BACHELOR

Eindige meetkunde en haar toepassingen in foutcorrectie

Landa, J.

Award date:
2014

Technische Universiteit
Eindhoven
University of Technology

Faculteit Wiskunde & Informatica
Den Dolech 2, 5612 AZ Eindhoven
Postbus 513, 5600 MB Eindhoven
www.tue.nl

Eindige meetkunde en haar toepassingen in foutcorrectie

Auteur
J. Landa

Bachelorscriptie
Technische Wiskunde

Begeleider
dr. A. Blokhuis

Datum
20 augustus 2014

Samenvatting

Codes bestaan uit een verzameling codewoorden ter lengte n, gemaakt uit een eindig alfabet. Deze kunnen worden beschouwd als posities in een eindige, discrete vectorruimte. Met de handige minimale Hamming afstand d kunnen deze onderscheiden worden. De dimensie van het opspansel van de codewoorden, k, staat direct voor het aantal datasymbolen dat per keer gecodeerd kan worden. De dimensie van de gehele vectorruimte, n, staat garant voor het aantal symbolen dat per keer verzonden wordt. Er worden dus $n - k$ redundante controlesymbolen toegevoegd om resistentie tegen ruis en fouten te bewerkstelligen, afhankelijk van de gebruikte codering. De handigste codes die gekozen kunnen worden zijn de codes waarvan de codewoorden in hun projectieve ruimte voldoen aan een nonsinguliere kwadratische vorm. Deze codes hebben dan de eigenschap dat ze de meeste data veilig kunnen verzenden met de parameters vast gekozen.

Technische Universiteit
Eindhoven
University of Technology

Inhoudsopgave

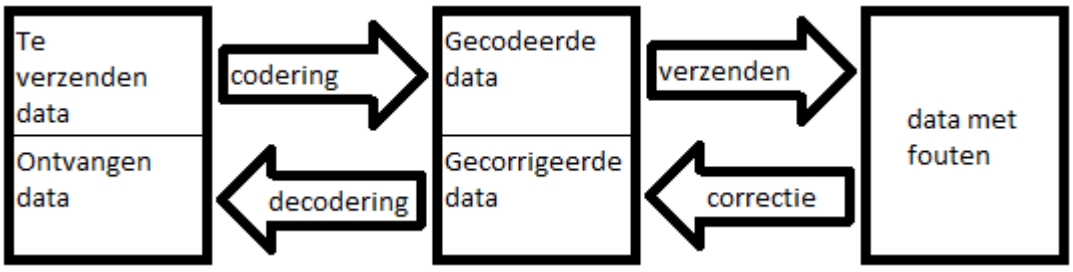
Titel
Eindige meetkunde en haar toepassingen in foutcorrectie

Inleiding

In het hedendaagse leven vindt er per dag zo veel datatransmissie plaats, dat het haast niet bij te houden is. Het vindt overigens niet alleen op grote schaal plaats. Denk bijvoorbeeld niet alleen aan het sturen van een e-mail of het bellen met je familie, maar ook het wegschrijven van data op een CD of harde schijf en het scannen van je boodschappen met een zelfscanner of van boeken bij de bibliotheek. Het is dus allicht van belang dat deze technologie goed werkt, om miscommunicatie te voorkomen.

Ten alle tijden moet rekening gehouden worden met ruis die onderweg de boodschap die verstuurd wordt kan verstoren en dusdanig onleesbaar maakt. Daarom schuilen achter de technologie die we gebruiken (computers, telefoons, scanners etc.) verscheidene manieren verscholen om de data te coderen, zodat deze bestand wordt tegen de eventuele ruis en/of leesfouten door de ontvanger. Deze tak van de wiskunde heet *coderingstheorie*. De ontvanger kan aan de hand van de code de boodschap corrigeren en vervolgens decoderen, zodat de originele boodschap overblijft. Dit zijn dan *foutcorrigerende codes*.

Om de codes goed te kunnen begrijpen heeft de lezer eerst wat kennis nodig van eindige meetkunde: de tak van de meetkunde die zich, zoals de naam al impliceert, bezighoudt met wiskundige structuren van eindige grootte. De schakel met coderingstheorie zal daarna verduidelijkt worden, door middel van het uilichten van een belangrijke bijdrager aan dit gebied. Aan het einde van dit geschrift staan praktijkvoorbeelden van in de technologie veel ingezette codes, ter illustratie van de vele gebieden waarop foutcorrigerende wordt ingezet.



Figuur 1: Het datatransmissie proces

1 Eindige Meetkunde

In dit hoofdstuk komt eindige projectieve meetkunde aan bod als basis voor het hoofdstuk erna, waarin foutcorrigerende codes behandeld worden. Zo zal duidelijk worden wat het verband is tussen deze twee principes en hoe dit gelegd is.

1.1 Het Galoislichaam $\mathbb{F}_q$

Op het gebied van algebra en meetkunde worden vele structuren bestudeerd, waaronder de algebraïsche structuren *groep*, *ring* en *lichaam*.

Definitie 1.1.1. *Een groep $(\mathcal{G}, \circ)$ is een verzameling elementen $\mathcal{G}$ met een binaire operator $\circ$ zodanig dat het resultaat van de operatie op twee elementen van $\mathcal{G}$ weer een nieuw element vormt. Hierbij horen de groepsaxioma's:*

G1 *$\mathcal{G}$ moet gesloten zijn onder $\circ$. Dit wil zeggen dat $\forall a, b \in \mathcal{G} : a \circ b \in \mathcal{G}$.*

G2 *$(\mathcal{G}, \circ)$ moet associatief zijn onder $\circ$, wat wil zeggen dat $\forall a, b, c \in \mathcal{R} : a \circ (b \circ c) = (a \circ b) \circ c$.*

G3 *$(\mathcal{G}, \circ)$ bevat het zogenaamde neutrale element met betrekking tot $\circ$, wat inhoudt dat $\exists e \in \mathcal{G} : \forall a \in \mathcal{G} : a \circ e = e \circ a = a$.*

G4 *Elk element is inverteerbaar met betrekking tot $\circ$. Er moet dan gelden dat $\forall a \in \mathcal{G}\ \exists b \in \mathcal{G} : a \circ b = b \circ a = e$.*

Definitie 1.1.2. *Een ring $(\mathcal{R}, +, \cdot)$ is een niet-lege verzameling $\mathcal{R}$ waarop optelling en vermenigvuldiging gedefinieerd zijn zodanig dat de axioma's voor ringen gelden:*

R1 *$(\mathcal{R}, +)$ een Abelse groep is, wat wil zeggen dat $(\mathcal{R}, +)$ een groep is en de optelling commutatief is in $\mathcal{R}$* (Niels Henrik Abel, 1802-1829).

R2 *De vermenigvuldiging associatief is.*

R3 *De linkse en rechtse distributiviteiten gelden, ergo $\forall a, b, c \in \mathcal{R}$*

- $a \cdot (b + c) = a \cdot b + a \cdot c$, *en*
- $(b + c) \cdot a = b \cdot a + c \cdot a$

Als een ring een neutraal element tegenover de vermenigvuldiging bevat, dan heet dit een *ring met identiteit*. Een ring heet *commutatief* als naast de optelling ook de vermenigvuldiging commutatief is. Een ring $\mathcal{R}$ is een *integriteitsgebied* als het een commutatieve ring met identiteit is en de eigenschap $(a \cdot b = 0) \Rightarrow (a = 0 \vee b = 0)$ voor alle $a, b \in \mathcal{R}$. Een ring heet een *delingsring* als alle elementen ongelijk aan 0 een groep onder de vermenigvuldiging vormen.

Definitie 1.1.3. *Een lichaam $\mathcal{L}$ is een commutatieve delingsring.*

Een lichaam is dus een ring waar de vermenigvuldiging commutatief is ($a \cdot b = b \cdot a$). Dit is nodig om een eenduidige multiplicatieve inverse te kunnen bepalen en dus om te kunnen delen. Het bestaan van zo een inverse voor elk niet-nul element verzekert het mogelijk zijn van delen.

Een deelverzameling van alle lichamen die in ieder geval voor dit project van belang zijn, zijn de lichamen met eindig veel elementen. Een bekend voorbeeld hiervan is het lichaam $\mathbb{Z}_p$. Dit lichaam bevat de getallen modulo p. p heet de *karakteristiek* van $\mathbb{Z}_p$. Schrijf $\mathrm{char}(\mathbb{Z}_p) = p$

Stelling 1.1.4. *$\mathbb{Z}_p$ is een lichaam dan en slechts dan als p een priemgetal is.*

Bewijs. Laat p een priemgetal zijn. p priem $\Leftrightarrow \forall x \in \mathbb{Z}_p : gcd(x, p) = 1 \Leftrightarrow$ er is een Diophantische vergelijking met getallen n en m zodanig dat $1 = mp + nx \Leftrightarrow \exists y \in \mathbb{Z}_p : n \equiv y \pmod p \Leftrightarrow yx = nx = 1 - mp = 1 \pmod p \Leftrightarrow yx = 1 \forall x \in \mathbb{Z}_p$ met $x \neq 0 \Leftrightarrow$ per definitie is $\mathbb{Z}_p$ een lichaam. □

Dit lichaam is te generaliseren naar een lichaam met $q = p^m$ elementen. q is dan de macht van een priemgetal. Gebruik het volgende lemma.

Lemma 1.1.5. *Zij $\mathcal{K}$ een lichaam met $\mathrm{char}(\mathcal{K}) = p$. Dan is de afbeelding $\sigma : x \mapsto x^p$ een isomorfisme van $\mathcal{K}$ naar het deellichaam $\mathcal{K}^p = \{x^p | x \in \mathcal{K}\}$.*

Bewijs. Merk op dat $\sigma(xy) = \sigma(x)\sigma(y)$ voor alle x, y en dat de binomiaalcoëfficient $\binom{p}{k}$ congruent 0 modulo p is als $0 < k < p$. Neem zulks een k en zie dat $\binom{p}{k} = \frac{p!}{k!(n-k!)} = p \cdot \frac{(p-1)!}{k!(p-k)!} \equiv 0 \mod p$. Als $k \in \{0, p\}$, dan is $\binom{p}{k} = \frac{p!}{p!} = 1$. Hieruit volgt dat voor alle x, y geldt dat $\sigma(x + y) = (x + y)^p = x^p + y^p = \sigma(x) + \sigma(y)$, dus is σ een automorfisme. σ is duidelijk injectief. □

Stelling 1.1.6. *Zij $\mathcal{K}$ een eindig lichaam.*

S1 *Laat $\mathrm{char}(\mathcal{K}) = p \neq 0$ een priemgetal zijn. Als m de uitbreidingsgraad van $[\mathcal{K} : \mathbb{Z}_p]$ is, dan is $|\mathcal{K}| = q = p^m$.*

S2 *Zij p een priemgetal en $q = p^m$ $(m \geq 1)$. Als Ω een algebraïsch gesloten lichaam is met $\mathrm{char}(\Omega) = p$, dan bestaat er een uniek deellichaam $\mathbb{F}_q$ van Ω met q elementen. Dit is de verzameling nulpunten van $X^q - X$.*

S3 *Alle eindige lichamen met $q = p^m$ elementen zijn isomorf aan $\mathbb{F}_q$.*

Bewijs. Als $\mathcal{K}$ eindig is, is $\mathbb{Q}$ niet bevat in $\mathcal{K}$, dus is de karakteristiek van $\mathcal{K}$ een priemgetal p. Als m de graad van uitbreiding is van $[\mathcal{K} : \mathbb{Z}_p]$, dan volgt hieruit dat $|\mathcal{K}| = q = p^m$. (S1).

Als Ω algebraïsch gesloten is en karakteristiek p heeft, dan vertelt het lemma dat $x \mapsto x^q$ een automorfisme is op Ω en dat het gelijk is aan de afbeelding σ^m. Dus vormen de elementen

$x \in \Omega$ invariant aan $x \mapsto x^q$ een deellichaam $\mathbb{F}_q$ van Ω. De afgeleide van $X^q - X$ is gelijk aan $qX^{q-1} - 1 = p \cdot p^{m-1}X^{q-1} - 1 = -1$ en is dus niet gelijk aan 0. Gegeven dat Ω algebraïsch gesloten is, voert dit tot het feit dat $X^q - X$ q nulpunten heeft, en dus $|\mathbb{F}_q| = q$. Als $\mathcal{K}$ een deellichaam is van Ω met q elementen dan bestaat de multiplicatieve groep van $\mathcal{K}$, $\mathcal{K}*$, $q-1$ elementen. Vanwege de kleine stelling van Fermat geldt dat $x^q = 1$ als $x \in \mathcal{K}*$ en dus dat $x^q = x$ als $x \in \mathcal{K}$. Dus geldt $\mathcal{K} \subset \mathbb{F}_q$. Maar omdat $|\mathbb{K}| = |\mathbb{F}_q|$. (S2).

Uit (S2), het feit dat alle lichamen met p^m elementen ingebed kunnen worden in Ω en dat Ω algebraïsch gesloten is, volgt (S3). □

Het laatste onderdeel van bovenstaande stelling helpt het begrijpen van $\mathbb{F}_q$. $\mathbb{Z}_p$ zijn de getallen $0, 1, 2, \ldots, p-1$, waarbij de operaties optellen en vermenigvuldigen aangevuld worden met de modulus p, zodat elk getal x staat voor de equivalentieklasse der getallen die rest x hebben na deling door p, wat het tot een lichaam maakt. Dit kan men ook met polynomen over $\mathbb{Z}_p$ van graad hoogstens $m-1$ doen. Deze verzameling bevat p^m elementen. Het polynoomequivalent van de priemeigenschap is de irreducibiliteit. Neem dus een irreducibel polynoom f_m van graad m over $\mathbb{Z}_p$ als modulus en de verzameling wordt de verzameling equivalentieklassen der polynomen die dezelfde rest na deling door f_m hebben. Aangezien $\mathbb{F}_q$ isomorf is met $\mathbb{Z}_p/f_m(x)$ helpt een denkwijze als deze om de structuur van $\mathbb{F}_q$ te doorgronden.

1.2 Vector- en projectieve ruimten over $\mathbb{F}_q$

Nu het eindige lichaam $\mathbb{F}_q$ goed gedefinieerd is, kunnen er lineaire vectorruimten over gemaakt worden.

Definitie 1.2.1. *Zij $V(n+1,q)$ de lineaire vectorruimte over $\mathbb{F}_q$ van dimensie $n+1$. Dan is $PG(n,q)$ de projectieve ruimte van $V(n+1,q)$. De deelruimten van $PG(n,q)$ van dimensie $0, \ldots, n-1$ stellen dan de deelruimten van $V(n+1,q)$ van dimensie $1, \ldots, n$ voor.*

Gevolg 1.2.2. *Het aantal 0- tot $n-1$-dimensionale deelruimten van $PG(n,q)$ moet gelijk zijn aan het aantal onafhankelijke 1- tot n-dimensionale deelruimten van $V(n+1,q)$. Dit zijn er dan:*

$$\binom{n+1}{k}_q := \frac{(q^{n+1}-1)(q^{n+1}-q)\cdots(q^{n+1}-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}$$

voor een k-dimensionale deelruimte van $V(n+1,q)$. ($k = \{1, \ldots, n+1\}$)

Bewijs. Het aantal k-dimensionale deelruimten uit $V(n+1,q)$ is het aantal verzamelingen met k onafhankelijke vectoren uit $V(n+1,q)$

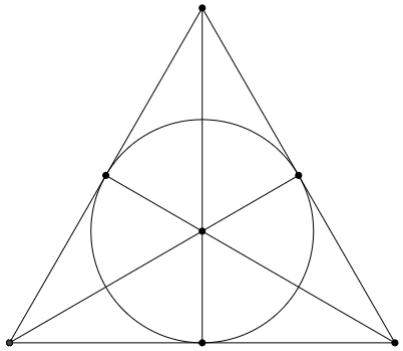
$$(q^{n+1}-1)(q^{n+1}-q)\cdots(q^{n+1}-q^{k-1})$$

gedeeld door het aantal verzamelingen met k onafhankelijke vectoren uit $V(k,q)$

$$(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})$$

□

Met projectieve ruimten wordt het dus gemakkelijk om afhankelijkheden van deelruimten te constateren.



Figuur 2: Het Fano-vlak

Voorbeeld 1.2.3. *De deelruimten van $PG(2,2)$ zijn de deelruimten van $V(3,2)$ maar dan een dimensie lager. $PG(2,2)$ heet ook het Fano-vlak (Gino Fano, 1871-1952).*

Onder de gekozen basis is elke hoek van het Fano-vlak (een punt) het opspansel van elke afzonderlijke basisvector (een lijn). Dit gaat analoog verder voor hogere dimensie deelruimten.

Zo is te zien dat tussen twee hoekpunten een lijn loopt, wat overeenkomt met het vlak door de oorsprong en de twee basisvectoren. In dat vlak liggen dus in ieder geval $\langle e_i\rangle$, $\langle e_j\rangle$ ($i \neq j$) en de nulvector. Let op: de nulvector is geen punt in het Fano-vlak, want de nulvector heeft al dimensie 0 in $V(3,2)$. Het derde punt op zo een lijn moet overeenkomen met een opspansel van een lineaire combinatie van e_i en e_j, namelijk $\langle e_i + e_j\rangle$. Het punt in het midden stelt dan $\langle \sum e_k\rangle$ voor.

1.3 Het projectieve vlak PG(2,q)

Beschouw nu de projectieve ruimte $PG(2,q)$. Dit wordt ook wel een *projectief vlak* genoemd. $PG(2,q)$ kan gezien worden als een incidentiestructuur.

Definitie 1.3.1. *Een incidentiestructuur is een triple (P,L,I) met P een verzameling punten, L een verzameling lijnen en $I \subseteq (P \times L)$ de incidentierelatie.*

De incidentiestructuur van $PG(2, q)$ wordt uitgelegd aan de hand van deze drie axioma's

A1 Door twee punten gaat één lijn.

A2 Twee lijnen snijden elkaar in één punt.

A3 Men kan vier punten kiezen zodanig dat elk drietal daarvan niet op één lijn ligt.

Stelling 1.3.2. *Op een lijn in $PG(2, q)$ liggen $q+1$ punten en door een punt in $PG(2, q)$ gaan $q+1$ lijnen.*

Bewijs. Zij P een punt en l een lijn, niet door P. Uit (A1) en (A2) volgt dat het aantal lijnen door P gelijk is aan het aantal punten op l, want gegeven de axioma's mag P met elk punt op l verbonden worden. Door (A3) mag men nu een ander punt $Q \neq P$ niet op l nemen. Ook dit punt mag met de punten van l verbonden worden. Het aantal punten op l is gelijk aan het aantal lijnen door P en het aantal lijnen door Q. Omdat P en Q willekeurig gekozen waren, volgt dat er $q+1$ punten op een lijn liggen en er $q+1$ lijnen door een punt gaan. □

$PG(2, q)$ is tevens een Desargues vlak (Girard Desargues, 1591-1661). Betekenis en gevolgen worden later uitgelegd.

2 Beniamino Segre

2.1 Achtergrond

Beniamino Segre werd op 16 februari 1903 geboren in het Italiaanse Turijn. Hij ging in Turijn naar school en won al op zijn zestiende, in 191, een studiebeurs voor de universiteit van Turijn, waar hij ging studeren. Hij werd daar onderwezen door onder andere Gino Fano en Corrado Segre (1863-1924), waar hij achterneef van was. In 1923 studeerde hij af aan de universiteit van Turijn met zijn scriptie, welke een meetkundig onderwerp had. Corrado was zijn begeleider. Opmerkelijk genoeg schreef hij na zijn studie een paper met een heel ander onderwerp, namelijk vloeistofdynamiek. Hij werd aangesteld als assistent-professor in Turijn, waar hij tot 1926 bleef. Nadat hij een tijd met Élie Cartan (1869-1951) in het academisch jaar 1926-1927 kreeg hij een Rockefeller beurs en werd de assistent van Francesco Severi (1879-1961) in Rome.

Figuur 3: Beniamino Segre

In 1931 had stonden er al 40 publicaties op zijn naam en werd hij 'Chair of Geometry' op de universiteit van Bologna. Een jaar later huwde hij Fernanda Coen, waar hij drie kinderen mee kreeg. Zijn relatie werd vaak beschreven als een relatie die zelfs tot het eind van hun levens de grootste genegenheden en begrip belichaamde. Vrienden konden hen apart niet meer voorstellen.

Toen in Italië in 1938 de fascistische regering opkwam, kwamen er wetten tegen de mensen met een Joodse achtergrond. De gevestigde orde kwam erachter dat Beniamino een Joodse achtergrond had, en dus werd hij van de universiteit gestuurd. Hij vluchtte met zijn vrouw en drie jonge kinderen naar Engeland. Ze woonden een tijdje in Londen en Cambridge, maar Italië was op dat moment een vijand van Engeland, dus werd Beniamino geïnterneerd op het eiland Man, tussen het Engelse vasteland en Ierland. Dit was waarschijnlijk de ongelukkigste periode in het leven van hun gezin. Zijn vrouw en kinderen woonden tijdelijk bij het gezin van Leonard Roth (1904-1968) toen het jongste kind van Beniamino en Fernanda stierf. Ook zonk er een schip wat krijgsgevangenen en geïnterneerden naar Canada moest brengen en zijn vrouw wist niet of hij op dat schip zat. Dit was gelukkig niet zo en uiteindelijk werden de geïnterneerden vrijgelaten. Hij werd herenigd met zijn vrouw en kinderen en samen

vestigden ze zich in Cambridge, waar hij er achter kwam dat zijn publicaties van de vroege oorlogsjaren ontbraken.

Toch stortte hij zich nog steeds graag op wiskunde en schreef een monografie dat in 1942 door de Oxford University Press werd uitgegeven. Dat leverde hem hetzelfde jaar nog een docentschap met Louis Mordell (1888-1972) in Manchester op. Het was in deze tijd dat hij grootse bijdragen leverde aan de algebraïsche meetkunde maar dat ook zijn interesses breder werden. Onder de invloed van Mordell en Kurt Mahler (1903-1988) onderzocht hij ook de Diophantische vergelijkingen en de algebraïsche variëteiten. In 1946 vertrok hij weer naar Bologna en volgde vier jaar later Severi op in Rome.

Zijn verzameling wetenschappelijke publicaties die over meetkunde en dergelijke gingen bereikte de driehonderd, de papers over andere onderwerpen niet meegeteld. Hij kreeg een goede reputatie door de goede kwaliteit van schrijven en wordt daar nog steeds om herinnerd. rond 1955 was hij vooral bezig met eindige meetkunde en heeft toen zijn beroemde stelling bewezen. In die tijd bundelde hij ook veel resultaten over dat onderwerp.

In 1973, toen Segre zeventig jaar was ging hij met pension en verliet hij zijn academische zetel. Hij ging met zijn vrouw in Frascati wonen, waar hij dichtbij zijn twee getrouwde kinderen en zijn kleinkinderen een gelukkig leven leidde. Drie jaar later, in 1976, stierf zijn vrouw. Dit was een klap waar hij nooit meer bovenop zou komen, aangezien hij een jaar later, op 2 oktober 1977, zelf overleed.

2.2 De stelling van Segre

De stelling die Segre bewees maakte hem tot belangrijke bijdrager aan de algebraïsche en combinatorieke meetkunde. Alvorens deze stelling aan bod komt, moeten eerst de begrippen *kegelsnede* en *ovaal* aan bod komen.

Definitie 2.2.1. *Een kegelsnede $\mathcal{C}$ in $PG(2,q)$ is een verzameling niet triviale nulpunten van een homogene tweedegraads vergelijking in drie variabelen.*

Schrijf $\mathcal{C} = \{\langle(x,y,z)\rangle | f(x,y,z) = 0\}$. Elk punt (x,y,z) kan genormaliseerd worden zodanig dat één van de coördinaten 1 is. Elk punt (x,y,z) is dus te classificeren als een punt van de vorm $(1,0,0)$, $(\alpha,1,0)$ of $(\beta,\gamma,1)$. Dit wordt toegepast in het volgende voorbeeld.

Voorbeeld 2.2.2. *Neem $f(x,y,z) = x^2 - yz$. Hieruit volgt dat $x^2 = yz$. Uit het bovengestelde volgt dat $z = 0$ of $z = 1$. Onderscheid de gevallen:*

- $z = 0 \Rightarrow (x,y,z)$ *is van de vorm* $(\alpha,1,0) \Rightarrow \alpha = 0$.
- $z = 1 \Rightarrow (x,y,z)$ *is van de vorm* $(\beta,\gamma,1) \Rightarrow \beta^2 = \gamma \Rightarrow (x,y,z) = (\beta,\beta^2,1)$, *met* $\beta \in \mathbb{F}_q$

$\Rightarrow \mathcal{C} = \{(\beta,\beta^2,1) | \beta \in \mathbb{F}_q\} \cup \{(0,1,0)\}$

Nu volgt de definitie van het begrip *ovaal*.

Definitie 2.2.3. *Een ovaal $\mathcal{O}$ in $PG(2,q)$ is een verzameling van $q+1$ punten zodanig dat elke lijn in het vlak hoogstens twee punten van $\mathcal{O}$ bevat.*

Bij ovalen (en dus ook kegelsneden) kan men 3 verschillende soorten lijnen onderscheiden:

- Een lijn die door twee punten van $\mathcal{O}$ gaat heet een *secant*.
- Een lijn die door één punt van $\mathcal{O}$ gaat heet een *tangent*.
- Een lijn die geen enkel punt $\mathcal{O}$ bevat heet een *passant*.

Stelling 2.2.4. *Elke kegelsnede is een ovaal in $PG(2,q)$.*

Gebruik voor het bewijs van de stelling het volgende lemma.

Lemma 2.2.5 (Chevalley-Warning)**.** *Zij $\mathbb{F}_q$ een eindig lichaam en $\{f_j\}_{j=1}^r \subset \mathbb{F}[X_1,\ldots,X_n]$ een stelsel polynomen zodanig dat*

$$n > \sum_{j=1}^{r} d_j$$

met d_j de totale graad van f_j (het maximum van de per term opgetelde exponenten). Dan zal gelden dat het aantal gemeenschappelijke oplossingen van het stelsel deelbaar is door de karakteristiek p van $\mathbb{F}$, dus is de kardinaliteit van de verzameling verdwijpunten van het stelsel congruent 0 modulo p.

Bewijs van het lemma: Als $i < p-1$ dan geldt

$$\sum_{x\in\mathbb{F}_q} x^i = 0$$

Dus verdwijnt de som van elk polynoom in $\mathbb{F}_q^n$van graad kleiner dan $n(p-1)$ ook. Het totale aantal gemeenschappelijke oplossingen modulo p van $f_1,\ldots,f_r = 0$ is gelijk aan

$$\sum_{x\in\mathbb{F}_q} (1-f_1^{p-1}(x))\cdots(1-f_r^{p-1}(x))$$

want elke term van de som 1 is voor een gemeenschappelijke oplossing, en anders 0. De eis de som van de totale graden van alle f_i kleiner is dan n heeft als gevolg dat het helemaal wegvalt door de eerste gelijkheid van het bewijs. Omdat $p \geq 2$ geldt ook nog dat als het stelsel polynomen de nuloplossing bevat, er ook nog minimaal één niet-triviale oplossing moet zijn. □

Hiermee kan Stelling 2.2.4 bewezen worden.

Bewijs van Stelling 2.2.4: Lemma 2.2.5 geeft ons het bestaan van niet triviale nulpunten van een kegelsnede. De kegelsnede is een tweedegraadsvergelijking, de lijnen in het projectief vlak hebben een eerstegraads vergelijking. Iedere lijn snijdt de kegelsnede dus in hoogstens twee punten. □

Gevolg 2.2.6. *Gegeven een kegelsnede $\mathcal{C}$ in $PG(2,q)$. Er zijn $\frac{1}{2}q^2+\frac{3}{2}q+1$ lijnen die op zijn minst door een punt van de kegelsnede gaan.*

Bewijs. Een kegelsnede in $PG(2,q)$ heeft $q+1$ punten. Beschouw één voor één de punten $c_i \in \mathcal{C}$.

c_1 Van de $q+1$ lijnen die door dit punt gaan, gaan er q ook door nog niet beschouwde punten en 1 door alleen c_1. Dit zijn $q+1$ nieuwe lijnen.

c_2 Van de $q+1$ lijnen die door dit punt gaan, gaan er $q-1$ ook door nog niet beschouwde punten, 1 door reeds beschouwde punten en 1 door alleen c_2. Dit zijn q nieuwe lijnen.

...

...

...

c_{q+1} Van de $q+1$ lijnen die door dit punt gaan, gaan er 0 ook door nog niet beschouwde punten, q door reeds beschouwde punten en 1 door alleen c_{q+1} Dit is 1 nieuwe lijn.

Bij elkaar opgeteld zijn dit $1+2+\cdots+q+(q+1)=\frac{q+2}{2}(q+1)=\frac{1}{2}q^2+\frac{3}{2}(q+1)$ lijnen. □

Met deze kennis kan de stelling van Segre bewezen worden. De stelling van Segre zegt:

Stelling 2.2.7 (Segre)**.** *Een ovaal in $PG(2,q)$ met q oneven is een kegelsnede.*

Deze stelling zal in de volgende sectie bewezen worden.

In het geval dat q even is gelden er hele andere dingen. Dan kunnen er namelijk verzamelingen van $q+2$ punten bestaan in $PG(2,q)$, waarvan elk drietal niet op een lijn ligt. Zulks een verzameling heet dan een *hyperovaal*. Deze werden voornamelijk onderzocht door de wiskundige Qvist in de jaren '50. Hij bewees onder andere dat de tangenten van de punten uit een hyperovaal samenkomen in een punt, de *nucleus*.

Voorbeeld 2.2.8. *In het Fanovlak bestaat er een hyperovaal, namelijk* $\{\langle e_1\rangle, \langle e_2\rangle, \langle e_3\rangle, \langle e_1+e_2+e_3+e_4\rangle\}$.

2.3 Twee varianten van het bewijs

Het bewijs van de stelling kent meerdere varianten. In deze sectie zullen 3 verschillende varianten aan bod komen, namelijk varianten van de wiskundigen Simeon Ball en Johan Jacob Seidel.

Simeon Ball bewijs. Zij $\langle x\rangle$, $\langle y\rangle$, $\langle z\rangle$ drie punten van een ovaal $\mathcal{O}$ in $PG(2,q)$. Kies basis $\{x,y,z\}$. $X=(X_1,X_2,X_3)$ ten opzichte van deze basis. Dan zijn de tangenten aan deze punten respectievelijk:

$$\begin{aligned}\alpha_{21}X_2+\alpha_{31}X_3&=0\\ \alpha_{12}X_1+\alpha_{32}X_3&=0\\ \alpha_{13}X_1+\alpha_{23}X_2&=0\end{aligned}$$

Zij $\langle s\rangle\in\mathcal{O}\setminus\{\langle x\rangle,\langle y\rangle,\langle z\rangle\}$. Definieer de lijn die $\langle z\rangle$ en $\langle s\rangle$ verbindt als $s_2X_1-s_1X_2=0$, met $s=(s_1,s_2,s_3)$ onder basis $\{x,y,z\}$.

Omdat $\mathcal{O}$ een ovaal is, bevat de verzameling

$$\{\frac{s_2}{s_1}|\langle s\rangle\in\mathcal{O}\setminus\{\langle x\rangle,\langle y\rangle,\langle z\rangle\}\}\cup\{-\frac{\alpha_{13}}{\alpha_{23}}\}$$

elk niet nul element van $\mathbb{F}_q$. Nu is af te leiden:

$$-\frac{\alpha_{13}}{\alpha_{23}}\prod_{\langle s\rangle\in\mathcal{O}\setminus\{\langle x\rangle,\langle y\rangle,\langle z\rangle\}}\frac{s_2}{s_1}=-1$$

Bekijk nu de tangentfuncties:

$$\begin{aligned}T_x(X)&=\alpha_{21}X_2+\alpha_{31}X_3\\ T_y(X)&=\alpha_{12}X_1+\alpha_{32}X_3\\ T_z(X)&=\alpha_{13}X_1+\alpha_{23}X_2\end{aligned}$$

Hieruit volgt dat:

$$
\begin{aligned}
T_x(y) &= \alpha_{21} \cdot 1 + \alpha_{31} \cdot 0 = \alpha_{21} \\
T_x(z) &= \alpha_{21} \cdot 0 + \alpha_{31} \cdot 1 = \alpha_{31} \\
T_y(x) &= \alpha_{12} \cdot 1 + \alpha_{32} \cdot 0 = \alpha_{12} \\
T_y(z) &= \alpha_{12} \cdot 0 + \alpha_{32} \cdot 1 = \alpha_{32} \\
T_z(x) &= \alpha_{13} \cdot 1 + \alpha_{23} \cdot 0 = \alpha_{13} \\
T_z(y) &= \alpha_{13} \cdot 0 + \alpha_{23} \cdot 1 = \alpha_{23}
\end{aligned}
$$

en dit impliceert:

$$
\begin{aligned}
T_x(y) \prod s_3 &= T_x(z) \prod s_2 \\
T_y(x) \prod s_3 &= T_y(z) \prod s_1 \\
T_z(x) \prod s_2 &= T_z(y) \prod s_1
\end{aligned}
$$

en dus geldt:

$$T_x(y)T_y(z)T_z(x) = T_x(z)T_y(x)T_z(y) \tag{2.3.1}$$

Zij $\langle u \rangle$, $\langle v \rangle$ en $\langle w \rangle$ drie punten uit $\mathcal{O} \setminus \langle x \rangle$. Door middel van interpolatie met een eerstegraads polynoom in de variabele X ziet men:

$$T_x(X) = T_x(u)\frac{\det(X, v, x)}{\det(u, v, x)} + T_x(v)\frac{\det(X, u, x)}{\det(v, u, x)}$$

Substitutie van $X = w$ en wat herschikking geeft:

$$T_x(w)\det(u, v, x) + T_x(v)\det(w, u, x) + T_x(u)\det(v, w, x) = 0 \tag{2.3.2}$$

Permuteer de rolllen van x,u,v,w en zie dat ook geldt:

$$
\begin{aligned}
T_u(w)\det(x, v, u) + T_u(v)\det(w, x, u) + T_u(x)\det(v, w, u) &= 0 \\
T_v(w)\det(x, u, v) + T_v(u)\det(w, x, v) + T_v(x)\det(u, w, v) &= 0 \\
T_w(u)\det(x, v, w) + T_w(v)\det(u, x, w) + T_w(x)\det(v, u, w) &= 0
\end{aligned}
$$

Door (2.3.1) en (2.3.2) volgt dat:

$$T_w(x)\det(u,v,x)+T_v(x)\frac{T_w(v)}{T_v(w)}\det(w,u,x)+T_u(x)\frac{T_w(u)}{T_u(w)}\det(v,w,x)=0$$

en dus is

$$\begin{aligned}&\det(u,v,x)(T_w(u)\det(x,v,w)+T_w(v)\det(u,x,w))+\\&\frac{T_w(v)}{T_v(w)}\det(w,u,x)(T_v(w)\det(x,u,v)+T_v(u)\det(w,x,v))-\\&\frac{T_w(u)}{T_u(w)}\det(v,w,x)(T_u(w)\det(x,v,u)+T_u(v)\det(w,x,u))=0\end{aligned}$$

Herschik de vergelijking (met (2.3.1) voor de derde coëfficient). Dit resulteert in:

$$\begin{aligned}&2T_w(u)\det(u,v,x)\det(x,v,w)+2T_w(v)\det(u,v,x)\det(u,x,w)+\\&2T_v(u)\frac{T_w(v)}{T_v(w)}\det(w,u,x)\det(w,x,v))=0\end{aligned}$$

In basis $\{u,v,w\}$ is te zien dat een willekeurig punt $\langle x\rangle\in\mathcal{O}$ voldoet aan de kegelsnedevergelijking:

$$2T_w(u)x_3x_1+2T_w(v)x_3x_2+2T_v(u)\frac{T_w(v)}{T_v(w)}x_2x_1=0$$

□

Johan Jacob Seidel gebruikt de stelling van Desargues als lemma om de stelling van Segre te bewijzen.

Lemma 2.3.1 (Desargues)**.** *In $PG(2,q)$ liggen de driehoek gevormd door 3 punten van de ovaal en de driehoek gevormd door de tangenten door deze punten perspectivisch.*

Bewijs van het lemma: Neem $\langle e_1\rangle$, $\langle e_2\rangle$, $\langle e_3\rangle$ en $\langle p\rangle$ op $\mathcal{O}$. Zij de vergelijkingen van de lijnen ten opzichte van deze basis $\langle e_i,p\rangle(i=1,2,3)$ gegeven door:

$$\begin{aligned}x_2&=\lambda_1x_3\\x_3&=\lambda_2x_1\\x_1&=\lambda_3x_2\end{aligned}$$

met $\lambda_1\lambda_2\lambda_3 = 1$. Zij ook de tangenten aan $\langle e_1\rangle$, $\langle e_2\rangle$ en $\langle e_3\rangle$ gegeven door:

$$t_1 : x_2 = \alpha_1 x_3$$
$$t_2 : x_3 = \alpha_2 x_1$$
$$t_3 : x_1 = \alpha_3 x_2$$

Laat $\langle p\rangle$ alle punten van $\mathcal{O} \setminus \{\langle e_1\rangle, \langle e_2\rangle, \langle e_3\rangle\}$ doorlopen. λ_i zal dan alle elementen van $\mathbb{F}_q$ behalve 0 en α_i doorlopen. Dus geldt:

$$\alpha_i \prod \lambda_i = -1$$

Als deze vergelijkingen voor alle i vermenigvuldigd worden, volgt dat:

$$\alpha_1\alpha_2\alpha_3 = -1$$

Dus de lijnen die $\langle e_i\rangle$ en $t_j \cap t_k$ verbinden ($\{i, j, k\} = \{1, 2, 3\}$) gaan door 1 punt. Noem dit punt U. Men kan de configuratie zo kiezen dat $U = \langle e_1 + e_2 + e_3\rangle$ is. Dan zal zijn:

$$\alpha_1 = \alpha_2 = \alpha_3 = -1$$

□

Met het bewijs van Lemma 2.3.1 gaat Johan Jacob Seidel verder met het bewijzen van de stelling van Segre.

Johan Jacob Seidel bewijs. Zoals in Lemma 2.3.1, kies $\langle e_1\rangle$, $\langle e_2\rangle$, $\langle e_3\rangle$, $\langle e_1 + e_2 + e_3\rangle$. De ovaal $\mathcal{O}$ en de kegelsnede

$$\mathcal{C} = \{\langle x\rangle \in PG(2, q) | x_1x_2 + x_2x_3 + x_3x_1 = 0\}$$

hebben in de punten $\langle e_i\rangle$ de tangent t_i met $\alpha_i = -1$. Voor het bewijzen van Segre moet alleen nog bewezen worden dat $(\langle p\rangle \in \mathcal{O}) \Rightarrow (\langle p\rangle \in \mathcal{C})$. Zij $P = \langle(p_1, p_2, p_3)\rangle$ een punt op $\mathcal{O}$ en $t : b_1x_1 + b_2x_2 + b_3x_3 = 0$ de tangent aan dit punt. Met het lemma van Desargues toegepast op de driehoek $\langle e_i\rangle$, $\langle e_j\rangle$, $\langle p\rangle$ is te zien dat

$$b_2(p_1+p_2) = b_3(p_1+p_3)$$
$$b_1(p_2+p_1) = b_3(p_2+p_3)$$
$$b_1(p_3+p_1) = b_2(p_3+p_2)$$

Men heeft de punten e_i, $i = 1,2,3$, met de tangent door e_i gelijk aan $x_j = -x_k$ (waarbij $\{i,j,k\} = \{1,2,3\}$. Verder is het punt $P = (p_1,p_2,p_3)$, met tangent $b_1x_1 + b_2x_2 + b_3x_3 = 0$. Bekijk de driehoek $\Delta(e_1,e_2,P)$. Hieruit gaan we een betrekking vinden, en door de indices $1,2,3$ cyclisch te permuteren, vinden we nog twee betrekkingen. We bepalen de snijpunten S_i van de tangenten $t_1 :\ x_2 = -x_3$, $t_2 :\ x_3 = -x_1$ en $t_p :\ b_1x_1 + b_2x_2 + b_3x_3 = 0$, met de overstaande zijden $\langle e_2,P\rangle$, $\langle P,e_1\rangle$ en $\langle e_1,e_2\rangle$:

$$S_1 = (p_1 : -p_3 : p_3), \quad S_2 = (-p_3 : p_2 : p_3), \quad S_3 = (b_2 : -b_1 : 0).$$

Zie dat $S_1 = P - (p_2+p_3)e_2$, $S_2 = P - (p_1+p_3)e_1 \hat{=} e_1 - \frac{1}{p_1+p_3}P$ en tenslotte $S_3 = -b_1e_2 + b_2e_1 \hat{=} e_2 - \frac{b_2}{b_1}e_1$. Uit het lemma van Desargues volgt nu dat het product van deze drie coëfficiënten gelijk is aan -1 of te wel, zonder alle mintekens:

$$(p_2+p_3)\cdot\frac{1}{p_1+p_3}\cdot\frac{b_2}{b_1} = 1 \ \text{ dus } \ \frac{b_1}{p_2+p_3} = \frac{b_2}{p_1+p_3}.$$

Door de coördinaten cyclisch te verwisselen ziet men dat deze verhouding ook gelijk is aan $\frac{b_3}{p_1+p_2}$. Dus:

$$(b_1 : b_2 : b_3) = (p_2+p_3 : p_3+p_1 : p_1+p_2)$$

Tenslotte geldt omdat $b_1p_1 + b_2p_2 + b_3p_3 = 0$ dat P voldoet aan de kwadratische vergelijking: $(p_2+p_3)p_1 + (p_3+p_1)p_2 + (p_1+p_2)p_3 = 0$, dus na uitwerken en delen door 2 is:

$$p_1p_2 + p_2p_3 + p_3p_1 = 0.$$

□

2.4 Constructie van het bewijs

De twee varianten van het bewijs gaan alletwee een andere kant op, alleen komen ze toch tot dezelfde conclusie: als q oneven is dan moet een ovaal een kegelsnede zijn in $PG(2,q)$.

Om de bewijzen goed te begrijpen is het belangrijk te weten wat het idee of de constructie van dit bewijs inhoudt.

De eerste stap van alle drie bewijzen is het laten zien dat de Desargues eigenschap geldt. Bij het bewijs van Johan Jacob Seidel staat het letterlijk als lemma, bij het bewijs van Simeon Ball komt het terug bij tangentenvergelijking (2.3.1).

De volgende denkstap om de stelling van Segre te bewijzen is door in te zien dat als er een ovaal door drie punten P_1, P_2, P_3 gaat, waarvan de tangenten aan P_1 en P_2 gelijk zijn aan de tangenten aan P_1 en P_2 van een kegelsnede door P_1, P_2, P_3, dan moet de tangent aan P_3 van de ovaal en de kegelsnede ook hetzelfde zijn. Een kegelsnede door P_1, P_2, P_3 is dan van de vorm $aXY + bYZ + cXZ = 0$, wat ook de uiteindelijke vorm van de vergelijking is bij alle drie de bewijzen.

Voor de laatste denkstap van het bewijs, herinner dat er $q + 1$ punten moeten zijn zowel bij de ovaal als de kegelsnede. In de drie bewijzen zijn er al drie punten gekozen en kan de vergelijking zo aangepast worden zodanig dat hij ook nog door elk ander punt kan gaan. Dit levert een stelsel polynoomvergelijkingen op, waarvan de affiene variëteit nog steeds een kegelsnedevergelijking is.

Zoals al eerder opgemerkt, kent de stelling veel bewijzen. Sommige wiskundigen als Johan Jacob Seidel hebben een voorkeur voor het vroeg beginnen in de notatie van kegelsnede-vergelijkingen en als basis voor het vlak de eerste drie gekozen punten te nemen. Simeon Ball daarentegen rekent met determinanten van matrices in de variabelen, waar uiteindelijk een kegelsnedevergelijking in een andere basis uitkomt.

3 Lineaire MDS Codes

Dit hoofdstuk is bedoeld om de definities van MDS codes toe te lichten en de schakel met de projectieve meetkunde aan te tonen. Een foutcorrigerende code heet ook wel een *blokcode*. Bevat in de verzameling blokcodes zit de verzameling met alle *maximum distance separable (MDS) codes*, een speciaal geval van de blokcode.

3.1 Blokcodes

Laat S met $|S| = s$ een eindige verzameling (het zogenaamde *alfabet* zijn en n een positief geheel getal. S^n is dan de verzameling n-dimensionale vectoren met coëfficiënten uit S. Deze vectoren hebben een zekere afstand van elkaar. De afstand die bij blokcodes gebruikt wordt is de *Hamming afstand*.

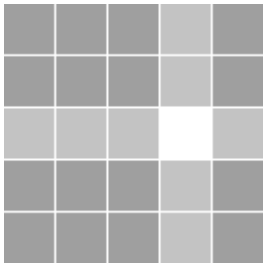
Definitie 3.1.1. *De Hamming-afstand* (Richard W. Hamming, 1915-1998) $d(x, y)$ *tussen twee vectoren* x *en* y *uit* S^n *is gelijk aan het aantal coëfficiënten waarin ze verschillen. Dus*

$$d(x, y) = |\{i|x_i \neq y_i\}|$$

Definieer de bol met straal $r \in \mathbb{N}_+$ *rond een punt* $x \in S^n$ *als*

$$B_r(x) = \{b \in S^n | d(x, b) < r\}$$

Voorbeeld 3.1.2. *Men kan de Hamming-afstand zich het best inbeelden door een twee- of driedimensionale kubus van blokjes in te denken, met één blokje voor elke* $x \in S^n$*. Als twee blokjes in eenzelfde rij of kolom liggen hebben ze afstand* 1*. Als ze minstens in een vlak liggen hebben ze afstand* 2*, enzovoorts.*



Figuur 4: Visualisatie van de Hamming-afstand

Stelling 3.1.3. *De Hamming afstand is een metriek, wat inhoudt dat:*

M1 $d(x,y) \geq 0$ *voor alle* $x, y \in S^n$

M2 $d(x,y) = 0 \Leftrightarrow x = y$ *voor alle* $x, y, \in S^n$

M3 $d(x,y) = d(y,x)$ *voor alle* $x, y \in S^n$

M4 $d(x,z) \leq d(x,y) + d(y,z)$ *voor alle* $x, y, z \in S^n$

Bewijs. Beschouw de definitie van de Hamming afstand als kardinaliteit van een verzameling, welke nooit negatief is. Dit voert tot (M1).

$d(x,y) = 0 \Leftrightarrow x$ en y verschillen op 0 indices $\Leftrightarrow x_i = y_i$ voor alle $i = 0, 1, \ldots, n \Leftrightarrow x = y$. (M2).

$d(x,y) = |\{i|x_i \neq y_i\}| = |\{i|y_i \neq x_i\}| = d(y,x)$. (M3).

$d(x,z)$ is het aantal coördinaten wat men moet veranderen om x in z te veranderen. Als men dit met een tussenstap doet (eerst verandert men x in y en daarna y in z), dan moet men minimaal in ieder geval coördinaten veranderen waar x en z verschillen plus de coördinaten waar x en y verschillen en vervolgens de coördinaten veranderen die van y z maken. (M4). □

Definitie 3.1.4. *De Hamming afstand van een $s \in S^n$ tot 0 heet het gewicht van s. Schrijf $wt(s)$. Dit kan gezien worden als het aantal coördinaten van s die ongelijk aan 0 zijn.*

Aan de hand van S^n kan de blokcode gedefinieerd worden.

Definitie 3.1.5. *Een blokcode ter lengte n is een verzameling $\mathcal{B} \subset S^n$. De elementen van $\mathcal{B}$ heten codewoorden. De minimale afand van een blokcode $\mathcal{B}$ is gelijk aan*

$$d := \min\{d(x,y)|x, y \in \mathcal{B}, x \neq y\}$$

Blokcodes worden in het algemeen gebruikt om boodschappen te reconstrueren die verzonden zijn over een kanaal waar de boodschap aan ruis is onderworpen, wat een vaak voorkomend verschijnsel is met de hedendaagse, draadloze communicatie. Een boodschap komt verwrongen aan, dus moet bepaald worden wat de originele boodschap is geweest. Stel iemand verzend de boodschap m, en de boodschap m' wordt ontvangen, dan zou een decoderingsalgoritme de $b \in \mathcal{B}$ kunnen bepalen zodanig dat $d(b,m') = \min\{d(x,m')|x \in \mathcal{B}\}$. Omdat m' dichter bij b moet liggen dan elk ander element uit $\mathcal{B}$, kan de code $\mathcal{B}$ hoogstens $\lfloor \frac{d-1}{2} \rfloor$ fouten corrigeren, mits het codewoord op hoogstens zoveel plaatsen verwrongen is. Maar hier meer over in Hoofdstuk 4.

Stelling 3.1.6. *Een blokcode $\mathcal{B} \subset S^n$ met minimale afstand d voldoet aan:*

$$|\mathcal{B}| \leq |S|^{n-d+1} = s^{n-d+1}$$

Bewijs. Neem $n - (d-1)$ coëfficiënten. Elke $b_1, b_2 \in \mathcal{B}$ moeten op deze plekken wel verschillen omdat anders $d(x,y) \geq d$. Dus $y \notin B_d(x)$. □

3.2 MDS codes

Een speciaal geval van de blokcode is de *maximum distance separable (MDS) code*.

Definitie 3.2.1. *Een MDS code is een blokcode $\mathcal{B} \subset S^n$ met minimale afstand d, waarvoor geldt dat $|\mathcal{B}| = |S|^{n-d+1} = s^{n-d+1}$.*

Een ander speciaal geval van de blokcode is de *lineaire blokcode*.

Definitie 3.2.2. *Een lineaire blokcode is een blokcode $\mathcal{B}$ die $\mathbb{F}_q$ als alfabet gebruikt zodat een lineaire combinatie van twee codewoorden ook weer een codewoord is. Laat de dimensie van het opspansel van de codewoorden k zijn, oftewel: $\mathcal{B}$ is een $[n, k, d]$-code in $\mathbb{F}_q$. Het aantal codewoorden in $\mathcal{B}$ is gelijk aan q^k en samen met Stelling 3.1.6 geeft ons dit de Singleton grens*

$$k \leq n - d + 1$$

welke behaald wordt bij MDS-codes, aldus Definitie 3.2.1.

Gevolg 3.2.3. *Het minimumgewicht van een lineaire code $\mathcal{B}$ is gelijk aan d, aangezien als $x, y \in \mathcal{B}$ met $x \neq y$ alle elementen van $\mathcal{B}$ afgaan, dan zal $x - y$ precies alle elementen van $\mathcal{B}$ afgaan die ongelijk aan 0 zijn.*

Deze twee verzamelingen van speciale blokcodes zijn niet disjunct. Dit houdt in dat er ook *lineaire MDS codes* bestaan.

Definitie 3.2.4. *Een lineaire MDS code is een $[n, n-d+1, d]$-code in $\mathbb{F}_q$. Deze bevatten dus q^{n-d+1} codewoorden*

Bij lineaire codes horen de zogenaamde *duale codes*.

Definitie 3.2.5. *De duale code van een lineaire code $\mathcal{B}$ is de verzameling*

$$\mathcal{B}^{\perp} = \{y \in S^n | x^T y = 0 \ \forall x \in \mathcal{B}\}$$

en is dus de annihilator van $\mathcal{B}$ met betrekking tot het inproduct.

Deze definitie van duale codes doet denken aan de nulruimte van een matrix. Dit is geen slecht idee, aangezien lineaire codes gegenereerd kunnen worden door middel van een *generatormatrix*.

Definitie 3.2.6. *Een generatormatrix G voor een lineaire $[n, k, d]$-code $\mathcal{B}$ ter lengte n is een $n \times k$-matrix zodanig dat $\{Gx | x \in \mathbb{F}_q^k\} = \mathcal{B}$. In de literatuur vindt men vaak de getranspo-neerde notatie. Vooral in de Verenigde Staten is het gebruikelijk om vectoren standaard als rijvectoren te zien. Daar gaat het dus om een $k \times n$-matrix zodanig dat $\{x^T G^T | x \in \mathbb{F}_q^k\} = \mathcal{B}$.*

Als men deze matrix vermenigvuldigt met een willekeurige vector, dan zal het beeld van die vector onder deze lineaire afbeelding altijd een element van $\mathcal{B}$ zijn. Zo verkrijgt men tevens makkelijk een basis voor $\mathcal{B}$. In de basis waarin na het kolomvegen de matrix in standaardvorm staat, is de matrix te schrijven als

$$G = \left(\frac{I_k}{A}\right)$$

met I_k de $k \times k$ identiteitsmatrix en A een $(n-k) \times k$-matrix afhankelijk van $\mathcal{B}$. Dat de kolommen onafhankelijk zijn moge duidelijk zijn.

Gevolg 3.2.7. *Een controlematrix H horende bij een generatormatrix G in standaardvorm is een matrix die de duale code van $\mathcal{B}$, $\mathcal{B}^\perp$, genereert. Dit is dus een $n \times (n-k)$-matrix en is van de vorm:*

$$H = \left(\frac{-A^T}{I_{n-k}}\right)$$

Hieruit is af te leiden dat de controlematrix van een lineaire code de generatormatrix is van de duale code en dat de controlematrix van de duale code weer de generatormatrix van de lineaire code is.

Stelling 3.2.8. *De duale code van een lineaire MDS code is een $[n, d-1, n-d+2]$-code.*

Bewijs. Omdat de duale code loodrecht op de originele lineaire MDS code staat (want de inproducten moeten gelijk aan 0 zijn volgens de definitie) moeten de dimensies van de code en zijn duale code samen tot n optellen. De dimensie van de duale lineaire MDS code is dus $n-(n-d+1) = d-1$. Nu moet er nog aangetoond worden dat de minimumafstand gelijk is aan $n-d+2$. Bij MDS codes wordt elke keer de minimumafstand behaald, dus is de minimumafstand gelijk aan het minimumgewicht.

$$\exists x \in \mathcal{B}^\perp : wt(x) \leq n-d+1$$

Omdat het een MDS code betreft, moet $\mathcal{B}$ q^{n-d+1} elementen bevatten, die elk afstand $n-d+1$ tot elkaar hebben. Ze verschillen dus op $n-d+1$ posities. Er zijn dus ook elementen uit $\mathcal{B}$ die precies op de plekken verschillen waar x_i niet nul is. Maar omdat $x \in \mathcal{B}^\perp$ en omdat x in een lineaire relatie staat tot de elementen van $\mathcal{B}$ die op dezelfde plekken als x ongelijk aan nul zijn, vormt dit een tegenspraak. □

3.3 Het MDS hoofdvermoeden

Het hoofdvermoeden over MDS codes brengt de codes samen met de eindige meetkunde uit Hoofdstuk 1. Dit doet hij met het begrip *boog*. Segre stelde hier uiteindelijk nog een belangrijke vraag over.

Definitie 3.3.1. *Een boog $\mathcal{A}$ is een verzameling van $p \geq r+1$ punten in $PG(r,q)$ met de eigenschap dat elk hypervlak op zijn hoogst r punten van $\mathcal{A}$ bevat. Dit heet ook wel een p-boog.*

Gevolg 3.3.2. *Door middel van een lineaire MDS code (een lineaire $[n, n-d+1, d]$-code) valt een n-boog te construeren in $PG(n-d,q)$. Van een n-boog in $PG(r,q)$ is een lineaire MDS $[n, r+1, n-r]$-code te construeren.*

Voorbeeld 3.3.3. *Voorbeeld 2.2.8 was een voorbeeld van een hyperovaal (en dus een boog met maximale lengte).*

$$\{\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_3 \\ \sum_{i=1}^{3} e_i \end{pmatrix} | e_i \in \mathbb{F}_2\} \subset PG(2,2)$$

is dus een lineaire MDS code. In het algemeen zal zelfs gelden dat

$$\{\begin{pmatrix} x_1 \\ \vdots \\ x_k \\ \sum_{i=1}^{k} x_i \end{pmatrix} | x_i \in \mathbb{F}_q\} \subset PG(k-1,q)$$

een lineaire MDS code ter lengte $k+1$ is.

In technische termen, betekent dit dat met de eigenschappen van een boog kunnen onderzocht kan worden hoeveel informatie over een kanaal verzonden kan worden. Als de minimumafstand d vast staat, wil men codes vinden met n zo groot mogelijk, zodat men meer informatie tegelijk kan verzenden. In de praktijk kan men niet willekeurig grote boodschappen in één keer verzenden zonder dat daar dan een meeschalende hoeveelheid ruis bij komt kijken. Meestal worden er dan een aantal bytes per keer verzonden. Als n dus vast staat, is het wenselijk om de minimumafstand d weer groot te houden, om zoveel mogelijk fouten te kunnen corrigeren. Omdat $r = n - d$ kan men de dimensie van de projectieve ruimte vaststellen en een zo groot mogelijke boog vinden. Op deze manier blijft r hetzelfde, maar kan er gemaximaliseerd worden in n en d.

Stelling 3.3.4. $\mathcal{A} \subset PG(r,q)$ *met*

$$\mathcal{A} = \{\langle(1, t, t^2, \ldots, t^r)\rangle | t \in \mathbb{F}_q\} \cup \{\langle(0, 0, \ldots, 0, 1)\rangle\}$$

is een boog met $q + 1$ *punten.*

Bewijs. Doorsnijd $\mathcal{A}$ met het hypervlak

$$\sum_{i=0}^{r} \alpha_i x_i = 0$$

Als $\alpha_r \neq 0$ dan heeft het hypervlak een punt gemeen met $\mathcal{A}$ voor elke t waarvoor geldt dat

$$\sum_{i=0}^{r} \alpha_i t^i = 0$$

welke hoogstens r oplossingen heeft, omdat het een polynoom in t is van graad r. Als α_r wel gelijk aan nul is, dan bevat het hypervlak het punt $\langle(0, 0, \ldots, 0, 1)\rangle$ van $\mathcal{A}$ en elk ander punt van $\mathcal{A}$ voor elke t die voldoet aan

$$\sum_{i=0}^{r-1} \alpha_i t^i = 0$$

welke op zijn hoogst $r - 1$ oplossingen heeft. □

Gevolg 3.3.5. *Voor elke priemmacht* q *en minimumafstand* d *is een lineaire MDS* $[q + 1, q + 2 - d, d]$*-code te maken. Als* $r = 2$*, is het makkelijk geschikte bogen te kiezen, aangezien dit gelijk de ovalen en dus de kegelsneden zijn.*

Op het gebied van de MDS-codes bestaat er een vermoeden over een mogelijke bovengrens voor de kardinaliteit van bogen in een projectieve ruimte. Dit wordt ook wel het *hoofdvermoeden over MDS-codes* genoemd.

Vermoeden (Hoofdvermoeden over MDS-codes)**.** *Zij* $\mathcal{A} \subset PG(r, q)$ *een boog met* $r \leq q - 1$*.*

1. *als* q *even is en* $r = 2$ *of* $q - 2$*, dan geldt* $|\mathcal{A}| \leq q + 2$
2. *als* q *oneven is of als* $3 \leq r \leq q - 3$ *dan geldt* $|\mathcal{A}| \leq q + 1$

Computerwerk heeft aangetoond dat het vermoeden in ieder geval geldt voor $q \leq 27$. In generalisatie van de stelling van Segre, is het vermoeden waar voor q priem. Zonder de specifieke voorwaarden geldt in het algemeen het volgende:

Stelling 3.3.6. *Zij $\mathcal{A} \subset PG(r,q)$ een boog. Dan heeft de kardinaliteit van de boog een zwakkere bovengrens, namelijk $|\mathcal{A}| \leq q + r$.*

Bewijs. Neem een willekeurig deelverzameling $\mathcal{U} \subset \mathcal{A}$ met $|\mathcal{U}| = r - 1$ Als de punten van $\mathcal{U}$ geen deelruimte van dimensie $r - 2$ opspannen dan kan men een hypervlak vinden die $r + 1$ punten van $\mathcal{A}$ bevat. Dit spreekt de definitie van een boog tegen. Als men wel een opgespande deelruimte van dimensie $r - 2$ heeft, dan is deze bevat in $q + 1$ hypervlakken. Die hypervlakken bevatten stuk voor stuk hoogstens één punt van $\mathcal{A} \setminus \mathcal{U}$, dus

$$|\mathcal{A}| \leq q + 1 + r - 1$$

□

Segre vroeg zich in 1955 het volgende af:

"Voor welke r en q bestaan er bogen van $q+1$ punten in $PG(r,q)$ die niet equivalent zijn aan de boog genoemd in Stelling 3.3.4?"

Segre wist dat als $r = 2$ of $r = q - 2$ en q even is, dat er andere voorbeelden waren. De bewijzen van deze gevonden voorbeelden kunnen, banaal gezegd, aan de lezer worden overgelaten.

Voorbeeld 3.3.7. *Segre bewees onder andere dat als q een even getal is dat geen kwadraat is dan is de verzameling*

$$\mathcal{A} = \{\langle(1, t, t^6)\rangle | t \in \mathbb{F}_q\} \cup \{\langle(0, 0, 1)\rangle\}$$

een boog met $q + 1$ punten en dus ook een ovaal. Deze boog is projectief gezien niet equivalent met de bogen uit Stelling 3.3.4.

Voorbeeld 3.3.8. *De wiskundige Glynn ontdekte in 1986 nog een niet klassieke boog. Als q oneven is, of als $3 \leq r \leq q - 3$. Deze heeft $q + 1$ punten.*

$$\mathcal{A} = \{\langle(1, t, t^2 + \alpha t^6, t^3, t^4)\rangle | t \in \mathbb{F}_9\} \cup \{\langle(0, 0, 0, 0, 1)\rangle\} \subset PG(4, 9)$$

is een voorbeeld hiervan. Deze boog heeft 10 punten. $\alpha \in \mathbb{F}_9$ is zodanig gekozen dat $\alpha^4 = -1$ geldt.

Betreffende het hoofdvermoeden en de vraag die Segre stelde, is er nog één stelling en een generalisatie daarvan. Segre zijn stelling volgt uit de eerste.

Stelling 3.3.9. *Zij $\mathcal{A} \subset PG(2,q)$ een boog en $\mathcal{A}^*$ de verzameling lijnen duaal aan $\mathcal{A}$. Zij*

$$k = \begin{cases} q+2-|\mathcal{A}| & \textit{als } q \textit{ even} \\ 2(q+2-|\mathcal{A}|) & \textit{als } q \textit{ oneven} \end{cases}$$

Dan is er een homogeen polynoom p (de termen hebben een gelijke som van exponenten) in drie variabelen van graad k zodanig dat de verzameling nulpunten van dat polynoom de verzameling punten bevat die precies op één lijn van $\mathcal{A}^*$ liggen.

Als q oneven is, dan geldt voor elk nulpunt op een lijn uit $\mathcal{A}^*$ zelfs dat p modulo die lijnvergelijking een nulpunt van graad 2 heeft in dat punt.

De generalisatie van $PG(2,q)$ naar $PG(r,q)$ is bewezen door Blokhuis, Bruen en Thas.

Stelling 3.3.10. *Zij $\mathcal{A} \subset PG(r,q)$ een boog en $\mathcal{A}^*$ de verzameling hypervlakken duaal aan $\mathcal{A}$. Zij*

$$k = \begin{cases} q+r-|\mathcal{A}| & \textit{als } q \textit{ even} \\ 2(q+r-|\mathcal{A}|) & \textit{als } q \textit{ oneven} \end{cases}$$

Dan is er een homogeen polynoom p (de termen hebben een gelijke som van exponenten) in $r+1$ variabelen van graad k zodanig dat de verzameling nulpunten van dat polynoom de verzameling punten bevat die precies op $r-1$ hypervlakken van $\mathcal{A}^*$ liggen.

Als q oneven is, dan geldt voor elk nulpunt welke op $r-1$ hypervlakken van $\mathcal{A}^*$ ligt, dat er een lijn is zodanig dat deze de doorsnede is van die $r-1$ hypervlakken en dat nulpunt. p modulo die lijnvergelijking heeft dan een nulpunt van graad 2 in dat punt.

4 Toepassingen

Dit hoofdstuk is vooral bedoeld om een schets van de praktijk te geven. Er zullen voorbeelden komen van blokcodes die in de technologie vaak worden gebruikt. Merk op dat in de praktijk, en dan vooral bij digitale doeleinden waar men met bits werkt, $q = p = 2$ is. Deze blokcode heet dan ook wel een binaire blokcode.

Bij foutcorrectie is het uiteraard van belang dat een code fouten kan corrigeren, maar alleen correctie is op zichzelf niet nuttig als men niet weet in wat voor gedaanten de ontvangen data moet zijn. Een code moet aan de hand van de binnengekomen data en ook fouten weten op te merken. Een code $\mathcal{B} \subset \mathbb{F}_q^n$ met minimumafstand d kan in de praktijk het volgende:

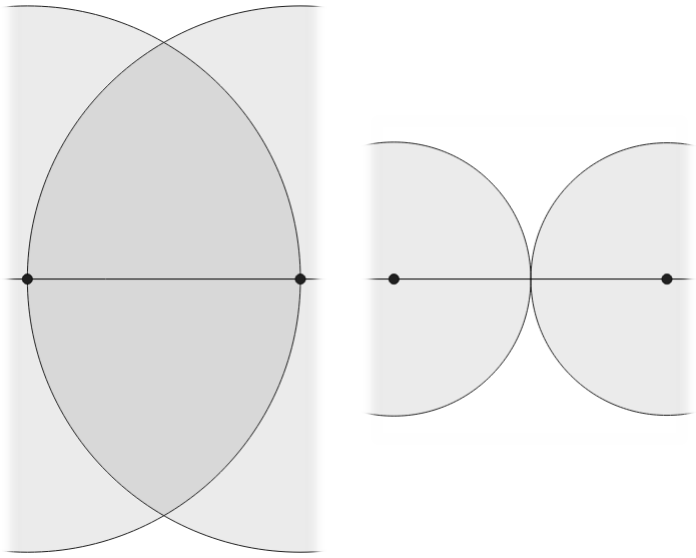
C1 $\mathcal{B}$ kan maximaal $d-1$ *fouten opmerken*. Ga na dat een codewoord b het enige codewoord is in de Hamming bol om b met straal $d-1$. Als de ontvangen boodschap geen element is van $\mathcal{B}$, kan de code dus $d-1$ fouten opmerken. Totale correctie is echter niet gegarandeerd.

C2 Zoals in Hoofdstuk 1 al vermeld staat, kan $\mathcal{B}$ elke deelverzameling van $\lfloor \frac{d-1}{2} \rfloor$ *fouten corrigeren*. De twee Hamming bollen op twee codewoorden met straal $\lfloor \frac{d-1}{2} \rfloor$ overlappen niet, dus als een codewoord ontvangen wordt en er worden hoogstens $\lfloor \frac{d-1}{2} \rfloor$ fouten opgemerkt, dan kan men de dichtste buur decodering gebruiken (Bijlage A) om al die fouten te corrigeren. Als er meer dan $\lfloor \frac{d-1}{2} \rfloor$ verstoringen zijn aangebracht, kan men nog lijstdecodering of hoogste waarschijnlijkheid decodering hanteren.

C3 $\mathcal{B}$ kan (hoogstens) $d-1$ *uitschrappingen corrigeren*. Stel dat er een paar posities van de boodschap niet leesbaar zijn. De desbetreffende posities zijn dus bekend. Er moet een invulling zijn voor de gewiste symbolen zodanig dat de afstand hoogstens $d-1$ is.

In de praktijk krijgt de dimensie van de opgespannen deelruimte k van een code meer betekenis, aangezien deze dimensie staat voor de lengte van de boodschap. De dimensie n van S^n, de ruimte van vectoren ter lengte n met elementen van het alfabet als coëfficienten, heet ook wel de bloklengte. In digitale termen: Een code heeft k databits en $n-k$ controlebits. Efficiënte codes houden dus $n-k$ klein, aangezien men zo weinig mogelijk extra symbolen wil toevoegen zodanig dat de k datasymbolen gecodeerd kunnen worden. De *rato* van een code wordt dan ook aangegeven door

$$R = \frac{k}{n}$$



Figuur 5: Het foutopmerkbereik (links) en foutcorrectiebereik (rechts)

4.1 Reed-Solomon Codes

Om de Reed-Solomon goed te kunnen begrijpen, is het belangrijk om te weten dat er *cyclische codes* bestaan.

Definitie 4.1.1. *Een cyclische blokcode $\mathcal{B}$ is een blokcode met de eigenschap dat alle cyclische permutaties van een codewoord ook weer codewoorden van $\mathcal{B}$ zijn.*

Voorbeeld 4.1.2. *Zij $\mathcal{B}$ een cyclische code ter lengte 4 met $q = p = 2$. Zij $1100 \in \mathcal{B}$. Hieruit volgt dat 0110, 0011 en 1001 ook tot $\mathcal{B}$ behoren.*

Het eerste voorbeeld van een in de praktijk vaak gebruikte code is de *Reed-Solomon code* (Irving S. Reed, 1923-2012 en Gustave Solomon, 1930-1996). Ze beschreven een systematische manier om codes te kiezen die willekeurige symboolfouten kunnen opsoren en corrigeren.

Definitie 4.1.3. *De Reed-Solomon Code, afgekort $RS(n, k)$, is een lineaire, cyclische, non-binaire $[n, k, n-k+1]$-code (en dus een MDS code) met $\mathbb{F}_q$ als alfabet die gebruik maakt van het toevoegen van $n-k$ redundante elementen zodanig dat de onjuistheid van $n-k$ posities opgemerkt kan worden, $\lfloor \frac{n-k}{2} \rfloor$ fouten gecorrigeerd kunnen worden en $n-k$ uitschrappingen hersteld kunnen worden. De code kan ook een combinatie van fouten en uitschrappingen herstellen.*

De Reed-Solomon Code vindt zijn toepassing op verscheidene technologische vlakken. Een CD-speler gebruikt bijvoorbeeld twee Reed-Solomon codes achter elkaar om de leesfouten die de laser maakt te kunnen corrigeren. Ook achter een DVD en de inmiddels verouderde, slechts industrieel gebruikte Digital Audio Tape (DAT) schuilt zulks een techniek.

Deze techniek heet *Cross-Interleaved Reed-Solomon Coding*, afgekort *CIRC*. Een vaak gebruikte technologie waarbij de Reed-Solomon codes ook gebruikt worden is de streepjescode. Bijna alle streepjescodes die gelezen worden door een lezer, zijn erop gebouwd zodanig dat een Reed-Solomon code ze kan onderscheiden. Ook de onbemande ruimtesondes van het Amerikaanse Voyager programma gebruikte Reed-Solomon codering voor de foto's die ze maakten.

De Reed-Solomon Code werkt conceptueel als een code die een boodschap $m \in \mathbb{F}_q^k$ codeert naar een $m' \in \mathbb{F}_q^n$ middels het invullen van n verschillende elementen van $\mathbb{F}_q$ in het polynoom. Merk op dat $n \leq q$ moet gelden. De coëfficienten van dit polynoom zij gelijk aan de coëfficienten van m.

$$p_m(x) = \sum_{i=1}^{k} m_i x^{i-1}$$

Voorbeeld 4.1.4. *Zij $q = p = 5$. Zij $n = q = 5$ en $k = 3$.Stel dat de boodschap $m = 341$ verzonden moet worden. Dan is*

$$p_m(x) = 3 + 4x + x^2$$

Nu kan men alle $n = q$ elementen van het alfabet $\mathbb{Z}_5$ op volgorde invullen en zien dat

$$m' = \begin{pmatrix} 3 + 4 \cdot 0 + 0^2 \\ 3 + 4 \cdot 1 + 1^2 \\ 3 + 4 \cdot 2 + 2^2 \\ 3 + 4 \cdot 3 + 3^2 \\ 3 + 4 \cdot 4 + 4^2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 0 \\ 4 \\ 0 \end{pmatrix}$$

Overigens vereist het niet veel werk om, gegeven dat het polynoom ingevuld wordt door $a_1, a_2, \ldots, a_n \in \mathbb{F}_q$, in te zien dat de $n \times k$ generatormatrix G gelijk is aan

$$G = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{k-1} \end{pmatrix}$$

Deze matrix heet ook wel de *Vandermonde matrix* (Alexandre-Théophile Vandermonde, 1735-1796).

4.2 Hamming Codes

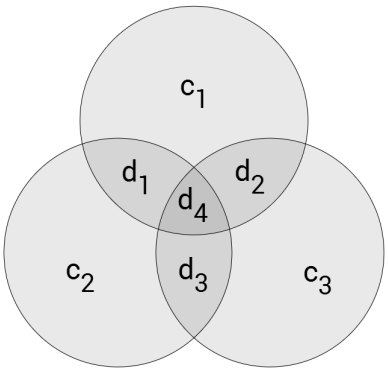
Een andere in de praktijk gebruikte code is de *Hamming code*.

Definitie 4.2.1. *De Hamming(7,4) code is een lineaire code die 4 bits codeert in 7 bits, door 3 controlebits toe te voegen. De grootte van het alfabet is dus $q = p = 2$. Als de ontvangen boodschap een fout in 1 bit heeft, kan deze altijd gecorrigeerd worden. De code kan ook alle fouten opmerken als er hoogstens 2 fouten in zitten. De minimumafstand van deze code is dus $d = 3$. Dat maakt deze code tot een $[7, 4, 3]$-code.*

De Hamming(7,4) code werd uitgevonden door Richard Hamming omdat hij zich ergerde aan het feit dat ponskaarten die destijds veel gebruikt werden in computers. Ook werden ze bijvoorbeeld vroeger gebruikt om orgelstukken op te slaan, waarvan de lange ponskaart door het orgel gehaald kon worden.

De code voegt de controlebits toe op posities die een macht van 2 zijn. In dit geval dus de posities 1, 2 en 4. Zij d_1, d_2, d_3 en d_4 de databits en c_1, c_2 en c_3 de controlebits. Het verzonden gecodeerde bericht is dus $c_1c_2d_1c_3d_2d_3d_4$. Verder geldt voor deze code dat

- c_1 de bits d_1, d_2 en d_4 controleert.
- c_2 de bits d_1, d_3 en d_4 controleert.
- c_3 de bits d_2, d_3 en d_4 controleert.



Figuur 6: Cirkeldiagram van de controlebits van de Hamming(7,4) code

In matrixvorm, met een 1 als c_i d_j controleert, is dit te schrijven als

$$C := \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Als deze rijen op posities 1, 2 en 4 worden ingevoerd tussen de rijen van de identiteitsmatrix, verkrijgt men de generatormatrix

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

De controlematrix H wordt op eenzelfde manier gemaakt, alleen worden nu de rijen van de identiteitsmatrix op posities 1, 2 en 4 ingevoerd tussen de rijen van C^T. Ga na dat

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Voorbeeld 4.2.2. *Zij $m = 1011$ de boodschap die verzonden moet worden. Om deze boodschap veilig te verzenden conform Hamming(7,4) codering, vermenigvuldigt men m met G.*

$$Gm = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

en dan ziet men dat 1011 gecodeerd wordt naar 0110011. De verzending vindt plaats en er worden 7 bits ontvangen. Er is dan een tweede eigenschap van de controlematrix die nu van pas komt, namelijk de mogelijkheid tot het vervaardigen van de syndroomvector, die in binaire representatie aangeeft op welke positie er een fout is gemaakt, mits er maar op hoogstens 1 positie een fout is opgetreden. Stel dat er geen fouten zijn gemaakt. Dan is inderdaad

$$H^T Gm = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Stel dat er wel een fout is gemaakt op slechts 1 positie. Dan is de ontvangen boodschap te schrijven als $Gm + e_i$ voor een zekere i. Vermenigvuldiging met H^T levert op:

$$H^T(Gm + e_i) = H^T Gm + H^T e_i = H^T e_i$$

en dit is precies de i^e kolom van H^T. Zoals ook aan H^T te zien is, is nu bevestigd dat H^T precies de getallen 1 tot 7 in binaire representatie van de kolommen bevat. De code kan dus zelf de positie aanwijzen waar er een fout is opgetreden en die daarna gelijk corrigeren.

Nadat de eventuele fout uit de ontvangen boodschap is gehaald, moet de boodschap nog gedecodeerd worden tot de originele 4 databits. Men dient alleen de drie controlebits ervan tussen te halen. De decoderingsmatrix die hoort bij deze operatie is

$$D = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In 1950 generaliseerde Richard Hamming zijn eigen Hamming(7,4) code naar de algemene Hamming code.

Definitie 4.2.3. *De generalisatie van de Hamming(7,4) code is de Hamming code. De generalisatiestap die gemaakt wordt is de variabiliteit van de parameters n en k. Zij $r \geq 2$. De Hamming code is dan een $[2^r - 1, 2^r - r - 1, 3]$-code. De rato van deze code is dan*

$$R = \frac{2^r - r - 1}{2^r - 1} = 1 - \frac{r}{2^r - 1}$$

4.3 Hadamard Codes

Het derde voorbeeld van een in de praktijk gebruikte, lineaire code is de *Hadamard Code* (Jacques Hadamard, 1865-1963). Deze code staat ook bekend als de *Walsch Code* (Joseph L. Walsh, 1895-1973) of *Walsh-Hadamard Code*. Jacques Hadamard heeft deze code niet

uitgevonden, maar de code wordt in het merendeel van de literatuur naar hem vernoemd omdat hij de generatormatrices voor deze codes (de Hadamard matrices) toevallig al veel eerder had geïntroduceerd en bestudeerd. De wiskundige Joseph L. Walsh gebruikte voor het eerst deze theorie voor een code.

Definitie 4.3.1. *De Hadamard Code is een lineaire, binaire foutcorrectiecode die k bits in blokken van 2^k bits codeert, met minimale afstand $d = 2^{k-1}$. Dat maakt de Hadamard Code een $[2^k, k, 2^{k-1}]$-code.*

De Hadamard Code wordt onder andere gecontrueerd via het alom bekende inproduct. zij $x \in \{0,1\}^k$ het te coderen codewoord. x wordt dan gecodeerd naar de rij der inproducten met alle $y \in \{0,1\}^k$, waarbij y in binaire representatie de getallen 0 tot $2^k - 1$ doorloopt.

$$Had(x) = \left(\left(\sum_{i=0}^{k} x_i y_i \right) \mod 2 \right)_{y \in \{0,1\}^k}$$

De generatormatrix van deze code bevat de oplopende binaire getallen als rijen. Echter, het is makkelijk te zien dat deze manier van coderen redelijk overdadig is.

Voorbeeld 4.3.2. *zij $k = 3$. Met deze parameter komt dit dus neer op een lineaire, binaire $[8,3,4]$-code. Nu is:*

$$Had(x) = Gx = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ x_3 \\ x_2 \\ x_2 + x_3 \\ x_1 \\ x_1 + x_3 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \end{pmatrix}$$

Het is te zien dat de eerste 4 posities van Gx hetzelfde zijn als de laatste 4, op de term x_1 na. Overigens bevat de eerste 4 posities geen informatie over x_1. Op deze manier kan x soms niet helemaal gedecodeerd worden.

Het is om deze reden dat de code efficiënter gemaakt werd, door alle y met een 0 als eerste bit weg te laten. Deze verbeterde code heet de *Punctured Hadamard Code*. De codewoorden verkrijgt men dan door

$$pHad(x) = \left(\left(\sum_{i=0}^{k} x_i y_i \right) \mod 2 \right)_{y \in \{1\} \times \{0,1\}^{k-1}}$$

De Punctured Hadamard Code is een $[2^{k-1}, k, 2^{k-2}]$-code. Deze code kan dus $2^{k-2} - 1$ fouten opmerken of uitschrappingen herstellen en $\lfloor \frac{2^{k-2}-1}{2} \rfloor = \lfloor 2^{k-3} - \frac{1}{2} \rfloor$ fouten corrigeren. De rato van deze code is

$$R = \frac{k}{2^{k-1}}$$

Hadamard Codes werden gebruikt in de jaren '70, toen de ruimtesonde Mariner 9 de eerste sonde werd die in een baan om een andere planeet werd gelanceerd en foto's in 64 grijswaarden (6 bits) naar de aarde stuurde.

Bibliografie

[1] A.E. Brouwer, *Ovals and conics in a finite projective plane*, http://www.win.tue.nl/~aeb/2WF02/ovals.pdf. Laatst geraadpleegd: 2 mei 2014.

[2] A.E. Brouwer, *Chevalley-Warning, quadrics*, http://www.win.tue.nl/~aeb/2WF02/ProjGeom.pdf. Laatst geraadpleegd: 2 mei 2014.

[3] S. Ball, *MDS Codes*, http://www.cimpa-icpam.org/archivesecoles/20140205111906/mds.pdf. Laatst geraadpleegd: 2 mei 2014.

[4] S. Ball & Zs. Weiner, *An Introduction to Finite Geometry*, 3rd edition, 2011.

[5] V. Chu, *History of Hamming Codes*, http://biobio.loc.edu/chu/web/Courses/Cosi460/hamming_codes.htm. Laatst geraadpleegd: 28 juli 2014.

[6] A.M. Cohen, F.G.M.T Cuypers & H.J.M. Sterk, *Algebra Interactive*, Springer-Verlag, Berlijn, 1999.

[7] D.A. Cox, J. Little & D. O'Shea, *Ideals, Varieties and Algorithms*, Springer-Verlag, Berlijn, 3rd edition, 2007.

[8] P. Garrett, *Check Matrices*, http://www.math.umn.edu/~garrett/coding/Overheads/15_check_mces.pdf. Laatst geraadpleegd: 16 juni 2014.

[9] K.A.S. Immink, *Reed-Solomon Codes and the Compact Disc*, in *Reed-Solomon Codes and their Applications*, bewerkt door V.K. Bhargava & S.B. Wicker.

[10] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, 1978.

[11] V.G. Maz'ya & T.O. Shaposhnikova, *Life and Work of Jacques Hadamard*, American Mathematical Society, 1998.

[12] T.K. Moon, *Error Correction Coding*, John Wiley & Sons, New Jersey, 2005.

[13] G.L. Mullen & D. Panario, *Handbook of Finite Fields*, CRC Press, London, Versiedatum: 15 mei 2013.

[14] J.J. O'Connor & E.F. Robertson, *Beniamino Segre*, http://www-history.mcs.st-andrews.ac.uk/history/Biographies/Segre_Beniamino.html. Laatst geraadpleegd: 2 mei 2014.

[15] I.S. Reed & G. Solomon, *Polynomial codes over certain finite fields*, Journal of the Society of Industrial and Applied Mathematics, Volume 8, 1960.

[16] J.J. Seidel, *Eindige Meetkunde*, Technische Hogeschool Eindhoven, 1982.

[17] J.P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973.

[18] A.A. Stoorvogel, *Matrixtheorie*, Technische Universiteit Eindhoven, 2004.

[19] B.M.M. de Weger, *Discrete Mathematics 2*, Technische Universiteit Eindhoven, versie 0.55, 2012.

A Algoritmen

Deze bijlage is bedoeld als begeleiding bij de verschillende behandelde onderwerpen. Dit is om de lezer te begeleiden in het begrijpen van de theorie van de vorige hoofdstukken. Omdat in het curriculum van de Bachelor Technische Wiskunde ook programmeren valt, zal voor elk algoritme naast de wiskundige procedure ook een Mathematica-code gegeven worden.

Opmerking: een streep tussen de code scheidt de verschillende invoervelden, zodat deze apart uitgevoerd kunnen worden. Dit om het onderscheid te maken tussen de parameterinvoer en de verschillende delen die tot uitvoer kunnen leiden.

Inverse in $\mathbb{F}_q$

In Hoofdstuk 1 staat hoe het eindige lichaam $\mathbb{Z}_p$ te generaliseren is naar een eindig lichaam $\mathbb{F}_q$ met $q = p^m$ een priemmacht. De truc om deze abstractiestap te begrijpen is om het principe van elementen in de vorm van getallen los te laten en het te zien als $\mathbb{Z}_p[X]/f_m(X)$, de verzameling restpolynomen van graad hoogstens $m-1$ over $\mathbb{Z}_p$, $f_m(X) \in \mathbb{Z}_p[X]$ irreducibel en van graad m. $f_m(X)$ is dan de modulus, welke min of meer de elementen koppelt aan hun inverse. De wortel α werd gebruikt om uit te breiden naar $\mathbb{F}_q$. Inverteren gaat op de volgende manier:

Procedure. *Neem een element $a \in \mathbb{F}_q$. Gezocht: een b uit $\mathbb{F}_q$ zodanig dat $ab = 1$. a en b zijn unieke restpolynomen. Zij $a = \sum_{i=0}^{m-1} a_i X^i$ en $b = \sum_{i=0}^{m-1} b_i X^i$. Merk op dat de $a_i \in \mathbb{Z}p$ bekend zijn en de $b_i \in \mathbb{Z}p$ niet. Vermenigvuldig a met b. ab is een polynoom over $\mathbb{Z}_p$ van graad hoogstens $2m-2$. Reduceer ab modulo $f_m(X)$. Dit kan door, elke keer als het resultaat nog geen polynoom van graad hoogstens $m-1$ is, de modulus zodanig te vermenigvuldigen en af te trekken van het product, dat de leidende term van het resterende product wegvalt. Het resultaat is een polynoom van hoogstens een graad lager. Het reductieproces zal dus eindigen.*

Mathematica-voorbeeld (Een inverse in $\mathbb{F}_q$)**.**

```
p = 2;
element = x^2;
modulus = 1 + x + x^3;
d = Exponent[modulus, x];
inverse = a + b*x + c*x^2;
```

```
If[PrimeQ[p],
 If[Exponent[element, x] < Exponent[modulus, x],
  If[IrreduciblePolynomialQ[modulus, Modulus -> p],
   If[Exponent[inverse, x] == Exponent[modulus, x] - 1,

    prod = element*inverse;
    prod = Expand[prod];
    prod = PolynomialMod[prod, modulus];
    eqlist = CoefficientList[prod, x];
    solvelist = ConstantArray[0, d - 1];
    solvelist = Prepend[solvelist, 1];
    s = Solve[eqlist == solvelist];
    solution = inverse /. s[[1]];
    solution = PolynomialMod[solution, p]

    , Print[inverse geen graad minder dan modulus]]
   , Print[modulus niet irreducibel]]
  , Print[element graad niet kleiner dan d]]
 , Print[p geen priem]]
```

In de bovenste regels kan de gebruiker de karakteristiek p, het te inverteren element en de irreducibele modulus definiëren. Er staat tevens een inverse klaar met onbekende coëfficienten die opgelost zullen worden. Er wordt gecontroleerd of p wel priem is, de modulus irreducibel is, de graad van het element wel strikt kleiner is dan die van de modulus en of de onbekende inverse een graad lager is dan die van de modulus. Als alles klopt wort vervolgens het element met zijn nog onbekende inverse vermenigvuldigd. Dit product wordt gereduceerd. Er wordt een stelsel vergelijkingen opgesteld die opgelost worden. Als laatste stap wordt de oplossing gesubstitueerd in de inverse, wat resulteert in een opgeloste inverse. De inverse is oplosbaar omdat er evenveel vergelijkingen en onbekenden zijn als de graad van de modulus.

In het voorbeeld wordt X^2 geïnverteerd. Zijn inverse is $X^2 + X + 1$, immers

$$X^2(X^2 + X + 1) = X^4 + X^3 + X^2 = (X + 1)(X^3 + X + 1) + 1$$

Generatormatrices

Definitie 3.2.6 beschrijft het principe van een generatormatrix G. Deze matrix valt onder andere te construeren door de elementen van $\mathcal{B}$ als kolommen te gebruiken. Omdat G vermenigvuldigd wordt met een willekeurige kolomvector in $\mathbb{F}_q^k$, zal de uitkomst altijd een lineaire combinatie zijn van de kolommen en laat nu net geëist zijn dat lineaire combinaties van codewoorden ook weer codewoorden zijn. Door overigens de kolommen te vegen, verschijnen er veel nulkolommen die wegvallen. Omdat het een [n,k,d]-code is blijft er een $n \times k$-matrix over.

Procedure. *Beschouw alle elementen $b_i \in \mathcal{B}$, een [n,k,d]-code. Laat nu $G = (b_1\ b_2\ \ldots)$ de matrix zijn met als kolommen alle b_i. Volgens de definities spant deze matrix een k-dimensionale deelruimte op. Als er geveegd wordt met de kolommen naar de standaardvorm, blijft er een matrix $(b'_1\ b'_2\ \ldots\ b'_k\ 0\ \ldots\ 0)$, met $\{b'_1, b'_2, \ldots, b'_k\}$ een basis voor $\{b_i | b_i \in \mathcal{B}\}$ zodanig dat*

$$G = \left(\frac{I_k}{A}\right)$$

Mathematica-voorbeeld (Een willekeurige lineaire code)**.**

```
p = 3;
n = 3;
k = 2;
gi = IdentityMatrix[k];
ga = Table[RandomInteger[p - 1], {n - k}, {k}];
g = Join[gi, ga]
```

```
Do[
 v = Table[Mod[Floor[i/(p^(k - x))], p], {x, k}];
 w = g.v;
 w = Table[Mod[w[[j]], p], {j, 1, n}];
 Print[w];
, {i, 0, (p^k) - 1}]
```

De gebruiker kan p, n en k naar behoeven aanpassen. Ook wordt er een willekeurige matrix aangemaakt conform de vorm van G, middels het samenvoegen van een k-identiteitsmatrix en een willekeurige matrix A met coëfficienten uit $\mathbb{Z}_p$.

Men kan de programmacode die G construeert en daarmee de code geneert, uitbreiden met zijn controlematrix H, die de duale code genereert.

Mathematica-voorbeeld (Een willekeurige lineaire code en zijn duale)**.**

```
p = 3;
n = 3;
k = 2;
gi = IdentityMatrix[k];
ga = Table[RandomInteger[p - 1], {n - k}, {k}];
g = Join[gi, ga]
hi = IdentityMatrix[n - k];
ha = -Transpose[ga];
h = Join[ha, hi]
```

```
Do[
 v = Table[Mod[Floor[i/(p^(k - x))], p], {x, k}];
 w = g.v;
 w = Table[Mod[w[[j]], p], {j, 1, n}];
 Print[w];
, {i, 0, (p^k) - 1}]
```

```
Do[
 t = Table[Mod[Floor[i/(p^(n - k - x))], p], {x, n - k}];
 u = h.t;
 u = Table[Mod[u[[j]], p], {j, 1, n}];
 Print[u];
, {i, 0, (p^(n - k)) - 1}]
```

Decoderen met dichtste buur

Een van de gegeven decoderingsalgoritmen uit Hoofdstuk 4 is het decoderingsalgoritme wat gebruik maakt van de Hamming afstand. Dit algoritme bepaalt het meest aannemelijke codewoord in $\mathcal{B}$ aan de hand van de verwrongen boodschap die ontvangen is. 'Meest aannemelijk' wordt hier gedefinieerd als 'met de minste afstand'. Voor de begrijpelijkheid wordt $q = p$ aangenomen. Als men $q = p = 2$ kiest, werkt het programma in werkelijkheid met bits, zoals in de praktijk.

Procedure. *Zij $x \in \mathbb{Z}_p^n$ de ontvangen boodschap. Uit alle $y \in \mathcal{B}$ moet de Hamming afstand $d(x, y)$ uitgerekend worden en vervolgens moet het element met de minste afstand aan de gebruiker worden gegeven.*

Mathematica-voorbeeld (Een decodering met minimumafstand)**.**

```
p = 2;
n = 8;
k = 3;
gi = IdentityMatrix[k];
ga = Table[RandomInteger[p - 1], {n - k}, {k}];
g = Join[gi, ga];
x = Table[RandomInteger[p - 1], {n}]
```

```
u = {};
Do[v = Table[Mod[Floor[i/(p^(k - l))], p], {l, k}];
 w = g.v;
 w = Table[Mod[w[[j]], p], {j, 1, n}];
 u = Append[u, w];
, {i, 0, (p^k) - 1}]
dtable = Table[HammingDistance[u[[m]], x], {m, 1, p^k}];
mintable = Position[dtable, Min[dtable]];
Do[
  Print[u[[mintable[[i, 1]]]]]
, {i, Length[mintable]}]
```

Opnieuw kan de gebruiker $q = p$, n en k naar behoeven instellen en een willekeurige lineaire code genereren middels een willeukeurige generatormatrix. Ook wordt er een $x \in \mathbb{Z}_p^n$ als ontvangen boodschap aangemaakt. Er wordt tevens een lijst aangemaakt met alle elementen van die code. Stuk voor stuk wordt de Hamming-afstand berekend tussen de elementen van de code en x. Het programma vindt de posities in de lijst waar de afstand het kleinst is en geeft deze terug aan de gebruiker.