RC 3446

Technische Hogeschool Eindhoven
Onderafdeling der Wiskunde
Rekencentrum

RC-Informatie nr. 22
6 bladzijden
datum: 28 oktober 1968

Algolprocedures voor het berekenen van een inwendig product in dubbele precisie.

1. Inleiding

De hier beschreven procedures DLINPROD en DLINPROD1 berekenen een expressie van de vorm

$$s = c + \sum_{i=i1}^{i2} ai \times bi \qquad (1)$$

(waarbij de expressies ai en bi in het algemeen van i afhangen) zodanig dat de partiele sommen met een relatieve nauwkeurigheid van ca. 1 op 10^{24} berekend en bewaard worden.

Bij de procedure DLINPROD wordt het eindresultaat afgerond op de normale nauwkeurigheid (ca 1 op 10^{12}) en daarna uitgevoerd, bij de procedure DLINPROD1 wordt het eindresultaat volledig uitgevoerd met behulp van twee reals. Voorts kan bij DLINPROD1 de grootheid c door middel van twee reals in dubbele nauwkeurigheid worden ingevoerd.

De procedures zijn vooral nuttig voor de nauwkeurige berekening van expressies van de vorm (1) indien $|c + \sum ai \times bi|$ zeer veel kleiner is dan $|c| + \sum |ai| \times |bi|$, zodat bij berekening zonder voorzorgen aanzienlijk cijferverlies kan optreden.
Daar de uitvoering van deze procedures enige malen meer tijd vergt dan die van de eenvoudige inwendig product procedure (zoals beschreven in RC-informatie nr. 18) moet gebruik in situaties waar de dubbele precisie niet essentieel is, dringend ontraden worden.

2. Gebruiksaanwijzing

2.1. real procedure DLINPROD (i, i1, i2, ai, bi, c)

2.1.1. Formele parameters

integer i	< variable >
	Wordt als Jensen parameter gebruikt.
	Behoeft bij een aanroep geen waarde te hebben.
	Na beeindiging undefined.
integer i1,i2	< expression >
	Onder en bovengrens van de sommatie-index.
	Worden by value aangeroepen.
real ai,bi	< expression >
	Factoren van het inwendig product.
real c	< expression >
	Getal waarbij het inwendige product opgeteld
	wordt. Wordt by value aangeroepen.
DLINPROD	Na beeindiging van de procedure heeft deze
	function designator als waarde het resultaat
	van de berekening.

2.2. procedure DLINPROD1 (i, i1, i2, ai, bi, c1, c0, s1, s0)

2.2.1. Formele parameters

integer i
integer i1,i2 } zelfde als in 2.1.1.
real ai,bi

real c1,c0	< expression >
	Het getal c waarbij het inwendig product
	wordt opgeteld is de exacte som van c1 en c0.
	c1 en c0 worden by value aangeroepen.
real s1,s0	< variable >
	Na beeindiging van de procedure bevatten s1 en s0
	samen het resultaat van de berekening.

3. Algol tekst

```
real procedure DLINPROD(i, i1, i2, ai, bi, c);
value i1, i2, c; integer i, i1, i2; real ai, bi, c;
begin real x, x1, x0, y, y1, y0, c0;
      c0 := 0;
      for i := i1 step 1 until i2 do
      begin x := ai; y := bi;
          if x ≠ 0 ∧ y ≠ 0 then
          begin x1 := 1048576 × x + x; x1 := x - x1 + x1; x0 := x - x1;
                y1 := 1048576 × y + y; y1 := y - y1 + y1; y0 := y - y1;
                x := x × y; y := x1 × y1 - x + x1 × y0 + x0 × y1 + x0 × y0;
                x1 := x + c;
                x0 := if abs(x) ≥ abs(c) then x - x1 + c + c0 + y else c - x1 + x + y + c0;
                c := x1 + x0; c0 := x1 - c + x0
          end
      end;
      DLINPROD := c
end DLINPROD;
```

```
procedure DLINPROD1(i, i1, i2, ai, bi, c1, c0, s1, s0);
value i1, i2, c1, c0; integer i, i1, i2; real ai, bi, c1, c0, s1, s0;
begin real x, x1, x0, y, y1, y0;
      x := c1 + c0; c0 := if abs(c1) ≥ abs(c0) then c1 - x +  c0 else c0 - x + c1; c1 := x;
      for i := i1 step 1 until i2 do
      begin x := ai; y := bi;
          if x ≠ 0 ∧ y ≠ 0 then
          begin x1 := 1048576 × x + x; x1 := x - x1 + x1; x0 := x - x1;
                y1 := 1048576 × y + y; y1 := y - y1 + y1; y0 := y - y1;
                x := x × y; y := x1 × y1 - x + x1 × y0 + x0 × y1 + x0 × y0;
                x1 := x + c1;
                x0 := if abs(x) ≥ abs(c1) then x - x1 + c1 + c0 + y else c1 - x1 + x + y + c0;
                c1 := x1 + x0; c0 := x1 - c1 + x0
          end
      end;
      s1 := c1; s0 := c0
end DLINPROD1;
```

4. Toelichting

Beide procedures berekenen de expressie (1).

Bij DLINPROD is c de waarde van de als actuele parameter meegegeven expressie c, bij DLINPROD1 is c de exacte som van de waarden van de expressies c1 en c0.

De producten van de waarden van de expressies ai en bi (die in het algemeen van de Jensen parameter i zullen afhangen) worden exact berekend en met een relatieve nauwkeurigheid van ca. 1 op 10^{24} bij de reeds aanwezige partiele som opgeteld.

De uiteindelijk verkregen som s wordt bij DLINPROD afgerond tot een real en aan de function designator DLINPROD toegekend. Bij DLINPROD1 wordt s zodanig aan het paar $\{ s1, s0 \}$ toegekend dat s1 de afronding van s tot een real is en s0 = s - s1.

Indien bij de evaluatie van de expressies c, c1, c0, ai en bi rekenkundige bewerkingen nodig zijn, dan worden deze uiteraard met de normale nauwkeurigheid uitgevoerd.

Als i2 < i1 dan is de som leeg en dan wordt als resultaat c afgeleverd.

5. Methode

De methode berust op de volgende, voor de EL X8 geldende feiten:

a. De waarde van een real is of 0 of van de vorm

$$m \times 2^p,$$

waarbij m en p geheel zijn, $0 < |m| < 2^{40}$.

b. Bij optelling of aftrekking van twee reals is het resultaat steeds de afronding naar de (een) dichtstbijzijnde door een real voorstelbare waarde.

Voor machines die de eigenschap b. hebben maar een andere mantisselengte kan de methode eenvoudig aangepast worden.

Voor verdere details betreffende methode en nauwkeurigheid zij verwezen naar een binnenkort te verschijnen rapport.

6. Nauwkeurigheid

Noemt men het door de procedure DLINPROD afgeleverde getal s, dan geldt

$$|s - c - \sum ai \times bi| \leq$$
$$2^{-40} \times |s| + 3 \times 2^{-80} \times (i2 - i1 + 1) \times (|c| + \sum |ai| \times |bi|)$$

Voor de door de procedure DLINPROD1 afgeleverde getallen s1 en s0 geldt (als $s = s1 + s0$, $c = c1 + c0$)

$$|s - c - \sum ai \times bi| \leq 3 \times 2^{-80} \times (i2 - i1 + 1) \times (|c| + \sum |ai| \times |bi|)$$

Opmerking

Berekening van de expressie (1) met de standaard-procedure inprod (zie RC informatie nr. 18), die de normale arithmetiek gebruikt, levert een resultaat s dat voldoet aan

$$|s - c - \sum_{i=i1}^{i2} ai \times bi| \leq 2^{-40} \times (i2 - i1 + 2) \times (|c| + \sum |ai| \times |bi|).$$