TAMPEREEN YLIOPISTO

Kandidaattitutkielma

Eemeli Lottonen

Algebrallista koodausteoriaa

Luonnontieteiden tiedekunta

Matematiikka

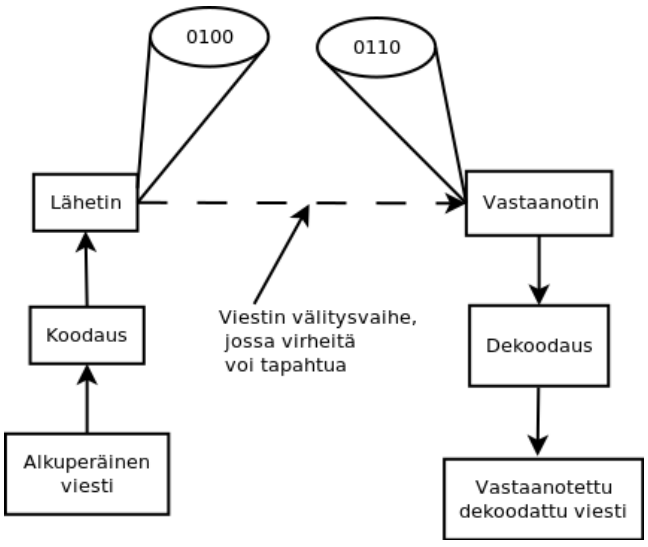
Tammikuu 2019

Sisältö

1 Johdanto

Koodausteoriassa käsitellään tiedonsiirron virhealttiutta ja erilaisia tapoja havaita ja korjata välityksessä tapahtuneita virheitä. Useat häiriötekijät voivat vaikuttaa viestiin, kun sitä yritetään välittää lähettimeltä vastaanottimelle. Virheitä voidaan estää lisäämällä viestiin lisäinformaatiota.

Kuvassa **??** on kuvattu viestin lähettämisen ja vastaanottamisen kaikki vaiheet. Ensin alkuperäinen viesti koodaataan eli lisätään lisäinformaatiota vastaanotinta varten. Sitten viesti etenee lähettimelle ja se lähetetään. Kun viesti on lähetetty, voi välityksen aikana tapahtua virheitä esimerkiksi huonon lähetyskanavan takia. Kun viesti saapuu vastaanottimelle, se dekoodataan eli käytetään lisäinformaatiota mahdollisten virheiden tulkitsemiseen ja samalla poistetaan viestin kannalta turha lisäinformaatio.



Kuva 1.1. Viestin eteneminen

Tässä tutkielmassa käsittelemme eri tapoja koodata ja dekoodata viestejä. Aloitamme hyvin yksinkertaisista esimerkeistä ja siirrymme sitten lähin naapuri dekoodaukseen. Lähin naapuri dekoodauksessa keskitymme ensin minimietäisyyden käyttöön ja sen jälkeen sivuluokkien käyttöön dekoodauksessa. Syvennymme sitten sivuluokkien käyttöön generoijamatriisien ja syndroomien kanssa.

Oletetaan, että lukija tuntee joukkojen rakenteet kuten ryhmät, kunnat. Loppu-

osassa tarvitaan myös tietoa sivuluokista, homomorfismista, isomorfismista ja niiden ominaisuuksista.

2 Lineaariset binäärikoodit

Useat kommunikointijärjestelmät käyttävät binäärijärjestelmää. Binäärijärjestelmässä mikä tahansa numero voidaan esittää muodossa, jossa on vain nollia ja ykkösiä. [**?**, s. 266] Esimerkiksi numero 9 on binäärijärjestelmässä bittijono 1001. Numeroita $n \in \{0,1\} = \mathbb{Z}_2$ kutsutaan biteiksi ja merkinnällä $Z_2 = \{0,1\}$ tarkoitetaan kuntaa, jonka yhteen- ja kertolasku ovat modulaarisia.

Oletetaan, että viesti koostuu k bitistä. Nyt viestiin voidaan lisätä korjausbittejä, joiden avulla voidaan havaita mahdolliset virheet. Muodostetaan siis koodisana, jonka pituus on n. Nyt koodisanan viestiosa koostuu k bitistä ja korjausosa $n - k$ bitistä. Tätä pituuden ja korjausbittien suhdetta $R = \frac{k}{n}$ kutsutaan informaatiosuhteeksi [**?**, s. 267].

Esimerkki 2.1. Muodostetaan kaikki mahdolliset 3-bittiset sanat. Ne ovat

$$\begin{array}{cccc} 000 & 001 & 011 & 111 \\ 110 & 100 & 101 & 010. \end{array}$$

Määrittelemme sanoihin neljännen bitin $x_4 = (x_1 + x_2 + x_3) \bmod 2$. Kaikkien 4-bittisten koodisanojen joukko eli koodi on siis seuraava:

$$C = \{\, x_1x_2x_3x_4 \mid x_i \in \{0,1\} = \mathbb{Z}_2, 1 \leq i \leq 3, x_4 = (x_1 + x_2 + x_3) \bmod 2 \,\}$$

Huomaamme, että kaikkien bittien summa $x_1 + x_2 + x_3 + x_4$ on aina parillinen, sillä jos kolmen ensimmäisen bitin summa $x_1 + x_2 + x_3$ on parillinen, niin $x_4 = 0$, tai jos summa on pariton, niin $x_4 = 1$. Tämän tiedon avulla vastaanottimessa voidaan päätellä, onko välityksessä tapahtunut virhe. Jos sanan kaikkien bittien summa ei ole parillinen, on viesti varmasti väärä. Virhe siis huomataan, mutta ei tiedetä, missä kohtaa se on tapahtunut. Kuitenkin, jos lähetyksessä on tapahtunut parillinen määrä virheitä, esimerkiksi kaksi, niin virhettä ei huomata ollenkaan.

Esimerkki 2.2. Koodataan sana 101 samaan tapaan kuin esimerkissä **??**. Lisätään sitten sanoihin neljäs bitti $x_4 = (x_1 + x_2 + x_3) \bmod 2$, joten koodattu sana on 1010. Jatketaan koodausta vielä toistamalla koodattu sana uudelleen eli sanan 101 lopullinen koodattu muoto on 10101010. Oletetaan, että välityksessä tapahtuu yksi virhe. Nyt vastaanottimessa viesti paloittellaan neljän bitin koodisanoihin. Esimerkin **??**

mukaan tiedetään, kumpi koodisanoista on väärä, ja täten osataan myös korjata virhe.

2.1 Tarvittavien käsitteiden määritelmiä

Määritelmä 2.1 (vrt. [**?**, s. 491])**.** Olkoon A jokin äärellinen joukko. Kutsumme joukkoa A *aakkostoksi*.

1. Alkiota $u \in A^n = \underbrace{A \times \cdots \times A}_{n \text{ kappaletta}}$ kutsutaan *sanaksi*. Sanan u *pituus* on n ja se muodostuu aakkoston A alkioista.

2. Osajoukkoa $C \subseteq A^n$ kutsutaan *koodiksi*.

3. Alkiota $u \in C \subseteq A^n$ kutsutaan *koodisanaksi*.

4. Jos joukko A on kunta, niin A^n on A-vektoriavaruus. Jos nyt $C \subseteq A^n$ ja koodi C on vektoriavaruuden A^n aliavaruus kutsutaan osajoukkoa C *lineaariseksi koodiksi*. Lisäksi, jos $\dim_A C = k$, kutsutaan osajoukkoa C (n, k)*-koodiksi*. Jos $A = \mathbb{Z}_2$, niin osajoukkoa C kutsutaan *lineaariseksi binäärikoodiksi*.

Esimerkki 2.3. Esimerkissä **??** esitimme $(4, 3)$-koodin

$$C = \{\, x_1x_2x_3x_4 \mid x_i \in \{0, 1\} = \mathbb{Z}_2, 1 \leq i \leq 3, x_4 = (x_1 + x_2 + x_3) \bmod 2 \,\} \subseteq A^4.$$

Silloin A^4 on vektoriavaruus, koska $A = \mathbb{Z}_2$ on kunta. Huomataan, että koodisana $0000 \in C$.

Olkoot koodisanat $u \in C$ ja $v \in C$. Nyt koodisanan u kaikkien komponenttien summa on 0 mod 2. Täten sama pätee myös koodisanalle $u + v$.

Olkoot sitten skalaari $a \in A = \mathbb{Z}_2$ ja koodisana $u \in C$. Nyt koodisanalle u pätee joko $au = 0$ tai $au = u$. Täten koodi C on vektoriavaruuden A^4 aliavaruus. Huomaamme myös, että koodin C dimensio on $\dim_A(C) = 3$.

2.2 Hamming-etäisyys ja -paino

Määritelmä 2.2 (vrt. [**?**, s. 492])**.** Olkoot u ja v sanoja vektoriavaruudesta A^n.

1. Sanojen u ja v komponenttien eroavaisuuksien lukumäärää kutsutaan *Hamming-etäisyydeksi* ja sitä merkitään notaatiolla $d(u, v)$.

2. Sanan u, niiden komponenttien lukumäärää, jotka eroavat nollasta kutsutaan *Hamming-painoksi* ja sitä merkitään notaatiolla $\mathrm{wt}(u)$.

3. Olkoon $r \geq 0$ reaaliluku. Nyt joukkoa

$$S_r(u) = \{ v \in A^n \mid d(u, v) \leq r \}$$

kutsutaan sanan u *r-palloksi*.

4. Olkoon C koodi vektoriavaruudessa A^n. Koodin C *minimietäisyys* on

$$d = \min\{ d(u, w) \mid u, w \in C, u \neq w \}.$$

Lause 2.1. *Olkoot u, v ja w sanoja vektoriavaruudessa A^n, missä $A = \mathbb{Z}_2$. Silloin seuraavat ehdot pätevät:*

1. $\mathrm{wt}(u) = d(u, 0)$

2. $d(u, v) = \mathrm{wt}(u - v)$

3. $d(u, v) = d(v, u)$

4. $d(u, v) = 0$*, jos* $u = v$

5. $d(u, w) \leq d(u, v) + d(v, w)$. *(kolmioepäyhtälö)*

Todistus (vrt. [**?**, s. 492]). Lauseen **??** kohta **??** seuraa suoraan määritelmän **??** kohdista **??** ja **??**.

Kohdassa **??** sanan $u - v$ komponentti kohdassa i on 1, jos sanat u ja v eroavat komponentissa i. Nyt määritelmän perusteella saadaan $\mathrm{wt}(u - v) = d(u, v)$.

Kohta **??** seuraa suoraan määritelmän **??** kohdasta **??**.

Kohdassa **??** sanojen u ja v etäisyys $d(u, v) = 0$, jos sanojen u ja v kaikki komponentit kohdassa i ovat samoja, kun $1 \leq i \leq n$.

Kohdan **??** todistus on hieman pidempi kuin muut. Olkoot $x = u - v$ ja $y = v - w$. Nyt lauseen **??** kohdan **??** nojalla epäyhtälön vasemmasta puolesta saadaan

$$d(u, w) = \mathrm{wt}(u - w) = \mathrm{wt}(x + y).$$

Samoin oikeasta puolesta saadaan

$$d(u, v) = \mathrm{wt}(u - v) = \mathrm{wt}(x) \text{ ja}$$
$$d(v, w) = \mathrm{wt}(v - w) = \mathrm{wt}(y).$$

Todistettava epäyhtälö on siis

$$\mathrm{wt}(x + y) \leq \mathrm{wt}(x) + \mathrm{wt}(y).$$

Olkoot x_i ja y_i sanojen x ja y kohdassa i olevia komponentteja. Olkoon myös sana $z = x + y$ ja sen kohdassa i oleva komponentti z_i. Huomaamme, että

$$\mathrm{wt}(x + y) = \sum_{i=1}^{n} z_i \qquad \mathrm{wt}(x) = \sum_{i=1}^{n} x_i \qquad \mathrm{wt}(y) = \sum_{i=1}^{n} y_i.$$

Riittää siis todistaa, että $z_i \leq x_i + y_i$ kaikilla $1 \leq i \leq n$. Tämä seuraa kuitenkin suoraan siitä, että $z_i = x_i + y_i \bmod 2$. □

Esimerkki 2.4. Tarkastellaan sitten yhtä etäisyyden käyttömahdollisuutta välityksessä. Oletetaan, että viestin välityksessä tapahtuu yksi virhe. Silloin vastaanotetun koodisanan ja oikean koodisanan etäisyys on 1. Määritellään seuraavaksi koodi C, jonka minimietäisyys on $d = 3$.

Esimerkissä **??** listasimme kaikki aakkoston A^3 sanat, missä $A = \mathbb{Z}_2$. Koodataan nämä sanat lisäämällä kolme uutta komponenttia. Määritellään koodi $C \subseteq A^6$ seuraavasti:

$$x_1x_2x_3x_4x_5x_6 \in C, \text{ jos}$$
$$x_4 = x_1 + x_2$$
$$x_5 = x_1 + x_3$$
$$x_6 = x_2 + x_3.$$

Listataan sitten uudelleen kaikki vektoriavaruuden A^3 sanat ja niiden koodisanat koodissa $C \subseteq A^6$. Saadaan taulukko

sana	koodisana
000	000000
001	001011
011	011110
111	111000
110	110011
100	100110
101	101101
010	010101

Tarkastellaan seuraavaksi sanojen etäisyyttä toisistaan. Valitaan ensin sanat, joiden etäisyys toisistaan on 1. Valitaan esimerkiksi sanat 110 ja 100. Niitä vastaavat koodisanat ovat $u = 110011 \in C$ ja $v = 100110 \in C$. Nyt alkuperäiset sanat eroavat vain yhdessä komponentissa, mutta niiden koodisanat eroavat kolmessa komponentissa eli $d(u, v) = 3$. Yleisesti, jos kaksi sanaa eroavat vain komponentissa i, komponentti i esiintyy kahdessa yhtälössä, jotka määrittelimme koodin C tarkastuskomponenteille x_4, x_5 ja x_6. Koodisanojen etäisyys on siis vähintään 3.

Tarkastellaan sitten sanoja, joiden etäisyys on 2. Valitaan sanat 110 ja 101. Olkoot niiden vastaavat koodisanat $u = 110011 \in C$ ja $v = 101101 \in C$, joiden etäisyys on $d(u, v) = 4$. Yleisesti, jos sanat eroavat komponenteissa i ja j, kahdessa yhtälöistä, jotka määrittelimme komponenteille $x_4x_5x_6$, esiintyy komponentti i, mutta ei komponenttia j, tai komponentti j, mutta ei komponenttia i.

Tarkastellaan sitten sanoja, joiden etäisyys toisistaan on 3. Jos jo sanojen etäisyys on 3, on koodisanojenkin etäisyys vähintään kolme, sillä sanat sisältyvät sellaisenaan koodisanaan. Täten olemme osoittaneet, että minimietäisyys tässä tapauksessa on $d = 3$.

Oletetaan, että lähetettävä sana on 001 ja sen vastaava koodisana $u = 011110 \in C$. Alussa oletettiin, että lähetyksessä tapahtuu yksi virhe. Kutsutaan tätä vastaanotettua koodisanaa koodisanaksi v. Koska koodin C minimietäisyys on $d = 3$, on lähetty koodisana u ainoa, jonka etäisyys vastaanotetusta koodisanasta v on $d(u, v) = 1$.

Määritelmä 2.3 (vrt. [**?**, s. 494])**.** Olkoot C (n, k)-koodi ja $u = x_1x_2x_3 \ldots x_k \ldots x_n \in C$ sen eräs koodisana. Nyt koodisanan u ensimmäiset k komponenttia muodostavat alkuperäisen viestin ja viimeisiä $n - k$ komponenttia kutsutaan *pariteetintarkastuskomponenteiksi*.

3 Virheenkorjaaminen

3.1 Minimietäisyys

Lause 3.1. *Olkoon koodi $C \subseteq A^n$ ja sen minimietäisyys d. Jos käytämme lähin naapuri dekoodausta, niin kaikille $t \in \mathbb{N}$ pätevät seuraavat ehdot:*

1. *Koodi C voi tunnistaa ainakin t kappaletta virheitä, jos $t + 1 \leq d$.*

2. *Koodi C voi korjata ainakin t kappaletta virheitä, jos $2t + 1 \leq d$.*

Todistus (vrt. [**?**, s. 494]). Aloitetaan lauseen **??** kohdasta **??**. Olkoot koodi $C \subseteq A^n$ ja sen minimietäisyys d. Valitaan sitten koodisana $u \in C$. Nyt pallo $S_t(u)$ sisältää kaikki mahdolliset vastaanotetut sanat, jos lähetetty koodisana oli u ja välityksessä tapahtui enintään t virhettä. Jos koodin C minimietäisyys $d < t$, niin pallo $S_t(u)$ sisältää vain koodisanan u. Täten jos virheitä tapahtuu enintään t kappaletta, niin vastaanotettu sana w ei ole koodisana. Näin voimme tunnistaa, jos virheitä on tapahtunut enintään t kappaletta.

Kohdassa **??** aloitetaan samalla tavalla. Olkoot koodi $C \subseteq A^n$ ja sen minimietäisyys d. Valitaan sitten kaksi koodisanaa $u, v \in C$ siten, että $u \neq v$. Jos minimietäisyydelle pätee $2t < d$, niin vastaanotettu sana w ei voi sisältyä molempiin palloihin $S_t(u)$ ja $S_t(v)$. Tämä voidaan osoittaa olettamalla ensin, että sana w kuuluu molempiin palloihin. Nyt kolmioepäyhtälöstä saadaan

$$d(u, v) \leq d(u, w) + d(w, v) \leq t + t < d,$$

joka on ristiriidassa minimietäisyyden määritelmän kanssa. Täten vastaanotettu sana w voidaan aina korjata lähimpään koodisanaan. □

Rajoitutaan loppuosassa binäärikoodeihin eli $A = \mathbb{Z}_2$.

Esimerkki 3.1. Tarkastellaan seuraavaksi esimerkin **??** $(6, 3)$-koodia C. Esimerkissä totesimme, että koodin C minimietäisyys $d = 3$. Nyt siis jos $t = 1$, koodi C osaa korjata lähetyksessä tapahtuneet virheet. Jos $t = 2$, koodi C tunnistaa virheet, mutta ei osaa korjata niitä.

Oletetaan, että lähetetty koodisana oli $111000 = u \in C$ ja vastaanotettu sana $011001 = w \in C$. Koska sana w ei ole koodisana, voimme tunnistaa, että lähetyksessä on tapahtunut virheitä. Nyt kuitenkin $d(u, w) = 2$ ja $d(u, v) = 2$, missä koodisana $v = 001011$. Emme voi siis tietää kumpi koodisanoista oli alun perin lähetetty.

3.2 Sivuluokkadekoodaus

Lause 3.2. *Olkoon C (n, k)-koodi, jonka minimietäisyys on $d \geq 2t + 1$. Nyt siis koodi C osaa korjata t virhettä käyttämällä lähin naapuri dekoodausta. Hamming-rajaksi kutsutaan seuraavaa epäyhtälöä:*

$$2^k = |C| \leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{t}},$$

$$\text{missä} \quad \binom{n}{b} = \frac{n!}{b!(n-b)!}$$

on binomikerroin ja $|C|$ on koodin C koodisanojen lukumäärä.

Todistus (vrt. [**?**, s.495]). Olkoon koodin $C \subseteq \mathbb{Z}_2^n$ dimensio $\dim_{\mathbb{Z}_2}(C) = k$. Täten koodin C koodisanojen lukumäärä on $|C| = 2^k$. Olkoon sitten koodisana $u \in C$. Nyt koodisana $w \in S_t(u)$, jos sanojen etäisyys $d(u, w) \leq t$. Koodisanassa u on n komponenttia, joten binomikertoimesta $n!/b!(n-b)$ saadaan sanat, jotka eroavat sanasta u b komponentissa. Nyt pallon $S_t(u)$ sanojen lukumäärä on siis

$$|S_t(u)| = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{t}.$$

Koska $2t + 1 \leq d$, niin ei ole olemassa alkiota $a \in S_t(u)$ ja $a \in S_t(v)$. Toisin sanoen palloissa ei ole yhteisiä alkioita eli ne ovat erilliset. Nyt siis alkioita kaikissa palloissa on yhteensä

$$\sum_{u \in C} |S_t(u)| = |S_t(u)|\,|C|\,.$$

Kuitenkin vektoriavaruudessa A^n alkioita on yhteensä 2^n kappaletta, koska $A = \mathbb{Z}_2$. Tästä seuraa seuraava epäyhtälö:

$$\begin{aligned} |S_t(u)|\,|C| &\leq 2^n \\ |C| &\leq \frac{2^n}{|S_t(u)|} \\ |C| &\leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{t}}. \end{aligned}$$

□

Olkoot C (n, k)-koodi ja joukko $A = \mathbb{Z}_2$. Nyt vektoriavaruus A^n on Abelin ryhmä yhteenlaskun suhteen, sillä vektoriavaruuden aksioomiin sisältyy kaikki Abelin ryhmän aksioomat. Abelin ryhmän A^n alkioiden lukumäärä on 2^n. Nyt koodi C on Abelin ryhmän A^n aliryhmä yhteenlaskun suhteen ja se sisältää 2^k koodisanaa. Koodilla C on nyt $\frac{|A^n|}{|C|} = \frac{2^n}{2^k} = 2^{n-k}$ sivuluokkaa vektoriavaruudessa A^n.

Seuraavaksi esittelemme sivuluokkadekoodauksen koodin C sivuluokkien avulla. Valitaan ensin jokaisesta sivuluokasta sana $u_i \in C$ siten, että sanalla u on pienin paino. Nyt jokainen sivuluokka on seuraavaa muotoa:

$$u_1 + C, u_2 + C, \ldots, u_N + C,$$

$$\text{missä } N = 2^{n-k}.$$

Valittuja sanoja u_i kutsutaan *sivuluokan johtajiksi*. Jos vastaanotetaan sana w, niin sanan w täytyy kuulua johonkin sivuluokkaan. Nyt siis sana $w \in u_i + C$, joten se voidaan dekoodata takaisin koodisanaksi vähentämällä siitä sana u_i. Dekoodattu sana on siis $w - u_i \in C$. Tämä dekoodaus voidaan helposti toteuttaa muodostamalla koodille C *standardikaavio*.

Esimerkki 3.2. Olkoon $(4, 2)$-koodi $C = \{\, 0000,\ 0110,\ 1011,\ 1101 \,\}$. Muodostetaan koodille C standardikaavio. Saadaan

0000	0110	1011	1101
1000	1110	0011	0101
0100	0010	1111	1001
0001	0111	1010	1100

Ensimmäisessä rivissä on kaikki koodin C koodisanat ja ensimmäisessä sarakkeessa kaikki sivuluokkien johtajat. Muut rivit on muodostettu valitsemalla vähiten painava sana u vektoriavaruudesta $A^4 = \mathbb{Z}_2^4$ ja sijoittamalla se rivin vasempaan reunaan. Sitten loput sanat ovat saatu laskemalla sana u yhteen koodin C koodisanojen kanssa.

Oletetaan, että vastaanotetaan sana 0010. Nyt sana 0010 esiintyy taulukon kolmannessa rivissä ja toisessa sarakkeessa. Taulukon kolmannen rivin vasemmasta reunasta nähdään sen sivuluokan johtaja 0100, joten vastaanotetun sanan sivuluokka on $0100 + C$. Täten vastaanotettu sana 0010 voidaan dekoodata $0010 - 0100 = 0110$.

Lause 3.3. *Sivuluokka dekoodaus on lähin naapuri dekoodausta.*

Todistus (vrt. [**?**, s. 496]). Olkoot C (n, k)-koodi ja vastaanotettu sana $w \in u + C$, missä sana u on sivuluokan johtaja. Nyt sana w voidaan dekoodata $w - u = v \in C$. Haluamme osoittaa, että epäyhtälö $d(w, v) \leq d(w, y)$ pätee jokaiselle $y \in C$. Käytetään ensin lauseen **??** kohtaa **??**, josta saadaan

$$d(w, v) = \mathrm{wt}(w - v) = \mathrm{wt}(u).$$

Koska sivuluokka voidaan merkitä sen minkä tahansa alkion avulla, jokaiselle $y \in C$ pätee $w - y \in w + C = u + C$. Tästä saadaan epäyhtälö

$$\mathrm{wt}(u) \leq \mathrm{wt}(w - y) = d(w, y).$$

□

4 Generoijamatriisit

Lineaariset binäärikoodit eli (n, k)-koodit koostuvat n bitin pituisista koodisanoista. Näitä koodisanoja on yhteensä 2^k kappaletta. Koodin kaikkien koodisanojen listaamisesta tulee nopeasti erittäin työlästä ja epäkäytännöllistä. Voimme kuitenkin tiivistää kaiken tarvittavan tiedon $k \times n$-matriisiin.

Esimerkki 4.1. Valitaan seuraava 3×6-matriisi:

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Nyt jokainen sana $x_1x_2x_3$ voidaan kirjoittaa 1×3-matriisina. Kun alkuperäinen koodattava sana kerrotaan matriisilla G, saadaan koodin C koodisana. Saamme koodisanaksi

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} G = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{bmatrix},$$

missä $x_4 = x_1 + x_2$, $x_5 = x_1 + x_3$ ja $x_6 = x_2 + x_3$. Koodi C on siis määritelty samoin kuin esimerkissä **??**. Esimerkiksi, jos halutaan lähettää sana 011, saamme koodisanaksi

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix} G = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix},$$

joka näkyy myös esimerkin **??** taulukosta. Täten koodi

$$C = \{\, wG \mid w \in A^k \,\}$$

ja matriisi G sisältää kaiken tarvittavan tiedon sanojen koodaamiseen.

Määritelmä 4.1 (vrt. [**?**, s. 497])**.** Olkoon G $k \times n$-generoijamatriisi. Nyt *generoijamatriisi* on

$$G = \begin{bmatrix} I_k & B \end{bmatrix},$$

missä I_k on $k \times k$-identiteettimatriisi ja B on $k \times (n - k)$-binäärimatriisi.

Esimerkki 4.2. Esimerkiksi esimerkin **??** generoijamatriisi on

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

$$\text{missä } I_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ ja } B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

4.1 Systemaattiset koodit

Määritelmä 4.2 (vrt. [**?**, s. 498])**.** Kutsumme (n, k)-koodia *systemaattiseksi koodiksi*, jos sana $u \in A^k$ muodostaa täsmälleen yhden koodisanan $v \in C$ ensimmäiset k komponenttia. Toisin sanoen koodissa C on olemassa vain yksi koodisana siten, että $u_i = v_i$, kaikilla $i \in [1, k]$.

Apulause 4.1. *Olkoot G ja G' ryhmiä ja funktio $\phi : G \to G'$ homomorfismi. Olkoon myös ydin $K = \ker(\phi) = \{ g \in G \mid \phi(g) = 0_{G'} \}$, missä $0_{G'}$ on ryhmän G' neutraalialkio. Nyt pätee isomorfia*

$$G/K \cong \phi(G).$$

Todistus (vrt. [**?**, s. 97]). Olkoon funktio $\chi : G/K \to \phi(G)$. Nyt joukon G/K mikä tahansa alkio on sivuluokka gK, jollain $g \in G$. Jokainen alkio $y \in \phi(G)$ on muotoa $\phi(g)$, jollain $g \in G$. Määrittelemme nyt funktion χ siten, että $\chi(gK) = \phi(g)$. Tarkastellaan ensin funktion χ on yksikäsitteisyyttä. Toisin sanoen, jos valitaan kaksi samaa alkiota $g_1K = g_2K$ ja siitä seuraa, että $\phi(g_1) = \phi(g_2)$, niin on funktio χ yksikäsitteinen. Oletetaan ensin, että $g_1K = g_2K$. Nyt, kun alkiot ovat samat niin $g_1g_2^{-1} \in K$, jolloin saamme

$$\phi(g_1g_2^{-1}) = 0_{G'}.$$

Koska funktio ϕ on ryhmähomomorfismi voimme kirjoittaa ylemmän yhtälön:

$$\begin{aligned} \phi(g_1)\phi(g_2)^{-1} &= 0_{G'} \\ \phi(g_1) &= \phi(g_2). \end{aligned}$$

Nyt siis funktio χ on yksikäsitteinen.

Olkoon g_1K ja g_2K kaksi alkiota joukosta G/K. Nyt saamme

$$\chi(g_1Kg_2K) = \chi(g_1g_2K) = \phi(g_1g_2) = \phi(g_1)\phi(g_2) = \chi(g_1K)\chi(g_2K),$$

joten funktio χ on ryhmähomomorfismi.

Olkoot g_1K ja g_2K kaksi alkiota joukosta G/K, joille pätee $\chi(g_1K) = \chi(g_2K)$. Nyt $\phi(g_1) = \phi(g_2)$, joten $\phi(g_1)\phi^{-1}(g_2) = 0_{G'}$. Saamme siis $\phi(g_1g_2^{-1}) = 0_{G'}$ eli $g_1g_2^{-1} \in K$, joten $g_1K = g_2K$. Olemme nyt osoittaneet, että funktio χ on injektio.

Olkoon y jokin joukon $\phi(G)$ alkio. Nyt alkio $y = \phi(x)$, jollakin alkiolla $x \in G$. Funktion χ määritelmän perusteella saamme

$$\chi(xK) = \phi(x)$$
$$\chi(xK) = y,$$

joten funktio χ on surjektio.

Täten $\chi : G/K \to \phi(G)$ on isomorfismi ja

$$G/K \cong K$$
$$G/\ker(\phi) \cong K.$$

□

Lause 4.2.

1. *Olkoon G $k\times n$-matriisi. Nyt koodi $C = \{ vG \mid v \in A^k \} \subseteq A^n$ on systemaattinen (n,k)-koodi.*

2. *Olkoon $C \subseteq A^n$ systemaattinen (n,k)-koodi. Nyt on olemassa $k \times n$-generoijamatriisi siten, että $C = \{ vG \mid v \in A^k \}$.*

Todistus (vrt. [**?**, s. 498]). Aloitetaan lauseen **??** kohdasta **??**. Olkoot matriisi G $k\times n$-generoijamatriisi ja koodi $C = \{ vG \mid v \in A^k \}$. Koska kaikki vektoriavaruudet A^r ovat Abelin ryhmiä yhteenlaskun suhteen, kun $r \geq 0$, voidaan määritellä funktio $\phi : A^k \to A^n$ siten, että $\phi(v) = Gv \in A^n$. Nyt tämä on ryhmähomomorfismi, sillä jokaiselle $u, v \in A^k$ pätee

$$\phi(v + u) = (v + u)G = vG + uG = \phi(v) + \phi(u).$$

Nyt siis koodi C on funktion ϕ kuvajoukko $\mathrm{Im}(\phi)$. Määritelmän **??** perusteella $G = \begin{bmatrix} I_k & B \end{bmatrix}$, joten

$$\phi(v) = \begin{bmatrix} v & vB \end{bmatrix} = \begin{bmatrix} u & uB \end{bmatrix} = \phi(u),$$

jos ja vain jos $u = v$. Täten sana $u \in A^k$ kuvautuu vain yhtenä koodisanana koodissa C ja siis funktio ϕ on injektio. Apulauseen **??** perusteella vektoriavaruus $A^k \cong C$ ja koodi C on ryhmän A^n aliryhmä. Täten määritelmän **??** kohdan **??** perusteella C on (n,k)-koodi. Koodi C on myös systemaattinen, sillä funktio ϕ on yksikäsitteinen.

Tarkastellaan sitten kohtaa **??**. Olkoon koodi C systemaattinen (n, k)-koodi. Muodostamme sitten generoijamatriisin G. Olkoon $\{e_i\}$ luonnollinen kanta vektoriavaruudelle A^k. Koska koodi C on systemaattinen, jokaiselle $1 \leq i \leq k$ on olemassa eri koodisana $c_i \in C$. Nyt siis koodisana $c_i = \begin{bmatrix} e_i & d_i \end{bmatrix}$, missä koodisanan c_i ensimmäiset k komponenttia ovat vektori e_i ja d_i on jokin vektori vektoriavaruudesta A^{n-k}. Olkoon nyt B $k \times (n-k)$-matriisi, missä on sanat d_i riveinä. Olkoon myös generoijamatriisi $G = \begin{bmatrix} I_k & B \end{bmatrix}$. Haluamme nyt osoittaa, että koodi $C = \{ vG \mid v \in A^k \}$. Koska

$$e_i B = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_i \\ \vdots \\ d_k \end{bmatrix} = \begin{bmatrix} d_i \end{bmatrix}$$

eli $e_i B$ on matriisin B rivi kohdassa i, kun $1 \leq i \leq k$. Nyt saamme kaikille $1 \leq i \leq k$

$$e_i G = e_i \begin{bmatrix} I_k & B \end{bmatrix} = \begin{bmatrix} e_i & e_i B \end{bmatrix} = \begin{bmatrix} e_i & d_i \end{bmatrix} = c_i \in C.$$

Nyt koska e_i oli luonnollinen kanta vektoriavaruudelle A^k, saamme

$$\{ vG \mid v \in A^k \} \subseteq C.$$

Olkoon $c \in C$, missä $c = \begin{bmatrix} u & w \end{bmatrix}$, $u \in A^k$ ja $w \in A^{n-k}$. Nyt saamme

$$uG = u \begin{bmatrix} I_k & B \end{bmatrix} = \begin{bmatrix} u & uB \end{bmatrix} = \begin{bmatrix} u & w' \end{bmatrix} \in C, \quad \text{missä } w' \in A^{n-k}.$$

Mutta koska koodi C on systemaattinen, saamme määritelmän **??** perusteella

$$\begin{bmatrix} u & w' \end{bmatrix} = \begin{bmatrix} u & w \end{bmatrix} = c = uG.$$

Täten olemme osoittaneet, että

$$C \subseteq \{ vG \mid v \in A^k \}.$$

Olemme osoittaneet nyt, että koodi $C = \{ vG \mid v \in A^k \}$. □

5 Syndroomat

Generoijamatriisit ovat käteviä sanoja koodatessa. Tarvitsemme vielä tavan dekoodata vastaanotetut sanat helposti. Tarkastellaan seuraavaksi syndrooma tapaa tähän tarkoitukseen.

Määritelmä 5.1 ([**?**, s. 499])**.** Olkoon C systemaattinen (n, k)-koodi, jolla on generoijamatriisi $G = \begin{bmatrix} I_k & B \end{bmatrix}$. Nyt $n \times (n-k)$-matriisia

$$H = \begin{bmatrix} B \\ I_{n-k} \end{bmatrix}$$

kutsutaan *pariteetintarkastusmatriisiksi*. Olkoon sana $w \in A^n$. Nyt sanan w *syndrooma* on $wH \in A^{n-k}$.

Esimerkki 5.1. Esimerkissä **??** pariteetintarkastusmatriisi oli

$$H = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Lause 5.1. *Olkoon C systemaattinen (n, k)-koodi, jolla on generoijamatriisi G ja pariteetintarkastusmatriisi H. Olkoon myös sana $v \in A^n$. Nyt sanan v syndrooma $vH = 0 \in A^{n-k}$, jos ja vain jos $v \in C$.*

Todistus (vrt. [**?**, s. 499]). Kuten lauseen **??** todistuksessa, käytetään tässäkin hyväksi tietoa, että kunnasta muodostettu vektoriavaruus on Abelin ryhmä yhteenlaskun suhteen. Määritellään funktio $\phi : A^n \to A^{n-k}$ siten, että $\phi(v) = vH$ jokaiselle $v \in A^n$. Nyt jokaiselle $u, v \in A^n$ pätee

$$\phi(u + v) = (v + u)H = uH + vH = \phi(u) + \phi(v)$$

Täten funktio ϕ on ryhmähomomorfismi. Funktio ϕ on myös surjektio, sillä olkoot $w \in A^{n-k}$ ja vektori $v = \begin{bmatrix} 0 & w \end{bmatrix} \in A^n$. Nyt vektorin v ensimmäiset k komponenttia

ovat nollia ja loput $n-k$ komponenttia vektorin w komponentteja. Nyt siis

$$\phi(v) = vH = \begin{bmatrix} 0 & w \end{bmatrix} \begin{bmatrix} B \\ I_{n-k} \end{bmatrix} = 0B + wI_{n-k} = w.$$

Täten löysimme jokaiselle $w \in A^{n-k}$ vektorin $v \in A^n$ siten, että $\phi(v) = w$. Nyt siis funktio ϕ on surjektio. Apulauseella **??** saamme

$$A^n/\ker(\phi) \cong A^{n-k}.$$

Täten $|\ker(\phi)| = |A|^k = |C|$. Osoitetaan vielä, että $C \subseteq \ker(\phi)$. Olkoon koodisana $v \in C$. Koska C on systemaattinen koodi, $v = wG$, jollekin sanalle $w \in A^k$. Täten $\phi(v) = vH = wGH = 0$, sillä

$$GH = \begin{bmatrix} I_k & B \end{bmatrix} \begin{bmatrix} B \\ I_{n-k} \end{bmatrix} = B + B = 0$$

Nyt siis koodisana $v \in \ker(\phi)$ ja koodi $C \subseteq \ker(\phi)$. Koska $|\ker(\phi)| = |C|$, niin on oltava $C = \ker(\phi)$. □

Lause 5.2. *Olkoon C systemaattinen (n,k)-koodi, jolla on generoijamatriisi G ja pariteetintarkastusmatriisi H. Nyt sanat u ja v ovat samassa sivuluokassa, jos ja vain jos sanoilla u ja v on sama syndrooma.*

Todistus (vrt. [**?**, s. 500]). Olkoot sanat $u, v \in A^n$. Aloitetaan olettamalla, että sivuluokat ovat samat. Silloin

$$\begin{aligned} u + C &= v + C \\ u + C - v &= C + v - v \\ (u - v) + C &= C, \end{aligned}$$

sillä ryhmä A^n on Abelin ryhmä eli yhteenlasku on vaihdannainen. Nyt siis

$$u + C = v + C, \quad \text{jos ja vain jos } u - v \in C.$$

Nyt lauseen **??** perusteella saamme, että $u - v \in C$, jos ja vain jos

$$\begin{aligned} (u - v)H &= 0 \\ uH &= vH. \end{aligned}$$

Täten sanojen u ja v syndroomat ovat samat. □

Esimerkki 5.2. Olkoon C esimerkin **??** $(4, 2)$-koodi, jonka pariteetintarkastusmatriisi H esitettiin esimerkissä **??**. Esitetään seuraavaksi jokaisen sivuluokan syndroomat. Ensimmäisellä rivillä on sivuluokan johtajat ja toisella niiden syndroomat.

v	0000	1000	0100	0001
vH	00	11	10	01

Oletetaan sitten, että vastaanotetaan sana $u = 0111$. Nyt sanan u syndrooma on $uH = 01$, joten u on sivuluokassa, jonka johtaja on sana $v = 0001$. Dekoodaamme sanan u siis $u - v = 0110 \in C$.

Osaamme nyt dekoodata viestejä syndroomien avulla. Tämä on paljon kätevämpää kuin muodostaa standardikaavio, kuten esimerkissä **??**.

Lähteet

[1] Gilbert, W. J. & Nicholson, W. K. *Modern Algebra with Applications. 2nd edition*. John Wiley & Sons, 2004.

[2] Papantonopoulou, A. *Algebra Pure & Applied*. Prentice-Hall, 2002.