Spline-functies

1 Definities

In het hoofdstuk over veelterminterpolatie hebben we gezien dat de interpolatiefout niet altijd kleiner wordt voor toenemende graad van de interpolerende veelterm. Om convergentie te garanderen zullen we functies benaderen niet met veeltermen maar met zogenaamde *spline-functies.*

Definitie 1 (spline functie) *Zij* $k \in \mathbb{N} = \{0, 1, 2, \ldots\}$ *en* $a = t_0 < t_1 < \cdots < t_n = b$ *met* $t_i \in \mathbb{R}$. *Dan is de functie* $s(x)$ *een* spline *van graad* k *(of van orde* $k+1$*) over het interval* $[a, b]$ *met als knooppunten* $t_0, t_1, \ldots, t_n$ *als en slechts als*

- $s(x)$ *over elk deelinterval* $[t_j, t_{j+1})$, $j = 0(1)n-2$ *en over het interval* $[t_{n-1}, t_n]$ *een veelterm is van graad* $\leq k$*;*
- *alle afgeleiden* $s^{(j)}(x)$, $j = 0(1)k-1$ *continu zijn over het interval* $[a, b]$.

Voor $k = 0$ is automatisch aan de tweede voorwaarde voldaan. Dus een spline van graad 0 is een stapfunctie. Een spline van graad 1 is een continue gebroken lijn.

Merk op dat de verzameling van alle veeltermen (over het interval $[a, b]$) van graad $\leq k$ een deelverzameling is van de verzameling van de splines van graad k.

Voor een gegeven stel knooppunten is de verzameling van de splines van graad k een vectorruimte. We zullen deze vectorruimte noteren als $\mathcal{V}_k(t_0, t_1, \ldots, t_n)$.

2 De voorstelling van splines

Volgens de definitie wordt de spline $s(x)$ in elk deelinterval $[t_j, t_{j+1})$ gegeven door een veelterm. D.w.z. dat we in elk deelinterval de spline kunnen voorstellen door een voorstelling te nemen van de corresponderende veelterm. Een voor de hand liggende voorstelling is de klassieke voorstelling

$$s(x) = p_j(x) = a_{j,k}x^k + a_{j,k-1}x^{k-1} + \cdots + a_{j,0}, \quad \forall x \in [t_j, t_{j+1}), \quad j = 0(1)n-1.$$

De continuïteitsvoorwaarden zorgen voor $k(n-1)$ lineair onafhankelijke vergelijkingen tussen de $n(k+1)$ coëfficiënten $a_{j,i}$ van de veeltermvoorstellingen. Hieruit volgt dat de dimensie van de vectorruimte $\mathcal{V}_k(t_0, t_1, \ldots, t_n)$ gelijk is aan $n(k+1) - k(n-1) = n + k$. In wat volgt gaan we op zoek naar een basis voor $\mathcal{V}_k(t_0, t_1, \ldots, t_n)$ wat ons een voorstelling oplevert waarbij de koördinaten in tegenstelling tot de coëfficiënten $a_{j,i}$ onafhankelijk van elkaar kunnen gekozen worden. Een eerste basis wordt gevormd door gebruik te maken van afgeknotte machtsfuncties. In de volgende sectie zullen we een basis opstellen met behulp van de zogenaamde B-splines.

Definitie 2 (afgeknotte machtsfuncties) *De afgeknotte machtsfunctie x_+^k van graad k wordt gedefiniëerd als*

$$\begin{aligned} x_+^k &= x^k, \qquad \textit{voor } x > 0 \\ &= 0, \qquad \textit{voor } x \leq 0 \end{aligned}$$

Merk op dat voor elk van de knooppunten t_j de functie $(t_j - x)_+^k$ ook een spline van graad k is. Het is nu eenvoudig na te gaan dat elke spline $s(x)$ van graad k met als knooppunten $t_0, t_1, \ldots, t_n$ kan voorgesteld worden als

$$s(x) = \sum_{i=0}^{k} a_j x^j + \sum_{j=1}^{n-1} c_j (t_j - x)_+^k.$$

Hieruit volgt dat de verzameling functies

$$\{1, x, x^2, \ldots, x^k, (t_1 - x)_+^k, \ldots, (t_{n-1} - x)_+^k\}$$

een basis vormt voor de vectorruimte $\mathcal{V}_k(t_0, t_1, \ldots, t_n)$. Het gebruik van deze basis zorgt echter voor numeriek onstabiele berekeningen. Daarom gaan we in de volgende sectie op zoek naar een andere basis gebaseerd op B-splines die resulteert in meer nauwkeurige resultaten.

3 B-splines

We hebben gedeelde differenties gebruikt bij de Newton-voorstelling van de interpolerende veelterm. We herhalen de definitie.

Definitie 3 (gedeelde differentie) *Zij gegeven de koppels $(x_i, f_i) \in \mathbb{R}^2$ voor $i = 0(1)n$ waarbij de x_i allemaal verschillend zijn van elkaar. Dan worden de gedeelde differenties op een recursieve manier gedefiniëerd als*

- $f[x_i] = f_i$, $i = 0(1)n$ *(gedeelde differentie van orde 0);*
- $f[x_i, x_{i+1}, \ldots, x_{i+k}] = \dfrac{f[x_i, x_{i+1}, \ldots, x_{i+k-1}] - f[x_{i+1}, x_{i+2}, \ldots, x_{i+k}]}{x_i - x_{i+k}}$,
  $k = 1(1)n$, $i = 0(1)n - k$ *(gedeelde differentie van orde k).*

We zullen de volgende twee eigenschappen van gedeelde differenties nodig hebben.

Eigenschap 1 (eigenschappen van gedeelde differenties)

1. *Met* $\pi(x) = \prod_{i=0}^{n}(x - x_i)$, *krijgen we*

$$f[x_0, x_1, \ldots, x_n] = \sum_{j=0}^{n} \frac{f_j}{\pi'(x_j)}. \tag{1}$$

2. *(formule van Leibnitz) Als* $f(x) = g(x)h(x)$, *geldt er*

$$f[x_0, x_1, \ldots, x_n] = \sum_{j=0}^{n} g[x_0, x_1, \ldots, x_j]\, h[x_j, x_{j+1}, \ldots, x_n]. \tag{2}$$

Wanneer we een functie $f(t, x)$ hebben in twee veranderlijken, gebruiken we de notatie $\Delta_x^k[x_i, x_{i+1}, \ldots, x_{i+k}]f(t, x)$ voor de gedeelde differentie voor de koppels $(x_j, f(x_j, t))$ en $\Delta_t^k[t_i, t_{i+1}, \ldots, t_{i+k}]f(t, x)$ voor de koppels $(t_i, f(x, t_i))$.
Merk op dat $\Delta_x^k[x_i, x_{i+1}, \ldots, x_{i+k}]f(t, x)$ een functie is in t en $\Delta_t^k[t_i, t_{i+1}, \ldots, t_{i+k}]f(t, x)$ een functie in x.

Definitie 4 (B-spline) *Een B-spline van graad k is elke functie*

$$M_{i,k}(x) = \Delta_t^{k+1}[t_i, t_{i+1}, \ldots, t_{i+k+1}](t - x)_+^k.$$

We zullen de B-splines normaliseren:

$$N_{i,k}(x) = (t_{i+k+1} - t_i)M_{i,k}(x).$$

Door deze normalisatie geldt

$$\sum_i N_{i,k}(x) \equiv 1.$$

Uit formule (1) volgt dat

$$M_{i,k}(x) = \sum_{j=0}^{k+1} \frac{(t_{i+j} - x)_+^k}{\pi'_{i,k+1}(t_{i+j})}$$

met $\pi_{i,k+1}(t) = \prod_{j=0}^{k+1}(t - t_{i+j})$.
Hieruit volgt dat B-splines ook splines zijn.

4 De B-spline voorstelling van spline-functies

Met de knooppunten $t_0, t_1, \ldots, t_n$ kunnen we $n-k$ verschillende B-splines van graad k definiëren, nl.

$$N_{i,k}(x) \quad \text{voor} \quad i = 0(1)n-k-1.$$

Deze $n-k$ B-splines moeten we nog aanvullen met $2k$ bijkomende splines om een basis voor $\mathcal{V}_k(t_0, t_1, \ldots, t_n)$ te bekomen. Daarom voegen we zowel links als rechts van het interval nog k knooppunten toe:

$$\begin{array}{l} t_{-k} < t_{-k+1} < \ldots < t_0 \\ t_n < t_{n+1} < \ldots < t_{n+k} \; . \end{array}$$

Dan kunnen we elke spline $s(x)$ van graad k met als knooppunten $t_0, t_1, \ldots, t_n$ op een unieke wijze schrijven als

$$s(x) = \sum_{j=-k}^{n-1} c_j N_{j,k}(x).$$

5 Eigenschappen van B-splines

De recursieve definitie van gedeelde differenties samen met formule (2) leidt tot de volgende recursiebetrekkingen voor B-splines.

Eigenschap 2 (recursiebetrekking)

$$\begin{aligned} M_{i,k}(x) &= \frac{x - t_i}{t_{i+k+1} - t_i} M_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_i} M_{i+1,k-1}(x) \; . \\ N_{i,k}(x) &= \frac{x - t_i}{t_{i+k} - t_i} N_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} N_{i+1,k-1}(x) \; . \end{aligned}$$

Voor het evalueren van splines is het van belang dat B-splines verschillend zijn van nul in een beperkt interval.

Eigenschap 3

$$M_{i,k}(x) = 0 \quad \textit{als} \quad x < t_i \quad \textit{of} \quad x \geq t_{i+k+1} \; .$$

Bewijs: Omdat $M_{i,k}(x)$ een gedeelde differentie is in de punten $t_i, t_{i+1}, \ldots, t_{i+k+1}$ kunnen we formule (1) toepassen:

$$M_{i,k}(x) = \sum_{j=0}^{k+1} \frac{(t_{i+j} - x)_+^k}{\pi'(t_{i+j})}$$

met

$$\pi(t) = \prod_{j=0}^{k+1} (t - t_{i+j}) \ .$$

Als $x \geq t_{i+k+1}$ zijn alle afgeknotte machtsfuncties $(t_{i+j} - x)_+^k$ gelijk aan nul. Als $x < t_i$ geldt dat

$$(t-x)_+^k = (t-x)^k \quad , \quad t = t_i, t_{i+1}, \ldots, t_{i+k+1} \ .$$

De B-spline $M_{i,k}(x)$ is gedefinieerd als

$$M_{i,k}(x) = \Delta_t^{k+1}(t_i, t_{i+1}, \ldots, t_{i+k+1})(t-x)_+^k \ ,$$

d.w.z. als een gedeelde differentie in de punten

$$t_i, t_{i+1}, \ldots, t_{i+k+1} \quad \text{voor de functie} \quad (t-x)_+^k \ .$$

Wanneer $x < t_i$ kunnen we $M_{i,k}(x)$ ook schrijven als

$$M_{i,k}(x) = \Delta_t^{k+1}(t_i, \ldots, t_{i+k+1})(t-x)^k \ ,$$

wat een gedeelde differentie is van orde $k+1$ van een veelterm van graad k. Hieruit volgt dat $M_{i,k}(x) = 0$ voor $x < t_i$. □

Eigenschap 4

$$M_{i,k}(x) > 0 \qquad \textit{als } t_i < x < t_{i+k+1}$$

Bewijs: We bewijzen de eigenschap door volledige inductie.

De eigenschap geldt voor $M_{i,0}$ want

$$\begin{aligned} M_{i,0}(x) &= \frac{1}{t_{i+1} - t_i} \qquad \text{voor} \quad t_i \leq x < t_{i+1} \\ &= 0 \qquad \text{voor} \quad x < t_i \quad \text{of} \quad x \geq t_{i+1} \ . \end{aligned}$$

Dus $M_{i,0}(x)$ is strikt positief tussen t_i en t_{i+1}.

De recursiebetrekking voor B-splines kan geschreven worden als

$$M_{i,k}(x) = \alpha \ M_{i,k-1}(x) + \beta \ M_{i+1,k-1}(x)$$

met

$$\alpha = \frac{x - t_i}{t_{i+k+1} - t_i} \quad \text{en} \quad \beta = \frac{t_{i+k+1} - x}{t_{i+k+1} - t_i} \ .$$

Als $t_i < x < t_{i+k+1}$, geldt er

$$\begin{aligned} \alpha &> 0 \\ \beta &> 0 \\ \alpha + \beta &= 1. \end{aligned}$$

M.a.w. $M_{i,k}$ is een strikt convexe combinatie van $M_{i,k-1}$ en $M_{i+1,k-1}$, dus ook strikt positief. $\square$

Uit eigenschap 3 volgt dat in elk punt x maximum $k+1$ B-splines van graad k verschillend zijn van nul.

De normalisatievoorwaarden kunnen dus ook geschreven worden als

$$\sum_i N_{i,k}(x) = \sum_{i=j-k}^{j} N_{i,k}(x) \equiv 1 \quad \text{als} \quad t_j \leq x < t_{j+1} \ .$$

6 Evalueren van spline-functies

We willen de waarde berekenen van de spline-functie $s(x)$ gegeven in zijn B-spline voorstelling

$$s(x) = \sum_{i=-k}^{n-1} c_i N_{i,k}(x) \ . \tag{3}$$

We nemen aan dat $t_j \leq x < t_{j+1}$.
Van de $n+k$ basisfuncties $N_{i,k}(x)$ in de som van formule (3) zijn er maximum $k+1$ verschillend van nul, nl.

$$N_{i,k}(x) \qquad , \qquad i = j - k(1)j \ .$$

Hieruit volgt dat

$$s(x) = \sum_{i=j-k}^{j} c_i N_{i,k}(x) \quad \text{als} \quad t_j \leq x < t_{j+1} \ .$$

Een eerste mogelijkheid om $s(x)$ te evalueren bestaat erin om de waarden $N_{i,k}(x)$, $i = j-k(1)j$ te berekenen door $k+1$ differentietabellen op te stellen. Voor elke $N_{i,k}(x)$ ziet zo'n differentietabel eruit als volgt:

$$\begin{array}{cccc} (t_i - x)_+^k & & & \\ (t_{i+1} - x)_+^k & \ddots & & \\ \vdots & \ddots & \ddots & \\ (t_{i+k+1} - x)_+^k & \dots & \dots & \frac{N_{i,k}(x)}{t_{i+k+1} - t_i} \end{array} \ .$$

Deze methode is echter numeriek onstabiel.

Een tweede methode maakt gebruik van de recursiebetrekking voor B-splines om $N_{i,k}(x)$, $i = j-k(1)j$ uit te rekenen. We weten dat er maximum slechts $l+1$ B-splines van graad l verschillend zijn van nul voor een vaste waarde van x. Hieruit volgt dat er een driehoekige tabel overblijft met volgende elementen:

$$\begin{array}{ccccc}
M_{j,0}(x) = 1/(t_{j+1}-t_j) & M_{j-1,1}(x) & M_{j-2,2}(x) & \dots & M_{j-k,k}(x) \\
& M_{j,1}(x) & M_{j-1,2}(x) & \dots & M_{j-k+1,k}(x) \\
& & M_{j,2}(x) & \dots & M_{j-k+2,k}(x) \\
& & & \ddots & \vdots \\
& & & & M_{j,k}(x)
\end{array}$$

Deze methode is numeriek stabieler want er worden altijd convexe combinaties van positieve getallen gevormd.

Een derde methode (algoritme van de Boor (1972)) berekent de driehoekige tabel

$$\begin{array}{llll}
c_j^{[0]}(x) & & & \\
c_{j-1}^{[0]}(x) & c_j^{[1]}(x) & & \\
\vdots & \vdots & \ddots & \\
c_{j-k}^{[0]}(x) & c_{j-(k-1)}^{[1]}(x) & \dots & c_j^{[k]}(x)
\end{array}$$

met

$$c_j^{[i]}(x) = \begin{cases} c_j & \text{als} \quad i = 0 \\ \dfrac{c_j^{[i-1]}(x)(x-t_j) + c_{j-1}^{[i-1]}(x)(t_{j+k+1-i}-x)}{t_{j+k+1-i}-t_j} & \text{als} \quad i > 0\,. \end{cases}$$

De waarde van $s(x)$ is dan

$$s(x) = \sum_{i=-k}^{n-1} c_i N_{i,k}(x) = c_j^{[k]}(x)\,.$$

De laatste methode kunnen we ook gebruiken om de waarde van $N_{i,k}(x)$ te berekenen door $c_i = 1$ te stellen en alle andere $c_j = 0$, $j \neq i$.

7 Interpolatie met kubische splines

In deze sectie zullen we de tabel van punten $\{(x_i, f_i)\}$, $i = 0(1)n$ interpoleren met een spline-functie $s(x)$ i.p.v. met een veelterm $y_n(x)$ van graad n. We zullen de knooppunten t_i gelijk nemen aan de x_i-waarden: $t_i = x_i$, $i = 0(1)n$. In sectie 2, hebben we gezien dat de dimensie van de vectorruimte

$\mathcal{V}_k(t_0, t_1, \ldots, t_n)$ van de splines van graad k gelijk is aan $n+k$. Dit betekent dat we naast de $n+1$ interpolatievoorwaarden

$$s(x_i) = f_i, \qquad i = 0(1)n \tag{4}$$

nog $k-1$ bijkomende voorwaarden moeten opleggen om de spline $s(x)$ uniek te kunnen bepalen.

Dikwijls wordt de graad k van de interpolerende spline-functie $s(x)$ gelijk aan drie genomen. Dit wil zeggen dat de spline, de eerste en de tweede afgeleide van de spline continue functies zijn. Het menselijk oog ervaart de spline-functie dan als een vloeiende kromme. De splines van graad drie worden *kubische splines* genoemd. Naast de $n+1$ interpolatievoorwaarden hebben we dan nog $k-1=2$ bijkomende voorwaarden nodig. Meestal wordt er een *natuurlijke* of een *periodieke* spline genomen.

- Een natuurlijke spline-functie voldoet aan de volgende twee voorwaarden:
$$s^{(2)}(t_0) = s^{(2)}(t_n) = 0.$$
- Een periodieke kubische spline-functie voldoet aan de volgende twee voorwaarden:
$$s^{(j)}(t_0) = s^{(j)}(t_n), \qquad \text{voor} \quad j = 1, 2.$$
Men veronderstelt dan ook meestal dat $f_0 = f_n$ zodat ook $s(t_0) = s(t_n)$.

We bepalen de coördinaten c_j van de interpolerende kubische spline $s(x)$ in zijn B-spline voorstelling

$$s(x) = \sum_{j=-3}^{n-1} c_j N_{j,3}(x)$$

door de interpolatievoorwaarden in te vullen:

$$\begin{aligned} s(x_i) &= \sum_{j=-3}^{n-1} c_j N_{j,3}(x_i) \\ &= \sum_{j=i-3}^{i-1} c_j N_{j,3}(x_i) \\ &= f_i. \end{aligned}$$

In wat volgt gebruiken we de notatie $N_j(x) = N_{j,3}(x)$. Nemen we de natuurlijke voorwaarden om de spline-functie enig te maken, dan krijgen we de volgende voorwaarden:

$$\begin{aligned} \sum_{j=-3}^{-1} c_j N_j^{(2)}(x_0) &= 0 \qquad \text{en} \\ \sum_{j=n-3}^{n-1} c_j N_j^{(2)}(x_n) &= 0. \end{aligned}$$

Dit leidt tot het volgende stelsel van $n+3$ lineaire vergelijkingen in de $n+3$ onbekenden $c_{-3}, c_{-2}, \ldots, c_{n-2}, c_{n-1}$:

$$\begin{bmatrix} N_{-3}^{(2)}(x_0) & N_{-2}^{(2)}(x_0) & N_{-1}^{(2)}(x_0) & & & & \\ N_{-3}(x_0) & N_{-2}(x_0) & N_{-1}(x_0) & & & & \\ & N_{-2}(x_1) & N_{-1}(x_1) & N_0(x_1) & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & N_{n-4}(x_{n-1}) & N_{n-3}(x_{n-1}) & N_{n-2}(x_{n-1}) & \\ & & & & N_{n-3}(x_n) & N_{n-2}(x_n) & N_{n-1}(x_n) \\ & & & & N_{n-3}^{(2)}(x_n) & N_{n-2}^{(2)}(x_n) & N_{n-1}^{(2)}(x_n) \end{bmatrix} \begin{bmatrix} c_{-3} \\ c_{-2} \\ c_{-1} \\ \vdots \\ c_{n-2} \\ c_{n-1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ f_0 \\ f_1 \\ \vdots \\ f_{n-1} \\ f_n \\ 0 \end{bmatrix}.$$

Het derde element uit de eerste rij kan gemakkelijk geëlimineerd worden m.b.v. de tweede rij. Analoog kan het op twee na laatste element op de laatste rij gemakkelijk geëlimineerd worden m.b.v. de voorlaatste rij. Het resulterende stelsel is een tridiagonaal-stelsel dat met $O(n)$ bewerkingen kan opgelost worden.

8 Oefeningen

- In dit hoofdstuk werden verschillende resultaten niet expliciet afgeleid. Probeer zelf enkele van deze resultaten te bewijzen, bijv. de recursiebetrekking voor B-splines.
- Bewijs dat de afgeleiden van een spline $s(x)$ van graad k met knooppunten $t_{-k} < t_{-k+1} < \cdots < t_{n-1} < t_n$ kunnen berekend worden m.b.v.

de volgende formule

$$s^{(\nu)}(x) = \prod_{i=1}^{\nu}(k+1-i) \sum_{j=-(k-\nu)}^{n-1} c_i^{(\nu)} N_{i,k-\nu}(x)$$

waarbij de $c_j^{(i)}$ voldoen aan de volgende recursiebetrekkingen

$$c_j^{(i)} = \begin{cases} c_j & \text{als } i = 0 \\ \dfrac{c_j^{(i-1)} - c_{j-1}^{(i-1)}}{t_{j+k+1-i} - t_j} & \text{als } i > 0 \ . \end{cases}$$

1 QR Factorisatie

1.1 Householder Transformatie

Gegeven een matrix A :

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Om nullen te maken in de eerste kolom vanaf de tweede rij, selecteer de kolomvector waar de nullen moeten komen. Als er nullen moeten komen vanaf de i+1'de rij, behoudt dan de elementen uit de kolomvector vanaf het i-de element (alle elementen die nul moeten worden en het element erboven). Noem deze vector x :

$$x = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}$$

Bereken v (is later nodig), $sgn(x_1)$ is hier de sign-functie van het eerste element uit x, in ons geval is die a_{11}

$$\mathbf{v} = \mathbf{x} + sgn(x_i)||x||.e_i$$

Hiermee berekenen we P_1

$$P_1 = I - \frac{2vv^T}{v^T v}$$

Deze P_1 heeft als eigenschap dat $P_1 x = (sgn(x_1)||x||, 0, ..., 0)^T$

Dus in ons voorbeeld wordt $\mathbf{P}x = (sgn(a_{11})||x||, 0, ..., 0)^T$

P_1A levert een nieuwe matrix op waarbij de gewenste elementen nul zijn :

$$P_1 A = \begin{pmatrix} sgn(a_{11})||x|| & * & * \\ 0 & b_{22} & b_{23} \\ 0 & b_{32} & b_{33} \end{pmatrix}$$

Hierbij zijn de * nieuwe waarden, maar onbelangrijk, en de b_{ij} nieuwe waarden die nog gebruikt worden.

Om in de volgende kolom nullen te creëren doen we hetzelfde, maar x is hier kleiner :

$$x = \begin{pmatrix} b_{22} \\ b_{32} \end{pmatrix}$$

We bereken v exact zoals hiervoor, en P_2' zoals P_1.

Stel we bekomen P_2' :

$$P_2' = \begin{pmatrix} p_{22} & p_{23} \\ p_{32} & p_{33} \end{pmatrix}$$

Om nu de benodigde P_2 te bekomen moet P_2' de juiste dimensie krijgen, die gebeurt als volgt :

$$P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & p_{22} & p_{23} \\ 0 & p_{32} & p_{33} \end{pmatrix}$$

Vermenigvuldigen we nu P_1A met P_2 dan bekomen we de gewenste bovendriehoeksmatrix :

$$P_2P_1A = \begin{pmatrix} sgn(a_{11})||x_1|| & * & * \\ 0 & sgn(b_{22})||x_2|| & * \\ 0 & 0 & c_{33} \end{pmatrix}$$

Stellen we nu :

$$Q^T = P_2P_1$$

Dan (De P's zijn symmetrisch):

$$Q = P_1P_2$$

Dan krijgen we de QR factorisatie van A als volgt :

$$R = P_2P_1A = Q^TA$$

$$QR = A$$

Deze procedure is volledig uitbreidbaar voor grotere (of kleinere) dimensies van A.

Oefeningen Numerieke Wiskunde

Foutenanalyse: twee verschillende aanpakken

Bij het doen van een formele foutenanalyse kan je twee methodes volgen om een eerste orde benadering te bekomen van de fout. Bij de eerste methode ga je waar nodig de stelling van Taylor in 1 veranderlijke toepassen om zo in een aantal stappen een eerste orde benadering te bekomen. Bij de tweede methode bereken je de eerste orde benadering in één keer via een Taylor benadering in meer veranderlijken.

Beide methoden geven hetzelfde resultaat. Welke methode het makkelijkst is, hangt af van probleem tot probleem en is ook afhankelijk van je persoonlijke voorkeur.

1 Voorbeeld

Als voorbeeld voeren we een foutenanalyse uit voor de berekening van

$$y = \sqrt{1+x} - 1,$$

waarbij men de berekeningen uitvoert volgens de uitdrukking hierboven. Men veronderstelt dat x exact voorgesteld kan worden op de machine.

De berekende waarde voor y is dan

$$\bar{y} = fl\left(fl(\sqrt{fl(1+x)}) - 1\right) = \left((\sqrt{(1+x)(1+\epsilon_1)})(1+\epsilon_2) - 1\right)(1+\epsilon_3) \tag{1}$$

Hierbij stellen $\epsilon_i,\ i = 1,2,3,$ de relatieve fouten voor die gemaakt worden bij afronding na iedere bewerking. We weten dat

$$|\epsilon_i| \leq \epsilon_{mach},\ i = 1,2,3.$$

2 Methode I

We herschrijven $\bar{y}$ als $\bar{y} \approx y(1+\delta)$ door in (1) gebruik te maken van de Taylor reeks rond nul

$$f(\epsilon) = f(0) + f'(0)\epsilon + f''(0)\frac{\epsilon^2}{2} + \ldots$$

en hogere orde termen te verwaarlozen. Bv.

$$\sqrt{1+\epsilon} \approx 1 + \frac{1}{2}\epsilon,$$

waarbij we veronderstellen dat ϵ klein genoeg is.

We krijgen dus

$$\begin{aligned}
\bar{y} &\approx \left((\sqrt{1+x})(1+\frac{1}{2}\epsilon_1+\epsilon_2) - 1\right)(1+\epsilon_3) \\
&= \left(y + (\sqrt{1+x})(\frac{1}{2}\epsilon_1+\epsilon_2)\right)(1+\epsilon_3) \\
&\approx y\left(1 + \frac{\sqrt{1+x}}{\sqrt{1+x}-1}(\frac{1}{2}\epsilon_1+\epsilon_2) + \epsilon_3\right),
\end{aligned}$$

waarbij we steeds de tweede orde termen verwaarloosd hebben.

3 Methode II

We interpreteren, in de uitdrukking voor $\bar{y}$, x als een parameter en $\epsilon_i,\ i = 1, 2, 3$, als variabelen,

$$\bar{y} = F(\epsilon_1, \epsilon_2, \epsilon_3).$$

Vermits de ϵ_i zeer klein zijn, laat $F(\epsilon_1, \epsilon_2, \epsilon_3)$ zich goed benaderen door de Taylorontwikkeling rond $(0, 0, 0)$ afgebroken na de lineaire termen

$$\bar{y} = F(\epsilon_1, \epsilon_2, \epsilon_3) \approx F(0, 0, 0) + \epsilon_1 \frac{\partial F}{\partial \epsilon_1}(0, 0, 0) + \epsilon_2 \frac{\partial F}{\partial \epsilon_2}(0, 0, 0) + \epsilon_3 \frac{\partial F}{\partial \epsilon_3}(0, 0, 0).$$

De constante term $F(0, 0, 0)$ in de lineaire benadering is de exacte waarde y. Voor de berekening van de partiële afgeleiden gaan we als volgt te werk

$$\begin{aligned}
F(\epsilon_1, 0, 0) &= \sqrt{(1+x)(1+\epsilon_1)} - 1 = \sqrt{1+x}(1+\epsilon_1)^{\frac{1}{2}} - 1 \\
\frac{\partial F}{\partial \epsilon_1}(\epsilon_1, 0, 0) &= \sqrt{1+x}\frac{1}{2}(1+\epsilon_1)^{-\frac{1}{2}} \\
\frac{\partial F}{\partial \epsilon_1}(0, 0, 0) &= \frac{1}{2}\sqrt{1+x}
\end{aligned}$$

Na uitwerking van de andere partiële afgeleiden vinden we de lineaire benadering voor $\bar{y}$ en een benaderende formule voor de absolute en de relatieve fout

$$\begin{aligned}
\bar{y} &\approx y + \frac{1}{2}\sqrt{1+x}\,\epsilon_1 + \sqrt{1+x}\,\epsilon_2 + y\epsilon_3 \\
\bar{y} - y &\approx \frac{1}{2}\sqrt{1+x}\,\epsilon_1 + \sqrt{1+x}\,\epsilon_2 + y\epsilon_3 \\
\frac{\bar{y} - y}{y} &\approx \frac{\sqrt{1+x}}{2(\sqrt{1+x} - 1)}\epsilon_1 + \frac{\sqrt{1+x}}{\sqrt{1+x} - 1}\epsilon_2 + \epsilon_3.
\end{aligned}$$

We krijgen uiteraard hetzelfde resultaat als met Methode I.

4 Oefeningen

Probeer beide methodes eens uit voor de volgende berekeningen:

1. $$y = \sqrt{e^x - 1}$$

2. $$y = (1 + x)^{\frac{1}{x}}$$

3. $$y = (1 + x^2)^{x^2}$$

DERIVATIVES AND INTEGRALS

Basic Differentiation Rules

1. $\frac{d}{dx}[cu] = cu'$
2. $\frac{d}{dx}[u \pm v] = u' \pm v'$
3. $\frac{d}{dx}[uv] = uv' + vu'$
4. $\frac{d}{dx}\left[\frac{u}{v}\right] = \frac{vu' - uv'}{v^2}$
5. $\frac{d}{dx}[c] = 0$
6. $\frac{d}{dx}[u^n] = nu^{n-1}u'$
7. $\frac{d}{dx}[x] = 1$
8. $\frac{d}{dx}[|u|] = \frac{u}{|u|}(u'), \quad u \neq 0$
9. $\frac{d}{dx}[\ln u] = \frac{u'}{u}$
10. $\frac{d}{dx}[e^u] = e^u u'$
11. $\frac{d}{dx}[\log_a u] = \frac{u'}{(\ln a)u}$
12. $\frac{d}{dx}[a^u] = (\ln a)a^u u'$
13. $\frac{d}{dx}[\sin u] = (\cos u)u'$
14. $\frac{d}{dx}[\cos u] = -(\sin u)u'$
15. $\frac{d}{dx}[\tan u] = (\sec^2 u)u'$
16. $\frac{d}{dx}[\cot u] = -(\csc^2 u)u'$
17. $\frac{d}{dx}[\sec u] = (\sec u \tan u)u'$
18. $\frac{d}{dx}[\csc u] = -(\csc u \cot u)u'$
19. $\frac{d}{dx}[\arcsin u] = \frac{u'}{\sqrt{1-u^2}}$
20. $\frac{d}{dx}[\arccos u] = \frac{-u'}{\sqrt{1-u^2}}$
21. $\frac{d}{dx}[\arctan u] = \frac{u'}{1+u^2}$
22. $\frac{d}{dx}[\operatorname{arccot} u] = \frac{-u'}{1+u^2}$
23. $\frac{d}{dx}[\operatorname{arcsec} u] = \frac{u'}{|u|\sqrt{u^2-1}}$
24. $\frac{d}{dx}[\operatorname{arccsc} u] = \frac{-u'}{|u|\sqrt{u^2-1}}$
25. $\frac{d}{dx}[\sinh u] = (\cosh u)u'$
26. $\frac{d}{dx}[\cosh u] = (\sinh u)u'$
27. $\frac{d}{dx}[\tanh u] = (\operatorname{sech}^2 u)u'$
28. $\frac{d}{dx}[\coth u] = -(\operatorname{csch}^2 u)u'$
29. $\frac{d}{dx}[\operatorname{sech} u] = -(\operatorname{sech} u \tanh u)u'$
30. $\frac{d}{dx}[\operatorname{csch} u] = -(\operatorname{csch} u \coth u)u'$
31. $\frac{d}{dx}[\sinh^{-1} u] = \frac{u'}{\sqrt{u^2+1}}$
32. $\frac{d}{dx}[\cosh^{-1} u] = \frac{u'}{\sqrt{u^2-1}}$
33. $\frac{d}{dx}[\tanh^{-1} u] = \frac{u'}{1-u^2}$
34. $\frac{d}{dx}[\coth^{-1} u] = \frac{u'}{1-u^2}$
35. $\frac{d}{dx}[\operatorname{sech}^{-1} u] = \frac{-u'}{u\sqrt{1-u^2}}$
36. $\frac{d}{dx}[\operatorname{csch}^{-1} u] = \frac{-u'}{|u|\sqrt{1+u^2}}$

Basic Integration Formulas

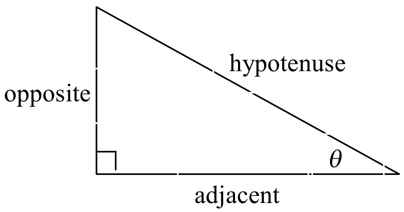
1. $\int kf(u)\,du = k\int f(u)\,du$
2. $\int [f(u) \pm g(u)]\,du = \int f(u)\,du \pm \int g(u)\,du$
3. $\int du = u + C$
4. $\int a^u\,du = \left(\frac{1}{\ln a}\right)a^u + C$
5. $\int e^u\,du = e^u + C$
6. $\int \sin u\,du = -\cos u + C$
7. $\int \cos u\,du = \sin u + C$
8. $\int \tan u\,du = -\ln|\cos u| + C$
9. $\int \cot u\,du = \ln|\sin u| + C$
10. $\int \sec u\,du = \ln|\sec u + \tan u| + C$
11. $\int \csc u\,du = -\ln|\csc u + \cot u| + C$
12. $\int \sec^2 u\,du = \tan u + C$
13. $\int \csc^2 u\,du = -\cot u + C$
14. $\int \sec u \tan u\,du = \sec u + C$
15. $\int \csc u \cot u\,du = -\csc u + C$
16. $\int \frac{du}{\sqrt{a^2-u^2}} = \arcsin\frac{u}{a} + C$
17. $\int \frac{du}{a^2+u^2} = \frac{1}{a}\arctan\frac{u}{a} + C$
18. $\int \frac{du}{u\sqrt{u^2-a^2}} = \frac{1}{a}\operatorname{arcsec}\frac{|u|}{a} + C$

Trig Cheat Sheet

Definition of the Trig Functions

Right triangle definition
For this definition we assume that

$0 < \theta < \frac{\pi}{2}$ or $0^\circ < \theta < 90^\circ$.

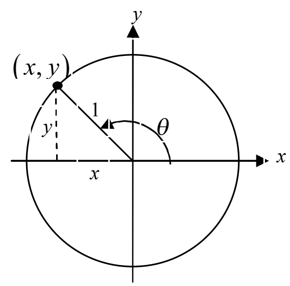


$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} \qquad \csc\theta = \frac{\text{hypotenuse}}{\text{opposite}}$$

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} \qquad \sec\theta = \frac{\text{hypotenuse}}{\text{adjacent}}$$

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} \qquad \cot\theta = \frac{\text{adjacent}}{\text{opposite}}$$

Unit circle definition
For this definition θ is any angle.



$$\sin\theta = \frac{y}{1} = y \qquad \csc\theta = \frac{1}{y}$$

$$\cos\theta = \frac{x}{1} = x \qquad \sec\theta = \frac{1}{x}$$

$$\tan\theta = \frac{y}{x} \qquad \cot\theta = \frac{x}{y}$$

Facts and Properties

Domain
The domain is all the values of θ that can be plugged into the function.

$\sin\theta$, θ can be any angle
$\cos\theta$, θ can be any angle
$\tan\theta$, $\theta \neq \left(n + \frac{1}{2}\right)\pi$, $n = 0, \pm 1, \pm 2, \ldots$
$\csc\theta$, $\theta \neq n\pi$, $n = 0, \pm 1, \pm 2, \ldots$
$\sec\theta$, $\theta \neq \left(n + \frac{1}{2}\right)\pi$, $n = 0, \pm 1, \pm 2, \ldots$
$\cot\theta$, $\theta \neq n\pi$, $n = 0, \pm 1, \pm 2, \ldots$

Range
The range is all possible values to get out of the function.

$-1 \le \sin\theta \le 1$ $\qquad$ $\csc\theta \ge 1$ and $\csc\theta \le -1$
$-1 \le \cos\theta \le 1$ $\qquad$ $\sec\theta \ge 1$ and $\sec\theta \le -1$
$-\infty < \tan\theta < \infty$ $\qquad$ $-\infty < \cot\theta < \infty$

Period
The period of a function is the number, T, such that $f(\theta + T) = f(\theta)$. So, if ω is a fixed number and θ is any angle we have the following periods.

$$\sin(\omega\theta) \rightarrow T = \frac{2\pi}{\omega}$$

$$\cos(\omega\theta) \rightarrow T = \frac{2\pi}{\omega}$$

$$\tan(\omega\theta) \rightarrow T = \frac{\pi}{\omega}$$

$$\csc(\omega\theta) \rightarrow T = \frac{2\pi}{\omega}$$

$$\sec(\omega\theta) \rightarrow T = \frac{2\pi}{\omega}$$

$$\cot(\omega\theta) \rightarrow T = \frac{\pi}{\omega}$$

Formulas and Identities

Tangent and Cotangent Identities

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \qquad \cot\theta = \frac{\cos\theta}{\sin\theta}$$

Reciprocal Identities

$$\csc\theta = \frac{1}{\sin\theta} \qquad \sin\theta = \frac{1}{\csc\theta}$$

$$\sec\theta = \frac{1}{\cos\theta} \qquad \cos\theta = \frac{1}{\sec\theta}$$

$$\cot\theta = \frac{1}{\tan\theta} \qquad \tan\theta = \frac{1}{\cot\theta}$$

Pythagorean Identities

$$\sin^2\theta + \cos^2\theta = 1$$

$$\tan^2\theta + 1 = \sec^2\theta$$

$$1 + \cot^2\theta = \csc^2\theta$$

Even/Odd Formulas

$$\sin(-\theta) = -\sin\theta \qquad \csc(-\theta) = -\csc\theta$$

$$\cos(-\theta) = \cos\theta \qquad \sec(-\theta) = \sec\theta$$

$$\tan(-\theta) = -\tan\theta \qquad \cot(-\theta) = -\cot\theta$$

Periodic Formulas
If n is an integer.

$$\sin(\theta + 2\pi n) = \sin\theta \qquad \csc(\theta + 2\pi n) = \csc\theta$$

$$\cos(\theta + 2\pi n) = \cos\theta \qquad \sec(\theta + 2\pi n) = \sec\theta$$

$$\tan(\theta + \pi n) = \tan\theta \qquad \cot(\theta + \pi n) = \cot\theta$$

Double Angle Formulas

$$\sin(2\theta) = 2\sin\theta\cos\theta$$

$$\begin{aligned}\cos(2\theta) &= \cos^2\theta - \sin^2\theta \\ &= 2\cos^2\theta - 1 \\ &= 1 - 2\sin^2\theta\end{aligned}$$

$$\tan(2\theta) = \frac{2\tan\theta}{1 - \tan^2\theta}$$

Degrees to Radians Formulas
If x is an angle in degrees and t is an angle in radians then

$$\frac{\pi}{180} = \frac{t}{x} \quad \Rightarrow \quad t = \frac{\pi x}{180} \quad \text{and} \quad x = \frac{180t}{\pi}$$

Half Angle Formulas

$$\sin^2\theta = \frac{1}{2}\left(1 - \cos(2\theta)\right)$$

$$\cos^2\theta = \frac{1}{2}\left(1 + \cos(2\theta)\right)$$

$$\tan^2\theta = \frac{1 - \cos(2\theta)}{1 + \cos(2\theta)}$$

Sum and Difference Formulas

$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$$

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$$

$$\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$$

Product to Sum Formulas

$$\sin\alpha\sin\beta = \frac{1}{2}\left[\cos(\alpha - \beta) - \cos(\alpha + \beta)\right]$$

$$\cos\alpha\cos\beta = \frac{1}{2}\left[\cos(\alpha - \beta) + \cos(\alpha + \beta)\right]$$

$$\sin\alpha\cos\beta = \frac{1}{2}\left[\sin(\alpha + \beta) + \sin(\alpha - \beta)\right]$$

$$\cos\alpha\sin\beta = \frac{1}{2}\left[\sin(\alpha + \beta) - \sin(\alpha - \beta)\right]$$

Sum to Product Formulas

$$\sin\alpha + \sin\beta = 2\sin\left(\frac{\alpha + \beta}{2}\right)\cos\left(\frac{\alpha - \beta}{2}\right)$$

$$\sin\alpha - \sin\beta = 2\cos\left(\frac{\alpha + \beta}{2}\right)\sin\left(\frac{\alpha - \beta}{2}\right)$$

$$\cos\alpha + \cos\beta = 2\cos\left(\frac{\alpha + \beta}{2}\right)\cos\left(\frac{\alpha - \beta}{2}\right)$$

$$\cos\alpha - \cos\beta = -2\sin\left(\frac{\alpha + \beta}{2}\right)\sin\left(\frac{\alpha - \beta}{2}\right)$$

Cofunction Formulas

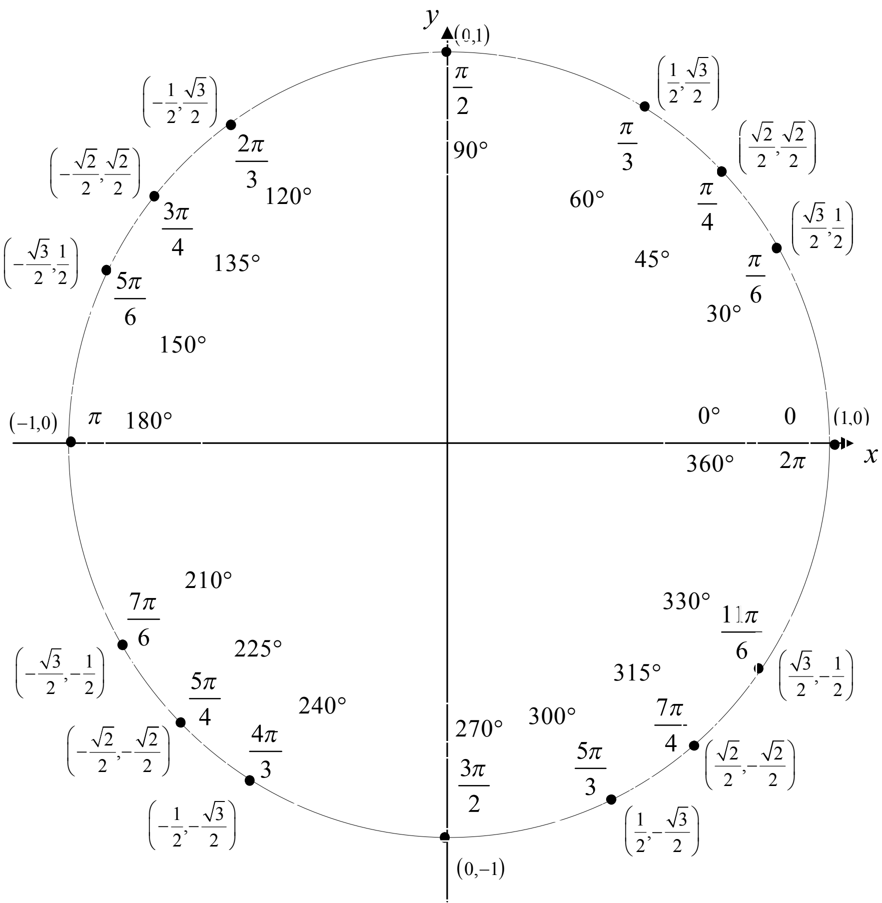
$$\sin\left(\frac{\pi}{2} - \theta\right) = \cos\theta \qquad \cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta$$

$$\csc\left(\frac{\pi}{2} - \theta\right) = \sec\theta \qquad \sec\left(\frac{\pi}{2} - \theta\right) = \csc\theta$$

$$\tan\left(\frac{\pi}{2} - \theta\right) = \cot\theta \qquad \cot\left(\frac{\pi}{2} - \theta\right) = \tan\theta$$

© 2005 Paul Dawkins

Unit Circle



For any ordered pair on the unit circle (x, y) : $\cos\theta = x$ and $\sin\theta = y$

Example

$$\cos\left(\frac{5\pi}{3}\right) = \frac{1}{2} \qquad \sin\left(\frac{5\pi}{3}\right) = -\frac{\sqrt{3}}{2}$$

Inverse Trig Functions

Definition

$y = \sin^{-1} x$ is equivalent to $x = \sin y$

$y = \cos^{-1} x$ is equivalent to $x = \cos y$

$y = \tan^{-1} x$ is equivalent to $x = \tan y$

Domain and Range

Function	Domain	Range
$y = \sin^{-1} x$	$-1 \le x \le 1$	$-\frac{\pi}{2} \le y \le \frac{\pi}{2}$
$y = \cos^{-1} x$	$-1 \le x \le 1$	$0 \le y \le \pi$
$y = \tan^{-1} x$	$-\infty < x < \infty$	$-\frac{\pi}{2} < y < \frac{\pi}{2}$

Inverse Properties

$\cos\left(\cos^{-1}(x)\right) = x \qquad \cos^{-1}\left(\cos(\theta)\right) = \theta$

$\sin\left(\sin^{-1}(x)\right) = x \qquad \sin^{-1}\left(\sin(\theta)\right) = \theta$

$\tan\left(\tan^{-1}(x)\right) = x \qquad \tan^{-1}\left(\tan(\theta)\right) = \theta$

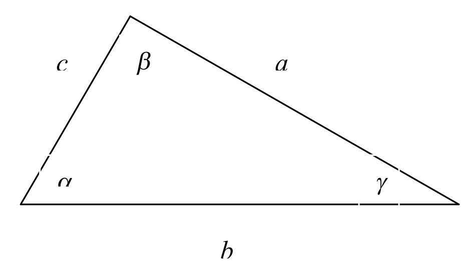
Alternate Notation

$\sin^{-1} x = \arcsin x$

$\cos^{-1} x = \arccos x$

$\tan^{-1} x = \arctan x$

Law of Sines, Cosines and Tangents



Law of Sines

$$\frac{\sin\alpha}{a} = \frac{\sin\beta}{b} = \frac{\sin\gamma}{c}$$

Law of Cosines

$$a^2 = b^2 + c^2 - 2bc\cos\alpha$$

$$b^2 = a^2 + c^2 - 2ac\cos\beta$$

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

Mollweide's Formula

$$\frac{a+b}{c} = \frac{\cos\frac{1}{2}(\alpha-\beta)}{\sin\frac{1}{2}\gamma}$$

Law of Tangents

$$\frac{a-b}{a+b} = \frac{\tan\frac{1}{2}(\alpha-\beta)}{\tan\frac{1}{2}(\alpha+\beta)}$$

$$\frac{b-c}{b+c} = \frac{\tan\frac{1}{2}(\beta-\gamma)}{\tan\frac{1}{2}(\beta+\gamma)}$$

$$\frac{a-c}{a+c} = \frac{\tan\frac{1}{2}(\alpha-\gamma)}{\tan\frac{1}{2}(\alpha+\gamma)}$$

© 2005 Paul Dawkins