Dit onderwerp is de laatste dekaden sterk ontwikkeld onder invloed van de theoretische informatica. In de natuur komen veel structuren voor die samengesteld zijn uit eindig veel basisstructuren, denk aan chemische verbindingen, kristallen, DNA-structuren. Ook (computer)talen vallen daaronder. In de informatica worden zulke structuren gesimuleerd en hun eigenschappen bestudeerd. Enkele belangrijke begrippen zullen we hier introduceren. Hoewel er ook veel theorie is over meerdimensionale structuren, bijvoorbeeld quasi-kristallen, zullen we ons hier beperken tot dimensie 1.

1 Woorden

De basisstructuren duiden we aan met letters (of cijfers) (waarbij gelijke letters gelijke basisstructuren representeren). De eindige verzameling van basisstructuren noemen we het *alfabet* $\mathcal{A}$. Voorbeelden van alfabetten zijn $\{0,1\}$ en $\{a,b,c,d,\}$. De structuren worden nu rijtjes letters, zogenaamde *woorden*. We zullen ons hoofdzakelijk beperken tot drie soorten woorden $w : I \to \mathcal{A}$:

eindige woorden: eindige rijtjes van letters
eenzijdige woorden: afbeeldingen van de natuurlijke getallen naar het alphabet
tweezijdige woorden: afbeeldingen van de gehele getallen naar het alphabet.

Een verzameling van woorden heet een *taal*. Ook het *lege woord* bestaande uit nul letters en gerepresenteerd door ϵ kan deel uitmaken van een taal; het is wat anders dan de lege verzameling. De verzameling van alle eindige woorden met letters uit $\mathcal{A}$ wordt genoteerd met $\mathcal{A}^*$. We kunnen er een monoïde van maken door als bewerking de *concatenatie* te nemen, d.w.z. de woorden achter elkaar te schrijven. De concatenatie van $V = v_1 \dots v_r$ en $W = w_1 \dots w_s$ is $VW = v_1 \dots v_r w_1 \dots w_s$. De operatie is associatief en heeft ϵ als eenheidselement. We noteren $\mathcal{A}^* \setminus \epsilon$ met $\mathcal{A}^+$.

Als $W = w_1 \dots w_s$ een woord is, dan heet s de lengte van W, notatie $|W|$. Het aantal letters a in W wordt aangeduid met $|W|_a$. Dus $|W| = \sum_{a \in \mathcal{A}} |W|_a$. Het woord $V = v_1 \dots v_r$ komt voor in het woord $u = (u_n)_{n \in I}$ als er een getal m bestaat zó dat $u_m = v_1, u_{m+1} = v_2, \dots, u_{m+r-1} = v_r$. Als V in u voorkomt, heet V een *deelwoord* van u. De verzameling van alle deelwoorden van u van lengte n wordt genoteerd met $\mathcal{L}_n(u)$. De verzameling van alle deelwoorden van u heet de taal $\mathcal{L}(u)$ van u. Dus $\mathcal{L}(u) \subseteq \mathcal{A}^*$.

Een oneindig woord u heet *recurrent* als elk eindig deelwoord van u oneindig vaak voorkomt. Het heet *uniform recurrent* als voor elk deelwoord W de gaten tussen opeenvolgende verschijningen van W begrensd zijn.

2 Substituties

Een *substitutie* is een afbeelding $\sigma : \mathcal{A} \to \mathcal{A}^+$. Een substitutie wordt een morfisme op $\mathcal{A}$ door concatenatie, $\sigma(WW') = \sigma(W)\sigma(W')$ en $\sigma(\epsilon) = \epsilon$. Als $|\sigma(a)|$ voor elke $a \in \mathcal{A}$ dezelfde waarde k heeft, heet σ *van constante lengte* k. Een *invariant woord* van een substitutie σ is een oneindig woord u met $\sigma(u) = u$. Als $\sigma(a)$ met a begint en tenminste twee letters telt, is er een uniek invariant woord dat begint met a, namelijk $\lim_{n\to\infty} \sigma^n(a) \in \mathcal{A}^{\mathbb{N}}$.

Voorbeelden. Zij $\mathcal{A} = \{a, b\}$.

1. *De Morsesubstitutie* gegeven door $\sigma(a) = ab, \sigma(b) = ba$ heeft als invariante woorden $abbabaabbaababba\ldots$ en $baababbaabbabaab\ldots$. Dit heten *Morsewoorden.*

2. *De Fibonaccisubstitutie* gegeven door $\sigma(a) = ab, \sigma(b) = a$ heeft als enig invariant woord $abaababaabaababaab\ldots$. Dit heet het *Fibonacciwoord.*

3. *De Cantorsubstitutie* gegeven door $\sigma(a) = aba, \sigma(b) = bbb$ heeft als invariante woorden $ababbbababbbbbbbbbaba\ldots$ en $bbbbbbbbbbbbbbb\ldots$. Het eerste woord heet het *Cantorwoord.*

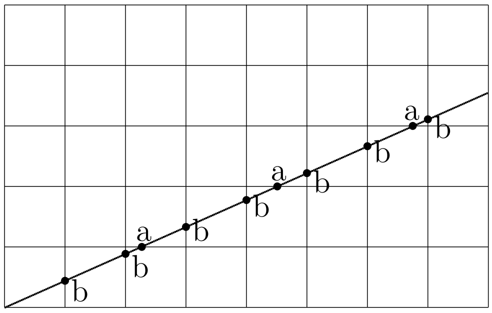
Zij $\mathcal{A} = \{a_1, \ldots, a_k\}$. Aan een substitutie σ wordt een *incidentiematrix* M_σ toegevoegd waarbij op plaats (i, j) de waarde $|\sigma(a_j)|_{a_i}$ staat. Een substitutie heet *unimodulair* als de determinant van de incidentiematrix gelijk is aan 1 of -1. Een substitutie heet *primitief* als er een k bestaat zó dat voor elk paar letters $a, b \in \mathcal{A}$ geldt $a \in \sigma^k(b)$.

Stelling 12.1. *Een invariant woord van een primitieve substitutie is uniform recurrent.*

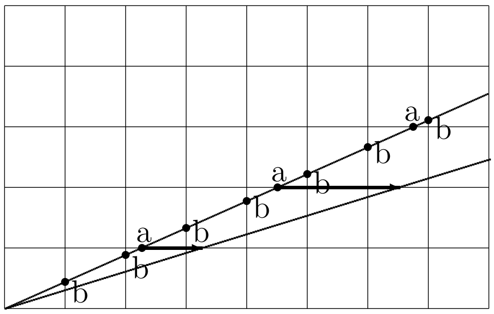
Bewijs. Zij $u = \sigma(u)$ een invariant woord van σ. Stel σ is k-primitief. Er geldt $u = \sigma^k(u) = \sigma^k(u_0)\sigma^k(u_1)\sigma^k(u_2)\ldots$. Zij $M = \max_{a\in\mathcal{A}}|\sigma(a)|$. Dan komt elke letter a in u voor met gaten begrensd door $2M^k$. Zij V een deelwoord van u. Dan bestaat een n zó dat V voorkomt in $\sigma^n(u_0)$. Voor elke $a \in \mathcal{A}$ komt V dan ook voor in $\sigma^n(\sigma^k(a))$. Dus komt V in u voor met gaten begrensd daar $2M^{k+n}$. $\square$

3 Snijwoorden

Een van de manieren om oneindige woorden te genereren is de volgende. Trek een halflijn in het x-y-vlak vanuit de oorsprong. Waar de halflijn een gehele x-waarde passeert noteren we een b, waar de halflijn een gehele y-waarde een a. In geval het tegelijk gebeurt noteren we consequent ab (of consequent ba). Schrijf vervolgens de letters in volgorde op. (Zie Figuur 1 waarin de gegenereerde rij gelijk is aan *bbabbabbab*....) Sommige substituties voeren de halflijn over in een andere halflijn. Zo voert de substitutie $a \to b, b \to a$ de halflijn over in de om de lijn $y = x$ gespiegelde halflijn $y = \alpha^{-1}x$ en de substitutie $a \to ba, b \to b$ de lijn $y = \alpha x$ over in $y = \frac{\alpha}{\alpha+1}x$. (Kijk naar Figuur 2.)



Figuur 1



Figuur 2

Beschouw nu de lijn $y = \alpha x$ met $\alpha = \frac{1}{2} + \frac{1}{2}\sqrt{5}$. Passen we eerst de substitutie $a \to ba, b \to b$ toe en vervolgens de substitutie $a \to b, b \to a$, dan is de substitutie $a \to ab, b \to a$ het resultaat. De halflijn $y = \alpha x$ gaat hierdoor over in de halflijn $y = \frac{\alpha+1}{\alpha}x$. Door de keuze van α is de coëfficiënt echter gelijk aan α en we zijn dus op de oorspronkelijke lijn terug. Dit betekent dat het door de halflijn gegenereerde oneindige woord invariant blijft onder de Fibonaccisubstitutie. Er is echter maar één woord dat invariant is onder deze substitutie, namelijk het Fibonacciwoord. Zo bewijs je dat het Fibonacciwoord het snijwoord is voor de halflijn $y = (\frac{1}{2} + \frac{1}{2}\sqrt{5})x$.

4 Rotatiewoorden

Het woord u is een *rotatiewoord* als er een getal $\alpha \in [0,1]$ en een reëel getal β bestaat zó dat

$$u_n = \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil \quad \text{voor alle } n \in \mathbb{N} \tag{1}$$

$$\text{of} \quad u_n = \lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor \quad \text{voor alle } n \in \mathbb{N}. \tag{2}$$

Dus u_n kan alleen de waarden 0 en 1 aannemen. We noemen het woord *irrationaal* als α irrationaal is. Het getal α heet de *hoek*, β het *beginpunt*. Zonder verlies van algemeenheid kunnen we $\beta \in [0,1)$ kiezen.

In geval (1) hebben we

$$\begin{aligned} u_n = 1 &\iff n\alpha + \beta + \alpha > \lfloor n\alpha + \beta \rfloor + 1 \\ &\iff n\alpha + \beta - \lfloor n\alpha + \beta \rfloor > 1 - \alpha \\ &\iff n\alpha + \beta \pmod 1 > 1 - \alpha \end{aligned}$$

en in geval (2)

$$u_n = 1 \iff n\alpha + \beta \pmod 1 \geq 1 - \alpha.$$

Dus u_n is bepaald door $n\alpha + \beta \pmod 1$. We beelden $[0,1)$ af op de eenheidscirkel door $x \mapsto \mathrm{e}^{2i\pi x}$. Een rotatiewoord ontstaat uit β door de functie $f : x \mapsto \alpha + x$ (mod 1), waarbij $u_n =$

$$\begin{cases} 0 & \text{als } f^n(\beta) \in I_0 \\ 1 & \text{als } f^n(\beta) \in I_1 \end{cases}$$

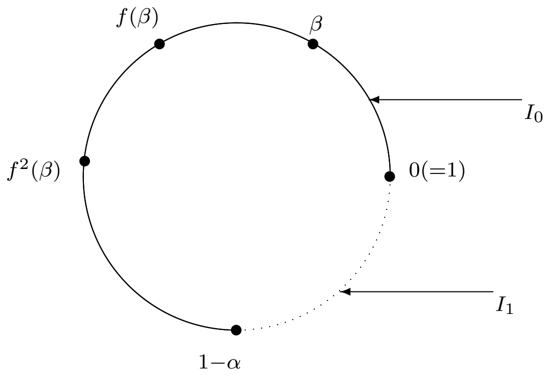
In geval (1) hebben we

$$\begin{cases} I_0 = (0, 1-\alpha] \\ I_1 = (1-\alpha, 1] \end{cases}$$

en in geval (2)

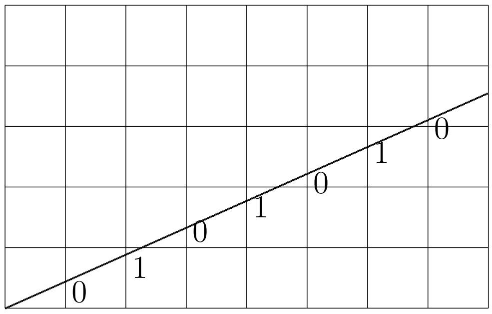
$$\begin{cases} I_0 = [0, 1-\alpha) \\ I_1 = [1-\alpha, 1). \end{cases}$$

De keuze is dus alleen relevant als $n\alpha + \beta$ een geheel getal is. Als α irrationaal is en $\beta = 0$, komt dit helemaal niet voor.



Figuur 3. De meetkundige interpretatie van een rotatiewoord.

Kijk naar Figuur 1. Elke a wordt voorafgegaan door een b. Vervang elke b direct gevolgd door een a door een 1 en elke andere b door een 0. Dit is duidelijk een omkeerbare transformatie.



Figuur 4. Het rotatiewoord voor de lijn $y = \alpha x + \beta$.

De letters corresponderen nu met de verticale lijnen. Het nieuwe woord heet het rotatiewoord met hoek α. We hebben 0 als $0 < n\alpha + \beta - \llcorner n\alpha + \beta \lrcorner \leq 1 - \alpha$ en anders 1, of 0 als $n\alpha + \beta - \llcorner n\alpha + \beta \lrcorner < 1 - \alpha$ en anders 1. Dus hebben we een rotatiewoord met hoek α gelijk aan de hoek van de lijn van het snijwoord. Dit levert een bijectie tussen snijwoorden en rotatiewoorden.

5 Sturmse woorden

Zij $u = (u_n)_{n=1}^{\infty}$ een oneindig woord over het alfabet $\mathcal{A}$. De *complexiteit* van u is de functie $p_u : \mathbb{N} \to \mathbb{N}$ met $p_u(n) = |\mathcal{L}_n(u)|$ voor $n = 1, 2, \ldots$. Dus $1 \leq p_u(n) \leq |\mathcal{A}|^n$ voor elke n.

Het woord u heet *periodiek* met periode T als $u_n = u_{n+T}$ voor elke n en *uiteindelijk periodiek* als $u_n = u_{n+T}$ voor elke $n > n_0$ voor zekere n_0. Merk op dat p_u begrensd is als u periodiek is. Het omgekeerde is ook waar. Er geldt nog iets veel sterkers:

Stelling 12.2. (Coven en Hedlund, 1973)
Als $p_u(n) \leq n$ voor zekere n, dan is u uiteindelijk periodiek.

Bewijs. Als $p_u(1) = 1$, dan is u constant. Als $p_u(1) > 1$ en $p_u(n) \leq n$, dan bestaat een kleinste k met $1 \leq k < n$ zó dat $p_u(k+1) = p_u(k)$. Omdat elke eindig deelwoord van u aan de rechterkant tot een langer deelwoord kan worden uitgebreid, is er een bijectie van deelwoorden van lengte k en deelwoorden van lengte $k+1$. Als $u_i \ldots u_{i+k-1} = u_j \ldots u_{j+k-1}$ dan volgt dat $u_{i+k} = u_{j+k}, u_{i+k+1} = u_{j+k+1}, \ldots$. Omdat $\mathcal{L}_k(u)$ eindig is, bestaan er i en j zó dat $u_i \ldots u_{i+k-1} = u_j \ldots u_{j+k-1}$. Dus geldt dan $u_{i+p} = u_{j+p}$ voor alle natuurlijke getallen p en heeft u uiteindelijk periode $j-i$. $\square$

Uit Stelling 12.2 volgt dat als u niet uiteindelijk periodiek is, er voor elke n geldt dat $p_u(n) \geq n+1$. Een woord u heet *sturms* als voor elke n geldt dat $p_u(n) = n+1$. Een sturms woord is dus "het meest regelmatige niet-periodieke woord".

Stelling 12.3. *Elke sturms woord is recurrent.*

Bewijs. Stel u is een sturms woord dat een deelwoord W van lengte n heeft van eindige lengte. Dan komt het deelwoord vanaf een zekere positie N niet meer voor. Definiëren we het woord v door van u de eerste N symbolen weg te laten, dan geldt $p_v(n) < p_u(n) \leq n+1$. Dus $p_v(n) \leq n$. Uit Stelling 12.2 volgt nu dat v en dus ook u uiteindelijk periodiek is. $\square$

Stelling 12.4.
a) *De taal van een rotatiewoord hangt alleen af van zijn hoek.*
b) *Elk irrationaal rotatiewoord is sturms.*
c) *Elk rationaal rotatiewoord is uiteindelijk periodiek.*

Bewijs. Kijk naar Figuur 3. Zij u een rotatiewoord met hoek α en beginpunt β. Stel we willen alle deelwoorden van lengte n kennen. Dan gaat het om het aantal verschillende woorden $u_m \dots u_{m+n-1}$ als m de positieve gehele getallen doorloopt. Voor de waarde van u_{m+k} is het essentieel te weten waar $(m+k)\alpha + \beta \pmod 1$ ligt ten opzichte van 0 en $1-\alpha$, d.w.z. $m\alpha + \beta \pmod 1$ ten opzichte van $-k\alpha \pmod 1$ en $-(k+1)\alpha \pmod 1$. Voor de verzameling deelwoorden zijn dus alleen de scheidingspunten $-k\alpha \pmod 1$ voor $k = 0, 1, \dots n$ relevant. Die scheidingspunten bepalen ten hoogste $n+1$ intervallen. Als α irrationaal is, zijn er precies $n+1$ scheidingspunten; als α rationaal is, zeg met noemer q, dan is het aantal scheidingspunten q als $n \geq q$. Als bekend is in welk interval β ligt, dan is het deelwoord uniek bepaald en verschillende intervallen leveren verschillende deelwoorden. De taal hangt dus alleen van α af. Als α irrationaal is, geldt voor elke n dat $p_u(n) = n+1$ en is u dus sturms. Als α rationaal is, geldt dat $p_u(n)$ begrensd is en is u volgens Stelling 12.2 uiteindelijk periodiek(, in feite zelfs zuiver periodiek). $\square$

6 Gebalanceerde woorden

Er is nog een andere manier om bijna-periodiciteit te meten, namelijk met gebalanceerdheid. Stel $\mathcal{A} = \{0, 1\}$. Een woord heet *gebalanceerd* als het aantal enen in elk paar deelwoor- den van gelijke lengte ten hoogste 1 verschilt. De definitie houdt in dat als u geba- lanceerd is, er voor het aantal enen in een deelwoord van lengte n maar twee opeenvolgende waarden mogelijk zijn.

Stelling 12.5. *Elk rotatiewoord is gebalanceerd.*

Bewijs. Zij $u = u_1u_2\ldots$ een woord met $u_n = \llcorner(n+1)\alpha+\beta\lrcorner - \llcorner n\alpha+\beta\lrcorner$ voor $n \in \mathbb{N}$. Dan is

$$|u_m u_{m+2} \ldots u_{m+n-1}|_1 = \llcorner(m+n)\alpha+\beta\lrcorner - \llcorner m\alpha+\beta\lrcorner =$$

$$((m+n)\alpha+\beta-\theta_1) - (m\alpha+\beta-\theta_2) = n\alpha+\theta_2-\theta_1$$

waarbij $0 \leq \theta_1 < 1$ en $0 \leq \theta_2 < 1$. Het aantal enen in een deelwoord van u van lengte n is dus $\llcorner n\alpha\lrcorner$ of $\ulcorner n\alpha\urcorner$. Dat kan dus hoogstens 1 uiteenlopen. Dus is u gebalanceerd. □

Gevolg. *Voor het aantal enen in een deelwoord V van lengte n van een rotatiewoord u met hoek α geldt $|\ |V|_1 - n\alpha\ | < 1$.*

Een iets zwakkere ongelijkheid geldt voor ieder gebalanceerd woord:

Stelling 12.6. *Stel $u = u_1u_2\ldots$ is een gebalanceerd woord over het alfabet $\{0,1\}$. Dan bestaat een getal $\alpha \in [0,1]$ zó dat voor elk deelwoord V van lengte n geldt dat $|\ |V|_1 - n\alpha\ | \leq 1$.*

Bewijs. Merk eerst op dat als de minimale waarde van $|V|_1$ met V van lengte n gelijk is aan a_n, de enige andere waarde $a_n + 1$ kan zijn.
We zullen bewijzen dat de rij $(a_n)_{n=1}^{\infty}$ een cauchyrij is. Zij $\epsilon > 0$. Stel n_0 is een geheel getal zó dat $n_0 \geq \frac{1}{\epsilon}$. Voor $m, n \geq n_0$ zij U een willekeurig deelwoord van u van lengte mn. Dan kan U in n woorden van lengte m opgesplitst worden, maar ook in m woorden van lengte n. Dus

$$ma_n \leq |U|_1 \leq m(a_n+1) \quad \text{en} \quad na_m \leq |U|_1 \leq n(a_m+1).$$

Na deling door mn krijgen we

$$\frac{a_n}{n} \leq \frac{|U|_1}{mn} \leq \frac{a_n+1}{n} \quad \text{en} \quad \frac{a_m}{m} \leq \frac{|U|_1}{mn} \leq \frac{a_m+1}{m}.$$

Dus $\frac{a_n}{n} - \frac{a_m}{m} \leq \frac{1}{m} \leq \epsilon$ en $\frac{a_m}{m} - \frac{a_n}{n} \leq \frac{1}{n} \leq \epsilon$. Hieruit volgt dat $(\frac{a_n}{n})$ een cauchyrij is. De rij heeft dus een limiet α.
Uit het voorgaande volgt dat voor elke n geldt dat $\frac{a_n}{n} \leq \alpha \leq \frac{a_n+1}{n}$, d.w.z. $a_n \leq n\alpha \leq a_n + 1$. Dus is $|\ |V|_1 - n\alpha\ | \leq 1$ voor elk deelwoord V van lengte n, waarbij het gelijkteken alleen kan voorkomen als α rationaal is. □

We willen aantonen dat als α irrationaal is, de gebalanceerde u in feite sturms is. Daartoe bewijzen we eerst het volgende lemma.

Lemma 12.1. *Stel $\mathcal{A} = \{0,1\}$. Als het woord u niet gebalanceerd is, dan bestaat een woord W zó dat zowel $0W0$ als $1W1$ deelwoorden van u zijn.*

Bewijs. Stel het woord u is niet gebalanceerd. Kies twee deelwoorden U en V van u van minimale lengte $n+1$ zó dat $|U|_1 - |V|_1 > 1$. Vanwege de minimaliteit van n

is de beginletter van U een 1 en de beginletter van V een 0. Dat geldt evenzo voor de eindletters. Dus $U = 1W1$ en $V = 0W'0$. Het is voldoende nog te bewijzen dat $W = W'$. Stel W en W' zijn verschillend. Kies de kleinste k waarvoor de k-de letter verschilt. Dan geldt

$$1 < |U|_1 - |V|_1 = (|1w_1w_2 \dots w_k|_1 - |0w_1w_2 \dots w'_k|_1) + (|U_{k+1} \dots 1|_1 - |V_{k+1} \dots 0|_1),$$

waarbij elke uitdrukking tussen haakjes vanwege de minimaliteit van n ten hoogste 1 is. Dus zijn beide uitdrukkingen tussen haakjes precies gelijk aan 1. Dat houdt echter in dat $w_k = w'_k$, een tegenspraak. Dus $W = W'$ en $0W0$ en $1W1$ komen beide in u voor. □

Stelling 12.7. *Zij u een gebalanceerd woord over twee letters dat niet uiteindelijk periodiek is. Dan is u sturms.*

Bewijs. Uit het ongerijmde. Stel u is een woord over $\{0, 1\}$ dat niet sturms en niet uiteindelijk periodiek is. Dan bestaat een kleinste getal n zó dat $p_u(n + 1) \geq n + 3$. Omdat $p_u = 2$ geldt $n \geq 1$. Omdat $p_u(n) = n + 1$, zijn er twee deelwoorden V en W van u van lengte n zó dat ook $V0, V1, W0, W1$ deelwoorden van u zijn. Laat de eerste letter van V en W weg. Dan krijgen we vier deelwoorden van u van lengte n voortkomend uit twee deelwoorden van lengte $n - 1$. Omdat $p_u(n) = p_u(n - 1) + 1$, is er maar één deelwoord van lengte $n - 1$ dat twee verschillende uitbreidingen met één letter naar rechts toestaat. Dus V en W verschillen alleen in de eerste letter. Dat betekent dat er een woord U van lengte $n - 1$ is zó dat $0U0, 0U1, 1U0$ en $1U1$ deelwoorden van u zijn. Hieruit volgt dat u niet gebalanceerd is. □

Vervolgens gaan we bewijzen dat elk sturms woord gebalanceerd is. We noemen een eindig woord $W = w_0w_1 \dots w_n$ een *palindroom* als $w_k = w_{n-k}$ voor $k = 0, 1, \dots, n$.

Stelling 12.8. *Elk sturms woord is gebalanceerd.*

Bewijs. Beschouw een sturms woord $u = u_0u_1 \dots$. Stel u bevat beide deelwoorden 00 en 11. Dan komt vanwege $p_u(2) = 3$ hetzij 01 hetzij 10 niet als deelwoord voor. Als 00 vóór 11 voorkomt, dan komt 01 voor en dus 10 niet. Maar dan volgen er na 11 alleen nog enen, in tegenspraak met Stelling 12.3. Als 11 vóór 00 voorkomt, dan is er een vergelijkbaar probleem. In een sturms woord komen dus niet beide deelwoorden 00 en 11 voor.

We bewijzen uit het ongerijmde dat u gebalanceerd is. Stel u is niet gebalanceerd. Dan volgt uit Lemma 12.1 dat er een deelwoord W is zo dat $0W0$ en $1W1$ in u voorkomen. We weten dat W niet leeg is. We beschouwen zo'n woord $W = w_0...w_n$ van minimale lengte $n + 1$.

- W is een palindroom.

Omdat 00 of 11 niet voorkomt, geldt $w_0 = w_n$. Zij k het kleinste natuurlijke getal met $w_k \neq w_{n-k}$, zeg $w_k = 0$ and $w_{n-k} = 1$. Dan komen $A := 0w_0...w_{k-1}0$ en

$B := 1w_{n-k+1}...w_n1$ beide in u voor. u. Er bestaat dus een W' zo dat $0W'0$ in A en $1W'1$ in B voorkomt. Dus zowel $0W'0$ als $1W'1$ komen voor in u. Uit deze tegenspraak volgt dat W een palindroom is.

• Precies drie van de vier woorden $0W0, 0W1, 1W0, 1W1$ zijn deelwoorden van u. Omdat $p_u(n+1) = n+2$ bestaan er woorden $A_1, ..., A_{n+1}$ van lengte $n+1$ en verschillend van W zó dat $\mathcal{L}_{n+1}(u) = \{A_1, ..., A_{n+1}, W\}$. Elk woord van lengte $n+2$ kan worden verkregen door een linkeruitbreiding van een woord uit of $\mathcal{L}_{n+1}(u)$. Dus bestaan $a_0, ..., a_{n+1} \in \{0,1\}$ zo dat $\mathcal{L}_{n+2}(u) = \{a_1A_1, ..., a_{n+1}A_{n+1}, 0W, 1W\}$, omdat $p_u(n+2) = p_u(n+1)+1$ en $0W$ en $1W$ in u voorkomen. Verder kan $\mathcal{L}_{n+3}(u)$ verkregen worden door elk woord uit $\mathcal{L}_{n+2}(u)$ een letter naar rechts uit te breiden. Omdat u sturms is, is er precies één woord van lengte $n+2$ dat rechts op twee manieren kan worden uitgebreid. Stel dit is het woord a_iA_i. Dan heeft ook A_i twee rechteruitbreidingen in u. Maar W is het woord van lengte $n+1$ met twee rechteruitbreidingen. Dus óf $0W$ óf $1W$ kan op twee manieren rechts worden uitgebreid tot een deelwoord van u. Vanwege de symmetrie mogen we veronderstellen dat het het woord $1W$ is. Dan komt $1W0$ wel, maar $0W1$ niet voor.

• $W0$ kan niet voorkomen vóór $W1$.
Stel $W0$ komt in u voor links van $W1$. Zij $W0 = u_iu_{i+1}\dots u_{i+n}0$ en $W1 = u_ju_{j+1}\dots u_{j+n}1$ met $i < j$ zó gekozen dat $j-i$ minimaal is. Dan is $u_mu_{m+1}\dots u_{m+n} \neq W$ voor $i < m < j$ en hebben deze woorden dus een unieke uitbreiding naar rechts. Dus $(u_mu_{m+1}\dots u_{m+n})_{m=i}^{j-1}$ zijn onderling verschillend waaruit wegens $P(n+1) = n+2$ volgt dat $j - i \leq n+2$. Omdat $0W1$ niet in u voorkomt, geldt $u_{j-1} = 1$ en dus $j-1 \neq i+n+1$. Uit de ongelijkheid $j \leq i+n+1$ volgt dat de deelwoorden $(W0 =)u_iu_{i+1}\dots u_{i+n}0$ en $(W1 =)u_ju_{j+1}\dots u_{j+n}1$ overlappen. Omdat W een palindroom is, volgt hieruit dat

$$1 = u_{j-1} = u_{i+(j-i-1)} = W_{j-i-1} = W_{n-j+i+1} = u_{n+i+1} = 0,$$

hetgeen een tegenspraak oplevert. Dus $W1$ komt niet in u voor nadat $W0$ voorgekomen is.

• Afronding van het bewijs.
We weten dat $W0$ ergens in u voorkomt en dat daarna $W1$ niet meer voorkomt. Dit betekent dat $W1$ maar eindig vaak in u voorkomt. Dan is u niet recurrent. Dit is in tegenspraak met Stelling 12.3. □

Stel $u = u_1u_2\dots$ is een oneindig woord op een alphabet $\mathcal{A}$. Voor $a \in \mathcal{A}$ definiëren we dan de *frequentie* of *dichtheid* van a als $\lim_{n\to\infty} \frac{|u_1u_2\dots u_n|_a}{n}$ mits de limiet bestaat. Uit het Gevolg van Stelling 12.5 volgt dat in een rotatiewoord op de letters 0 en 1 met hoek α de dichtheid van de 1 gelijk is aan α en dus de dichtheid van de 0 gelijk is aan $1-\alpha$. Merk op dat de dichtheid in dit geval gelijk is aan $\lim_{n\to\infty} n^{-1}\sum_{m=1}^{n} u_m$. In het geval dat we met een woord op de letters 0 en 1 te maken hebben, noemen we de dichtheid van 1 de *dichtheid van de rij*. Volgens Stelling 12.6 heeft elk gebalanceerd woord een dichtheid en volgens Stelling 12.8 dus ook elk sturms woord.

Stelling 12.9 *Elk sturms woord heeft een irrationale dichtheid.*

Bewijs. Stel u is sturms en heeft dichtheid p/q met p en q gehele getallen met $q > 1$. Dan is

$$na_q \leq\mid u_1...u_{qn} \mid_1 \leq n(a_q + 1).$$

Deel door qn en neem de limiet voor $n \to \infty$:

$$\frac{a_q}{q} \leq \frac{p}{q} \leq \frac{a_q + 1}{q}.$$

Dus $a_q = p$ of $p - 1$. Stel $a_q = p$. Stel er bestaat een deelwoord van u van lengte q met $p + 1$ enen. Dan zijn er vanwege Stelling 12.3 oneindig veel van. Dus kunnen we twee deelwoorden vinden op plaatsen die congruent modulo q zijn. Dus bestaat voor zekere h een deelwoord V van lengte hq zo dat $|V|_1 \geq hp + 2$. Omdat u gebalanceerd is, heeft elk woord ter lengte hq tenminste $hp + 1$ enen. Hieruit volgt dat de dichtheid $\geq \frac{hp+1}{hq} > \frac{p}{q}$ is. Dit is uitgesloten. Als $a_q = p - 1$, dan zijn er vanwege het recurrent zijn oneindig veel deelwoorden van lengte q met $p - 1$ enen en krijgen we op soortgelijke wijze een tegenspraak. □

We gaan tenslotte bewijzen dat elk sturms woord een rotatiewoord is. Stel $u = u_1, u_2 \ldots$ is een oneindig woord op het alfabet $\mathcal{A} = \{0, 1\}$. Dan kunnen we het woord ook vastleggen door de rij (s_i) te noteren waarbij i de positie is van de i-de 1 in u. De volgende resultaten tonen aan dat de rij (s_i) voor een sturms woord heel speciaal is.

Lemma 12.2. *Zij γ een reëel getal en u een sturms woord met dichtheid α. Dan geldt:*

$$\begin{array}{lll} & s_i \leq \llcorner \frac{i}{\alpha} + \gamma \lrcorner & \text{voor elke } i \in \mathbb{N} \\ \text{of} & s_i \geq \llcorner \frac{i}{\alpha} + \gamma \lrcorner & \text{voor elke } i \in \mathbb{N}. \end{array}$$

Bewijs. Stel Lemma 12.2 is onjuist. Dan bestaan $\alpha \in (0, 1)$, $\gamma \in \mathbb{R}$ en $i, j \in \mathbb{N}$ zo dat

$$s_i < \llcorner \frac{i}{\alpha} + \gamma \lrcorner \quad \text{en} \quad s_j > \llcorner \frac{j}{\alpha} + \gamma \lrcorner. \tag{3}$$

Veronderstel eerst dat $i < j$. Dan volgt

$$\llcorner \frac{j}{\alpha} + \gamma \lrcorner - \llcorner \frac{i}{\alpha} + \gamma \lrcorner > \left(\frac{j}{\alpha} + \gamma - 1\right) - \left(\frac{i}{\alpha} + \gamma\right) = \frac{j - i}{\alpha} - 1.$$

Wegens (3) geldt

$$s_j - s_i \geq \llcorner \frac{j}{\alpha} + \gamma \lrcorner - \llcorner \frac{i}{\alpha} + \gamma \lrcorner + 2 > \frac{j - i}{\alpha} + 1. \tag{4}$$

We tellen het aantal enen tussen s_i en s_j op twee verschillende manieren. Enerzijds is dit, vanwege de definitie van de rij s gelijk aan $j - i - 1$. Anderzijds weten we uit Stelling 12.9 dat u een irrationale dichtheid α heeft. Volgens Stelling 12.8 is u gebalanceerd en daarom volgt uit Stelling 12.6 en de laatste regel van haar bewijs dat voor elk deelwoord V van u van lengte n geldt dat $|\,|V|_1 - n\alpha\,| < 1$. In het bijzonder moeten er in u tussen positie s_i en positie s_j tenminste $\alpha(s_j - s_i - 1) - 1$ enen zijn en volgens (4) is dit aantal groter dan $j - i - 1$. Dit levert een tegenspraak in het geval $i < j$. Stel nu $i > j$. Dan is $s_i > s_j$ en hebben we

$$\llcorner \frac{i}{\alpha} + \gamma \lrcorner - \llcorner \frac{j}{\alpha} + \gamma \lrcorner < \frac{i}{\alpha} + \gamma - \frac{j}{\alpha} + \gamma + 1 = \frac{i-j}{\alpha} + 1.$$

Dus, vanwege (3),

$$s_i - s_j \leq \llcorner \frac{i}{\alpha} + \gamma \lrcorner - \llcorner \frac{j}{\alpha} + \gamma \lrcorner - 2 < \frac{i-j}{\alpha} - 1$$

vergelijkbaar met (4). Met een analoog argument als in het andere geval krijgen we een tegenspraak voor het geval $i > j$. □

Stelling 12.10 *Zij u een sturms woord met dichtheid α. Dan bestaat een reëel getal δ zó dat*

$$s_i = \llcorner \frac{i}{\alpha} + \delta \lrcorner \quad \text{voor alle} \quad i \in \mathbb{N} \qquad \text{of} \qquad s_i = \ulcorner \frac{i}{\alpha} + \delta \urcorner \quad \text{voor alle} \quad i \in \mathbb{N}.$$

Bewijs. Er bestaat een getal γ zo dat $\gamma > s_1 - \frac{1}{\alpha}$. Uit Lemma 12.2 volgt dat $\{\gamma : s_i \leq \llcorner \frac{i}{\alpha} + \gamma \lrcorner \text{ voor alle } i\}$ een naar beneden begrensde, niet-lege verzameling is. Beschouw het infimum en noem het δ. Dan is $s_i \leq \llcorner \frac{i}{\alpha} + \delta \lrcorner$ voor elk positief getal i, maar voor elke $\epsilon > 0$ is er een i zo dat $s_i > \llcorner \frac{i}{\alpha} + \delta - \epsilon \lrcorner$. Uit Lemma 12.2 volgt dat voor elke $\epsilon > 0$ geldt dat $s_i \geq \llcorner \frac{i}{\alpha} + \delta - \epsilon \lrcorner$ voor elke i. Door beide ongelijkheden te combineren krijgen we $s_i = \llcorner \frac{i}{\alpha} + \delta \lrcorner$ voor alle i met $\frac{i}{\alpha} + \delta \notin \mathbb{Z}$, en $s_i = \frac{i}{\alpha} + \delta$ of $s_i = \frac{i}{\alpha} + \delta - 1$ als $\frac{i}{\alpha} + \delta \in \mathbb{Z}$. Omdat α irrationaal is, is er ten hoogste één zo'n i. In het eerste geval is $s_i = \llcorner \frac{i}{\alpha} + \delta \lrcorner$ voor alle i, in het tweede geval is $s_i = \ulcorner \frac{i}{\alpha} + \delta - 1 \urcorner$ voor alle i. □

Stelling 12.11 *Elk sturms woord is een rotatiewoord.*

Bewijs. Zij u een sturms woord. Dan heeft u volgens Stelling 12.9 een irrationale dichtheid α. Volgens Stelling 12.11 bestaat een reëel getal δ zo dat

$$s_i = \llcorner \frac{i}{\alpha} + \delta \lrcorner \quad \text{voor alle} \quad i \in \mathbb{N} \qquad \text{of} \qquad s_i = \ulcorner \frac{i}{\alpha} + \delta \urcorner \quad \text{voor alle} \quad i \in \mathbb{N}.$$

Het is geen beperking van algemeenheid te veronderstellen dat we het eerste geval hebben. Definieer $\beta = -\delta\alpha$. Dan geldt $u_n = 1$ als en slechts als n van de vorm $\llcorner \frac{i}{\alpha} - \frac{\beta}{\alpha} \lrcorner$ is voor geschikte i. Dit is het geval als en alleen als $n \leq \frac{i-\beta}{\alpha} < n + 1$ en dit is equivalent met $i - \alpha < \alpha n + \beta \leq i$. Het laatste wil zeggen dat het fractionele deel van $\alpha n + \beta$ in ligt tussen $1 - \alpha$ en 1. Hieruit blijkt dat het sturmse woord inderdaad een rotatiewoord is, en wel met hoek α. □

We concluderen uit de voorgaande stellingen dat de klasse van sturmse woorden identiek is met de klasse van irrationale gebalanceerde woorden en met de klasse van irrationale rotatiewoorden. Elk rationaal rotatiewoord is gebalanceerd, maar er zijn gebalanceerde woorden met rationale hoek die geen rotatiewoord vormen. Voorbeelden zijn: 0000000100000... en 0010010010010001001001001... . Er bestaan ook woorden met rationale dichtheid die voldoen aan $p_u(n) \leq n + 1$ voor alle n en geen rotatiewoord zijn. Voorbeelden zijn: 00000000001111111111111... en 001001001001000100010001... . Al deze woorden zijn uiteindelijk periodiek, maar niet zuiver periodiek.

7 *Automaten en getalsystemen*

Aan elke substitutie σ kan men een automaat S koppelen die die substitutie karakteriseert. De automaten die hierbij ontstaan zijn aangeklede gerichte grafen. Als knopen (*staten* genoemd) neemt men de letters uit het alfabet, zeg $\mathcal{A} = \{a, b, c, ..., n\}$. Als voor $x \in \mathcal{A}$ de k-de letter van $\sigma(x)$ gelijk is aan de letter y, dan trekt men een boog van letter x naar letter y en geeft deze boog label $k - 1$. Dit wordt gedaan voor alle x, y en k. We beschouwen a als de startletter. Als dan een eindige cijfercode wordt opgegeven, kan men deze proberen te volgen door in a te starten en telkens de boog te volgen die het volgende cijfer aangeeft. Als de wandeling tot het einde voortgezet kan worden noteren we de letter van de staat waar de wandeling eindigt; anders heet de cijfercode verboden. Uit de automaat kan men meestal eenvoudig aflezen welke cijfercombinaties verboden zijn. We werken dit hier niet verder uit, maar geven enkele voorbeelden.

De Morsesubstitutie $a \to ab, b \to ba$.
Geen enkele combinatie bestaande uit nullen en enen is verboden.

De Cantorsubstitutie $a \to aba, b \to bbb$.
Geen enkel combinatie bestaande uit nullen, enen en tweeën is verboden.

De Fibonaccisubstitutie $a \to ab, b \to a$.
Het is duidelijk dat er geen twee opeenvolgende enen kunnen optreden.

De Tribonaccisubstitutie $a \to ab, b \to ac, c \to a$.
Alleen cijfercodes bestaande uit nullen en enen zijn toegestaan. De enige verboden cijfercodes zijn die met drie opeenvolgende enen.

Het zal duidelijk zijn dat als we een substitutie van constante lengte k hebben, alle cijfercombinaties met cijfers $0, 1, \ldots k-1$ toegestaan zijn. Andere substituties werken met net zoveel cijfers als het langste substitutiewoord telt, maar kennen daarbij verboden combinaties.

8 *Getalsystemen*

We beschouwen een substitutie σ met een invariant woord beginnend met a. We construeren de daarbij behorende automaat. Vervolgens gaan we de niet-negatieve getallen aftellen in het bijbehorende getalsysteem. Is het woord verboden, dan slaan we het over. Anders noteren we de eindletter. Zo ontstaat een oneindig woord. We gaan in elk van de eerder genoemde substituties na wat het woord is.

De Morsesubstitutie $a \to ab, b \to ba$.
Het resulterende woord is het Morsewoord $abbabaabbaababba$....

De Cantorsubstitutie $a \to aba, b \to bbb$.
Geen enkel combinatie bestaande uit nullen, enen en tweeën is verboden. Het resulterende woord is het Cantorwoord $ababbbababbbbbbbbbababbbaba$....

De Fibonaccisubstitutie $a \to ab, b \to a$.
Het is duidelijk dat er geen twee opeenvolgende enen kunnen optreden. Het resulterende woord is het Fibonacciwoord $abaababaabaabab$....

De Tribonaccisubstitutie $a \to ab, b \to ac, c \to a$.
Alleen cijfercodes bestaande uit nullen en enen zijn toegestaan. De enige verboden cijfercodes zijn die met drie opeenvolgende enen. Het resulterende woord is het Tribonacciwoord $abacabaabacababacaba$....

Bij elke substitutie krijgt men altijd het invariante woord. Dit volgt uit de constructie (opgave). Als de substitutie van constante lengte k is, telt met af in het k-tallig stelsel. In de andere gevallen induceert de substitutie een ingewikkelder getalsysteem.

De getalsystemen worden pas interessant, als de representaties van de getallen gevonden kunnen worden met een eenvoudig algoritme, bij voorkeur het gretige algoritme. Merk op dat, omdat er een invariant woord is dat met a begint, de combinaties 1, 10, 100, 1000, ... toegestaan zijn. Bepaal de waarden (posities) van deze cijfercombinaties en noem ze $N_0, N_1, N_2, N_3, ...$ Om de gretige representatie van een getal n te bepalen trek je van n de grootste N_i af die niet groter is dan n, vervang je n door $n - N_i$ en zoek je weer de grootste N_j om af te trekken. Als je de representatie van n krijgt door achtereenvolgens op te schrijven hoe vaak je $N_i, N_{i-1}, N_{i-2}, \dots$ afgetrokken hebt, dan heb je een bruikbaar getalsysteem.

Hieraan voldoet de Fibonaccisubstitutie. Daarvoor geldt $N_i = F_{i+2}$ als $F_0, F_1, F_2, ... = 0, 1, 1, ...$ de rij van Fibonacci is. Bijvoorbeeld vind je in het Fibonaccigetalsysteem voor 45 de representatie 10010100, want 45=34+8+3. Er kunnen nooit twee enen achter elkaar komen, omdat de som van twee opeenvolgende Fibonaccigetallen gelijk is aan de volgende en dus eerder afgetrokken had moeten worden. Uit de bijbehorende automaat is direct uit de representatie 10010100 af te lezen dat de 46ste letter van het Fibonacciwoord gelijk is aan a.

Bij de substitutie σ gegeven door $a \to aab, b \to c, c \to aac$ werkt het gretige algoritme niet naar behoren. Er geldt $N_0 = 1, N_1 = 3, N_2 = 7, N_3 = 17, N_4 = 43$. Beschouw het getal 41. Met het gretige algoritme vinden we $41 = 2100$. Dit is echter een verboden combinatie. Het zal duidelijk zijn dat hier nog heel wat theorie achter zit waar we hier niet op ingaan.

Literatuur bij Hoofdstuk 12

M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics and Applications, Addison-Wesley, 1983.

M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002.

Alex Heinis, Arithmetics and Combinatorics of Words of Low Complexity,
proefschrift, Univ. Leiden, 2001, Hoofdstuk 2.
www.math.leidenuniv.nl/~tijdeman/

P. Fogg, Substitutions in Dynamics, Arithmetics and Combinatorics, Lecture Notes in Mathematics 1794, Springer-Verlag, 2002.

9 Vraagstukken

1 a) Bepaal de incidentiematrices van de Morsesubstitutie, de Fibonaccisubstitutie en de Cantorsubstitutie.
1 b) Welke van de matrices zijn unimodulair?
1 c) Druk $M\sigma^n$ uit in M_σ.

2. Beschouw het invariante woord u van de Tribonacci substitutie

$$\sigma : a \to ab, b \to ac, c \to a.$$

2 a) Schrijf $\sigma^5(a)$ uit.
2 b) Bewijs dat voor elke $n > 3$ geldt dat $\sigma^n(a) = \sigma^{n-1}(a)\sigma^{n-2}(a)\sigma^{n-3}(a)$.
2 c) Bewijs dat u uniform recurrent is.

3. Zij $\alpha = 1 + \sqrt{2}$.
3 a) Voor welke substitutie is het snijwoord van de lijn $y = \alpha x$ invariant?
3 b) Wat zijn de eerste 10 letters van dit snijwoord?
3 c) Wat zijn de eerste 5 letters van het corrresponderende rotatiewoord?
3 d) Welke substitutie voert een rotatiewoord over in het corresponderende snijwoord?

4 a) Bewijs dat het Cantorwoord, het Fibonacciwoord en het Morsewoord niet uiteindelijk periodiek zijn.
4 b) Bewijs dat het Cantorwoord en het Morsewoord niet sturms en niet gebalanceerd zijn.
4 c) Bewijs dat het Cantorwoord niet uniform recurrent is.
4 d) Bewijs dat het Fibonacciwoord en het Morsewoord wel uniform recurrent zijn.

5 a) Bereken $p_u(n)$ voor $n = 1, 2, 3, 4, 5$ als u het Morsewoord is.
5 b) Bewijs dat voor deelwoorden V en W van u van gelijke lengte geldt dat $|V|_1 - |W|_1 \leq 2$.

6) Zij $u = u_0u_1\dots$ een oneindig woord op twee letters, 1 en 2. We definiëren $P(0) = (0, 0)$ en $P(n+1) - P(n) = (1, 0)$ als $u_n = 1$ en $P(n+1) - P(n) = (0, 1)$ als $u_n = 2$.
6 a) In welke (minimale) strip ligt de rij punten $(P(n))_{n=0}^{\infty}$ als u het Fibonacciwoord is op de letters 1 en 2?
6 b) In welke (minimale) strip ligt de rij punten $(P(n))_{n=0}^{\infty}$ als u het Morsewoord is op de letters 1 en 2?

7 a) Bereken de eigenwaarden van de incidentiematrices van de Morsesubstitutie, de Fibonaccisubstitutie en de Cantorsubstitutie.
7 b) Bereken de eigenvectoren die bij de grootste eigenwaarden behoren.
7 c) Ga na dat de grootste eigenwaarde de vergrotingsfactor $\lim_{n\to\infty} \frac{\sigma^{n+1}(a)}{\sigma^n(a)}$ is.
7 d) Ga na dat de coördinaten van de corresponderende eigenvector de dichtheidsverhoudingen aangeven van de letters in het invariante woord.

Tentameneisen voor Combinatoriek Alle vraagstukken, definities, stellingen, enz., met uitzondering van blz. 65 en secties 12.7 en 12.8. Alle bewijzen uit Hoofdstukken 2, 4, 6 - 10, 12 met uitzondering van de bewijzen van Stellingen 2.4, 6.5, 9.2, 9.3, 10.3 en van alle bewijzen vanaf Stelling 12.8.