CHALMERS

GÖTEBORGS UNIVERSITET

Felkorrigerande koder

Reed-Solomon-kodernas algebraiska egenskaper och tillämpning för Quick Response-koder

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Klara Carlström
Fredrik Davidsson
Victor Jonsson
Ahmed Mohamadi

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2018

Felkorrigerande koder

Reed-Solomon-kodernas algebraiska egenskaper och tillämpning för Quick Response-koder

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Fredrik Davidsson

Kandidatarbete i matematik inom civilingenjörsprogrammet Automation och Mekatronik vid Chalmers
Ahmed Mohamadi

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers
Klara Carlström Victor Jonsson

Handledare: Jan Stevens

Examinator: Maria Roginskaya/Marina Axelson-Fisk

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2018

Populärvetenskaplig presentation

En QR-kod (Quick Response) är en slags tvådimensionell streckkod som från början togs fram av fordonsindustrin i Japan på 1990-talet för att hjälpa dem att hålla ordning på sitt lager. QR-kodens användningsområde har med tiden vuxit och används idag i många olika sammanhang såsom reklambranchen, bankväsendet och identitetshandlingar för att nämna några. Då QR-koder dyker upp över allt i vardagen där de kan utsättas för flera olika påfrestningar, såsom kladd på själva koden eller repor på avläsarens kameralins, så måste koden vara tålig och kunna återställa information som kanske gått förlorad.

Att en kod kan reparera sig själv låter ju lite konstigt men det är här de så kallade Reed-Solomon-koderna (RS-koder) kommer in. Dessa matematiska koder, framtagna 1960 av Irving S. Reed och Gustave Solomon är så kallade felkorrigerande koder som används inom all möjlig informationsteknik där fel lätt uppstår, såsom repor på CD-(Compact Disc) eller DVD-skivor (Digital Video Disc). Det som gör RS-koder speciella är att de är väldigt effektiva mot så kallade klusterfel, det vill säga när felen ligger nära varandra och inte är slumpmässiga, denna typ av fel är just vad repor på CD- och DVD-skivor är. Framför allt är det just den typ av fel som vanligen uppstår när det finns något förhinder med en QR-kod som till exempel att det är något i vägen för mobiltelefonkameran när den försöker avläsa koden.

Att en QR-kod bygger på RS-koder är just det som gör det möjligt för en mobiltelefon att avläsa koden utan att förutsättningarna är helt perfekta. För att förstå hur en RS-kod fungerar så måste man bryta ned den i dess beståndsdelar och för att kunna göra detta krävs att man lär sig lite algebra och inledande kodningsteori.

Kodningsteori är en relativt liten och relativt ung gren inom matematiken, den första artikeln i det som skulle komma att bli kodningsteori publicerades år 1948 och titulerades ”A Mathematical Theory of Communication” författad av Claude Shannon. Denna artikel publicerades i tidsskriften *Bell System Technical Journal* och handlade om problemet hur man bäst kodar ett meddelande som en sändare vill skicka. Den person som banat vägen för kodningsteori är dock matematikern Richard Hamming, genom sitt arbete inom felkorrigerande koder på 1940-talet och framåt; flertalet koncept inom denna falang av matematiken har namngivits efter honom.

Kodningsteori är som namnet antyder en gren inom matematiken som handlar om koder och hur felkorrigerande kodning effektivast kan göras. Koder och kodning appliceras inom många olika områden såsom IT (informationsteknik), elektroteknik och datateknik. Kodning brukar delas in i fyra kategorier och dessa är datakomprimering, felkorrigering, kryptografi och linjekodning. Detta arbete handlar mer än något annat om felkorrigerande koder då det är den typen av kodning som huvudsakligen måste förstås för att kunna förstå uppbyggnaden av en QR-kod.

Sammandrag

Att skydda ett meddelande från fel som kan uppstå under en överföringsprocess är något som måste göras vid all möjlig data- och informationskommunikation. Reed-Solomon-koder är en klass av felkorrigerande koder som gör just detta. Det här arbetetet innehåller en matematisk härledning av Reed-Solomon-kodernas optimala egenskaper samt en implementering av meddelandet KODNINGSTEORI i form av en QR-kod (Quick Response). Vi förklarar begrepp såsom kod och felkorrigering och studerar algebraiska begrepp inom ring-och kroppteori samt cykliska polynomkoder och primitiva polynom. Vi bevisar matematiskt att Reed-Solomon-koderna är optimala i den mening att de uppfyller Singletons gräns. Den QR-kod som har implementeras har kapacitet att korrigera upp till 15 % felaktig indata med hjälp av en Reed-Solomon-kod.

Abstract

Protecting a message from error during a transfer process is something that has to be done during all kinds of data and information communication. Reed-Solomon codes are a class of error correcting codes that accomplish this. This thesis contains a mathematical deduction of the Reed-Solomon codes' optimal properties as well as an implementation of the message KODNINGSTEORI ("Coding theory" in Swedish) in the form of a QR (Quick Response) code. We explain concepts such as code and error correction and study algebraic concepts within ring and field theory as well as cyclic polynomial codes and primitive polynomials. We prove mathematically that the Reed-Solomon codes are optimal as they meet the Singleton bound. The QR code that has been implemented has the capacity to correct up to 15 % incoming errors with the help of a Reed-Solomon code.

Förord

Författarna, som studerar på programmen Teknisk matematik samt Automation och mekatronik vid Chalmers tekniska högskola och på Matematikprogrammet vid Göteborgs universitet är tacksamma till docent Jan Stevens för hans handledning och respons under arbetets gång.

Under arbetet har en loggbok förts över samtliga författares individuella bidrag. Denna loggbok är inte inkluderad här men nedan redovisas gruppmedlemmarnas individuella bidrag till respektive textavsnitt i rapporten, samt den huvudsakliga ansvarsfördelningen under vilken arbetet har genomförts.

I listan nedan anges vem eller vilka av författarna, utan inbördes ordning, som huvudsakligen bidragit till författandet av respektive avsnitt.

- Populärvetenskaplig presentation: Victor
- Abstract, sammandrag: Victor
- Förord: Klara
- 1: Klara
- 2: Klara
- 3.1-3.3: Fredrik
- 3.4: Klara
- 4: Klara
- 5.1-5.2: Ahmed/Victor
- 5.3: Victor
- 5.4-5.5: Ahmed
- Bilaga A: Fredrik
- Bilaga B: Victor

Utöver uppdelning i huvudsakligt författande av text har arbetet genomförts under nedanstående ansvarsfördelning (utan inbördes ordning):

- Informationssökning och studie avseende QR-koden samt implementering av denna: Ahmed
- Stuide av teorin om algebraiska koder: Klara och Fredrik
- Studie av sambandet mellan implementeringen av en QR-kod och den algebraiska teorin om felkorrigerande koder: Victor och Klara
- Huvudsakligt redaktionellt ansvar: Klara

Etiska aspekter

Författarna har gemensamt gjort bedömningen att inga nämnvärda etiska aspekter har varit angelägna att beakta i och med arbetet med den här rapporten.

Innehåll

1. Inledning

I dagens samhälle överförs information över många olika typer av kommunikationskanaler, exempelvis genom kopparledningar, via optiska fiber, och olika typer av trådlös överföring. Gemensamt för alla fysiska överföringskanaler är förekomsten av brus under överföringen, det vill säga uppkomsten av extra signaler som gör att den ursprungliga informationen modifieras. Med ökad användning av elektronisk utrustning och datorer i samhället följer ett växande behov av att kunna överföra stora mängder av digital information korrekt och effektivt.

För att handskas med uppkomsten av brus som kan modifiera den överförda informationen används felkorrigerande koder. Dessa härleds med hjälp av ett område inom matematiken som kallas för kodningsteori. Sedan kodningsteorins begynnelse på 1940-talet har många olika slags koder utvecklats, bland annat för att tillfredställa en växande mängd tillämpningar. *Reed-Solomon-koder* (RS-koder) är en klass av felkorrigerande koder som har spelat en huvudroll inom utvecklingen av telekommunikation, och möjliggjort såväl överföring av bilder från yttre rymden som lagring av information på CD-skivor (Compact Disc).

En specifik tillämpning av Reed-Solomon-koder vars utbredda användning har vuxit lavinartat de senaste åren är lagring och kommunikation av information med så kallade QR-koder. Akronymen QR står för *Quick Response* och koden används till allt från ursprungsmärkning av varor till marknadsföring av företag och validering av elektroniska biljetter. Användningen av Reed-Solomon-kod i en QR-kod möjliggör att QR-koden kan avläsas trots att exempelvis fläckar har uppstått eller att en del av QR-koden saknas eller har ersatts med en företagslogotyp.

Trots många konkreta tekniska tillämpningar är kodningsteori en matematisk teori och många av verktygen som används för att förklara de olika kodernas egenskaper är av abstrakt karaktär. Reed-Solomon-koderna presenterades först av Irving S. Reed och Gustave Solomon i artikeln *Polynomial Codes Over Certain Finite Fields* år 1960 [6] och koderna är just polynomkoder över ändliga kroppar. Det finns huvudsakligen två olika varianter av Reed-Solomon-koder vilka brukar kallas för den ursprungliga varianten respektive BCH-varianten (efter R. Bose, D.K. Ray-Chaudhuri, och A. Hocquenghem). Den här rapporten redogör för konstruktionen av den populära BCH-varianten av Reed-Solomon-koderna (när hänvisning görs till Reed-Solomon-koderna är det således BCH-varianten som avses) genom en teoretisk undersökning av kodernas matematiska uppbyggnad samt genom en implementering av en QR-kod från grunden.

Den befintliga litteraturen om konstruktionen av QR-koder är knapphändig och sällan av matematisk karaktär, trots att konstruktionen av QR-koder bygger på algebra. Det huvudsakliga syftet med arbetet är därför att beskriva konstruktionen av en QR-kod ur ett matematiskt perspektiv, med särskilt fokus på användningen av felkorrigerande koder.

1.1. Arbetets uppbyggnad

Rapporten utgörs av fyra huvudsakliga delar, motsvarande fyra olika kapitel; i den första delen introduceras konceptet kodningsteori genom exempel på några enkla feldetekterande och felkorrigerande koder. Den andra delen innehåller några definitioner och grundläggande resultat inom algebra som behövs för att förstå teorin om konstruktionen av algebraiska koder. Den tredje delen utgörs av en redogörelse för den algebraiska konstruktionen av

Reed-Solomon-koder. Till sist visas i rapportens fjärde och avslutande del hur en Reed-Solomon-kod används för att implementera en QR-kod från grunden. En stor del av arbetet med rapporten har bestått i att erhålla en förståelse av den grundläggande algebra som kodningsteori bygger på. En avvägning har gjorts över vilka definitioner och resultat som har varit nödvändiga att inkludera i rapporten; övriga grundläggande, relevanta resultat inom algebra återfinns i bilaga A.

2. Introduktion till kodningsteori

I de här kapitlet introduceras grundläggande kodningsteori samt begreppet felkorrigerande kod. Först ges några enkla exempel på koder och sedan introduceras blockkoder. De algebraiska strukturerna *grupp*, *ring*, och *kropp* definieras och i kapitlets avslutande del introduceras slutligen ett antal begrepp som används för att jämföra olika koders egenskaper.

2.1. Feldetekterande och felkorrigerande koder

I de flesta digitala kommunikationssystem översätts information till binära sekvenser, det vill säga följder av ettor och nollor. Exempelvis kan ett meddelande X översättas till 0 och ett meddelande Y till 1. Om det finns brus i kanalen kan konsekvensen bli att en etta omvandlas till en nolla eller tvärtom. Detta är problematisk eftersom då man tar emot exempelvis meddelande X inte vet huruvida det ursprungligen var X eller i själva verket Y som skickades.

När en binär sekvens tas emot vill man kunna upptäcka och i vissa fall rätta eventuella fel som uppstått under överföringen. Ett sätt att uppnå detta är att koda meddelandet som ska skickas med hjälp av antingen en *feldetekterande* eller *felkorrigerande* kod. En av de enklaste varianterna av en feldetekterande kod är användning av en så kallad *jämn paritetsbit*. En paritetsbit är en binär siffra som visar om antalet ettor i en binär sekvens är jämnt eller udda. När jämn paritetbit används så räknas antalet ettor i sekvensen och om antalet är udda så läggs en etta till i slutet av sekvensen. Om antalet är jämnt så läggs en nolla till istället. Följdaktligen har sekvenser innehållande en jämn paritetsbit alltid ett jämnt antal ettor. Om ett udda antal bitar har omvandlats under överföringen av en sekvens innehållande en paritetsbit så kan det alltså upptäckas. Det framgår inte vilka bitar som är fel när meddelandet tas emot men det kan upptäckas att något fel har uppstått i sekvensen och det kan då begäras att meddelandet skickas igen.

Denna metod används i sammanhang där det är möjligt att skicka samma meddelande flera gånger på begäran, till exempel i vissa hårdvaruapplikationer. I många sammanhang är detta dock inte möjligt. I exempelvis överföringar från satelliter och rymdsonder skulle den extra utrustningen som krävs för att kunna lagra och återsända meddelanden utgöra onödig vikt. Då används istället felkorrigerande koder som inte kräver att information som blivit skadad av brus skickas igen utan möjliggör korrigering av de fel som uppstått [4]. Felkorrigerande koder används även bland annat på digitala lagringsenheter som CD- och DVD-skivor (Digital Video Disc) och inte minst i QR-koder [1][7].

Ett mycket enkelt exempel på en felkorrigerande kod är *repetitionskoden*. Repetitionskoden upprepar meddelandet som ska skickas flera gånger; om två meddelanden X och Y ska skickas så översätts de alltså först till binär form, 0 respektive 1, men istället för att bara skicka 0 eller 1 så skickas något av kodorden 000 respektive 111 (det här är alltså en

repetitionskod som upprepar meddelandet tre gånger):

$$\begin{aligned} \text{X} &\mapsto 0 \mapsto 000, \\ \text{Y} &\mapsto 1 \mapsto 111. \end{aligned}$$

Under antagandet att högst ett fel kan uppstå under överföringen, det vill säga högst en etta kan omvandlas till en nolla eller tvärtom, kan den här repetitionskoden korrigera det eventuella felet som uppstår. Om meddelande X skickas kan mottagaren nämligen nås av någon av följande sekvenser: $000, 100, 010, 001$; under antagandet att det finns högst ett fel samt att 000 och 111 är de enda kodorden som finns är det tydligt att samtliga fyra sekvenser härstammar från meddelandet X. Det samma gäller naturligtvis om Y skickas. Den här repetitionskoden sägs därför *korrigera ett fel*.

I tabell 1 illustreras skillnaden mellan en kod med jämn paritetsbit och repetitionskoden när de används för att koda meddelanden av längd 2. Paritetsbitkoden resulterar i kodord av längd 3 och repetitionskoden har kodord av längd 6. Att en kostnad i form av extra bitar följer av att koden ska kunna korrigera fel gäller i allmänhet; avvägningen mellan att uppnå en hög felkorrigeringskapacitet och att hålla nere kodordens längd är ett centralt problem när felkorrigerande koder ska väljas [5].

Tabell 1: I tabellen visas ett exempel på en kod med jämn paritetsbit (andra kolumnen) respektive en repetitionskod som upprepar meddelandet tre gånger (tredje kolumnen). Meddelanden som kodas är av längd 2 (första kolumnen).

Meddelande	**Kodord med paritetsbit**	**Kodord i repetitionskoden**
00	000	00 00 00
10	101	10 10 10
01	011	01 01 01
11	110	11 11 11

2.2. Blockkoder

En kod kallas för en blockkod om den kodade informationen kan delas in i block med n symboler vardera, och dessa block kan avkodas oberoende av varandra. Dessa block utgör kodorden och n är ordlängden (eller bara längden av koden). För att koda ett meddelande som ska överföras över någon kanal med hjälp av en blockkod delas meddelandet upp i block bestående av k *informationssymboler* vardera, och kodas genom att $n - k$ *kontrollsymboler* läggs till varje block. Detta resulterar i ett kodord bestående av n *kodsymboler*; en sådan kod kallas för en (n, k)-kod och k kallas för kodens dimension. Både repetitionskoden och koden med paritetsbit ovan är blockkoder. Repetitionskoden i tabell 1 är en $(6, 2)$-kod och exemplet med paritetsbit i samma tabell utgör en $(3, 2)$-kod.

Det gäller generellt för en blockkod att information kodas med hjälp av ett alfabet Q med q olika symboler. Vi låter Q^n beteckna mängden av alla sekvenser $(x_1, x_2, ..., x_n)$ sådana att $x_1, x_2, ..., x_n \in Q$. Vi antar att informationen vi vill koda utgörs av en mängd X av meddelanden och vi låter C beteckna mängden av alla kodord. Mängden C är då enligt

resonemanget ovan en delmängd till Q^n och formellt kan en blockkod betraktas som en funktion

$$X \to C \subseteq Q^n.$$

Hädanefter i den här rapporten, när en allmän felkorrigerande *kod* nämns, avses en delmängd C till någon mängd Q^n där Q utgör kodens alfabet. I fallen med repetitionskoden samt paritetsbitkoden tidigare i kapitelt utgörs alfabetet av $Q = \mathbb{Z}_2 = \{0, 1\}$ och koderna är delmängder till $Q^n = \mathbb{Z}_2^n$ vilket är mängden av alla binära sekvenser av längden n; både repetitionskoden och paritetsbitkoden kallas således för binära koder.

När ett ord tas emot i änden av en kommunikationskanal kontrollerar avkodaren om det motsvarar ett kodord. Om en felkorrigerande kod används och det visar sig att det mottagna ordet inte är ett kodord avkodas ordet till det kodord som med störst sannolikhet var det som skickades, precis som i exemplet med repetitionskoden. I en (n, k)-kod är det ursprungliga meddelandet som nämnt k symboler långt; om meddelandena utgörs av symboler från ett alfabet med storleken q finns alltså ur en mottagares perspektiv q^k möjliga meddelanden eftersom varje informationssymbol kan vara någon av de q symbolerna i alfabetet. Om vi antar att vi har en kod som gör att varje meddelande kodas till ett unikt kodord består koden alltså av q^k kodord. De kodade ord som når mottagaren har å andra sidan n symboler så om godtyckliga symboler kan omvandlas på grund av störningar under överföringen så finns q^n olika ord som kan nå mottagaren, varav endast q^k stycken är giltiga kodord.

Inom kodningsteori studeras framför allt koder som har en stark *algebraisk struktur*. En algebraisk struktur är en mängd med en eller flera tillhörande operatorer som verkar under en kombination av olika matematiska grundantaganden- eller axiom. Mängden $\mathbb{Z}_2 = \{0, 1\}$ tillsammans med operationen addition modulo 2 utgör en algebraisk struktur som kallas för en *grupp*. Denna struktur samt strukturerna *ring* och *kropp* är centrala i arbetet med algebraiska koder.

Definition 2.2.1. En *grupp* är en algebraisk struktur bestående av en mängd element G och en operation $\star$ sådan att gruppen är sluten under operationen och följande tre axiom är uppfyllda.

(i)	$a \star (b \star c) = (a \star b) \star c \quad , \forall\, a, b, c \ \in G$	(associativitet),
(ii)	$\exists \ \ e \in G : a \star e = e \star a = a \quad , \forall a \in G$	(neutralt element e),
(iii)	$\forall\, a \in G \quad \exists\, b \in G : a \star b = b \star a = e$	(existens av invers).

Om det för varje $a, b \in G$ gäller att $a \star b = b \star a$, så sägs G vara en *abelsk* grupp.

Definition 2.2.2. En *ring* är en algebraisk struktur bestående av en mängd element R och två operationer, addition $(a + b)$ och multiplikation (ab), sådana att följande tre villkor är uppfyllda:

(i)	R med avseende på addition är en abelsk grupp,
(ii)	Multiplikationen är associativ,
(iii)	$a(b + c) = ab + ac$ och $(a + b)c = ac + bc \ \forall\ a, b, c \in R$.

Om det för varje $a, b \in R$ gäller att $ab = ba$, så sägs R vara en *kommutativ* ring.

Definition 2.2.3. En *kropp* är en kommutativ ring vars mängd av nollskilda element bildar en grupp med avseende på multiplikation. En ändlig kropp med q element betecknas som $\mathbb{F}_q$.

Algebraiska koder konstrueras över ett alfabet Q som utgörs av den ändliga kroppen $\mathbb{F}_q$, så att kodorden är sekvenser av symboler i $\mathbb{F}_q$. En kod med längden n över $\mathbb{F}_q$ utgör en delmängd till det n-dimensionella linjära rummet $\mathbb{F}_q^n$; kodorden kan således betraktas som vektorer i $\mathbb{F}_q^n$. Addition över $\mathbb{F}_q^n$ sker modulo q och utförs i övrigt som vanlig vektoraddition.

Definition 2.2.4. En kod C över $\mathbb{F}_q$ är *linjär* om C är ett linjärt underrum till $\mathbb{F}_q^n$.

Linjäritet är en viktig egenskap av de flesta koder som studeras inom kodningsteori och innebär att varje linjärkombination av kodord också är ett kodord.

2.3. Grundläggande egenskaper av koder

En central fråga inom kodningsteori är hur en kod konstrueras — och framför allt hur en bra kod konstrueras. För att ta reda på detta krävs kunskap om vad som menas med en ”bra kod”. Om en felkorrigerande kod eftersöks kan ett första steg vara att kräva att koden ska kunna korrigera ett visst antal fel. Repetitionskoden som konstruerades tidigare i kapitlet hade förmåga att rätta ett uppkommet fel, medan paritetsbitskoden kunde upptäcka när ett udda antal fel hade uppstått men inte avgöra vilken bit som blivit fel och därmed inte rätta felet. Frågan är vad repetitionskoden har för egenskap som gör att den kan rätta ett fel. Det handlar helt enkelt om att de existerande kodorden, 000 och 111, skiljer sig såpass mycket åt så att givet att endast ett fel kan ha uppstått så går det att ta reda på vilket kodord den mottagna vektorn härstammar från. Detta brukar inom kodningsteori uttryckas som att avståndet mellan kodorden är tillräckligt stort.

Definition 2.3.1. Om $\mathbf{x} \in Q^n$, $\mathbf{y} \in Q^n$, så är *avståndet* $d(\mathbf{x}, \mathbf{y})$ mellan $\mathbf{x}$ och $\mathbf{y}$

$$d(\mathbf{x}, \mathbf{y}) := |\{i \,|1 \leq i \leq n,\, x_i \neq y_i\}|.$$

En kod som bara består av ett enda kodord kallas för en trivial kod. Avståndet $d(\mathbf{x}, \mathbf{y})$ kallas för *Hamming-avståndet* och är en egenskap av ett par av vektorer.

Definition 2.3.2. *Det minsta avståndet* av en (icke-trivial) kod C är

$$d(C) = \min\{d(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x}, \boldsymbol{y} \in C, \boldsymbol{x} \neq \boldsymbol{y}\}.$$

För kodord $\boldsymbol{x}, \boldsymbol{y} \in Q^n$ representerar Hamming-avståndet $d(\boldsymbol{x}, \boldsymbol{y})$ ett faktiskt avstånd mellan elementen $\boldsymbol{x}$ och $\boldsymbol{y}$ i den mening att det utgör en metrik över Q^n. Hamming-avståndet $d(\boldsymbol{x}, \boldsymbol{y})$ är nämligen en funktion $d : Q^n \times Q^n \to \mathbb{R}$ sådan att, för alla element $\boldsymbol{x}, \boldsymbol{y} \in Q^n$, följande villkor är uppfyllda:

$$\begin{array}{lll} (i) & \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) \geq 0 & \text{(icke-negativitet)}, \\ (ii) & \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{d}(\boldsymbol{y}, \boldsymbol{x}) & \text{(symmetri)}, \\ (iii) & \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y} & \text{(Leibniz lag)}, \\ (iv) & \mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) \leq \mathrm{d}(\boldsymbol{x}, \boldsymbol{z}) + \mathrm{d}(\boldsymbol{z}, \boldsymbol{y}) & \text{(triangelolikheten)}. \end{array} \tag{1}$$

Att de tre första villkoren är uppfyllda följer trivialt ur definitionen av $d(\boldsymbol{x}, \boldsymbol{y})$; triangelolikheten bevisar vi nu.

Bevis. (Triangelolikheten för avståndet) Vi ska bevisa att olikheten $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$ gäller för alla element $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in Q^n$. Vi gör det genom att observera att vänsterledet $d(\boldsymbol{x}, \boldsymbol{y})$ är antalet i sådana att $x_i \neq y_i$ och kan betraktas som antalet komponenter i $\boldsymbol{x}$ som måste ändras för att få vektorn $\boldsymbol{y}$. Vi kan vidare se det som att en omvandling av $\boldsymbol{x}$ till $\boldsymbol{y}$ kan göras antingen i ett steg, genom att ändra alla x_i sådana att $x_i \neq y_i$ till y_i, eller i två steg genom att först ändra några komponenter i $\boldsymbol{x}$ så att resultatet blir en

vektor $\boldsymbol{z}$ och sedan ändra de komponenter z_j i $\boldsymbol{z}$ sådana att $z_j \neq y_j$ till y_j. Det totala antalet komponenter som behöver ändras i metoden med två steg är då precis högerledet $d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$ (vi kallar den för HL-metoden) och antalet ändringar i metoden med ett steg är vänsterledet (vi kallar den för VL-metoden). Vi kan nu observera att om $\boldsymbol{z}$ är en vektor som erhålls genom att några, men inte alla, av de x_i sådana att $x_i \neq y_i$ ändras till y_i så är likhet i triangelolikheten uppfylld, för då är metoden med ett steg och metoden med två steg ekvivalenta. Om $\boldsymbol{z}$ inte är en sådan vektor så betyder det antingen att $\boldsymbol{z} = \boldsymbol{x}$ (vilket är trivialt) eller att några komponenter som ändras för att gå från $\boldsymbol{x}$ till $\boldsymbol{z}$ måste ändras igen för att komma till $\boldsymbol{y}$, vilket gör att *en* (eller ingen) ändring i VL-metoden svarar mot *två* ändringar i HL-metoden; att en ändring behövs i HL-metoden innebär dock alltid att precis en ändring krävs även VL-metoden och ingen ändring i HL-metoden motsvarar ingen ändring i VL-metoden; HL är således större än eller lika med VL.

□

En annan användbar egenskap av koder är *minsta vikten* av en kod, som definieras nedan.

Definition 2.3.3. Med *vikten* av ett kodord $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ menas

$$w(\boldsymbol{x}) = \text{antalet } i \text{ sådana att } x_i \neq 0.$$

Med *minsta vikten av en kod* C menas den minsta vikten av alla nollskilda kodord i C.

För repetitionskoden C, som korrigerar ett fel, så gäller att $d(C) = 3$ då kodorden 000 och 111, vilka utgör hela koden, skiljer sig på tre ställen. Denna insikt kan formuleras i allmänna termer; en 1-felkorrigerande kod är en mängd av vektorer C i Q^n sådan att avståndet mellan olika kodord i C är minst lika med 3, det vill säga $d(C) \geq 3$. Sats 2.3.1 etablerar ett generellt samband mellan antalet fel som en kod korrigerar och kodens minsta avstånd; först definieras vad som menas med att en kod *korrigerar t fel*.

Definition 2.3.4. En kod C sägs *detektera t fel* om det, då t eller färre fel har inträffat under överföringen, kan upptäckas att ett mottaget ord inte är ett kodord. En kod C sägs vidare *korrigera t fel* om det mottagna ordet, då t eller färre fel har inträffat under överföringen, kan avkodas till det avsända ordet.

Sats 2.3.1. *En kod C korrigerar t fel om och endast om $d(C) > 2t$.*

Bevis. Antag att koden korrigerar t fel och att det finns två kodord $\boldsymbol{c_1}$ och $\boldsymbol{c_2}$ som har avstånd $2t$ eller mindre, så att $d(\boldsymbol{c_1}, \boldsymbol{c_2}) \leq 2t$. Under antagandet att t eller färre fel kan uppstå finns då ett mottaget ord $\hat{\boldsymbol{c}}$ som skiljer sig från både $\boldsymbol{c_1}$ och $\boldsymbol{c_2}$ i t eller färre komponenter. Ordet $\hat{\boldsymbol{c}}$ skulle alltså kunna härstamma från antingen $\boldsymbol{c_1}$ eller $\boldsymbol{c_2}$ och skulle inte bli korrekt avkodat i båda fallen.

Vi antar omvänt att avståndet mellan kodord i C är minst $2t$. Under antagandet att t eller färre fel kan uppstå så finns inget mottaget ord $\hat{\boldsymbol{c}}$ som skiljer sig från två kodord på t eller färre positioner. Det betyder att om t eller färre fel har uppstått så avkodas det mottagna ordet alltid korrekt om det avkodas till det närmaste kodordet. Således korrigerar koden t fel. □

För att Sats 2.3.1 ska vara användbar behövs kunskap om hur en kods minsta avstånd på ett effektivt sätt kan beräknas.

Sats 2.3.2. *Låt C vara en linjär kod. Då gäller att $d(C) = w(C)$.*

Bevis. Låt $w(C) = w(\boldsymbol{x})$ för $\boldsymbol{x}$ i C så att kodens minsta vikt är samma som vikten av kodordet $\boldsymbol{x}$. Eftersom ett kodords vikt är ekvivalent med kodordets avstånd till nollvektorn måste det gälla att

$$w(C) = w(\boldsymbol{x}) = d(\boldsymbol{x}, 0) \geq d(C)\,,$$

eftersom $d(C)$ är det minsta avståndet mellan kodord i C.

Vi betraktar nu godtyckliga kodord $\boldsymbol{x}, \boldsymbol{y} \in C$ så att $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ och $\boldsymbol{y} = (y_1, y_2, ..., y_n)$. Avståndet $d(\boldsymbol{x}, \boldsymbol{y})$ är antalet positioner på vilka $\boldsymbol{x}$ och $\boldsymbol{y}$ skiljer sig åt, det vill säga antalet j sådana att $x_j \neq y_j$. Vidare är $w(\boldsymbol{x} - \boldsymbol{y})$ antalet nollskilda positioner i ordet $\boldsymbol{x} - \boldsymbol{y}$, det vill säga antalet j sådana att $x_j - y_j \neq 0$. Eftersom $x_j \neq y_j$ är ekvivalent med $x_j - y_j \neq 0$ så gäller alltså att $d(\boldsymbol{x}, \boldsymbol{y}) = w(\boldsymbol{x} - \boldsymbol{y})$.

Låt nu $d(C) = d(\boldsymbol{x}, \boldsymbol{y})$ för $\boldsymbol{x}, \boldsymbol{y}$ i C så att kodens minsta avstånd är samma som avståndet mellan kodorden $\boldsymbol{x}$ och $\boldsymbol{y}$. Då är,

$$d(C) = d(\boldsymbol{x}, \boldsymbol{y}) = w(\boldsymbol{x} - \boldsymbol{y}) \geq w(C)\,,$$

eftersom C är en linjär kod och således $\boldsymbol{x} - \boldsymbol{y} \in C$. Vi har alltså att $w(C) \geq d(C)$ samt $w(C) \leq d(C)$ och det följer att $w(C) = d(C)$. □

Det här kapitlet inleddes med frågan om hur en ”bra kod” konstrueras. Problemet att konstruera ”bra koder” går till stor del ut på att konstruera koder som kan överföra information så snabbt och så säkert som möjligt. I det här sammanhanget används begreppet *hastighet*.

Definition 2.3.5. Om C är en kod i Q^n med $|C|$ kodord så kallas talet $R = \frac{\log_q |C|}{n}$ för kodens *hastighet*. Om koden är linjär och av dimensionen k är således $R = \frac{k}{n}$.

Repetitonskoden i tabell 1 har hastigheten $\frac{1}{3}$ jämfört med paritetsbitkoden i samma tabell som har hastigheten $\frac{2}{3}$. Paritetsbitkoden har alltså högre hastighet, men en lägre felkorrigeringskapacitet.

C.E. Shannon bevisade 1948 att varje överföringskanal har en definitiv kapacitet K sådan att det för varje hastighet $R < K$ existerar koder med hastigheten R för vilka sannolikheten för felaktig avkodning kan göras godtyckligt liten. Med andra ord, genom att öka kodordslängden n samtidigt som hastigheten R hålls under kanalens kapacitet kan sannolikheten för felaktig avkodning göras godtyckligt liten. Shannon sats ges i sin fullständiga formulering nedan där R är kanalens hastighet, och kapaciteten $K = 1 + p \log p + q \log q$. M_n är antalet kodord i koden med längd n, p är sannolikheten att en mottagen symbol är felaktig, $P^*(M_n, n, p)$ är minsta värdet av sannolikheten att respektive kodord i koden avkodas felaktigt, och $q := 1 - p$.

Sats 2.3.3 (Shannons sats)**.** *Om $0 < R < 1 + p\,log\,p + q\,log\,q$, och $M_n := q^{\lfloor Rn \rfloor}$ så gäller att $P^*(M_n, n, p) \to 0$ om $n \to \infty$.*

Vi ska inte gå närmare in på hur termerna i Shannons sats härleds eller på beviset av satsen men intressant att notera är att Shannons teori förklarar att en godtyckligt liten sannolikhet för felaktig avkodning är *möjlig* att uppnå; teorin säger dock ingenting om hur detta görs i praktiken. Ett av satsens villkor är att kodens längd $n \to \infty$, vilket är i enlighet med att effektiva koder i praktiken har väldigt stora n. Effektiva koder innehåller ofta runt 2^{100} meddelanden samt ofta lika många möjliga mottagna ord. Att effektivt kunna hantera

koder i den storleksordningen är en anledning att intressera sig för klasser av koder som har en stark algebraisk struktur. I den meningen utgör Shannons sats i sig en motivation till att studera algebraisk kodningsteori.

Med lite närmare kunskap om hur en kods kapacitet att korrigera fel hänger ihop med kodens algebraiska struktur kan konstruktionen av koder studeras närmare. För det krävs vissa kunskaper i grundläggande algebra, vilket nästa kapitel ämnar att förmedla.

För en mer djupgående genomgång av grundläggande kodningsteori än vad som presenterades här hänvisas den intresserade läsaren till boken *Introduction to Coding Theory* av J.H. van Lint [5].

3. Grundläggande algebra

I det här kapitlet presenteras den algebra som är nödvändig för att förstå konstruktionen av Reed-Solomon-koder och andra cykliska polynomkoder. För den som vill ha tillgång till de algebraiska definitioner och satser som detta kapitel bygger på återfinns dessa i bilaga A. Många satser har vi valt att inte visa och hänvisar istället till [2] Moden Algebra An Introduction av John R. Durbin.

3.1. Ringar och kroppar

Det går att visa att en av två isomorfa grupper är kommutativa om bara en av dem är kommutativ. På samma sätt gäller för två isomorfa ringar att de är kommutativa om den ena är kommutativ. Andra egenskaper som delas av isomorfin inkluderar existensen av en etta, existens av nolldelare, om ena är ett integritetsområde och om den andra är en ring. Ett sätt att visa att två ringar inte är isomorfa är att försöka hitta en egenskap som den andra inte har. Det finns en användbar definition som kan bestämma vad som är unikt för en ring för heltalen. Definitionen är väldigt användbar när kroppar studeras.

Definition 3.1.1. Låt R vara en ring. Om det finns ett positivt heltal n så att $na = 0$ för varje $a \in R$ så finns ett minsta sådant heltal som kallas för *karaktäristiken* av R. Om det inte existerar ett sådant positivt heltal så sägs R ha karaktäristik 0.

Exempel 3.1.1. Ringen av heltalen har karaktäristik 0, för det finns inget positivt heltal n sådan att $n \cdot 1 = 0$. Av samma anledning har ringarna av rationella och reella tal karaktäristik 0.

I bilaga A visas att om en ring $\mathbb{Z}_n$ är ett integritetsområde så är n ett primtal. Det betyder att om $\mathbb{Z}_n$ är ett integritetsområde så måste den ha en karaktäristik som är ett primtal. I texten kommer vi betrakta endast kommutativa ringar för annars måste vi skilja på höger-, vänster- och tvåsidiga-ideal.

Definition 3.1.2. En delring I av en ring R sägs vara ett *ideal* i R om $ar \in I$ och $ra \in I$ för varje $a \in I$ och alla $r \in R$.

Eftersom R anses vara en kommutativ ring så räcker det att visa antingen att $ar \in I$ eller $ra \in I$ då båda uttrycken är ekvivalenta.

Ett ideal är speciell delmängd i ringar. Tag exempelvis jämna tal: resultaten från addition och subtraktion av jämna tal bevarar sin jämnhet och multiplikation av ett jämnt tal med ett annat tal resulterar alltid ett nytt jämnt tal.

Definition 3.1.3. Varje element a i en ring R genererar ett ideal (a), ett så kallat *huvudideal*, som är det minsta ideal som innehåller a. Eftersom R är kommutativ så är

$$(a) = \{ra : r \in R\}.$$

Exempel 3.1.2. Låt S vara ett ideal i R sådant att $a \in S$. Antag S är ett mindre ideal än (a), då finns det ett element $b \in (a)$ så att $b \notin S$, där $b = ar$ för något $r \in R$. Vi har att $b \notin S$ och då gäller att $ar \notin S$ men eftersom $a \in S$ och $r \in R$ så betyder det ju att S inte är ett ideal. Alltså finns inget ideal mindre än (a) och därför måste (a) vara det minsta ideal i R som innehåller a.

Sats 3.1.1. *Låt I vara ett ideal i den kommutativa ringen R och låt R/I beskriva mängden av alla sidoklasser till I som en delgrupp med avseende på addition till R. För $I + a \in R/I$ och $I + b \in R/I$ låt*

$$(I + a) + (I + b) = I + (a + b)$$

och

$$(I + a)(I + b) = I + (ab).$$

Med dessa operationer så är R/I en ring- en så kallad restklassring.

Bevis. I beviset av satsen om kvotgrupper (A.5.1) så bildar R/I en grupp med den generella multiplikationen, byter vi bara till addition istället så uppfylls detta. Så allt som behövs visa är att multiplikationen är väldefinierad, det vill säga för $I + a_1 = I + a_2$ och $I + b_1 = I + b_2$ så vill vi visa att $I + a_1b_1 = I + a_2b_2$. Eftersom $I + a_1 = I + a_2$ då finns $n_1 \in I$ så att $a_1 = n_1 + a_2$. På samma sätt gäller för $I + b_1 = I + b_2$ så finns $n_2 \in I$ sådan att $b_1 = n_2 + b_2$. Vilket implicerar att

$$a_1b_1 = (n_1 + a_2)(n_2 + b_2) = n_1n_2 + n_1b_2 + a_2n_2 + a_2b_2$$

där $n_1n_2 + n_1b_2 + a_2n_2 \in I$ ty, $n_1, n_2 \in I$. I är ett ideal i R så det följer att a_1b_1 kan skrivas på formen $a_1b_1 = n_3 + a_2b_2$ där $n_3 \in I$, alltså är $I + a_1b_1 = I + a_2b_2$. □

3.2. Polynom

Ett *polynom* med koefficienter i en kommutativ ring R kan vi se som ett uttryck

$$a_0 + a_1x + ... + a_nx^n \tag{2}$$

där $a_0, a_1, ..., a_n \in R$.

Definition 3.2.1. Låt R vara en kommutativ ring. Ett *polynom i ett obestämt x över R* är ett uttryck av formen (2) där *koefficienterna* $a_0, a_1, ..., a_n$ är element i R. Om $a_n \neq 0$ då sägs heltalet n vara *polynomgraden* och a_n är dess *ledande koefficient*. Ett polynom över en kropp sägs vara *moniskt* om den ledande koefficienten är en etta i kroppen. Två polynom i x sägs vara *lika* om och endast om koefficienterna till varje potens av x är lika. Vi skriver att mängden av polynom i variabeln x över R är *polynomringen* $R[x]$.

Sats 3.2.1 (Divisionsalgoritmen)**.** *Antag att $f(x)$ och $g(x)$ är polynom över en kropp $\mathbb{F}$, där $g(x) \neq 0$. Då existerar det två entydliga polynom: kvot- $q(x)$ och rest-polynomet $r(x)$ över $\mathbb{F}$ sådan att*

$$f(x) = g(x)q(x) + r(x), \text{ där } r(x) = 0 \text{ eller } \deg r(x) < \deg g(x).$$

Bevis. Låt $f(x) = a_m x^m + ... + a_1 x + a_0$ och $g(x) = b_n x^n + ... + b_1 x + b_0$. Eftersom $g(x) \neq 0$ så kan vi anta att $b_n \neq 0$ sådan att $\deg g(x) = n$. Om $f(x) = 0$ då följer det att $q(x) = 0$ och $r(x) = 0$. Vi kan därför anta att $a_m \neq 0$ sådan att $\deg f(x) = m$.
Vi vill visa existensen av $q(x)$ och $r(x)$ med hjälp av induktion på m. Antag att $m < n$ då kan vi välja $q(x) = 0$ och $r(x) = f(x)$ och vi får

$$f(x) = g(x) \cdot 0 + f(x)$$

vilket uppfyller hypotesen. Antag att $m \geq n$. Om $m = 0$ så är $f(x) = a_0$ och $g(x) = b_0$. Väljer vi $q(x) = b_0^{-1} a_0$ och $r(x) = 0$ får vi att

$$a_0 = b_0 \cdot b_0^{-1} a_0 + 0$$

vilket uppfyller hypotesen. Antag nu att satsen gäller för $\deg f(x) < m$. Vi vill visa att detta även gäller för $\deg f(x) = m$. Multiplicera faktorn $g(x)$ med $a_m b_n^{-1} x^{m-n}$ och vi får

$$\begin{aligned} a_m b_n^{-1} x^{m-n} g(x) &= a_m b_n^{-1} x^{m-n} (b_n x^n + ... + b_1 x + b_0) \\ &= a_m x^m + ... + a_m b_n^{-1} b_1 x^{m-(n-1)} + a_m b_n^{-1} b_0 x^{m-n}. \end{aligned}$$

Eftersom den ledande termen i polynomet ovan finns i $f(x)$ så gäller att polynomet

$$f_1(x) = f(x) - a_m b_n^{-1} x^{m-n} g(x)$$

är av lägre grad än m. Då följer det av induktionens hypotes att det existerar polynom $q_1(x)$ och $r_1(x)$ sådan att

$$f_1(x) = g(x) q_1(x) + r_1(x) \text{ där } r_1(x) = 0 \text{ eller } \deg r_1(x) < \deg g(x).$$

Detta implicerar att

$$\begin{aligned} f(x) - a_m b_n^{-1} x^{m-n} g(x) &= g(x) q_1(x) + r_1(x) \\ f(x) &= g(x)(a_m b_n^{-1} x^{m-n} + q_1(x)) + r_1(x) \end{aligned}$$

med $q(x) = a_m b_n^{-1} x^{m-n} + q_1(x)$ och $r(x) = r_1(x)$, vilket visar existensen av $q(x)$ och $r(x)$. För att visa entydligheten av polynomen $q(x)$ och $r(x)$, antag att det existerar polynom $q^\star(x)$ och $r^\star(x)$ över $\mathbb{F}$ sådan att

$$f(x) = g(x) q^\star(x) + r^\star(x) \text{ där } r^\star(x) = 0 \text{ eller } \deg r^\star(x) < \deg g(x).$$

Då har vi att

$$g(x) q(x) + r(x) = g(x) q^\star(x) + r^\star(x)$$

och

$$g(x)(q(x) - q^\star(x)) = r^\star(x) - r(x). \tag{3}$$

Om $q(x) - q^\star(x) \neq 0$ ovan så gäller att polynomgraden på vänsterledet av (3) är av högre eller lika grad som graden av $g(x)$. Men eftersom $\deg r_1(x)$ och $\deg r(x)$ är antingen noll eller av mindre grad än graden av $g(x)$ så måste högerledet av (3) vara av lägre grad än $g(x)$. Alltså måste $q(x) - q^\star(x) = 0$ för annars har vi att polynomgraden i vänsterledet och högerledet i (3) är olika, vilket är en motsägelse. Därför gäller att $q(x) - q^\star(x) = 0$ och då följer det att $r^\star(x) - r(x) = 0$. Alltså är $r(x) = r^\star(x)$ och $q(x) = q^\star(x)$, vilket visar entydligheten. □

Sats 3.2.2 (Faktorsatsen)**.** *Om $f(x) \in \mathbb{F}[x]$ och $\mathbb{F}$ är en kropp så gäller*

$$(x-c) \mid f(x) \Leftrightarrow f(c) = 0$$

för varje $c \in \mathbb{F}$.

Bevis. Antag att $(x-c) \mid f(x)$. Då finns $g(x) \in \mathbb{F}[x]$ så att $f(x) = (x-c)g(x)$ och därmed är $f(c) = (c-c)g(c) = 0$.
Antag att $f(c) = 0$. Då följer det från divisionalgoritmen för $\deg(x-c) = 1$ att $r(x) = 0$ eller $\deg r(x) = 0$. Alltså finns ett polynom $q(x) \in \mathbb{F}[x]$ så att

$$f(x) = (x-c)q(x) + r \text{ där } r \in \mathbb{F}.$$

Eftersom $f(c) = 0$ så är $0 = (c-c)q(c) + r$ och därmed är $r = 0$. □

Elementet c i $\mathbb{F}$ kallas för ett *nollställe* till ett polynom $f(x) \in \mathbb{F}[x]$ om $f(c) = 0$. Så från faktorsatsen gäller att c är ett nollställe till ett polynom $f(x)$ om och endast om $x - c$ är en faktor till $f(x)$.

Sats 3.2.3. *Ett polynom av grad n över en kropp $\mathbb{F}$ har högst n nollställen i $\mathbb{F}$.*

Bevis. Vi vill med hjälp av induktion på grad n bevisa satsen.
Ett polynom av grad 0 består bara av en konstant och därför saknar nollställe.
Antag satsen gäller för polynom av grad $n-1$ och låt $f(x) \in \mathbb{F}[x]$ vara ett polynom av grad n. Om $f(x)$ saknar nollställen så är satsen uppfylld. Om $f(x)$ har nollställen, låt c vara ett nollställe. Då gäller enligt Faktorsatsen att

$$f(x) = (x-c)g(x)$$

där $\deg g(x) = n-1$.

Eftersom $\mathbb{F}$ saknar nolldelare så gäller att $f(d) = 0$ om och endast om $(d-c) = 0$ eller $g(d) = 0$. Därför måste något nollställe till $f(x)$ vara antingen lika med c eller vara ett nollställe till $g(x)$. Enligt hypotesen så har alltså $g(x)$ högst $n-1$ nollställen, så $f(x)$ måste ha högst n nollställen. □

Definition 3.2.2. Antag att $a(x)$ och $b(x)$ är nollskilda polynom över en kropp $\mathbb{F}$ då finns ett unikt moniskt polynom $d(x)$ över $\mathbb{F}$ så att

(a) $d(x) \mid a(x)$ och $d(x) \mid b(x)$,

(b) om $c(x)$ är ett polynom sådan att $c(x) \mid a(x)$ och $c(x) \mid b(x)$ då gäller att $c(x) \mid d(x)$.

Polynomet $d(x)$ kallas för *största gemensamma delare* till $a(x)$ och $b(x)$. Vi betecknar detta som $\gcd(a(x), b(x)) = d(x)$.

Sats 3.2.4. *Låt att $a(x)$ och $b(x)$ vara nollskilda polynom över en kropp $\mathbb{F}$ och låt $d(x) = \gcd(a(x), b(x))$. Då existerar två polynom $u(x)$ och $v(x)$ över $\mathbb{F}$ sådan att*

$$d(x) = a(x)u(x) + b(x)v(x).$$

Definition 3.2.3. Låt $\mathbb{F}$ vara en kropp. Ett icke-konstant polynom $f(x)$ i polynomringen $R[x]$ sägs vara *irreducibelt över R* eller *irreducibel i $R[x]$* om det inte finns två polynom $g(x)$ och $h(x)$ i $R[x]$ av högre grad än noll så att $f(x) = g(x)h(x)$. Om $f(x) \in R[x]$ är icke-konstant polynom som inte är irreducibel över R så kallas det *reducibelt över $\mathbb{F}$* eller *reducibel i $R[x]$*.

Definition 3.2.4. Om $\mathbb{E}$ och $\mathbb{F}$ är kroppar så sägs $\mathbb{E}$ vara en *kroppsutvidning* av $\mathbb{F}$ om $\mathbb{E}$ innehåller en delkropp som är isomorf med $\mathbb{F}$.

Låt $\mathbb{E}$ vara en kropputvidning av en kropp $\mathbb{F}$. Antag att $\mathbb{F} \subseteq \mathbb{E}$ och S är en delmängd av $\mathbb{E}$. Då existerar det minst en delkropp av $\mathbb{E}$ som innehåller både $\mathbb{F}$ och S, nämligen $\mathbb{E}$ själv. Snittet av alla delkroppar av $\mathbb{E}$ som innehåller både $\mathbb{F}$ och S är delkroppar av $\mathbb{E}$. Detta skrivs

$$\mathbb{F}(S).$$

Om $S \subseteq \mathbb{F}$ då är $\mathbb{F}(S) = \mathbb{F}$. Om $S = \{a_1, a_2, ..., a_n\}$ då skrivs $\mathbb{F}(S)$ som $\mathbb{F}(a_1, a_2, ..., a_n)$. Ett enkelt exempel är $\mathbb{R}(i) = \mathbb{C}$.

Definition 3.2.5. Antag $\mathbb{E}$ är en kroppsutvidning av en kropp $\mathbb{F}$. Ett element $a \in \mathbb{E}$ sägs vara *algebraisk över* $\mathbb{F}$ om a är en lösning till något polynom

$$a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n = 0, \quad a_0, a_1, ..., a_{n-1}, a_n \in \mathbb{F}.$$

Exempel 3.2.1. $\sqrt{2}$ är algebraisk över $\mathbb{Q}$ eftersom $\sqrt{2}$ är en lösning till polynomet $x^2 - 2 = 0$.

Definition 3.2.6. Om $\mathbb{F}(a) = \mathbb{E}$ för något element $a \in \mathbb{E}$, då sägs $\mathbb{E}$ vara en *enkel utvidning* av $\mathbb{F}$.

Vi kan klassificera en enkel utvidning av $\mathbb{F}$ genom att använda oss av polynomringen över $\mathbb{F}$ och

$$\mathbb{F}[a] = \{a_0 + a_1 a + \cdots + a_n a^n : a_0, a_1, ..., a_n \in \mathbb{F}\}$$

som är ringen som genereras av $\mathbb{F}$ och a. Skillnaden hos polynomen i $\mathbb{F}[x]$ och $\mathbb{F}[a]$ är att polynomen i $\mathbb{F}[x]$ är lika endast om koefficienterna är lika. Medan om a är algebraisk över $\mathbb{F}$ så kan två polynom vara lika utan att koefficienterna av a är lika.

Exempel 3.2.2.

$$\begin{array}{rl} 1 + 3\sqrt{2} = \text{-}1 + 3\sqrt{2} + \sqrt{2}^2 & \text{i } \mathbb{Q}[\sqrt{2}], \text{ men} \\ 1 + 3x \neq \text{-}1 + 3x + x^2 & \text{i } \mathbb{Q}[x]. \end{array}$$

Sats 3.2.5. *Antag $\mathbb{F}$ är en kropp och att $p(x) \in \mathbb{F}[x]$. Då gäller att $\mathbb{F}[x]/(p(x))$ är en kropp om och endast om $p(x)$ är irreducibel över $\mathbb{F}$.*

Bevis. För enkelhetens skull, låt $I = (p(x))$ vara huvudidealet under hela beviset. För att få en motsägelse, antag att $p(x)$ är reducibel över $\mathbb{F}$. Då finns $a(x), b(x) \in \mathbb{F}[x]$ sådan att

$$p(x) = a(x)b(x) \text{ där } \deg a(x) < \deg p(x) \text{ och } \deg b(x) < \deg p(x).$$

Då $I = (p(x))$ är ett huvudideal och $\deg a(x) < \deg p(x)$ så gäller att $a(x) \notin I$, med samma argument gäller att $b(x) \notin I$. Då följer det att $I + a(x) \neq 0$ och $I + b(x) \neq 0$ i $F[x]/I$. Vi har att

$$(I + a(x))(I + b(x)) = I + a(x)b(x) = I + p(x) = I$$

vilket visar att $I + a(x)$ och $I + b(x)$ är nolldelare i $F[x]/I$, alltså är $F[x]/I$ inte en kropp. Inte ens ett integritetsområde.
Antag nu istället att $p(x)$ är irreducibel. Vi vill visa att $\mathbb{F}[x]/I$ är en kropp, vi har kommutativitet över multiplikation:
Låt $I + a(x), I + b(x) \in \mathbb{F}[x]/I$ då gäller

$$(I + a(x))(I + b(x)) = I + a(x)b(x) = I + b(x)a(x) = (I + b(x))(I + a(x))$$

Vi har att I + 1 är en etta ty, för något element $I + c(x) \in \mathbb{F}[x]/I$ där $1 \in \mathbb{F}[x]$ så gäller

$(I+1)(I+c(x)) = I + 1c(x) = I + c(x)$ och $(I+c(x))(I+1) = I + c(x)1 = I + c(x)$.

Invers med avseende på multiplikation:
Antag att $I + f(x) \neq 0$ då gäller $f(x) \notin I$ så $f(x)$ kan inte skrivas som en multipel av $p(x)$ (då $I = (p(x))$). Eftersom $p(x)$ är irreducibel så är $\gcd(f(x), p(x)) = 1$, då finns $u(x), v(x) \in \mathbb{F}[x]$ så att $1 = p(x)u(x) + f(x)v(x)$, vilket ger

$$1 - f(x)v(x) = u(x)p(x) \in I \Rightarrow I + 1 = I + f(x)v(x) = (I + v(x))(I + f(x))$$

alltså är $I + v(x)$ den multiplikativa inversen till $I + f(x)$. □

Sats 3.2.6. *Om $\mathbb{E}$ är en enkel utvidning av $\mathbb{F}$, där $\mathbb{E} = \mathbb{F}(a)$ och a algebraisk över $\mathbb{F}$ så gäller att*

$$\mathbb{E} \cong \mathbb{F}[x]/(p(x))$$

där $p(x)$ irreducibelt polynom över $\mathbb{F}$ och $(p(x))$ är ett ideal som innehåller alla $f(x) \in \mathbb{F}[x]$ sådan att $f(a) = 0$.

Bevis. Definiera $\theta : \mathbb{F}[x] \to \mathbb{F}[a]$ så att

$$\theta(a_0 + a_1x + \cdots + a_nx^n) = a_0 + a_1a + \cdots + a_na^n.$$

Det går att verifera att θ är en ringhomomorfi, vilket skulle då implicera enligt Fundamentala homomorfisatsen för ringar A.3.1 att $\mathbb{F}[x]/I \cong \mathbb{F}[a]$, där $I = \ker\theta$ är ett ideal i $\mathbb{F}[x]$. Eftersom a är algebraisk över $\mathbb{F}$ så gäller att $I \neq (0)$. Notera att I innehåller varje polynom som har nollställe i a, ty

$$f(x) \in I \Leftrightarrow \theta(f(x)) = f(a) = 0.$$

Enligt Theorem A.5.3 så är varje ideal i $\mathbb{F}[x]$ ett huvudideal, så $I = (p(x))$ för något $p(x) \in \mathbb{F}[x]$. Eftersom $\mathbb{F}[a] \subseteq \mathbb{E}$ och $\mathbb{E}$ saknar nolldelare och då kan $\mathbb{F}[a]$ inte ha nolldelare därmed kan inte $\mathbb{F}[x]/(p(x))$ ha nolldelare, då följer det att $p(x)$ är irreducibel över $\mathbb{F}$ enligt Sats 3.2.5.

□

Definition 3.2.7. En kropp $\mathbb{E}$ är en *splittringskropp* av ett icke-konstant polynom $p(x)$ över en kropp $\mathbb{F}$ om $\mathbb{E}$ är en utvidning av $\mathbb{F}$ sådan att

(*i*) $p(x)$ kan skrivas som den linjära faktoriseringen $p(x) = a(x - c_1)(x - c_2)\cdots(x - c_n)$
(*ii*) $\mathbb{E} = \mathbb{F}(c_1, c_2, ..., c_n)$

där $c_1, c_2, ..., c_n$ är nollställen till $p(x)$ i $\mathbb{E}$.

Sats 3.2.7. *Om $\mathbb{E}$ och $\mathbb{E}'$ är splittringskroppar av ett polynom $p(x)$ över en kropp $\mathbb{F}$ då existerar en isomorfi $\theta : \mathbb{E} \to \mathbb{E}'$ sådan att $\theta(a) = a$ för varje $a \in \mathbb{F}$.*

3.3. Ändliga kroppar

Återigen så hänvisar vi till bilaga A om begrepp verkar otydliga. Vi går genom några enklare och sedan visar viktiga satser här som är väldigt centrala för arbetet.

Definition 3.3.1. För något element $a \in G$ så är mängden

$$\{a^n : n \in \mathbb{Z}\}$$

en delgrupp till G. Denna delgrupp kallas för *delgruppen genererad av a* och betecknas med $\langle a \rangle$. Om H är en grupp och $H = \langle a \rangle$ för något element $a \in H$ så kallas H för en *cyklisk* grupp.

Sats 3.3.1 (Lagranges sats)**.** *Om H är en delgrupp av en ändlig grupp G så gäller att ordningen av H är delare till ordningen av G.*

Bevis. Eftersom högersidoklasserna av H är ekvivalensklasser så gäller att högersidoklasserna av H bildar en partition av G, således måste två högersidoklasser i H antingen vara lika eller disjunkta. Eftersom G är ändlig så finns det bara ändligt många sidoklasser. Välj ett element från varje sidoklass och låt de valda elementen vara $a_1, a_2, ..., a_k$. Då gäller

$$G = Ha_1 \cup Ha_2 \cup \cdots \cup Ha_k.$$

Varje sidoklass Ha_i innehåller H element enligt Sats A.1.4 och det finns inga fler element än i en annan sidoklass, så det följer att $|G| = |H|k$. Alltså är $|H|$ delare till $|G|$. □

Sats 3.3.2. *Låt G vara en cyklisk grupp av ordning n där $G = \langle a \rangle = \{e, a, a^2, ..., a^{n-1}\}$.*

1. *Varje delgrupp av G är cyklisk.*
2. *För varje positiv delare d till n så har G exakt en delgrupp av ordning d.*

Sats 3.3.3. *Om G är en ändlig grupp så att för varje positivt heltal n så finns det högst n element av $x \in G$ så att $x^n = e$, där e är neutrala elementet i G, då gäller att G är cyklisk.*

Bevis. Låt H vara en cyklisk grupp så att $|H| = |G| = m$. Enligt Laranges sats så gäller att varje element i G genererar en cyklisk delgrupp av ordning d för någon delare d till n. För varje x i denna delgrupp så är $x^d = e$. Därför gäller enligt hypotes att delgruppen innehåller alla lösningar i G på formen $x^d = e$.

Enligt Sats 3.3.2 så finns det exakt en cyklisk delgrupp av ordning d för varje delare d till m. Alltså måste H ha minst lika många element av ordning d som G har. Eftersom $|H| = |G|$, så måste H ha lika många element av ordning d för varje delare d till m. Eftersom H är cyklisk och har ett element av ordning m så måste G med ha det vilket implicerar att G cyklisk. □

Sats 3.3.4. *Om $\mathbb{F}$ är en kropp av karaktäristik p, då gäller att $(a+b)^p = a^p + b^p$ för varje $a, b \in \mathbb{F}$.*

Bevis. Binomialsatsen ger att

$$(a+b)^p = a^p + \binom{p}{p-1}a^{p-1}b + \binom{p}{p-2}a^{p-2}b + ... + b^p$$

eftersom p är ett primtal så gäller att p delar $\binom{p}{k}$ för $1 \leq k \leq p-1$. Eftersom $\mathbb{F}$ har karaktäristik p så är alla termer är 0 förutom den första och sista termen. Vilket ger att $(a+b)^p = a^p + b^p$. □

Sats 3.3.5. *Om p är ett primtal och n är ett positivt heltal så existerar en kropp med p^n element och varje par av sådana kroppar är isomorfa.*

Bevis. Låt $q = p^n$ för något positivt heltal n. Vi vill visa att en kropp $\mathbb{F}$ är av ordning q om $\mathbb{F}$ är en splittringskropp av polynomet $x^q - x$ över $\mathbb{Z}_p$. Med hjälp av Theorem 3.2.7 kan vi visa att varje par kroppar av ordning q är isomorfa. Vi kan då visa att splittringskroppen av $x^q - x$ över $\mathbb{Z}_p$ är då av ordning q vilket i sin tur visar existensen av en kropp med p^n element.

Antag att $|\mathbb{F}| = q$. Då är prima delkroppen av $\mathbb{F}$ isomorf med $\mathbb{Z}_p$. Alla nollskilda element i

$\mathbb{F}$ bildar en multiplikativ grupp av ordning $q-1$ så $x^{q-1}=1$ för varje nollskilda $x \in \mathbb{F}$ enligt Lagranges sats. Därför gäller att $x^q = x$ för varje $x \in \mathbb{F}$. Om $\mathbb{F} = \{a_1, a_2, ..., a_q\}$ då är $(x-a_1)(x-a_2)\cdots(x-a_q)$ en faktor till $x^q - x$ enligt faktorsatsen då varje a_k är distinkta. Således är $x^q - x = (x-a_1)(x-a_2)\cdots(x-a_q)$ och $\mathbb{F}$ är en splittringskropp av polynomet $x^q - x$ över $\mathbb{Z}_p$.

Låt $\mathbb{E}$ vara en splittringskropp av $f(x) = x^q - x$ över $\mathbb{Z}_p$. Derivatan av $f(x)$ är $f'(x) = qx^{q-1} - 1 = -1$, eftersom $\mathbb{E}$ har karakteristik p och $p \mid q$. Således är $f'(c) \neq 0$ för varje $c \in \mathbb{E}$ och alltså saknar $f(x)$ multipla nollställen i $\mathbb{E}$.

Vi vill visa att de q distinkta nollställen $c_1, c_2, ..., c_q$ till $x^q - x$ bildar en delkropp av $\mathbb{E}$ för då implicerar det att $\mathbb{E} = \{c_1, c_2, ..., c_q\}$ och $|\mathbb{E}| = q$. Antag att a och b är nollställen till $x^q - x$ i $\mathbb{E}$. Då $\mathbb{E}$ är av karaktäristik q har vi att $(a+b)^p = a^p + b^p$, $(a+b)^{p^2} = (a^p + b^p)^p = a^{p^2} + b^{p^2}$ och så vidare, vilket implicerar att $(a+b)^q = a^q + b^q$. Eftersom a och b är nollställen till $x^q - x$ så följer det att $a+b$ också är ett nollställe. Dessutom har vi $(ab)^q - (ab) = a^q b^q - ab = ab - ab = 0$, så ab är också ett nollställe. Vi har även att $(a^{-1})^q - a^{-1} = a^{-q} - a^{-1} = a^{-1} - a^{-1} = 0$. Alltså gäller att nollställerna till $x^q - x$ bildar en delkropp över $\mathbb{E}$. □

Satsen ovan visar att det finns väsentligen en unik kropp av ordning p^n. Denna kropp kallas för *Galoiskroppen* av ordning p^n som vi betecknar $\mathbb{F}_{p^n}$.

Sats 3.3.6. *Multiplikativa gruppen av en ändlig kropp är cyklisk.*

Bevis. Enligt Sats 3.2.3 så har en kropp $\mathbb{F}$ högst n lösningar i $x^n = e$ för varje $n \geq 1$. Eftersom $\mathbb{F}$ är ändlig så följer det av Sats 3.3.3 att multiplikativa gruppen av en ändlig kropp är cyklisk. □

Definition 3.3.2. Om g och h är element till abelska grupper av ordning a respektive b. Då finns det ett element av ordning $\text{lcm}(a,b)$, där lcm står för minsta gemensamma multipel.

Sats 3.3.7. *Om högsta ordningen av elementen i en abelsk grupp G är r så är $x^r = e$ för varje $x \in G$.*

Bevis. Låt $g \in G$ vara ett element av högsta ordningen r. Om h är ett element of ordning t så finns det ett element av ordning $\text{lcm}(r,t)$ enligt definition 3.3.2. Eftersom $\text{lcm}(r,t) \leq r$ så gäller att $t \mid r$. Därför måste $h^r = e$. □

Sats 3.3.8. *Låt $\mathbb{F}_q^*$ vara mängden av nollskilda element i en Galoiskropp $\mathbb{F}_q$. Då är $(\mathbb{F}_q^*, \cdot)$ en cyklisk grupp av ordning $q-1$.*

Bevis. Låt r vara högsta ordningen av elementen i $(\mathbb{F}_q^*, \cdot)$. Då gäller enligt Sats 3.3.7 att

$$x^r - 1 = 0 \text{ för varje } x \in F_q^*.$$

Således är varje nollskilt element i Galoiskroppen F_q ett nollställe till polynomet $x^r - 1$. Och enligt Sats 3.2.3 så har polynom av grad r högst r nollställen över vilken kropp som helst. Alltså gäller att $r \geq q-1$. Men enligt Lagranges sats så $r \mid (q-1)$ alltså måste $r = q-1$.

$(F_q^*, \cdot)$ måste alltså vara en grupp av ordning $q-1$ som innehåller alla element av ordning $q-1$ och således måste vara cyklisk. □

Definition 3.3.3. Låt $\mathbb{E}$ vara en kropputvidning av $\mathbb{F}$ och tag ett element $\alpha \in \mathbb{E}$. *Minimalpolynomet* av α är ett moniskt polynom av lägsta polynomgrad i polynomringen $\mathbb{F}[x]$ där α är ett nollställe. Detta α existerar när α är algebraisk över $\mathbb{F}$.

Definition 3.3.4. Ett *primitivt polynom* är ett polynom som genererar alla element av en kroppsutvidning. Primitiva polynom är också irreducibla polynom. För någon potens n av primtalet p så finns det ett primitivt polynom av polynomgrad n över $\mathbb{F}_{p^n}$.

Exempel 3.3.1. För $\mathbb{F}_4 = \mathbb{Z}_2(\alpha)$ så är multiplikativa gruppen av nollskilda element $\mathbb{F}_4^*$ en cyklisk grupp av ordning 3 där elementen α och $\alpha + 1$ är de primitiva elementen.

3.4. Konstruktion av ändliga kroppar

En algebraisk kod definieras över en ändlig kropp $\mathbb{F}_q$, vilket betyder att informations- och kodsymbolerna är element i $\mathbb{F}_q$. Att ändliga kroppar av samma storlek är isomorfa, vilket etablerades i Sats 3.3.5, betyder att när ordningen, q, av den ändliga kroppen $\mathbb{F}_q$ är specificerad så kan den ändliga kroppen väljas hur som helst, så länge den har q element och uppfyller kriterierna för en kropp med tillhörande operationer. Detta betyder att för att konstruera en algebraisk kod, vilket innefattar att utföra operationer på elementen i $\mathbb{F}_q$, så räcker det inte att specificera ordningen q utan kroppen som ska användas behöver så kallat *konstrueras*; att konstruera kroppen innebär att specificera vilka element som ingår, samt hur de adderas och multipliceras med varandra.

Både Sats 3.3.5 och Sats 3.2.5 är centrala i det här sammanhanget; den första etablerar att $\mathbb{F}[x]/(p(x))$ vars element utgörs av alla sidoklasser till polynomet $p(x)$ över $\mathbb{F}$ är en kropp om och endast om $p(x)$ är irreducibel över $\mathbb{F}$, och den andra fastslår att kroppen $\mathbb{F}[x]/(p(x))$ är isomorf med alla andra kroppar av samma ordning. Det betyder att om $p(x)$ är ett polynom av grad m som är irreducibelt över $\mathbb{F}_p$ så är kroppen $\mathbb{F}_{p^m}$ isomorf med kroppen $\mathbb{F}_p[x]/(p(x))$; eftersom elementen i $\mathbb{F}_p[x]/(p(x))$ utgörs av alla polynom av grad $< m$ med koefficienter i $\mathbb{F}_p$ är ordningen av kroppen nämligen p^m.

Vidare kan kroppen $\mathbb{F}_p[x]/(p(x))$ betraktas som en kroppsutvidgning av $\mathbb{F}_q$ och enligt Sats A.5.4 finns därför ett element $\beta \in \mathbb{F}_q$ så att $p(\beta) = 0$. Enligt Sats 3.2.6 gäller också att $F_p(\beta) \cong \mathbb{F}_p[x]/(p(x))$ och att varje element i $\mathbb{F}_p(\beta)$ på ett unikt sätt kan skrivas på formen

$$v_0 + v_1\beta + v_2\beta^2 + ... + v_{n-1}\beta^{n-1},\ v_i \in \mathbb{F}_p. \tag{4}$$

Kroppen $\mathbb{F}_{p^m}$ kan således ses som ett linjärt rum $(\mathbb{F}_p)^m$ med basen $\{1, \beta, \beta^2, ..., \beta^{n-1}\}$.

Genom att specificera ett irreducibelt polynom av grad m kan alltså en konkret representation av elementen i $\mathbb{F}_{p^m}$ erhållas; varje element representeras som ett polynom i elementet β av grad $< m$ med koefficienter i $\mathbb{F}_p$. Med denna representation kan två element lätt adderas med varandra eftersom detta helt enkelt sker genom att koefficienterna av lika potenser av x adderas modulo p. Multiplikation av element är däremot med denna representation mödosamt eftersom det kräver upprepat utnyttjande av att $p(\beta) = 0$. Därför nöjer vi oss inte med den här representationen utan använder också det faktum att den multiplikativa gruppen $\mathbb{F}_q^*$ är cyklisk och således genereras av ett primitivt element α så att $\mathbb{F}_q = \{0, 1, \alpha, \alpha^2, ..., \alpha^{q-2}\}$ där $\alpha^{q-1} = 1$. Med den här representationen kan multiplikation av element lätt utföras genom att addera exponenter modulo $q - 1$.

Eftersom båda operationerna, multiplikation och addition, behöver kunna utföras på

elementen i kroppen $\mathbb{F}_q$ vill man kunna gå mellan polynomrepresentationen $F_q = \{v_0 + v_1\beta + v_2\beta^2 + ... + v_{n-1}\beta^{n-1} |\, v_i \in \mathbb{F}_p,\, \beta \in \mathbb{F}_q,\, p(\beta) = 0\}$ och den exponentiella representationen $\mathbb{F}_q = \{0, 1, \alpha, \alpha^2, ..., \alpha^{q-2}\}$. Särskilt lämpligt är det därför om polynomet $p(x)$ inte bara är irreducibelt över $\mathbb{F}_p$ utan även primitivt i $\mathbb{F}_q$, det vill säga om $p(\alpha) = 0$; då kan nämligen varje nollskilt element i $\mathbb{F}_{p^m}$ på ett entydigt sätt skrivas både som en potens av α och som ett polynom i α av grad $< m$.

De primitiva elementen i kroppen $\mathbb{F}_q$ hittas genom att låta elementen i kroppen representeras på formen (4) och sedan beräkna successiva potenser av alla element och notera vilka element som genererar $\mathbb{F}_q^*$. Ett primitivt polynom av grad m kan sedan hittas genom att konstruera ett irreducibelt polynom som har m stycken primitiva element som nollställen.

Konstruktionen av $\mathbb{F}_{2^8}$ illustreras i exemplet nedan.

Exempel 3.4.1. Polynomet $p(x) = x^8 + x^4 + x^3 + x^2 + 1$ är irreducibelt över $\mathbb{F}_p$ så vi kan konstruera kroppen $\mathbb{F}_{2^8}$ genom isomorfin $\mathbb{F}_{2^8} \cong \mathbb{F}_2[x]/(x^8 + x^4 + x^3 + x^2 + 1)$. Polynomet $p(x) = x^8 + x^4 + x^3 + x^2 + 1$ är även primitivt över $\mathbb{F}_2$ och $\alpha \in \mathbb{F}_{2^8}$ är ett primitivt element så att $\alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1 = 0$. Alla element i $\mathbb{F}_{2^8}$ kan således skrivas som en linjärkombination av elementen $1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7$ alternativt som en enda potens av α. Motsvarigheten mellan de två representationerna erhålls genom att reducera varje potens av α modulo $\alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1$ eftersom $\alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1 = 0$. De två representationerna visas i tabell 2, för 14 av de 256 elementen i $\mathbb{F}_{2^8}$. I kolumnen längst till höger uttrycks dessutom respektive linjärkombination som en binär sekvens av längd 8.

Tabell 2: Tabellen visar tre olika representationer av de första 14 elementen i $\mathbb{F}_{2^8}$.

0	$=$														$=$	$(0\,0\,0\,0\,0\,0\,0\,0)$
1	$=$													1	$=$	$(0\,0\,0\,0\,0\,0\,0\,1)$
α	$=$											α			$=$	$(0\,0\,0\,0\,0\,0\,1\,0)$
α^2	$=$								α^2						$=$	$(0\,0\,0\,0\,0\,1\,0\,0)$
α^3	$=$						α^3								$=$	$(0\,0\,0\,0\,1\,0\,0\,0)$
α^4	$=$					α^4									$=$	$(0\,0\,0\,1\,0\,0\,0\,0)$
α^5	$=$				α^5										$=$	$(0\,0\,1\,0\,0\,0\,0\,0)$
α^6	$=$			α^6											$=$	$(0\,1\,0\,0\,0\,0\,0\,0)$
α^7	$=$	α^7													$=$	$(1\,0\,0\,0\,0\,0\,0\,0)$
α^8	$=$					α^4	$+$	α^3	$+$	α^2	$+$			1	$=$	$(0\,0\,0\,1\,1\,1\,0\,1)$
α^9	$=$				α^5	$+$	α^4	$+$	α^3	$+$		α			$=$	$(0\,0\,1\,1\,1\,0\,1\,0)$
α^{10}	$=$			α^6	$+$	α^5	$+$	α^4	$+$		α^2				$=$	$(0\,1\,1\,1\,0\,1\,0\,0)$
α^{11}	$=$	α^7	$+$	α^6	$+$	α^5	$+$		α^3						$=$	$(1\,1\,1\,0\,1\,0\,0\,0)$
α^{12}	$=$	α^7	$+$	α^6	$+$				α^3	$+$	α^2	$+$		1	$=$	$(1\,1\,0\,0\,1\,1\,0\,1)$

De två representationerna i exemplet, som alltså erhölls tack vare att $p(x)$ var ett primitivt polynom över $\mathbb{F}_2$, innebär att elementen i kroppen $\mathbb{F}_{2^8}$ kan representeras av alla möjliga bitsträngar av längden 8, vilket är användbart om binär data ska kodas samt gör det enkelt att utföra addition av element i kroppen eftersom dessa adderas som binära strängar, det vill säga bitvis modulo 2. Den exponentiella representationen är användbar då element i kroppen ska multipliceras eftersom detta helt enkelt görs genom att addera elementens exponenter modulo 255.

Hur övergången mellan de två representationerna går till illustreras för den mindre kroppen $\mathbb{F}_{2^3}$ i exemplet nedan.

Exempel 3.4.2. Polynomet $p(x) = x^3 + x + 1$ är ett primitivt polynom i $\mathbb{F}_{2^3}$. Det betyder att varje nollskilt element kan skrivas som någon linjärkombination av elementen i mängden $\{1, \alpha, \alpha^2\}$ för något element $\alpha \in \mathbb{F}_{2^3}$ sådant att $p(\alpha) = 0$. Vi låter α vara primitivt. (Att α är primitivt ses genom att successivt beräkna potenser av α i termer av $1, \alpha, \alpha^2$ och notera att $\alpha^7 \equiv 1$ samt att α genererar $\mathbb{F}_{2^3}^*$). Således har elementen i $\mathbb{F}_{2^3}$ följande motsvarande representationer (den exponentiella i vänsterleden och polynomrepresentationen i högerleden):

$$\begin{aligned} 0 &= 0, \\ 1 &= 1, \\ \alpha^1 &= \alpha, \\ \alpha^2 &= \alpha^2, \\ \alpha^3 &= \alpha + 1, \\ \alpha^4 &= \alpha^2 + \alpha, \\ \alpha^5 &= \alpha^2 + \alpha + 1, \\ \alpha^6 &= \alpha^2 + 1. \end{aligned}$$

Addition av potenser av α utförs nu med hjälp av polynomrepresentationen, så att addition av α^3 och α^6 ger $\alpha^3 + \alpha^6 = (\alpha + 1) + (\alpha^2 + 1) = (\alpha^2 + \alpha) = \alpha^4$, med resultatet åter på exponentiell form.

För vidare studie av ändliga kroppar rekommenderas *Modern Algebra with Applications* av William J. Gilbert och W. Keith Nicholson [4], eller *Modern Algebra An Introduction* av John R. Durbin [2].

4. Cykliska polynomkoder

I det här kapitlet introduceras en klass av koder som kallas för *polynomkoder*, samt *cykliska koder* som utgör en underklass till dessa. Det så kallade *generatorpolynomet* utgör en viktig komponent i polynomkoder och introduceras i början av kapitelet varefter det studeras mer ingående för cykliska koder i avsnitt 4.2. I avsnitt 4.3 definieras slutligen BCH-varianten av Reed-Solomon-koderna, följt av en redogörelse för deras felkorrigerande egenskaper.

4.1. Polynomkoder

Polynomkoder är (n, k)-koder i vilka varje kodord $\boldsymbol{c} = \{c_0, c_1, c_2, ..., c_{n-1}\}$ översätts till ett kodpolynom genom att man låter symbolerna i kodordet utgöra koefficienterna i ett polynom av grad $n - 1$ på följande vis:

$$\boldsymbol{c} = (c_0, c_1, c_2, ..., c_{n-1}) \rightarrow c_0 + c_1 x + c_2 x^2 + ... + c_{n-1} x^{n-1}.$$

På samma sätt som ett kodord har ett motsvarande kodpolynom har varje block av informationssymboler ett motsvarande informationspolynom. Ett block $\boldsymbol{m} = (m_0, m_1, ..., m_{k-1})$ bestående av k stycken informationssymboler representeras alltså genom sitt informationspolynom $m(x) = m_0 + m_1 x + ... + m_{k-1} x^{k-1}$.

Enligt den allra enklaste kodningsmetoden kodas informationspolynomet genom multiplikation med ett polynom $g(x)$ av grad $n - k$ för att generera kodpolynomet $c(x)$, enligt

(5), och det resulterande kodordet utgörs av kodpolynomets koefficienter. Polynomet $g(x)$ kallas för kodens *generatorpolynom* och är specifikt för varje kod.

$$c(x) = m(x)g(x) \tag{5}$$

Den enkla kodningsmetoden beskriven ovan resulterar i att informationssymbolerna blir blandade med kontrollsymbolerna i kodordet. En annan kodningsmetod utförs enligt

$$c(x) = m(x) \cdot x^{n-k} - r(x), \tag{6}$$

där $r(x)$ är resten vid division av $m(x) \cdot x^{n-k}$ med $g(x)$. Först multipliceras $m(x)$ med x^{n-k} vilket resulterar i att informationssymbolerna i sekvensen $(m_0, m_1, ..., m_{k-1})$ "flyttas" $n-k$ steg till vänster eller, med andra ord, att $n-k$ nollor läggs till i slutet av sekvensen. Sedan delas $m(x) \cdot x^{n-k}$ med generatorpolynomet $g(x)$ varefter resten $r(x)$ subtraheras från $m(x) \cdot x^{n-k}$ för att bilda kodordet.

Även med den sista metoden är det resulterande kodordet $c(x)$ en multipel av generatorpolynomet; enligt divisionsalgoritmen gäller nämligen att

$$m(x) \cdot x^{n-k} = q(x)g(x) + r(x),$$

för några unika polynom $q(x)$ och $r(x)$ sådana att $r(x) = 0$ eller $\deg r(x) < \deg g(x)$, där $r(x)$ är just resten då $m(x) \cdot x^{n-k}$ divideras med $g(x)$.

Eftersom $g(x)$ är av grad $n-k$ och $\deg r(x) < \deg g(x)$ är $\deg r(x) < n-k$; addition av $m(x) \cdot x^{n-k}$ med $r(x)$ resulterar således i att kodordet har informationssymboler i de k första positionerna och $n-k$ kontrollsymboler i de efterföljande positionerna, till skillnad från metoden i (5) som resulterar i blandade symboler.

En komplett definition av en polynomkod lyder som följer.

Definition 4.1.1. Låt $g(x)$ vara ett polynom $a_0 + a_1x + ... + a_{n-k}x^{n-k}$ av grad $n-k$ med koefficienter $a_0, a_1, ..., a_{n-k}$ som tillhör den ändliga kroppen $\mathbb{F}_q$. *Polynomkoden* av längden n som *genereras* av $g(x)$ över $\mathbb{F}_q$ är den kod vars kodord utgörs av de polynom av grad $< n$ över $\mathbb{F}_q$ som är delbara med $g(x)$. Polynomet $g(x)$ kallas för kodens *generatorpolynom.*

Vi har hittills betraktat kodord i en linjär kod C av längd n över en kropp $\mathbb{F}_q$ som vektorer i det linjära rummet $\mathbb{F}_q^n$. Den här tolkningen har gett koden viss algebraisk struktur eftersom ett linjärt rum är en sluten mängd med två operationer- vektoraddition och skalärmultiplikation- vilket gör att en linjärkombination av kodord också är ett kodord. När polynomkoder betraktas introduceras, som vi såg i föregående avsnitt, ytterligare en operation, nämligen parvis multiplikation av element. Därför är det användbart att även introducera ytterligare algebraisk struktur. Ett viktigt verktyg i samband med studie av polynomkoder är således isomorfin mellan det linjära rummet $\mathbb{F}_q^n$ och restklassringen

$$\mathbb{F}_q[x]/(x^n - 1) = \{a_0 + a_1x + ... + a_{n-1}x^{n-1} |\, a_i \in \mathbb{F}_q,\, 0 \leq i < n\},$$

som består av alla polynom av grad $n-1$ med koefficienter i $\mathbb{F}_q$, eller med andra ord alla polynom i $\mathbb{F}_q[x]$ reducerade modulo $(x^n - 1)$.

Isomorfin mellan $\mathbb{F}_q^n$ och $\mathbb{F}_q[x]/(x^n - 1)$ innebär att varje element i $\mathbb{F}_q^n$ kan representeras som något element i $\mathbb{F}_q[x]/(x^n - 1)$; varje element har en unik representant. Detta är

vad som gör det lämpligt att låta varje kodord av längd n i en linjär kod representeras som ett polynom av grad $n-1$. Sambandet kan beskrivas enligt:

$$(c_0, ..., c_{n-1}) = \mathbf{c} \in \mathbb{F}_q^n \mapsto \phi(\mathbf{c}) = c_0 + c_1 x + ... + c_{n-1}x^{n-1} \in \mathbb{F}_q[x]/(x^n - 1),$$

där avbildningen $\phi : \mathbb{F}_q^n \to \mathbb{F}_q[x]/(x^n - 1)$ utgör isomorfin.

En linjär kod C av längd n kan således betraktas inte bara som ett underrum till det linjära rummet $\mathbb{F}_q^n$ utan även som en delmängd till $\mathbb{F}_q[x]/(x^n - 1)$.

4.2. Cykliska koder

En viktig egenskap av vissa linjära koder är att de är *cykliska*.

Definition 4.2.1. En linjär (n, k)-kod C sägs vara *cyklisk* om det för varje kodord $(c_0, c_1, ..., c_{n-1}) \in C$ även gäller att

$$(c_{n-1}, c_0, c_1, ..., c_{n-2}) \in C.$$

Cykliska koder utgör en speciell delmängd till polynomkoder, vilket följande sats implicerar.

Sats 4.2.1. *Den linjära koden $C \subset \mathbb{F}_q^n$ är cyklisk om och endast om C är ett ideal i $\mathbb{F}_q[x]/(x^n - 1)$.*

Bevis. Antag att C är ett ideal i $\mathbb{F}_q[x]/(x^n - 1)$, vi vill visa att C är cyklisk. Eftersom C är ett ideal så gäller för något fixt kodord

$$c(x) = c_0 + c_1 x + c_2 x^2 + ... + c_{n-1}x^{n-1} \in \mathbb{F}_q[x]/(x^n - 1)$$

att $x \cdot c(x)$ också är ett kordord. Vi har att

$$\begin{aligned} x \cdot c(x) &= x(c_0 + c_1 x + ... + c_{n-2}x^{n-2} + c_{n-1}x^{n-1}) \\ &= c_0 x + c_1 x^2 + ... + c_{n-2}x^{n-1} + c_{n-1}x^n \\ &= c_0 x + c_1 x^2 + ... + c_{n-2}x^{n-1} + c_{n-1} \\ &= c_{n-1} + c_0 x + c_1 x^2 + ... + c_{n-2}x^{n-1}. \end{aligned}$$

Från isomorfin får vi att

$$(c_{n-1}, c_0, ..., c_{n-2}) \in C.$$

vilket visar från definition (4.2.1) att C är cyklisk.

Antag nu istället att C är cyklisk. Vi vill visa att C är ett ideal i $\mathbb{Z}_n[x]/(x^n - 1)$. Eftersom C är cyklisk så gäller för varje kodord $c(x)$ att $x \cdot c(x)$ också är ett kodord i C. På samma sätt gäller för varje k att $x^k c(x)$ också är ett kodord till C. Då C är linjär så följer det för varje kodord $a(x)$ att $a(x)c(x) \in C$. Alltså gäller att C är ett ideal. □

Satsen implicerar, eftersom $\mathbb{F}_q[x]/(x^n - 1)$ är en huvudidealring och varje ideal i $\mathbb{F}_q[x]/(x^n - 1)$ därmed är ett huvudideal, att varje cyklisk kod C utgör ett huvudideal. Därmed finns ett moniskt polynom $g(x)$ av lägsta grad i C som genererar C. Detta polynom är vad som kallas för generatorpolynomet; varje cyklisk kod är således en polynomkod. Vidare gäller att generatorpolynomet $g(x)$ till en cyklisk kod är delare till $x^n - 1$ eftersom det annars skulle gälla att gcd $(g(x), x^n - 1) = a(x) \in C$ där deg $a(x) <$ deg $g(x)$ vilket motsäger att $g(x)$ är av lägsta grad i C.

Varje cyklisk kod av längd n har alltså ett generatorpolynom som delar $x^n - 1$. Om $x^n - 1$ faktoriseras i faktorer som är irreducibla över $\mathbb{F}_q$, enligt $x^n - 1 = f_1(x)f_2(x), ..., f_t(x)$, kan således alla cykliska koder av längd n genereras genom att på alla möjliga sätt välja en av de 2^t faktorerna av $x^n - 1$ som generatorpolynom ($x^n - 1$ har 2^t olika faktorer eftersom det har t irreducibla faktorer och varje irreducibel faktor $f_i(x)$ antingen tillåts att ingå i produkten eller inte). Sedan definieras respektive kod som mängden av alla multiplar av det valda generatorpolynomet modulo $(x^n - 1)$, enligt följande exempel.

Exempel 4.2.1. Antag att en kod av längd $n = 9$ ska definieras över kroppen $\mathbb{F}_2$. Kroppen $\mathbb{F}_2$ utgörs av elementen $\{0, 1\}$ och vi använder oss av isomorfin $\mathbb{F}_2^9 \cong \mathbb{F}_2[x]/(x^9 - 1)$ samt resonemanget om faktoriseringen i föregående stycke. Faktorisering av $x^9 - 1$ i irreducibla faktorer över $\mathbb{F}_2$ ger $x^9 - 1 = (x - 1)(x^2 + x + 1)(x^6 + x^3 + 1)$. Det finns alltså 3 irreducibla faktorer och därmed 2^3 sätt att välja faktorer som ska utgöra kodens generatorpolynom; eftersom olika generatorpolynom definierar olika koder finns det med andra ord sammanlagt 8 cykliska koder av längd 9 över $\mathbb{F}_2$. Vi kan till exempel välja $g(x) = (x - 1)$ vilket resulterar i en feldetekterande $(9, 8)$-kod vars kodord är alla möjliga ord av längd 9 med jämn vikt (det vill säga en kod med jämn paritetsbit liknande den i tabell 1). Om vi väljer $g(x) = (x^2 + x + 1)(x^6 + x^3 + 1) = x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ får vi en $(9, 1)$-kod vars meddelanden är antingen 0 eller 1 och vars enda kodord är **0** och **1**; vi har alltså genererat repetitionskoden av längd 9.

Om en cyklisk kod kan definieras som mängden av alla multiplar av något polynom så är det intuitivt att koden även kan specificeras genom att kräva att alla dess kodpolynom har vissa specifika nollställen. Om $\alpha \in \mathbb{F}_q$, för $q = p^m$ där p är ett primtal och m ett positivt heltal, så är *minimalpolynomet* av α över $\mathbb{F}_p$ det irreducibla polynom $f(x) \in \mathbb{F}_p[x]$ som uppfyller $f(\alpha) = 0$. Om vi nu återigen betraktar faktoriseringen $x^n - 1 = f_1(x)f_2(x), ..., f_t(x)$ och låter α_i vara ett nollställe till det irreducibla polynomet $f_i(x)$ i $\mathbb{F}_q$ så är allså $f_i(x)$ minimalpolynomet av α_i och således utgörs koden som genereras av $f_i(x)$ av mängden av polynom $c(x)$ för vilka $c(\alpha_i) = 0$. En kod, C, kan definieras på det här sättet; det vill säga genom att ta en mängd $\alpha_1, \alpha_2, ..., \alpha_s$ och låta ett kodpolynom $c(x) \in C$ om och endast om $c(\alpha_i) = 0$ för alla $i = 1, 2, ..., s$. För att alla nollställen ska komma med följer det att kodens generatorpolynom är den minsta gemensamma multipeln av minimalpolynomen av $\alpha_1, \alpha_2, ..., \alpha_s$. Det är på ett sätt liknande det här som de så kallade BCH-koderna definieras, vilket vi ska se i avsnitt 4.3.

Generatorpolynomets utformning, för en cyklisk kod i $\mathbb{F}_q[x]/(x^n - 1)$, har alltså att göra med faktoriseringen av $x^n - 1$ samt nollställena till de irreducibla faktorer $f_i(x)$ som ingår i generatorpolynomet. Nollställena till faktorerna $f_i(x)$ är naturligtvis nollställen till $x^n - 1$ så hur många irreducibla faktorer som finns över F_q beror på hur många nollställen $x^n - 1$ har i $\mathbb{F}_q$. Om $n = q - 1$ så är $x^n - 1 = x^{q-1} - 1$ och således är den $q - 1$: *te* enhetsroten α i F_q ett nollställe till $x^n - 1$ över F_q. Följdaktligen är samtliga element i den multiplikativa gruppen av $\mathbb{F}_q$, det vill säga α^i för $0 \leq i \leq q - 2$, nollställen till $x^n - 1$ eftersom $(\alpha^i)^n = (\alpha^n)^i = 1^i = 1$ för alla i. Eftersom ett polynom av grad n har högst n stycken nollställen utgör elementen i den multiplikativa gruppen av F_q alla nollställen till $x^n - 1$, vilken således kan faktoriseras i linjära faktorer över F_q, så att $x^n - 1 = (x - \alpha)(x - \alpha^2)...(x - \alpha^{n-1})(x - \alpha^n)$. Eftersom α är primitiv i $\mathbb{F}_q$ så är alla potenser $\alpha, \alpha^2, ..., \alpha^{n-1}, \alpha^n$ olika och således är de linjära faktorerna olika. Observationerna i det här stycket är centrala i samband med konstruktionen av Reed-Solomon-koderna, vilket framgår i avsnitt 4.3.

4.3. BCH-koder och Reed-Solomon-koder

BCH-koder (namngivna efter dess uppfinnare R. Bose, D.K. Ray-Chaudhuri, och A. Hocquenghem) är en klass av cykliska polynomkoder vars generatorpolynom konstrueras baserat på hur många fel koden ska kunna korrigera. Enligt Sats 2.3.1 korrigerar en linjär kod, C, t stycken fel om kodens minsta avstånd $d(C) > 2t$. När en BCH-kod av längd n konstrueras bestäms generatorpolynomet av vad koden *minst* får ha för minsta avstånd.

Definition 4.3.1. En *BCH-kod med avstånd d minst lika med δ* är en cyklisk kod av längd n över $\mathbb{F}_q$ vars generatorpolynom $g(x)$ är den minsta gemensamma multipeln av minimalpolynomen av $\alpha^l, \alpha^{l+1}, ..., \alpha^{l+\delta-2}$, för något l där α är en primitiv n*:te* enhetsrot. Om $n = q^m - 1$, för ett positivt heltal m, och följdaktligen α är ett primitivt element i $\mathbb{F}_{q^m}$, så kallas BCH-koden för *primitiv*.

Reed-Solomon-koderna är också en klass av cykliska polynomkoder, närmare bestämt en underklass till BCH-koderna. Varje Reed-Solomon-kod karaktäriseras av tre parametrar (som vanligt med blockkoder), nämligen kodens längd n, meddelandenas längd k, och alfabetets storlek q (det vill säga den ändliga kroppens ordning). För dessa parametrar gäller att $k \leq n \leq q$. Varje parameterkombination ger upphov till ett visst minsta avstånd. En Reed-Solomon-kod över $\mathbb{F}_q$ är en BCH-kod med parametrarna $n = q - 1$ och $l = 1$, enligt följande definition.

Definition 4.3.2. En *Reed-Solomon-kod* är en primitiv BCH-kod av längden $n = q - 1$ över $\mathbb{F}_q$ som har generatorpolynomet

$$g(x) = \prod_{i=1}^{\delta-1}(x - \alpha^i),$$

där α är ett primitivt element i $\mathbb{F}_q$.

En Reed-Solomon-kod som korrigerar t stycken fel kan konstrueras genom att låta $\delta - 1 = 2t$ i definitionen. Nedan bevisas att detta gäller så länge $t < p^{m-1}$ där p är ett primtal sådant att $q = p^m$.

Sats 4.3.1. *Om $t < p^{m-1}$ så är minsta avståndet mellan kodorden i Reed-Solomon-koden med längd $n = p^m - 1$ över $\mathbb{F}_{p^m}$ minst $2t + 1$.*

Bevis. Antag att det omvända gäller, det vill säga att koden innehåller ett kodpolynom med färre än $2t + 1$ nollskilda termer (vi minns från Sats 2.3.2 att minsta avtåndet är lika med minsta vikten för linjära koder),

$$c(x) = c_1x^{r_1} + c_2x^{r_2} + ... + c_{2t}x^{r_{2t}},\ r_1 < ... < r_{2t}.$$

Eftersom $c(x)$ är av grad $2t$ och är en multipel av generatorpolynomet $g(x)$ har $c(x)$ de $2t$ n:te enhetsrötterna $\alpha, \alpha^2, ..., \alpha^{2t}$ som nollställen. Därför gäller, för alla i som uppfyller $1 \leq i \leq 2t$,

$$c(\alpha^i) = c_1\alpha^{ir_1} + c_2\alpha^{ir_2} + ... + c_{2t}\alpha^{ir_{2t}} = \alpha^{ir_1}(c_1 + c_2\alpha^{ir_2 - ir_1} + ... + c_{2t}\alpha^{r_{2t} - r_1}) = 0.$$

Genom införande av $s_i = r_i - r_1$ uppfyller koefficienterna $c_1, ..., c_{2t}$ följande linjära ekvationssystem:

$$\begin{array}{ccccccccc} c_1 & + & c_2\alpha^{s_2} & + & ... & + & c_{2t}\alpha^{s_{2t}} & = & 0 \\ c_1 & + & c_2\alpha^{2s_2} & + & ... & + & c_{2t}\alpha^{2s_{2t}} & = & 0 \\ . & & & & & & & . & \\ . & & & & & & & . & \\ . & & & & & & & . & \\ c_1 & + & c_2\alpha^{2ts_2} & + & ... & + & c_{2t}\alpha^{2ts_{2t}} & = & 0. \end{array}$$

Det linjära ekvationssystemet har koefficientmatrisen:

$$\begin{pmatrix} 1 & \alpha^{s_2} & ... & \alpha^{s_{2t}} \\ 1 & \alpha^{2s_2} & ... & \alpha^{2s_{2t}} \\ . & & & . \\ . & & & . \\ 1 & \alpha^{2ts_2} & ... & \alpha^{2ts_{2t}} \end{pmatrix}$$

Denna matris är på samma form som den kända Vandermonde-matrisen som beskriver geometriska följder och dess determinant är således Vandermonde-determinanten

$$\prod_{2t \geq i > j \geq 2} (\alpha^{s_i} - \alpha^{s_j}) \neq 0.$$

Determinanten är nollskild eftersom antagandet $t < p^{m-1} \implies 2t < 2 \cdot p^{m-1} \leq p^m$ vilket betyder att $\alpha, \alpha^2, ..., \alpha^{2t}$ är olika (eftersom α är primitiv i $\mathbb{F}_{2^m}$). Att determinanten är nollskild implicerar att kolonnerna i koefficientmatrisen är linjärt oberoende och således har de linjära ekvationerna den unika triviala lösningen $c_1 = c_2 = ... = c_{2t} = 0$. Antagandet att $c(x)$ har färre än $2t + 1$ nollskilda termer implicerar alltså att $c(x) = 0$. Koden har således inga nollskilda kodord med färre än $2t + 1$ nollskilda termer, och därmed är kodens minsta avstånd minst $2t + 1$. □

4.4. Reed-Solomon-kodernas egenskaper för felkorrigering

Reed-Solomon-kodernas minsta avstånd, d, har det största värdet som är möjligt för linjära (n, k)-koder att uppnå. Detta värde kallas för Singletons gräns och innebär att likhet uppfylls i olikheten

$$A_q(n, d) \leq q^{n-d+1}, \tag{7}$$

där $A_q(n, d)$ är det maximala antalet kodord i en kod av längd n vars minsta avstånd är d och vars alfabet utgörs av q somboler. Olikheten Singletons gräns gäller för alla blockkoder med dessa parametrar. För en linjär blockkod C över en ändlig kropp $\mathbb{F}_q$ med meddelanden av längd k är det maximala antalet kodord vidare lika med q^k. Singletons gräns kan alltså för linjära blockkoder skrivas som

$$q^k \leq q^{n-d+1},$$

vilket är ekvivalent med

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

Den sista olikheten uttrycker på ett tydligare sätt Singletons gräns som en begränsning på kodens minsta avstånd.

Koder som uppfyller likhet i olikheterna ovan, det vill säga vars minsta avstånd är lika med Singletons gräns, kallas för MDS-koder (Maximum Distance Separable); Reed-Solomon-koderna är således MDS-koder och är i den meningen optimala. Nedan visar vi att olikheten Singletons gräns (7) måste gälla för alla blockkoder.

Bevis. Betrakta en godtycklig blockkod C över ett alfabet med storleken q och minsta avståndet d. Vi antar att kodorden har genererats genom kodningsmetoden (6) som beskrevs i avsnitt 4.1, så att kontrollsymbolerna utgör de sista $n - k$ symbolerna i varje kodord. Om vi raderar de sista $d - 1$ symbolerna i varje kodord (detta kallas för att punktera koden) så får vi en ny kod, men de resulterande kodorden är fortfarande parvis olika eftersom

de ursprungliga kodorden skiljer sig på minst d ställen. Alltså har den nya koden samma storlek som den ursprungliga. Varje nytt kodord har längden $n - d + 1$ så det kan finnas som flest q^{n-d+1} kodord i den nya, och således även den ursprungliga, koden. Eftersom koden C var godtycklig måste olikheten $A_q(n, d) \leq q^{n-d+1}$ gälla för den största möjliga koden med samma parametrar som C. □

Sats 4.4.1. *Reed-Solomon-koderna som definieras enligt definition 4.3.2 uppfyller likhet i Singletons gräns.*

Bevis. Av konstruktionen av en Reed-Solomon-kod som anges i definition 4.3.2, med längd n och generatorpolynom $g(x)$, följer att kodens dimension är $k = n - \deg g(x) = n - (\delta - 1)$. Vidare är kodens minsta avstånd d minst lika med δ (enligt Sats 4.3.1) så att $d \geq \delta = n-k+1$ men enligt Singletons gräns är $d \leq n - k + 1$ så därför är $d = n - k + 1$. □

Minsta avståndet för en Reed-Solomon-kod är alltså $d = n - k + 1$ och vi minns från kapitel 2 att en kod korrigerar t fel om minsta avståndet $d > 2t$. För minsta avståndet av en Reed-Solomon-kod gäller att $d > n - k = 2\frac{n-k}{2}$ och en Reed-Solomon-kod korrigerar således $\frac{n-k}{2}$ fel.

Utöver att Reed-Solomon-koderna är MDS-koder är de särskilt lämpade för korrigering av så kallade *burst*-fel, det vill säga fel som inte uppstår slumpmässigt i koden utan förekommer i kluster. CD-skivor och QR-koder är exempel på applikationer där burst-fel är förekommande, i form av exempelvis repor eller smuts. Anledningen till att Reed-Solomon-koderna är bra på att korrigera burst-fel är att eftersom symbolerna i koden utgörs av sekvenser av bitar så spelar det ingen roll ur felkorrigeringsperspektiv hur många bitar i en symbol som blivit fel; eftersom Reed-Solomon-koderna har förmåga att korrigera ett visst antal felaktiga symboler så uppfattas multipla bitfel i praktiken som ett enda fel, så länge de felaktiga bitarna tillhör samma symbol.

För intressant läsning och en mer ingående studie av Reed-Solomon-koderna och dess egenskaper och tillämpningar än vad som får plats här hänvisas läsaren till boken *Reed-Solomon Codes and Their Applications* av Stephen B. Wicker och Vijay K. Bhargava [6]. För vidare läsning om koders begränsningar rekommenderas återigen *Introduction to Coding Theory* av J.H. van Lint [5].

4.5. Förkortning av Reed-Solomon-koden

Som framgår av definition 4.3.2 har en Reed-Solomon-kod över $\mathbb{F}_q$ alltid längd $n = q - 1$ efter konstruktionen. Så pass långa kodord behövs dock inte i alla sammanhang och i de fallen kan koden *förkortas*. Det går till så att informationssekvenserna fylls ut med binära nollor så att kodningen trots allt resulterar i kodord med längd $n = 255$. Efter kodningen tags sedan de extra nollorna bort så att kortare kodord erhålls, för att sedan återigen läggas till kodorden när dessa ska avkodas. Om en (n_0, k_0)-kod behövs så konstrueras alltså en $(255, k_0+(255-n_0))$-kod varefter $(255-n_0)$ binära nollor avlägsnas från samtliga kodord [3].

En $(26, 16)$-kod kan således konstrueras genom att informationssekvenserna av 16 informationssymboler fylls ut med 229 binära nollor. Kodningen resulterar sedan i ett kodord av längd $n = 255$, men efter att de extra nollorna har tagits bort har de resulterande kodorden längden $n = 26$.

5. Konstruktion av en QR-kod

Det här kapitlet redogör för implementeringen av ordet KODNINGSTOERI i form av en QR-kod med hjälp av en Reed-Solomon-kod.

5.1. Bakgrund och specifikationer

QR-koder används idag i många olika sammanhang såsom i biblioteket, reklambranchen, och transportföretag. En QR-kod är en tvådimensionell kod som konstruerades av det japanska företaget Denso Wave med ambitionen att sammanställa stora mängder av information i mindre utrymme med hjälp av binära koder. Binära koder eller binära tal är ett data-och programmeringsspråk som refererar till talen 0 och 1 vilket används i datakommunikation. En QR-kod läses och avkodas vanligvis med hjälp av applikationer i smarta enheter som mobiltelefoner.

Beroende på vilken typ av information som ska kodas används olika så kallade *former*; det fyra formerna som finns är numerisk form, alfanumerisk form, byte, och kanji. Kanji är ett av de japanska skriftspråken och det består av 3000 tecken vars ursprung förklaras av namnet som just betyder ”skrivtecken från Kina”. De olika formerna använder olika standarder för konstruktionen av själva QR-koden. En QR-koder kan göras i en av 40 olika storlekar, eller *versioner*, varvid den minsta, version 1, består av 21×21 pixlar och den största, version 40, består av 177×177 pixlar. Varje version har en maximal kapacitet på antalet informationssymboler som får plats; kapaciteten är beroende av vilken form och vilken felkorrigeringsnivå som används. En QR-kod kan ha en av de fyra felkorrigeringsnivåerna L, M, Q och H, vilka motsvarar 7 %, 15 %, 25 % respektive 30 % felkorrigering. QR-koden som har implementerats i det här arbetet är av version 1 (det vill säga 21×21 pixlar) och felkorrigeringsnivå M, vilket motsvarar 15 % felkorrigering.

Resten av kapitlet ägnas åt en beskrivning av konstruktionen av en QR-kod med ordet KODNINGSTEORI. Källor för konstruktionen av QR-koder är begränsade och majoriteten av studien inom det här arbetet avseende QR-koden är baserad på hemsidan `thonky` [7]. Vid konstruktionen av QR-koden med ordet KODNINGSTEORI används QR-kod- och symbolikspecifikationer enligt gällande standarder; QR-koden har antagits som en ISO-standard (ISO/IEC 18004) [8]. Kodningen med en Reed-Solomon-kod bygger på kunskaper från tidigare kapitel, samt tekniker inom data- och informationsteknik.

5.2. Analys och data

Att koda data med en viss specificerad felkorrigeringsnivå är i stora drag en process i två steg. Först delas meddelandet upp i informationssymboler som utgörs av 8 bitar vardera och sedan läggs konstrollsymboler till med hjälp av en RS-kod med vissa parametrar. I avsnitt 4.4 visades att en RS-kod av längd n och dimension k korrigerar $\frac{n-k}{2}$ fel. När en QR-kod implementeras bestäms längden n av storleken på QR-koden, så att olika *versioner* har olika värden på n. För version 1 som har använts i det här arbetet är $n = 26$. Dimensionen k är sedan en designparameter som bestäms av den felkorrigeringsnivå som ska uppnås. I det här arbetet implementeras en QR-kod som rättar 15 % felaktiga kodsymboler och för den felkorrigeringsnivån behövs $k = 16$. Antalet bitar som utgör det ursprungliga meddelandet är inte alltid delbart med 8, och dessutom måste meddelandet som nämnt ha en viss längd för att den specificerade felkorrigeringsnivån ska kunna uppnås; därför utökas processen till ytterligare några steg där extra utfyllnadsbitar läggs till det ursprungliga meddelandet.

Processen ska nu beskrivas i sin helhet.

I textmeddelandet $X = \text{KODNINGSTEORI}$ som implementeras är bokstäverna versaler eftersom det är vad som finns i den alfanumeriska formen. Det första som måste göras är att omvandla texten till binärkod med hjälp av den alfanumeriska tabellen i bilaga B.1 och en teknik som vi illustrerar nu. Tecknen i textmeddelandet delas in i par så långt det är möjligt (KO, DN, IN, GS, TE, OR och I) och sedan multipliceras det första tecknets alfanumeriska värde (se tabell i bilaga B.1) med 45, som är det totala antalet tecken i den alfanumeriska formen; därefter adderas värdet som motsvarar det andra tecknet. Resultatet görs slutligen om till binär form:

$$\begin{aligned} &\text{K} = 20,\ \text{O} = 24, \\ &(45 * 20) + 24 = 924, \\ &924 = 1110011100. \end{aligned} \tag{8}$$

Det största värdet som operationerna i (8) kan resultera i är $(45 * 44) + 44 = 2024$ (eftersom det största alfanumeriska värdet är 44) vilket på binär form är en 11-bitarssträng. De binära talen motsvarande respektive teckenpar skrivs därför som 11-bitarssträngar (det vill säga extra nollor läggs till i början av strängen om 11 bitar egentligen inte behövs).

I fallet då man bara har en teckenkombination av udda längd, likt bokstaven I i ordet KODNINGSTEORI, så omvandlas denna till sitt alfanumeriska värde på binär form varefter strängen utökas till en 6-bitarssträng (om den inte redan är 6 bitar lång), eftersom det största alfanumeriska värdet 44 kräver 6 bitar.

När samma procedur som i (8) har upprepats för alla par i ordet KODNINGSTEORI och de binära strängarna har utökats till 6 respektive 11 bitar blir resultatet:

$$\begin{aligned} &\{\text{KO, DN, IN, GS, TE, OR, I}\} = \{01110011100,\ 01001100000, \\ &01101000001,\ 01011101100,\ 10100100111,\ 10001010011,\ 010010\}. \end{aligned}$$

Nästa steg är att välja en *formindikator* samt en *teckenräknarindikator*, vilka sedan placeras framför meddelandet. En formindikator är en indikator på vilken typ av tecken som ska kodas så att en QR-läsare kan ställa in sig på vilken form av kod det är; den alfanumeriska formindikatorn är 0010 och resterande indikatorer ses i tabell 3. En teckenräknarindikator är en indikator som visar hur många tecken som ska kodas, och denna indikator har olika längd beroende på vilken version QR-koden har. KODNINGSTEORI består av 13 tecken och QR-koden som ska konstrueras är av version 1; enligt tabell 3 ska alltså teckenräknarindikatorn vara 9 bitar lång. Det betyder att talet 13 ska översättas till en 9 bitar lång sträng; resultatet blir 000001101.

Tabell 3: I tabellen visas de formindikatorer som motsvarar de olika teckenformerna samt hur långa teckenräknarindikatorerna ska vara beroende på teckenform och version.

Teckentyp	Formindikator	Version 1-9	Version 10-26	Version 27-40
Numerisk	0001	10	12	14
Alfanumerisk	0010	9	11	13
Byte	0100	8	16	16
Kanji	1000	8	10	12

Textmeddelandet KODNINGSTEORI består av 13 tecken och ligger i det ändliga kroppen $\mathbb{F}_{2^8}$. Enligt QR-kodsspecfikationerna behöver en QR-kod av version 1 och felkorrigeringnivå M, innehålla 238 bitar totalt sett varav 30 tillhör formatdelen. I nuläget så har vi strängen,

$$\begin{aligned} X = \ & 0010\ 000001101\ 01110011100\ 01001100000 \\ & 01101000001\ 01011101100\ 10100100111\ 10001010011\ 010010\ , \end{aligned} \tag{9}$$

vilket är en 85 bitar lång sträng där de kvarvarande bitarna utav de 238 kommer fyllas upp med uppfyllnad och bland annat felkorrigerande kodord och formatering.

En QR-kod av version 1 med felkorrigeringsnivå M kodas med en RS-kod med parametrarna $n = 26$, $k = 16$ (parametervärden för de olika versionerna och felkorrigeringsnivåerna specificeras i tabell i bilaga B.2). Ordet KODNINGSTEORI ska alltså delas upp i 16 stycken 8 bitarssträngar. Meddelandet KODNINGSTEORI innehåller dock inte så många bitar och därför måste meddelandet fyllas ut tills den har rätt längd.

Om fyra eller färre utfyllnadsbitar behövs för att uppnå $16 \cdot 8 = 128$ bitar så läggs de helt enkelt till i slutet av meddelandet i (9). Saknas det fler än fyra bitar, såsom i fallet med KODNINGSTEORI, så läggs en sträng med fyra nollor 0000 till i slutet av meddelandekodet i ekvation (9). Är strängen inte delbar med 8 så läggs ytterligare en sträng med nollor till vars längd är differensen mellan den nuvarande längden på meddelandet och närmsta större decimala tal som är delbart med 8. I fallet med KODNINGSTEORI ska alltså en 7-bitarssträng med nollor läggas till, eftersom 96 är det decimala tal delbart med 8 som ligger närmast den nuvarande längden 89.

Är meddelandekoden inte lång nog (det vill säga mindre än 128) efter dessa steg så läggs slutligen den 16 bitar långa bitsträngen 11101100 00010001 till upprepade gånger i slutet av meddelandet tills det är 128 bitar långt. I fallet med KODNINGSTEORI måste denna sträng läggas till två gånger för att nå $k = 128$ bitar.

När meddelandet nu har fyllts ut och blivit 128 bitar långt kan det delas upp i 16 stycken 8-bitarssträngar; resultatet är således

$$\begin{aligned} X = \ & 00100000\ 01101011\ 10011100\ 01001100 \\ & 00001101\ 00000101\ 01110110\ 01010010 \\ & 01111000\ 10100110\ 10010000\ 00000000 \\ & 11101100\ 00010001\ 11101100\ 00010001\ . \end{aligned}$$

5.3. Felkorrigerande kodord

Nästa steg är att lägga till felkorrigerande kontrollsymboler till meddelandet. Antalet kontrollsymboler som ska läggas till är $n - k$, vilket med $n = 26$ och $k = 16$ ger $n - k = 10$. QR-koder använder RS-koder för att konstruera dessa felkorrigerande kontrollsymboler; första steget för att koda ett meddelande är att konstruera ett meddelandepolynom

$$p(x) = m_0 + m_1 x + ... + m_{k-1} x^{k-1},$$

där $x_i \in \mathbb{F}_q$. Koefficienterna $(m_{k-1}, m_{k-2}, ..., m_0)$ fås genom att översätta 8-bitarssträngarna som utgör meddelandet X (det vill säga informationssymbolerna i meddelandet KODNINGSTEORI med indikatorer och utfyllnadsbitar) från föregående avsnitt till decimala

tal; 8-bitarssträngarna i ordet X motsvarar de decimala talen:

$$(m_{15}, m_{14}, ..., m_0) = (32, 107, 156, 76, 13, 5, 118, 82, 120, 166, 144, 0, 236, 17, 236, 17)\,. \quad (10)$$

Nästa steg är att konstruera ett generatorpolynom

$$g(x) = (x - \alpha^0)(x - \alpha^1)...(x - \alpha^{n-k-1}), \quad (11)$$

där $\alpha \in \mathbb{F}_q$ är ett primitivt element. För att generera de felkorrigerande kontrollsymbolerna divideras meddelandepolynomet $p(x)$ multiplicerat med en skalfaktor x^{n-k} med generatorpolynomet $g(x)$ enligt kodningsmetoden (6) som beskrevs i avsnitt 4.1. Denna division resulterar i ett restpolynom, vars grad är högst $n - k - 1$,

$$r(x) = r_0 + r_1 x + ... + r_{n-k-1} x^{n-k-1}$$

där $x_i \in \mathbb{F}_q$ och vars koefficienter $(r_{n-k-1}, r_{n-k-2}, ..., r_0)$ är de sökta felkorrigerande kontrollsymbolerna. Dessa felkorrigerande kontrollsymboler tillsammans med meddelandet X utgör det slutliga kodordet:

$$C = (m_{k-1}, m_{k-2}, ..., m_0, r_{n-k-1}, r_{n-k-2}, ..., r_0) \in \mathbb{F}_q^n.$$

Målet är alltså att i fallet med KODNINGSTEORI bestämma $(r_0, r_1, ..., r_9)$; för att kunna utföra divisionen $\frac{x^{n-k}p(x)}{g(x)}$ måste först generatorpolynomet $g(x)$ konstrueras. När QR-koder konstrueras används RS-koder över den ändliga kroppen $\mathbb{F}_{2^8}$ eftersom informatons- och kodsymbolerna utgörs av 8-bitarssträngar och det finns 2^8 möjliga sådana. I exempel 3.4.1 visades hur denna kropp kan konstrueras med hjälp av isomorfin $\mathbb{F}_{2^8} \cong \mathbb{Z}_2[x]/(1 + x^2 + x^3 + x^4 + x^8)$ och det är precis denna isomorfi som används för QR-koder. Bitsträngarna av längd 8 i meddelandet X utgör informationssymbolerna, enligt (10). Enligt tabellen i exempel 3.4.1 har varje element i $\mathbb{F}_{2^8}$ en entydig representaion som en 8-bitarssträng, vilket är precis vad som används. Som visades i exempel 3.4.1 kan elementen också representeras i termer av det primitiva elementet α. När vi gör beräkningarna nedan använder vi α-notationen.

För att illustrera hur ett generatorpolynom konstrueras låter vi $n - k = 3$ i (11) och får

$$\begin{aligned} g_3(x) &= (x - 1)(x - \alpha^1)(x - \alpha^2) \\ &= x^3 + \left(\alpha^0 + \alpha^1 + \alpha^2\right) x^2 + \left(\alpha^1 + \alpha^2 + \alpha^3\right) x + \alpha^2. \end{aligned}$$

För att kunna addera α-potenserna som tillhör samma potens av x så räknar vi på samma sätt som i exempel 3.4.2, men i kroppen $\mathbb{F}_{2^8}$ istället för $\mathbb{F}_{2^3}$, och får på så vis det resulterande generatorpolynomet för $n - k = 3$,

$$g_3(x) = x^3 + \alpha^{198} x^2 + \alpha^{199} x + \alpha^2.$$

I fallet med $n - k = 10$ för att koda meddelandet KODNINGSTEORI fås på samma sätt generatorpolynomet

$$g(x) = x^{10} + \alpha^{251} x^9 + \alpha^{67} x^8 + \alpha^{46} x^7 + \alpha^{61} x^6 + \alpha^{118} x^5 + \alpha^{70} x^4 + \alpha^{64} x^3 + \alpha^{94} x^2 + \alpha^{32} x + \alpha^{45}.$$

När generatorpolynomet är konstruerat kan divisionen $\frac{x^{n-k}p(x)}{g(x)}$ utföras och slutligen fås resten

$$r(x) = 82x^9 + 88x^8 + 66x^7 + 171x^6 + 69x^5 + 173x^4 + 42x^3 + 99x^2 + 234x + 81.$$

När beräkningarna utförs används $\alpha = 2$. Koefficienterna i detta restpolynom motsvarar kontrollsymbolerna:

$$(r_9, r_8, ..., r_0) = (82, 88, 66, 171, 69, 173, 42, 99, 232, 81)\,,$$

och följdaktligen blir det slutgiltiga kodordet

$$\begin{aligned} C = (m_{15}, m_{14}, ..., m_0, r_9, r_8, ..., r_0) = (32, 107, 156, 76, 13, 5, 118, 82, 120, \\ 166, 144, 0, 236, 17, 236, 17, 82, 88, 66, 171, 69, 173, 42, 99, 232, 81) \end{aligned}$$

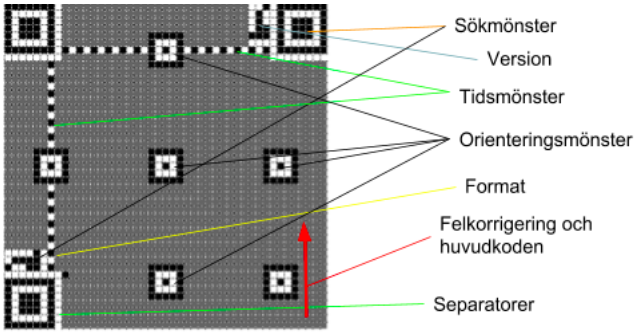
Nu återstår bara att konvertera de decimala talen till 8-bitarssträngar; det slutgiltliga $8 \cdot 26 = 208$ bitar långa kodordet som motsvarar meddelandet $X =$ KODNINGSTEORI och $n - k = 10$ kontrollsymboler är:

$$\begin{aligned} C = \; & 00100000\ 01101011\ 10011100\ 01001100\ 00001101 \\ & 00000101\ 01110110\ 01010010\ 01111000\ 10100110 \\ & 10010000\ 00000000\ 11101100\ 00010001\ 11101100 \\ & 00010001\ 01010010\ 01011000\ 01000010\ 10101011 \\ & 01000101\ 10101101\ 00101010\ 01100011\ 11101000 \\ & 01010001\ . \end{aligned} \tag{12}$$

5.4. Maskering

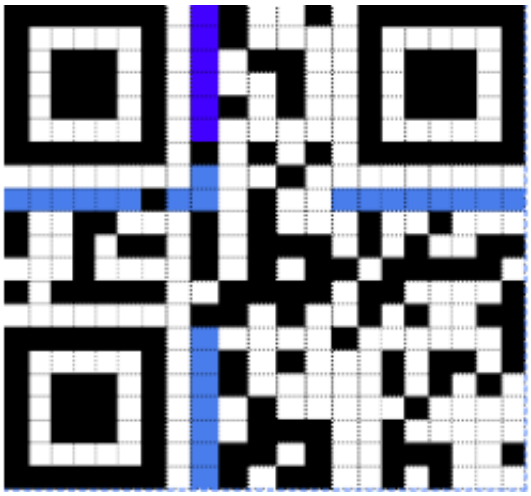
QR-koder består av olika funktionsmönster som har vissa specifikationer vilket underlättar för QR-kod-avläsaren att identifiera koderna och sedan läsa dem. De olika funktionsmönstren illustreras i figur 1 och beskriv i följande lista.

- Sökmönster är de tre kvadratformade mönster som i sig innehåller två mindre kvadratformade mönster vardera. QR-läsaren letar alltid först upp sökmönstret för att kunna identifiera mönstret som en QR-kod och sedan hitta och läsa själva koden.
- Separatorerna är de vita linjerna runt sökmönstren som separerar sökmönstret från resten av QR-koden.
- Tidsmönster är en linje som kopplar ihop alla sökmönstrena både vertikalt och horisontellt. Tidsmönstrets huvuduppgift är att underlätta för QR-läsaren att hitta positionen av alla elementen i en QR-kod.
- Orienteringsmönster används om en QR-kod är version 2 eller större. Orienteringsmönstret hjälper QR-läsaren vid avkodning.
- Format innehåller information om datamaskningsmönstret och kodens feltolerans.
- Version innehåller information om QR-kodens version.



Figur 1: Figuren illustrerar de olika delarna som bygger upp en QR-kod.

Informationssymbolerna och kontrollsymbolerna placeras ut nedifrån och upp. Bitarna läggs med startelement längst ned till höger och sedan fortsätter det ett steg till vänster och sedan ett steg upp. Nollor färgas med vit färg och ettor med svart färg. I figur 2 visas resultatet efter att kodordet C i (12) har placerats i QR-koden.

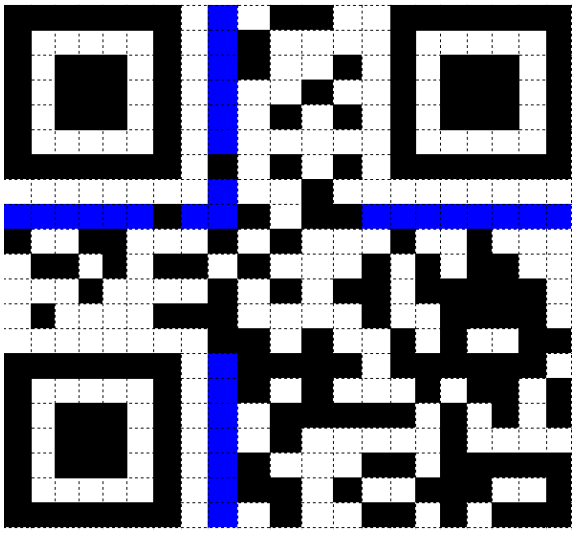


Figur 2: QR-kod med kodordet inplacerat.

Efter att datakoden har placerats måste koden maskeras så att QR-kod-avläsaren lätt och snabbt kan avläsa den. Maskering kan defineras som en binärkodsskiftning vilket betyder att en nolla omvandlas till en etta och tvärtom. Maskering görs enligt standarden för QR-koder med hjälp av de 8 olika ekvationerna som presenteras i tabell 4.
Efter att koden har maskerats på 8 olika sätt, väljs den QR-kod som har det minsta kompakta binärmönstret. För QR-koden i det här arbetet väljs $i \equiv 0 \pmod 2$ och resultatet visas i figur 3. Detta betyder att om $i \equiv 0 \pmod 2$ så skiftar binärkoden. De blåfärgade bitmodulerna i bilden visar format- och versionsinformation så att QR-avläsaren kan förstå vilket maskeringsformat och vilken felkorrigeringsnivå som används för just den specifika QR-koden.

Tabell 4: Tabellen visar ekvationer för olika maskeringsnummer (i = radnummer, j = kolumn-nummer)

Masknummer	Ekvation
0	$i + j \equiv 0 \pmod 2$
1	$i \equiv 0 \pmod 2$
2	$j \equiv 0 \pmod 3$
3	$i + j \equiv 0 \pmod 3$
4	$\lfloor \frac{i}{2} \rfloor + \lfloor \frac{j}{3} \rfloor \equiv 0 \pmod 2$
5	$i \cdot j \pmod 2 + i \cdot j \pmod 3 \equiv 0$
6	$i \cdot j \pmod 2 + i \cdot j \pmod 3 \equiv 0 \pmod 2$
7	$i + j \pmod 2 + i \cdot j \pmod 3 \equiv 0 \pmod 2$



Figur 3: Figuren illustrerar den resulterande QR-koden efter maskering.

5.5. Format

I det här avsnittet används BCH-koder över $\mathbb{Z}_2$ vilket är en annan felkorrigeringsteknik. Formatdelen i en QR-kod har oavsett version och felkorrigeringsnivå längden $n = 15$. För en QR-kod av version 1 med felkorrigeringnivå M behöver formatdelen, enligt standard för QR-koder, innehålla $k = 5$ informationssymboler och $n - k = 10$ kontrollsymboler; det betyder att ett informationspolynom av grad 4 och ett generatorpolynom av grad $n - k = 10$ behöver konstrueras. Enligt sats 3.2.5 och exempel 4.2.1, faktorisering av tre irreducibla polynomen genererar generator polynomet,

$$\begin{aligned} g(x) &= (x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)(x^2 + x + 1) \\ &= (x^{10} + x^8 + x^5 + x^4 + x^2 + x + 1). \end{aligned}$$

Generatorpolynomet ligger i den ändliga kroppen $\mathbb{F}_2$ som utgörs av elementen $\{0, 1\}$. Som visas har vi valt att använda generatorpolynomet ovan vilket korrigerar upp till $t = 3$ fel och har minsta distansen $d = 7$. Informationssymbolerna ges av tabell 5 vilken visar binära tal för varje felkorrigeringsnivå, samt tabell 4 som visar de binära talen för ekvationen som användes för maskeringsnummer 1 (001).
Uppsättningen av de två binära talen ger informationssymbolerna 00001, vilket betyder

Tabell 5: Presentation av felkorrigeringsnivå i binära tal.

Felkorrigeringsnivå	Binär
L	01
M	00
Q	11
H	10

att koefficienterna i informationspolynomet är

$$(m_0, m_1, ..., m_4) = (0, 0, 0, 0, 1)\,,$$

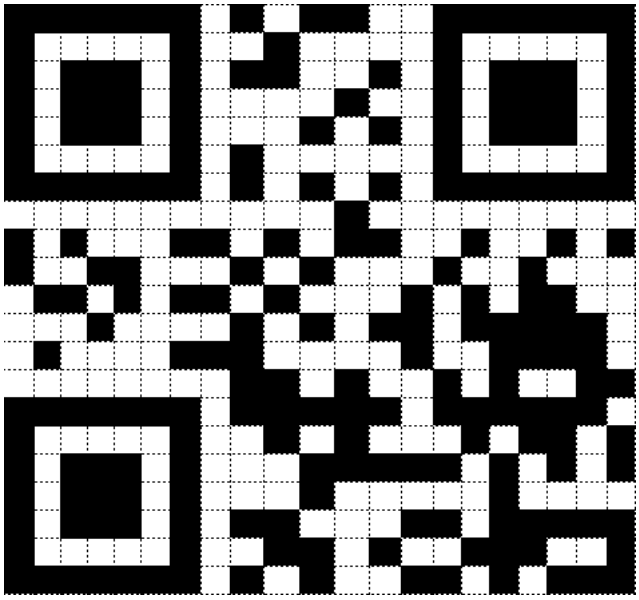
och informationspolynomet $p(x)$ är således

$$p(x) = x^4.$$

När generatorpolynomet och informationspolynomet är konstruerade kan divisionen med rest utföras enligt metoden (6) i avsnitt 4.1. Divisionen ger restkoden 0100110111 och uppsättningen med informationssymbolerna ger kodordet 000010100110111. Enligt QR-kod-specifikationen, när divisionen har gjots, måste ytterligare division utföras mellan restkoden och talet 101010000010010 vilket kallas för maskeringsmönstret. Det slutgiltiga meddelandet innehållande formatinformation bestående av 15 bitar blir,

$$101000100100101.$$

Koden kan nu placeras med start från separatorn upp till vänster och då placeras från vänster till höger och sedan fortsätter upp åt. Sedan börjar vi med separatorn ner till vänster och fortsätter med separatorn upp til höger i QR-koden. Placering av kontrollkoden resulterar i den slutliga QR-koden som visas i figur 4 .



Figur 4: Den slutliga QR-koden, av verson 1 med 15 % felkorrigeringsnivå, som vid avläsning visar ordet KODNINGSTEORI.

Referenser

[1] Vijay K. Bhargava och Stephen B. Wicker, *Reed-Solomon Codes and Their Applications*, 1 uppl., IEEE, New York, 1994.

[2] John R. Durbin, *Modern Algebra An Introduction*, 6 uppl., John Wiley & Sons, Hoboken, 2008.

[3] Ming-Hua Chang och Ta-Hsiang Hu. *Decoding Shortened Reed Solomon Codes at Bit Level*, WSEAS Transactions on Communications. Vol. 9 Nr. 11. Taiwan, 2010.

[4] William J. Gilbert och W. Keith Nicholson, *Modern Algebra with Applications*, 2 rev. uppl., John Wiley & Sons, Hoboken, 2004.

[5] J.H. van Lint, *Introduction to Coding Theory*, 3 rev. uppl., Springer, Berlin, 1999.

[6] I.S. Reed och G. Solomon, *Polynomial Codes Over Certain Finite Fields*, Journal of the Society for Industrial and Applied Mathematics, Vol. 8 Nr. 2, USA, 1960.

[7] Thonky.com, 2015: *QR Code Tutorial*, `www.thonky.com/qr-code-tutorial/introduction` (Hämtad 14 maj, 2018).

[8] QRCode.com, 2018: *QR Code Standardization*, `www.qrcode.com/en/about/standards.html` (Hämtad 14 maj, 2018).

Bilaga A Algebraiska strukturer

En mängd inom den abstrakta algebran med en eller flera operatorer som verkar i en kombination av olika matematiska grundantaganden (axiom) utgör en så kallad *algebraisk struktur*. Addition och multiplikation på tal är typiska exempel av operatorer som kombinerar varje par av element hos en mängd för att bilda ett tredje element. De operatorer vi är bekanta med följer algebraiska axiom så som *associativitet* $a + (b + c) = (a + b) + c$ för addition eller $a(bc) = (ab)c$ för multiplikation. En annan lag i kombination av addition och multiplikation kallas för den *distributiva* lagen $a(b + c) = ab + ac$.

Vi kommer beskriva olika algebraiska strukturer med hjälp av definitioner och satser så att vi kan visa de viktiga satser som används inom kodningsteorin.

A.1 Grupper och ringar

Definition A.1.1. En *grupp* är en algebraisk struktur bestående av en mängd element G och en operation $\star$ sådan att gruppen är sluten under operationen och följande tre axiom är uppfyllda.

(i)	$a \star (b \star c) = (a \star b) \star c \quad , \forall\, a, b, c \ \in G$	(associativitet),
(ii)	$\exists \ \ e \in G : a \star e = e \star a = a \quad , \forall a \in G$	(neutralt element e),
(iii)	$\forall\, a \in G \quad \exists\, b \in G : a \star b = b \star a = e$	(existens av invers).

Om det för varje $a, b \in G$ gäller att $a \star b = b \star a$, så sägs G vara en *abelsk* eller kommutativ grupp.

Exempel A.1.1. Antag att $\star : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{Z}_+$ definierat som $m \star n = m^n$, det vill säga om för varje par $(a, b) \in \mathbb{Z}_+ \times \mathbb{Z}_+$ så bildar de ett nytt element $a \star b = c$ där $c \in Z_+$. Se upp att denna operator inte är kommutativ eftersom

$$9 = 3^2 = 3 \star 2 \neq 2 \star 3 = 2^3 = 8$$

precis som med division eller subtraton, ordningen av elementen kan spela roll. Notera att $\mathbb{Z}_+$ med avseende på $\star$ inte bildar en grupp. Enligt axiomen ovan saknar vi inverst element. För något element $a \in \mathbb{Z}$ får vi att

$$a \star e = a^e = 1 \Leftrightarrow e = 0 \notin \mathbb{Z}_+.$$

Även om vi kan bilda en ny mängd med nollan så kan vi ändå inte bilda en grupp. Antag att $S = \{\mathbb{Z}_+, 0\}$ med samma operator så har vi för något godtyckligt element $a \in S$ att

$$1 = a^0 = a \star 0 \neq 0 \star a = 0^a = 0.$$

Exempel A.1.2. Tag gruppen $\mathbb{Z}$ med avseende på addition. Inversen till varje element $a \in \mathbb{Z}$ motsvarar till $-a \in \mathbb{Z}$ och det neutrala elementet är 0, ty

$$a + (-a) = (-a) + a = 0.$$

Exempel A.1.3. Tag mängden $\mathbb{Z}$ med avseende på multiplikation. Inversen $a^{-1} \in \mathbb{Z}$ till varje element $a \in \mathbb{Z}$ saknas. Vi kan alltså inte ha en grupp med avseende på multiplikation enbart hos heltalen.

Mängden av alla jämna heltal är en delmängd till mängden av alla heltal. Båda mängderna är grupper med avseende på addition, därför måste gruppen av alla jämna heltal vara vara en *delgrupp* till gruppen av alla heltal. Vi formulerar detta som en definition.

Definition A.1.2. En delmängd H till en grupp G med en operator $\star$ sägs vara en *delgrupp* om H också är en grupp med avseende på samma operation på G.

Om G är en grupp med operatorn $\star$ där H är en delgrupp till G och $a, b \in H$ så gäller att $a \star b \in H$. Det vill säga att H måste vara sluten med avseende på operatorn $\star$. Speciellt gäller att $a \star a \in H$ för varje $a \in H$.
Vi vill ha ett enklare sätt att bestämma om en delmängd är en delgrupp till en grupp. Det kan vi göra med hjälp av följande sats:

Sats A.1.1. *Låt G vara en grupp med operatorn $\star$ och låt H vara en delgrupp till G, då gäller*

1. *Om f är ett neutralt element till H och e är ett neutralt element till G då gäller att $e = f$.*

2. *Om $a \in H$ då gäller att inversen till a i H har samma invers till a som i G.*

Sats A.1.2. *Låt G vara en grupp med operationen $\star$ och låt H vara en delmängd till G. Då är H en delgrupp till G om och endast om*

1. *$H \neq \emptyset$,*

2. *om $a \in H$ och $b \in H$ så gäller att $a \star b \in H$,*

3. *om $a \in H$ då finns $a^{-1} \in H$.*

Nedan definerar vi vad en sidoklass är. En sidoklass är en relation hos grupper som vi kommer använda för att visa satserna om så kallade *kvotringar* och *kvotgrupper*.

Sats A.1.3. *Låt H vara en delgrupp till en grupp G och definiera relationen $\sim$ på G som*

$$a \sim b \Leftrightarrow ab^{-1} \in H$$

då är $\sim$ en ekvivalensrelation på G.

Bevis. Reflexitivitet:
Om $a \in G$ så gäller att $a \sim a$ ty,

$$aa^{-1} = e \in H$$

Symmetri:
Om $a \sim b$ så är $ab^{-1} \in H$. Det följer att $ba^{-1} = (ab)^{-1} \in H$ för att varje element i en grupp har en invers, alltså är $b \sim a$.

Transitivitet:
Om $a \sim b$ och $b \sim c$ då gäller att

$$ab^{-1} \in H \text{ och } bc^{-1} \in H$$

vi har att

$$ac^{-1} = a(b^{-1}b)c^{-1} = (ab^{-1})(bc^{-1}) \in H$$

ty, H är sluten under sin operation och då gäller att $a \sim c$.

□

Ekvivalensklasserna för denna ekvivalensrelation är vad som kallas för *höger-sidoklasser* till H.

Sats A.1.4. *Om H är en ändlig delgrupp av en grupp G och $a \in G$ då gäller $|H| = |Ha|$.*

Sats A.1.5 (Lagranges sats)**.** *Om H är en delgrupp av en ändlig grupp G så gäller att ordningen av H är delare till ordningen av G.*

Bevis. Högersidoklasserna av H är ekvivalensklasser. Då gäller att högersidoklasserna av H bildar en partition av G, således måste två högersidoklasser i H antingen vara lika eller disjunkta. Eftersom G är ändlig så finns det bara ändligt många sidoklasser. Välj ett element från varje sidoklass och låt de valda elementen vara $a_1, a_2, ..., a_k$. Då gäller

$$G = Ha_1 \cup Ha_2 \cup \cdots \cup Ha_k.$$

Varje sidoklass Ha_i innehåller H element enligt Sats A.1.4 och det finns inga fler element än i en annan sidoklass, så det följer att $|G| = |H|k$. Alltså är $|H|$ delare till $|G|$. □

A.2 Homomorfier över grupper

En *homomorfi* inom algebran är en avbildning mellan två algebraiska strukturer som bevarar sin struktur. Homomorfi är ett grekiskt ord där *homo* betyder samma och *morfe* betyder form eller ett utseende.

Definition A.2.1. Antag G är en grupp med operatorn $\star$ och H är en grupp med operatorn $\circ$, då sägs en avbildning $\theta : G \to H$ vara en *homomorfi* om

$$\theta(a \star b) = \theta(a) \circ \theta(b)$$

för varje $a, b \in G$.

Innan vi tar exempel så bör vi definiera surjektivitet och injektivitet hos en avbildning.

Definition A.2.2. Låt S och T vara två mängder. En avbildning $\pi : S \to T$ sägs vara surjektiv om dess bildmängd

$$\pi(S) = \{\pi(x) : x \in S\} = T.$$

Det vill säga, π är *surjektiv* om för varje $y \in T$ så finns minst ett element $x \in S$ så att $\pi(x) = y$.

Definition A.2.3. Låt S och T vara två mängder. En avbildning $\pi : S \to T$ sägs vara *injektiv* om

$$x_1 \neq x_2 \Rightarrow \pi(x_1) \neq \pi(x_2), \quad x_1, x_2 \in S.$$

Om en avbildning som skulle vara både surjektiv och injektiv så säger vi att avbildningen är en *bijektion*.

Exempel A.2.1. Definiera $\theta : \mathbb{Z} \to \mathbb{Z}$ så att $\theta(a) = 2a$ för varje $a \in \mathbb{Z}$. Vi kan se att θ är en homomorfi

$$\begin{aligned}\theta(a+b) &= 2(a+b) \\ &= 2a + 2b \\ &= \theta(a) + \theta(b)\end{aligned}$$

för varje $a, b \in \mathbb{Z}$. Så, θ är en homomorfi och vi kan se att den är injektiv ty, antag att $\theta(a) = \theta(b)$ då gäller ju att

$$0 = \theta(a) - \theta(b) = 2a - 2b = 2(a-b) = \theta(a-b) \Rightarrow a = b.$$

θ är inte surjektiv eftersom bildmängden bara innehåller alla jämna tal

$$\theta(\mathbb{Z}) = \{..., -4, -2, 0, 2, 4, ...\} \neq \{..., -2, -1, 0, 1, 2, ...\} = \mathbb{Z}.$$

Om en avbildning θ är en bijektiv homomorfi så säger vi att det är en *isomorfi*. Vi skriver detta som en definition.

Definition A.2.4. Låt G vara en grupp med avseende på $\star$ och H en grupp med operatorn $\circ$. Vi säger att en avbildning $\theta : G \to H$ är en *isomorfi* om den är bijektiv och om

$$\theta(a \star b) = \theta(a) \circ \theta(b)$$

för varje element $a, b \in G$. Om det finns en isomorfi från G till H så säger vi att de är *isomorfa*. Vi betecknar detta som

$$G \cong H.$$

A.3 Ringar

En ring inom abstrakt algebra är en algebraisk struktur. Denna struktur är en mängd utrustad med en generalisering av addition och multiplikation. Denna generalisering gör att vi kan använda aritmetik på icke-numeriska objekt så som polynom, matriser, serier eller funktioner.

Definition A.3.1. Vi säger att mängden R är en ring med två operatorer på R, *addition* $(a + b)$ och *multiplikation* (ab) så att

i. R med avseende på addition är en abelsk grupp.

ii. Multiplikationen är assosciativ.

iii. $a(b + c) = ab + ac$ och $(a + b)c = ac + bc$ för varje $a, b, c \in R$.

I mer detalj måste båda operatorerna uppfylla följande axiomer:

$$a + (b + c) = (a + b) + c \quad \forall a, b, c \in R,$$

det måste finnas ett element $0 \in R$ så att

$$a + 0 = 0 + a = a \quad \forall a \in R$$

för varje $a \in R$ så finns ett element $-a \in R$ så att $a + (-a) = (-a) + a = 0$

$$\begin{aligned} a+b &= b+a && \forall a,b \in R, \\ a(bc) &= (ab)c && \forall a,b,c \in R, \\ a(b+c) = ab+ac \text{ och } (a+b)c &= ac+bc && \forall a,b,c \in R \end{aligned}$$

Gruppen som formas av R med avseende på addition kallas den *additiva* gruppen av R. Det neutrala elementet $0 \in R$ kallas för *nollan* till ringen. Innan vi går till ett exempel på en ring så gör vi följande definition.

Definition A.3.2. För $[a] \in \mathbb{Z}_n$ och $[b] \in \mathbb{Z}_n$ (där $\mathbb{Z}_n$ är heltalen $\mod n$) definiera $[a] \oplus [b]$ genom

$$[a] \oplus [b] = [a+b]$$

och $[a] \odot [b]$ som

$$[a] \odot [b] = [ab]$$

Exempel A.3.1. För varje positivt heltal n så bildar $\mathbb{Z}_n$ en ring med avseende på operatorerna $\oplus$ och $\odot$.

Bevis. Om vi definierar $\oplus$ för $[a] \in \mathbb{Z}_n$ och $[b] \in \mathbb{Z}_n$ som $[a] \oplus [b] = [a+b]$ så har vi att Z_n bildar en abelsk grupp med avseende på $\oplus$:

$$\begin{aligned} [a] \oplus ([b] \oplus [c]) &= [a] \oplus [b+c] && \text{definition av } \oplus \\ &= [a+(b+c)] && \text{definition av } \oplus \\ &= [(a+b)+c] && \text{associativitet av } + \\ & [a+b] \oplus [c] && \text{definition av } \oplus \\ & ([a] \oplus [b]) \oplus [c] && \text{definition av } \oplus. \end{aligned}$$

Detta visar att $\oplus$ är associativ. Vi har att det neutrala elementet är $[0]$:

$$\begin{aligned} [0] \oplus [a] &= [0+a] = [a] \\ [a] \oplus [0] &= [a+0] = [a] \end{aligned}$$

Enligt definition av $\oplus$ så har vi att

$$[-a] \oplus [a] = [(-a)+a] = [0] = [a+(-a)] = [a] \oplus [-a] \quad \text{för något } [a] \in \mathbb{Z}_n$$

vilket visar att $[-a]$ är inversen till $[a]$ i $\mathbb{Z}_n$. Att paret $(\mathbb{Z}_n, \oplus)$ är en abelsk grupp följer av att addition är kommutativ

$$[a] \oplus [b] = [a+b] = [b+a] = [b] \oplus [a].$$

På samma sätt för $\odot$ så följer att denna operator är associativ och kommutativ men har $[1]$ som sitt neutrala element. Notera att $\odot$ aldrig bildar en grupp med $\mathbb{Z}_n$ men det var inget krav att vara en ring heller. Allt som saknas nu är att visa den distributiva lagen och vi är klara.

$$\begin{aligned} [a] \odot ([b] \oplus [c]) &= [a] \oplus [b+c] && \text{definition av } \oplus \\ &= [a(b+c)] && \text{definition av } \odot \\ &= [ab+ac] && \text{distributivitet av } + \text{ och } \cdot \\ & [ab] \oplus [ac] && \text{definition av } \oplus \\ & ([a] \odot [b]) \oplus ([a] \odot [c]) && \text{definition av } \odot. \end{aligned}$$

Alltså bildar $\mathbb{Z}_n$ en ring med avseende på operatorerna $\oplus$ och $\odot$. □

Definition A.3.3. Antag att R och S är två ringar, då säger vi att den avbildningen $\theta : R \to S$ är en (ring) *homomorfi* om

$$\theta(a+b) = \theta(a) + \theta(b)$$

och

$$\theta(ab) = \theta(a)\theta(b)$$

för varje element $a, b \in R$.

Om θ ovan är surjektiv och injektiv så säger vi att θ är en *ring-isomorfi*. Vi har att *nollan* till en ring R, betecknas 0_r, avbildar $\theta(0_r) = 0_s$ där 0_s är nollan till ringen S.
I följande exempel, låt n vara ett fixt positivt heltal och låt k vara något heltal. Vi betecknar som vanligt $[k]$ som kongruensklassen modulo n.

Exempel A.3.2. För något positivt heltal n, definiera $\theta : \mathbb{Z} \to \mathbb{Z}_n$ så att $\theta(a) = [a]$ för varje $a \in \mathbb{Z}$. Då har vi att

$$\theta(a+b) = [a+b] = [a] \oplus [b] = \theta(a) \oplus \theta(b)$$

för varje $a, b \in \mathbb{Z}$. Så θ är en homomorfi (över addition) men inte en isomorfi eftersom θ inte är injektiv. Med avseende på multiplikation får vi på samma sätt att

$$\theta(ab) = [a \cdot b] = [a] \odot [b] = \theta(a) \odot \theta(b)$$

för varje $a, b \in \mathbb{Z}$. Så, θ är en ringhomomorfi.

Definition A.3.4. Om $\theta : R \to S$ är en ringhomomorfi så betecknas *kärnan till* θ som mängden

$$\ker \theta = \{r \in R : \theta(r) = 0_s\}.$$

Exempel A.3.3. Definiera $\theta : \mathbb{Z}_6 \to \mathbb{Z}_3$ genom att $\theta([a]_6) = [a]_3$. Vi vill bestämma kärnan till θ men först måste vi kontrollera om θ är en ringhomomorfi. Antag att $[a]_6, [b]_6 \in \mathbb{Z}_6$ vi får att

$$\begin{aligned}\theta([a]_6 \oplus [b]_6) &= \theta([a+b]_6) \\ &= [a+b]_3 \\ &= [a]_3 \oplus [b]_3 \\ &= \theta([a]_6) \oplus \theta([b]_6)\end{aligned}$$

och på samma sätt visas för multiplikation

$$\begin{aligned}\theta([a]_6 \odot [b]_6) &= \theta([a \cdot b]_6) \\ &= [a \cdot b]_3 \\ &= [a]_3 \odot [b]_3 \\ &= \theta([a]_6) \odot \theta([a]_6)\end{aligned}$$

Så θ är en ringhomomorfi. Vi har att $\ker \theta = \{[0]_3, [3]_3\}$ eftersom $\theta([3]_6) = [3]_3 = [0]_3$ och $\theta([0]_6) = [0]_3$.

Vi har att $\ker \theta$ är en additiv delgrupp av ringen R och att θ är en ringhomomorfi ger oss samma egenskaper som uppfyller de speciella kraven hos grupphomomorfier (A.2.1), att de är så kallade *normala*.

Definition A.3.5. En delgrupp N till en grupp G sägs vara en *normal delgrupp* till G om

$$gng^{-1} \in N \quad \forall g \in G \text{ och } \forall n \in N$$

detta skriver vi som $N \triangleleft G$.

Kärnan är speciell till skillnad från andra delringar, de är så kallat ett *ideal*.

Definition A.3.6. En delring I av en ring R sägs vara ett *ideal* i R om $ar \in I$ och $ra \in I$ för varje $a \in I$ och alla $r \in R$.

En viktig punkt i definitionen ovan är att under produkten ar och ra så kan r vara vilket element som helst i R, r är inte begränsad till I. Skulle R vara kommutativ så är förhållandet mellan $ar \in I$ och $ra \in I$ ekvivalenta.
Låt R vara en kommutativ ring med *etta* 1_R och låt $a \in R$ då beskrivs mängden (a) som mängden av alla multiplar till a av elementen ur R:

$$(a) = \{ra : r \in R\}.$$

Vi vill visa att (a) är ett ideal till R. Vi vill först visa att (a) en delgrupp över den additiva gruppen till R:

i Om $r, s \in R$ sådana att $ra, sa \in (a)$ då gäller att $ra + sa = (r + s)a \in (a)$.

ii $0 = 0a \in (a)$

iii Ett negativt element ra i (a) är $(-r)a$, detta existerar också i (a).

Alltså är (a) en kommutativ grupp med avseende på addition i R. Om $ra \in (a)$ och $s \in R$ så får vi att $s(ra) = (sr)a \in (a)$. Alltså är (a) ett ideal till R.

Sats A.3.1 (Fundamentala homomorfisatsen för ringar)**.** *Låt R och S vara ringar och låt $\theta : R \to S$ var en homomorfi med* $\ker\theta = I$*. Då gäller att avbildningen $\phi : R/I \to S$ definierad som*

$$\phi(I + a) = \theta(a) \text{ för varje } I + a \in R/I$$

är en isomorfi. Alltså är

$$R/I \cong S.$$

A.4 Kroppar

Innan vi går igenom vad en kropp är i algebraiska strukturer så måste det nämnas en viktig klass till ringar, så kallade *integritetsområde*. För att uppfylla detta måste vi sakna en egenskap med avseende på multiplikation: *nolldelare*. Antag att vi har en ring över heltalen. Om a och b är två heltal till ringen så att $ab = 0$ så gäller ju att antingen $a = 0$ eller $b = 0$. Detta gäller inte i alla ringar. I $\mathbb{Z}_6$ exempelvis har vi att $[2] \odot [3] = [0]$ men varken $[2]$ eller $[3]$ är $[0]$.

Definition A.4.1. Ett element $a \neq 0$ i en kommutativ ring R sägs vara en *nolldelare* i R om det finns ett element $b \neq 0$ i R så att $ab = 0$.

Bland heltalen $\mathbb{Z}$ har vi inga nolldelare men både $[2]$ och $[3]$ är nolldelare i $\mathbb{Z}_6$. Notera också i definitionen att ringen måste vara kommutativ med avseende på multiplikation; tag exempelvis matrismultiplikation då har vi ingen kommunikativitet för om $A \neq \mathbf{0}$ och $B \neq \mathbf{0}$ är godtyckliga matriser så kan $AB = \mathbf{0}$ men $BA \neq \mathbf{0}$, detta kallas för en ensidig nolldelare.

Exempel A.4.1. Låt $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ och $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Då är ju $AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ men $BA \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Definition A.4.2. En kommutativ ring R med etta $1_R \neq 0$ som saknar nolldelare sägs vara ett *integritetsområde*.

Notera att om en ring saknar nolldelare så är det samma som att säga att mängden av nollskilda element är sluten under multiplikation.
Anledningen till att $\mathbb{Z}_6$ inte är ett integritetsområde beror på att 6 inte är ett primtal. Utan primtalsordning på elementen så ser vi för ett icke-primtal $n = rs$ där $r, s > 1$ så gäller

$$[r] \odot [s] = [rs] = [n] = [0]$$

där $[r] \neq [0]$ och $s \neq [0]$ i $\mathbb{Z}_n$. Å andra sidan om n är ett primtal så kan vi visa att $\mathbb{Z}_n$ är ett integritetsområde. Vi gör detta som ett exempel.

Exempel A.4.2. Antag p är ett primtal, då gäller att $\mathbb{Z}_p$ är ett integritetsområde.

Bevis. Sedan tidigare vet vi att $\mathbb{Z}_p$ är en kommutativ ring. Vi har att ettan till ringen är $[1] \neq [0]$. Antag nu att $[a], [b] \in \mathbb{Z}_p$ så att $[a] \odot [b] = [0]$; då följer att $[ab] = [0]$ ty, $[a] \odot [b] = [ab]$. Detta är ekvivalent med att $ab = kp$ för något heltal k. Eftersom p är ett primtal så delar p antingen a eller b vilket endast gäller då $[a] = [0]$ eller $[b] = [0]$. Alltså saknar $\mathbb{Z}_p$ nolldelare och måste vara ett integritetsområde. □

Eftersom integritetsområden saknar nolldelare så är mängden av alla nollskilda element sluten under multiplikation. Därför är det naturligt att fråga sig själv om integritetsområdet bildar en grupp med avseende på multiplikation. Eftersom ringar kräver att multiplikationen skall vara associativ så måste operatorn vara associativ. I ett integritetsområde finns en etta så multiplikationen har ett neutralt element. Alltså gäller (med avseende på multiplikation) att mängden av nollskilda element till ett integritetsområde är en grupp om varje element till mängden har en invers med avseende på denna operator. Ett integritetsområde över heltalen visar att de enda talen som har invers över multiplikation är -1 och 1 medan resten av inverserna till elementen är rationella tal. Så denna mängd kan inte bilda en grupp med avseende på multiplikation. Integritetsområden som uppfyller gruppaxiomen (A.1.1) med avseende på multiplikation är den klass av algebraiska strukturer vi kallar för *kroppar*.
Vi har följande relationen av klasserna hos ringar

$$\text{kroppar} \subset \text{integritetsområden} \subset \text{kommutativa ringar} \subset \text{ringar}.$$

Om vi istället har ändligt många element hos en ring då följer det att kroppar och integritetsområden är av samma klass. Detta kan visas som sats.

Sats A.4.1. *Varje ändligt integritetsområde är en kropp.*

Tidigare exempel så visade vi att $\mathbb{Z}_n$ är ett integritetsområde om n är ett primtal. Då $\mathbb{Z}_n$ är ändlig så följer därför från sats ovan att $\mathbb{Z}_n$ är en kropp.

A.5 Kvotringar

Om I är ett ideal i en ring R, då bildar I en delgrupp av den additiva gruppen av R och denna grupp är normal.

Exempel A.5.1. Om N är en delgrupp av en abelsk grupp G där $n \in N$ och $g \in G$ då gäller att

$$gng^{-1} = ngg^{-1} = ne = n \in N.$$

Sats A.5.1. *Antag N är en normal delgrupp av G och låt G/N beskriva alla sidoklasser* (A.1.3) *till N i G. För*

$$Na \in G/N \text{ och } Nb \in G/N \text{ låt } (Na)(Nb) = N(ab)$$

med denna operation är G/N en grupp och kallas kvotgruppen (eller faktorgruppen) till G genom N.

Bevis. Vi vill visa att operatonen på G/N är väldefinierad. Det vill säga, om $Na_1 = Na_2$ och $Nb_1 = Nb_2$ så gäller att $N(a_1b_1) = N(a_2b_2)$. Eftersom $Na_1 = Na_2$ finns $n_1 \in N$ sådan att $a_1 = n_1a_2$. På samma sätt gäller för $Nb_1 = Nb_2$ att det finns $n_2 \in N$ sådan att $b_1 = n_2b_2$. Vi har att

$$a_1b_1 = n_1a_2n_2b_2$$

och eftersom $N \triangleleft G$ finns $n_3 \in N$ sådan att $a_2n_2a_2^{-1} = n_3$ detta ger att $a_2n_2 = n_3a_2$ så vi får att

$$a_1b_1 = n_1a_2n_2b_2 = n1n_3a_2b_2$$

där produkten"$n_1n_2 \in N$. Detta visar att $N(a_1b_1) = N(a_2b_2)$. Alltså är operationen på G/N väldefinierad.
Vi vill visa att G/N är en grupp:

Associativitet:
Låt $a, b, c \in G$ då är

$$Na(NbNc) = Na(N(bc)) = N(a(bc)) = N((ab)c) = N(ab)Nc = (NaNb)Nc.$$

Neutralt element:
Låt $a \in G$ då är

$$NaNe = N(ae) = Na \text{ och } NeNa = N(ea) = Na$$

Vilket visar att Ne är ett neutralt element enligt gruppaxiomen.

Inverst element:
Vi vill visa att Na^{-1} är invers till Na för något $a \in G$:

$$NaNa^{-1} = N(aa^{-1}) = Ne \text{ och } Na^{-1}Na = N(a^{-1}a) = Ne.$$

□

Sats A.5.2. *Antag att $\mathbb{F}$ är en kropp och $p(x)$ är ett polynom av grad n över $\mathbb{F}$ och $I = (p(x))$ är ett ideal av $\mathbb{F}[x]$. Då gäller att varje element i $\mathbb{F}[x]/I$ kan skrivas entydligt på formen*

$$I + (b_0 + b_1x + ... + b_{n-1}x^{n-1}), \text{ där } b_0, b_1, ..., b_{n-1} \in \mathbb{F}. \tag{13}$$

Dessutom är delkroppen $\{I + b : b \in \mathbb{F}\}$ av $\mathbb{F}[x]/I$ isomorf med $\mathbb{F}$.

Bevis. Om $I + f(x) \in \mathbb{F}[x]/I$, då följer det av divisionsalgoritmen, för $q(x), r(x) \in \mathbb{F}[x]$ att

$$f(x) = p(x)q(x) + r(x) \text{ där } r(x) = 0 \text{ eller } \deg r(x) < \deg p(x).$$

Eftersom $f(x) - r(x) = p(x)q(x) \in I$ så har vi att $I + f(x) = I + r(x)$. Det visar att varje element av $\mathbb{F}[x]/I$ kan skrivas på minst ett sätt på formen (13). Å andra sidan, antag att

$$I + (b_0 + b_1x + ... + b_{n-1}x^{n-1}) = I + (c_0 + c_1x + ... + c_{n-1}x^{n-1})$$

då gäller att

$$I + (b_0 - c_0) + (b_1 - c_1)x + ... + (b_{n-1} - c_{n-1})x^{n-1} \in I.$$

Så $p(x) \mid (b_0 - c_0) + (b_1 - c_1)x + ... + (b_{n-1} - c_{n-1})x^{n-1}$, vilket implicerar att

$$(b_0 - c_0) + (b_1 - c_1)x + ... + (b_{n-1} - c_{n-1})x^{n-1} = 0,$$

eftersom $\deg p(x) = n > n - 1$. Så $b_0 = c_0, b_1 = c_1, ..., b_{n-1} = c_{n-1}$, vilket visar entydligheten.
Låt $S = \{I + b : b \in \mathbb{F}\}$. Vi vill visa att denna mängd är isomorf med $\mathbb{F}$. Definiera $\theta : \mathbb{F} \to S$ som $\theta(b) = I + b$. Vi har för $a, b \in \mathbb{F}$ att

$$\theta(a + b) = I + (a + b) = I + a + I + b = \theta(a) + \theta(b)$$

och

$$\theta(ab) = I + (ab) = (I + a)(I + b) = \theta(a)\theta(b).$$

Vi har att θ är surjektiv: $\theta(\mathbb{F}) = \{\theta(a) : a \in \mathbb{F}\} = S$. För att visa att θ är injektiv, antag att $a = b$ där $a, b \in \mathbb{F}$. Vi vill visa att $\theta(a) = \theta(b)$:

$$\theta(\underbrace{a - b}_{=0}) = I + (a - b) = I + a + I + (-b) = \theta(a) - \theta(b) = 0.$$

Alltså är θ en isomorfi. □

Exempel A.5.2. $\mathbb{R}[x]/(1 + x^2) \cong \mathbb{C}$.

Sats A.5.3. *Om $\mathbb{F}$ är en kropp så gäller att varje ideal till polynomringen $\mathbb{F}[x]$ är ett huvudideal.*

Bevis. □

Definition A.5.1. Mängden av polynom med koefficienter i R bildar en ring, *$R[x]$-polynomringen över R.*

A.5.1 Kroppsutvidgningar

Sats A.5.4. *Antag att $\mathbb{F}$ är en kropp och att $p(x) \in \mathbb{F}[x]$ är irreducibel över $\mathbb{F}$. Då gäller att $\mathbb{F}[x]/(p(x))$ är en kroppsutvidning av $\mathbb{F}$ och $p(x)$ har ett nollställe i $\mathbb{F}[x]/(p(x))$.*

Bevis. Sätt $I = (p(x))$. Eftersom $p(x)$ är irreducibel följer det av Theorem 3.2.5 att $\mathbb{F}[x]/I$ är en kropp. Enligt Theorem A.5.2 visade vi att för varje $I + b \in \mathbb{F}[x]/I$ där $b \in \mathbb{F}$ att det finns en delkropp till $\mathbb{F}[x]/I$ som är isomorf med $\mathbb{F}$. Därför är $\mathbb{F}[x]/I$ en kroppsutvidning av $\mathbb{F}$.
Antag att $p(x) = a_0 + a_1x + ... + a_nx^n$ och låt $\alpha = I + x \in \mathbb{F}[x]/I$. Vi har

$$p(\alpha) = a_0 + a_1(I + x) + ... + a_n(I + x_n)^n = I + (a_0 + a_1x + ... + a_nx^n) = I$$

där I är en nolla i $\mathbb{F}[x]/I$. Alltså är α ett nollställe till $p(x)$ i $\mathbb{F}[x]/I$. □

A.6 Cykliska koder

Ett viktigt verktyg av cykliska koder är att multiplarna av $I = x^n - 1$ bildar ett huvudideal till polynomringen $\mathbb{F}_q[x]$. Vi får att

$$\mathbb{F}_q[x]/I \cong \mathbb{F}_q^n$$

där restklassringen $\mathbb{F}_q[x]/I$ består av mängden polynom

$$\{a_0 + a_1x + \cdots + a_{n-1}x^{n-1} : a_i \in \mathbb{F}_q, 0 \leq i < n\}.$$

Vi har följande

$$(a_0, ..., a_{n-1}) = \mathbf{a} \in \mathbb{F}_q^n \mapsto c(\mathbf{a}) = a_0 + a_1x + ... + a_{n-1}x^{n-1} \in \mathbb{F}_q[x]/I \tag{14}$$

där $c : \mathbb{F}_q^n \to \mathbb{F}_q[x]/I$ är en isomorfi. Vi betecknar kodordet $\mathbf{c}$ som $c(x)$ i ovan.

Sats A.6.1. *Den linjära koden $C \subset \mathbb{F}_q^n$ är cyklisk om och endast om C är ett ideal till $\mathbb{F}_q[x]/(x^n - 1)$.*

Bevis. Antag att C är ett ideal i $\mathbb{F}_q[x]/(x^n - 1)$, vi vill visa att C är cyklisk. Eftersom C är ett ideal så gäller för något fixt kodord

$$c(x) = c_0 + c_1x + c_2x^2 + ... + c_{n-1}x^{n-1} \in \mathbb{F}_q[x]/(x^n - 1)$$

att $xc(x)$ också är ett kordord. Vi har att

$$\begin{aligned} xc(x) &= x(c_0 + c_1x + ... + c_{n-2}x^{n-2} + c_{n-1}x^{n-1}) \\ &= c_0x + c_1x^2 + ... + c_{n-2}x^{n-1} + c_{n-1}x^n \\ &= c_0x + c_1x^2 + ... + c_{n-2}x^{n-1} + c_{n-1} \\ &= c_{n-1} + c_0x + c_1x^2 + ... + c_{n-2}x^{n-1}. \end{aligned}$$

Från isomorfin får vi att

$$(c_{n-1}, c_0, ..., c_{n-2}) \in C.$$

vilket visar från definition (4.2.1) att C är cyklisk.

Antag nu istället att C är cyklisk. Vi vill visa att C är ett ideal i $\mathbb{Z}_n[x]/(x^n - 1)$. Eftersom C är cyklisk så gäller för varje kodord $c(x)$ att $xc(x)$ också är ett kodord till C. På samma sätt gäller för varje k att $x^kc(x)$ också är ett kodord till C. Då C är linjär så följer det för två kodord $c(x)$ och $a(x)$ att $a(x)c(x) \in C$ för varje polynom $a(x)$ också är ett kodord. Alltså gäller att C är ett ideal. □

Bilaga B Tabeller

B.1 Alfanumerisk tabell

Tabell 6: I denna tabell listas de tecken som den alfanumeriska formen inkuderar samt de decimala värden som de motsvarar.

Värde	Korresponderande tecken
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F
16	G
17	H
18	I
19	J
20	K
21	L
22	M
23	N
24	O
25	P
26	Q
27	R
28	S
29	T
30	U
31	V
32	W
33	X
34	Y
35	Z
36	(space)
37	$
38	%
39	*
40	+
41	-
42	.
43	/
44	:

B.2 Parameterspecifikation för QR-koder

Tabell 7: I denna tabell listas bland annat antalet informationssymboler och antalet kontrollsymboler som behövs för respektive version och felkorrigeringsnivå för QR-koder.

Version och felkorrigeringsnivå	Informations-symboler (k)	Kontrollsymboler per block ($n-k$)	Antal block	Informations-symboler per block	Totalt antal kodord
1-L	19	7	1	19	$(19{\cdot}1) = 19$
1-M	16	10	1	16	$(16{\cdot}1) = 16$
1-Q	13	13	1	13	$(13{\cdot}1) = 13$
1-H	9	17	1	9	$(9{\cdot}1) = 9$
2-L	34	10	1	34	$(34{\cdot}1) = 34$
2-M	28	16	1	28	$(28{\cdot}1) = 28$
2-Q	22	22	1	22	$(22{\cdot}1) = 22$
2-H	16	28	1	16	$(16{\cdot}1) = 16$
3-L	55	15	1	55	$(55{\cdot}1) = 55$
3-M	44	26	1	44	$(44{\cdot}1) = 44$
3-Q	34	18	2	17	$(17{\cdot}2) = 34$
3-H	26	22	2	13	$(13{\cdot}2) = 26$