiv 0$ e $b_0(x) \equiv 1$.

$F_n(x)$ e $b_n(x)$ sono quindi due soluzioni linearmente indipendenti della (2.2). La (3.4) si rileva quindi utile per il calcolo delle $b_n(x)$, risultando numericamente stabile.

4. – Consideriamo la (2.8), e in particolare l'espressione per l'n-esima componente del vettore f

$$F_n = c_1 b_{n-1} + c_2 b_{n-2} + \ldots + c_{n-1} b_1 + c_n . \tag{4.1}$$

Sostituendo la (2.6) in quest'ultima, e richiamando la (2.9), otteniamo

$$F_n = d_{n-1} F_{-1} + b_n F_0 , \tag{4.2}$$

dove $d_{n-1} = d_{n-1}(x)$ è un polinomio di grado $n - 1$. È ora banale provare che anche d_{n-1} soddisfa la (2.2), e in particolare

$$d_n + \frac{x}{n+1} d_{n-1} - \frac{1}{2(n+1)} d_{n-2} = 0 \qquad (n = 1, 2, \ldots), \tag{4.3}$$

con $d_{-1} = 1$ e $d_0 = 0$.

Abbiamo risolto numericamente il sistema (2.5) per $x = .1(.1)2.$ e $N = 1, 2, \ldots, 50$. In tutti i casi l'accuratezza della soluzione $F_n(x)$ $(n = 1, 2, \ldots, N)$ è risultata inferiore a quella ottenuta con la (2.2); la discrepanza aumenta col crescere di n: mentre per i primi valori di n i risultati si equivalgono, per $n = 50$ abbiamo notato molto spesso una differenza di 6-7 cifre significative.

Il confronto numerico tra la (4.2) e la (2.2), per gli stessi valori di x e N della prova precedente, ha invece mostrato che dal punto di vista della precisione queste ultime due sono essenzialmente equivalenti.

5. – Dal lavoro di Gautschi [**3**]$_2$ appare evidente che l'applicazione della relazione ricorsiva (2.2) a ritroso converge molto lentamente quando l'argomento x è vicino a zero. V. Dose e C. Semini [**2**] propongono allora una variazione di questo metodo che dovrebbe ovviare al suddetto inconveniente.

Richiamando dapprima la descrizione fatta in [**2**], scriviamo

$$F_n = S_n - Y_n\,, \tag{5.1}$$

dove

$$S_n = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{(t-x)^n}{n!} \exp[-t^2]\,dt \tag{5.2}$$

e

$$Y_n = \frac{2}{\sqrt{\pi}} \int_0^x \frac{(t-x)^n}{n!} \exp[-t^2]\,dt\,. \tag{5.3}$$

Definiamo poi

$$R(Z_n) = Z_n + \frac{x}{n} Z_{n-1} - \frac{1}{2n} Z_{n-2} \qquad (n = 1, 2, \ldots),$$

e integriamo la (5.2) per parti; si verifica facilmente che S_n, e quindi Y_n, entrambe soddisfano la relazione

$$R(Z_n) = \frac{1}{\sqrt{\pi}} \frac{(-x)^{n-1}}{n!} \qquad (n = 1, 2, \ldots). \tag{5.4}$$

Formando l'espressione $2(n-1)[nR(Z_n) + xR(Z_{n-1})]$, dalla (5.4) deduciamo che S_n e Y_n soddisfano anche la seguente relazione ricorsiva omogenea:

$$2n(n-1)Z_n + 4(n-1)xZ_{n-1} + (2x^2 - n + 1)Z_{n-2} - xZ_{n-3} = 0 \quad (n = 2, 3, \ldots). \tag{5.5}$$

Ricordando la (2.2), è poi immediato osservare che $b_n(x)$, definita dalla (3.4), è una terza soluzione della (5.5), linearmente indipendente da Y_n e S_n.

Per quanto riguarda il comportamento di Y_n quando $n \to \infty$, richiamando il metodo di Laplace (vedi per es. [4], p. 81) abbiamo

$$(5.6) \qquad Y_n \sim \frac{2x}{\sqrt{\pi}} \frac{(-x)^n}{nn!}, \qquad n \to \infty .$$

È inoltre noto [3]$_2$ che

$$(5.7) \qquad F_n \sim \frac{\exp[-1/2\,x^2]}{2^n\,\Gamma(n/2+1)} \exp[-\sqrt{2nx}] , \qquad n \to \infty .$$

Dall'esame di queste due ultime relazioni, segue allora che Y_n tende a zero, quando $n \to \infty$, molto più velocemente di F_n, e quindi di S_n. Questo spiega come mai Y_n può essere calcolata in modo efficiente ricorrendo alla (5.5), applicata a ritroso, come proposto in [2]. Ciononostante il calcolo di S_n sembra presentare le stesse difficoltà di quello delle F_n.

Dose e Semini suggeriscono di calcolare S_n attraverso la sua rappresentazione polinomiale

$$(5.8) \qquad S_n = \frac{1}{\sqrt{\pi}} \sum_{k=0}^{n} \frac{\Gamma((k+1)/2)}{k!(n-k)!} (-x)^{n-k} ;$$

tuttavia quest'ultima appare soggetta a errori di cancellazione considerevoli.

Il metodo è stato esaminato per $x = 0.1(0.1)1.5$ e $n = 1(1)50$. Nella tabella seguente riportiamo il numero di cifre significative perse nel calcolo di S_n, per alcuni valori di x e n.

n	Cifre significative perse		
	$x = 0.1$	$x = 0.5$	$x = 1$
5	0.5	1	2
10	0.5	1.5	3.5
15	1	1.5	4.5
20	1	2.5	5.5
25	1.5	2.5	6
30	1.5	2.5	6
35	1.5	3	6.5
40	1.5	3.5	7.5
50	1.5	3.5	8.5

In particolare, dal confronto di questo metodo con la (2.2), per i suddetti valori di x e n, abbiamo rilevato che la precisione del primo non è mai risultata superiore a quella del secondo. Per esempio, per $x = 1.5$ e $n = 50$ abbiamo riscontrato una differenza di ben 4 cifre. Per valori di x più piccoli, invece, la discrepanza non è mai stata superiore a 1 o 2 cifre.

6. – La precisione dei due metodi proposti non è mai risultata superiore a quella ottenuta applicando la (2.2) direttamente. Per la sua semplicità, pensiamo quindi che quest'ultima sia senz'altro da preferirsi, laddove la (2.2) stessa a ritroso mostra una convergenza troppo lenta.

Tutti i calcoli sono stati fatti su un CDC 6500.

Ringrazio il mio direttore di ricerca, Prof. W. Gautschi, per la sua costante guida e per i suoi preziosi suggerimenti.

Bibliografia.

[1] A. ASCARI, *Sul calcolo numerico degli integrali iterati della funzione degli errori*, Riv. Mat. Univ. Parma (2) **11** (1970), 103-110.

[2] V. DOSE and C. SEMINI, *Expansion of Gaussian functions in hydrogen eigenfunctions*, Helv. Phys. Acta **47** (1974), 307-320.

[3] W. GAUTSCHI: $[\cdot]_1$ *Error function and Fresnel integrals*, Appl. Math. Ser. 55, Washington, D. C. (1964), 295-329; $[\cdot]_2$ *Recursive computation of the repeated integrals of the error function*, Math. Comp. **15** (1961), 227-232; $[\cdot]_3$ *Computational aspects of three-term recurrence relations*, SIAM Rev. **9** (1967), 24-82; $[\cdot]_4$ *Evaluation of the repeated integrals of the coerror function* (In corso di pubblicazione su ACM Transactions on Mathematical Software); $[\cdot]_5$ *Algorithm. Repeated integrals of the coerror function* (In corso di pubblicazione su ACM Transactions on Mathematical Software).

[4] F. W. J. OLVER, *Asymptotics and special functions*, Academic Press, New York and London (1974).

Sommario.

Due metodi recenti per il calcolo numerico degli integrali iterati della funzione degli errori vengono esaminati e confrontati con l'applicazione diretta della relazione ricorsiva che tali integrali soddisfano. Si mostra che i risultati ottenuti con i primi due non migliorano quelli derivati con la suddetta relazione, e quindi quest'ultima, per la sua semplicità, è senz'altro da preferirsi.

Abstract.

Two recent methods for the computation of the iterated integrals of the error function are examined and compared with the use of the three-term recurrence relation satisfied by these integrals. We show that the results derived from the proposed methods do not improve upon those obtained by the recurrence relation, and hence the latter is to be preferred.

* * *