Tampereen yliopisto

Juulia Kela

ALGEBRALLISTA KOODAUSTEORIAA

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Lokakuu 2019

Tiivistelmä

Juulia Kela: Algebrallista koodausteoriaa
Kandidaattitutkielma
Tampereen yliopisto
Matematiikan ja tilastotieteen tutkinto-ohjelma
Lokakuu 2019

Tutkimuksen tarkoituksena on löytää tehokas tapa lähetetyn informaation virheiden paikantamiseen ja korjataamiseen algebrallisilla menetelmillä. Aluksi käydään läpi algebrassa ja koodausteoriassa keskeisiä määrielmiä ja termejä ja siitä edetään virheiden määrän todennäköisyyksiin ja kuinka monta virhettä voidaan korjata. Tutkimuksessa päädyttiin tehokkaaseen sivuluokkadekoodaukseen, joka käyttää virheiden paikantamiseen ja korjaamiseen apunaan lineaarikoodeja ja tarkistusmatriiseja. Tutkimus on toteutettu käyttämällä kahta eri kirjallista lähdettä.

Avainsanat: koodausteoria, algebra, koodisana, koodaus ja dekoodaus

Sisältö

1 Johdanto

Nykypäivänä teknologia on isossa osassa elämäämme ja kommunikointiamme. Tietoa siirtyy jatkuvasti ympäri maailmaa muun muassa puhelimien, radioiden, televisioiden sekä tietokoneiden välityksellä.

Joskus kuitenkin tiedonsiirrossa tapahtuu virheitä. Ongelmaksi tässä tilanteessa syntyy se, että tiedon vastaanottaja ei voi tietää, onko virhettä tapahtunut, missä virhe on tapahtunut tai millainen tieto oikeasti pitäisi vastaanottaa. Pahimmassa tapauksessa virhettä ei huomata ja vastaanotettu informaatio on väärää. Virheiden löytämisen ja korjaamisen avuksi on kehitelty koodausteoria, joka mahdollistaa virheen löytämisen ja korjaamisen vastaanotetulla tiedolla mahdollisimman helposti ja nopeasti.

Tässä tutkielmassa käsitellään algebrallista koodausteoriaa, jossa perehdytään virheiden löytämiseen ja korjaamiseen algebrallisilla menetelmillä. Ensimmäisessä luvussa tutustumme keskeisiin termeihin ja määritelmiin, luvussa kolme virheiden todennäköisyyksin ja luvussa kolme virheiden löytämiseen ja korjaamiseen.

Lukijalle oletetaan tunnetuksi joukko-opin, algebran, lineaarialgebran ja todennäköisyyslaskennan perusteet. Päälähteinä työssä on käytetty lähteitä [2] ja [1].

2 Peruskäsitteitä

Tässä luvussa esitetään tarvittavia algebran määritelmiä sekä koodausteoriaan liittyvää termistöä.

Määritelmä 2.1. Ryhmä $(G, \circ)$ on epätyhjä joukko, jossa on määritelty binäärioperaatio $\circ$, joka toteuttaa seuraavat ehdot:

1. $a \circ b \in G, \forall a, b \in G$ [sulkeutuvuus],
2. $(a \circ b) \circ c = a \circ (b \circ c), \forall a, b, c \in G$ [assosiatiivisuus],
3. $\exists e \in G \colon \forall a \in G \colon a \circ e = e \circ a = a$ [neutraalialkio],
4. $\forall a \in G \colon \exists a' \in G \colon a \circ a' = a' \circ a = e$ [käänteisalkio]

[2, s. 324–325].

Esimerkki 2.2. Tiedetään, että $(B^m, \oplus)$ on ryhmä, jossa $B^m = B \times B \times \cdots \times B$ (m-kertaa), kun $B \in \mathbb{Z}_2$, ja määritellään operaatio $\oplus$ siten, että $(x_1, x_2, \ldots, x_m) \oplus (y_1, y_2, \ldots, y_m) = ((x_1 + y_1) \bmod 2, (x_2 + y_2) \bmod 2, \ldots, (x_m + y_m) \bmod 2)$. Siis $B = \{0, 1\}$, ja yhteenlasku saadaan taulukosta

$\oplus$	0	1
0	0	1
1	1	0

[2, s. 375–376].

Määritelmä 2.3. Ryhmä H on ryhmän $(G, \circ)$ aliryhmä, kun voimassa on seuraavat ehdot:

1. Neutraalialkio $e \in G$ kuuluu myös joukkoon H,
2. $a \circ b \in H, \forall a, b \in H$,
3. $a^{-1} \in H, \forall a \in H$.

[2, s. 333].

Määritelmä 2.4. Olkoon $R \subseteq A \times B$ relaatio joukosta A joukkoon B. Arvojoukoksi $Ran(R)$ kutsutaan joukkoa, joka koostuu joukon B alkioista, joihin relaatio R kuvautuu. [2, s. 93]

Määritelmä 2.5. Valitaan luvut $n > m$ sekä injektio $e\colon B^m \to B^n$. Tällöin funktiota e kutsutaan (m,n)-**koodausfunktioksi**. [2, s. 376].

Määritelmä 2.6. Jos $b \in B^m$ ja funktio e on (m,n)-koodausfunktio, niin silloin alkiota b kutsutaan **sanaksi** ja alkiota $e(b)$ kutsutaan **koodisanaksi**. [2, s. 376].

Esimerkki 2.7. Olkoon funktio $e\colon B^m \to B^{m+1}$ parillisuustarkistuskoodausfunktio. Jos annamme funktiolle sanan $b = b_1 b_2 \cdots b_m$, niin silloin $e(b) = b_1 b_2 \cdots b_m b_{m+1}$ siten, että

$$b_{m+1} = \begin{cases} 1, & \text{jos lukuja 1 on pariton määrä,} \\ 0, & \text{jos lukuja 1 on parillinen määrä.} \end{cases}$$

Nyt jos annamme funktiolle sanat 010, 111 ja 011, niin saamme silloin $e(010) = 0101$, $e(111) = 1111$ ja $e(011) = 0110$.

Määritelmä 2.8. Kun koodisana $x = e(b)$ on lähetetty, niin vastaanotetaan koodisana x_t. Funktiota $d\colon B^n \to B^m$ kutsutaan koodausfunktiota e vastaavaksi (n,m)-**dekoodausfunktioksi**, jos $d(x_t) = b' \in B^m$. Lisäksi jos lähetyksessä ei ole tapahtunut virhettä niin silloin $b' = b$. [2, s. 389].

Esimerkki 2.9. Olkoot lähetettävät sanat 110, 101 ja 001. Käytetään toistomenetelmää, jossa $(m,3m)$-koodausfunktio $e\colon \mathbb{Z}_2^m \to \mathbb{Z}_2^{3m}$ on sellainen, että $b = b_1 b_2 \cdots b_m \in \mathbb{Z}_2^m$ ja $x = e(b) = b_1 b_2 \cdots b_m b_1 b_2 \cdots b_m b_1 b_2 \cdots b_m \in \mathbb{Z}_2^m$. Koodausfunktiota e vastaava $(3m,m)$-dekoodausfunktio on tällöin $d\colon \mathbb{Z}_2^{3m} \to \mathbb{Z}_2^m$ ja $x_t = b_1 b_2 \cdots b_{3m}$. Dekoodausfunktio vertailee nyt jokaista koodisanan x_t bittejä i, $i+m$ ja $i+2m$ $\forall i \in \{1,\ldots,m\}$ keskenään. Nyt siis $m = 3$. Syötetään sanat koodausfunktioon, jolloin saadaan seuraavat koodisanat

$$e(110) = 110110110,$$
$$e(101) = 101101101,$$
$$e(001) = 001001001.$$

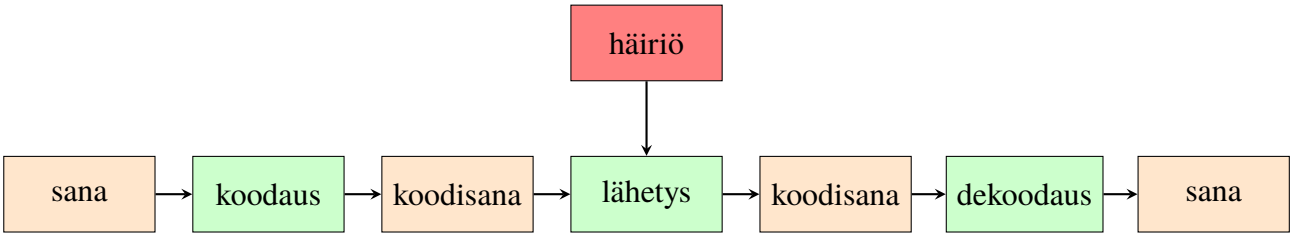
Lähetetään koodisanat tiedonsiirtokanavaa pitkin vastaanottajalle, joka vastaanottaa koodisanat

$$110110110,$$
$$101111101,$$
$$001001101.$$

Vastaanotetut koodisanat syötetään dekoodausfunktioon $d\colon \mathbb{Z}_2^{3m} \to \mathbb{Z}_2^m$, jolloin

$$d(110110110) = 110,$$
$$d(101111101) = 101,$$
$$d(001001101) = 001.$$

Määritelmä 2.10. Koodausfunktio $e\colon B^m \to B^n$ ja dekoodausfunktio $d\colon B^n \to B^m$ muodostavat yhdessä **koodin** C. [1, s. 121].



Määritelmä 2.11. Olkoon $x \in B^n$. Silloin koodisanan x nollasta poikeavien numeroiden lukumäärää kutsutaan **Hamming-painoksi** ja sitä merkitään notaatiolla $|x|$. [2, s. 377].

Esimerkki 2.12. Olkoot koodisanat 1001, 1011, 0111 ja 0010. Nyt Hamming-painot ovat $|1001| = 2$, $|1011| = 3$, $|0111| = 3$ ja $|0010| = 1$.

Määritelmä 2.13. Olkoot x ja y sanoja joukossa B^m. **Hamming-etäisyys** $\delta(x, y)$ sanojen x ja y välillä on tällöin $|x \oplus y|$ [2, s. 378].

Lause 2.14. *Olkoot x, y, $z \in B^n$. Silloin Hamming-etäisyydellä on seuraavat ominaisuudet*

1. $|x| = \delta(x, \overline{0})$, *jossa* $\overline{0} = (0, 0, \ldots, 0) \in B^n$,
2. $\delta(x, y) \geq 0$,
3. $\delta(x, y) = 0$, *jos ja vain jos* $x = y$,
4. $\delta(x, y) = \delta(y, x)$,
5. $\delta(x, y) \leq \delta(x, z) + \delta(z, y)$.

[1, s. 122].

Todistus. Olkoot $x, y, z \in B^n$ sekä $i = 1, \ldots, n$.

1. $|x| = |x \oplus \overline{0}| = \delta(x, \overline{0})$,

2. $\delta(x, y) = |x \oplus y| = |(x_1, x_2, \ldots, x_n) \oplus (y_1, y_2, \ldots, y_n)|$
   $= |((x_1 + y_1) \bmod 2, (x_2 + y_2) \bmod 2, \ldots, (x_n + y_n) \bmod 2)| \geq 0$
   koska lukumäärä ei voi olla negatiivinen.

3. ($\Leftarrow$)

   Valitaan $x = y$.
   Nyt $\delta(x, y) = \delta(x, x) = |x \oplus x| = |(x_1, x_2, \ldots, x_n) \oplus (x_1, x_2, \ldots, x_n)|$
   $= |((x_1 + x_1) \bmod 2, (x_2 + x_2) \bmod 2, \ldots, (x_n + x_n) \bmod 2)| = |\overline{0}| = 0$.

   ($\Rightarrow$)

   Olkoon $\delta(x, y) = 0$.
   Silloin $x_i \oplus y_i = 0$ aina, kun $i = 1, 2, ..., n$. Siis $x_i = y_i$. Näin ollen $x = y$.

4. $\delta(x, y) = |x \oplus y| = |(x_1, x_2, \ldots, x_n) \oplus (y_1, y_2, \ldots, y_n)|$
   $= |((x_1 + y_1) \bmod 2, (x_2 + y_2) \bmod 2, \ldots, (x_n + y_n) \bmod 2)|$
   $= |((y_1 + x_1) \bmod 2, (y_2 + x_2) \bmod 2, \ldots, (y_n + x_n) \bmod 2)|$
   $= |(y_1, y_2, \ldots, y_n) \oplus (x_1, x_2, \ldots, x_n)|$
   $= \delta(y, x)$.

5. $\delta(x, y) = |x \oplus y| = |x \oplus 0 \oplus y| = |x \oplus z \oplus z \oplus y|$
   $\leq |x \oplus z| + |z \oplus y| = \delta(x, z) + \delta(z, y)$.

□

Esimerkki 2.15. Tarkastellaan koodisanoja $x = 011111$, $y = 010111$ ja $z = 000100$. Koodisanojen x ja y Hamming-etäisyys on silloin $|x \oplus y| = |011111 \oplus 010111| = |001000| = 1$, ja koodisanojen y ja z Hamming-etäisyys on silloin $|y \oplus z| = |010111 \oplus 000100| = |010011| = 3$.

Määritelmä 2.16. Olkoon koodausfunktio $e\colon B^m \to B^n$. Funktion e **minimietäisyys** $\delta_{\min}$ on pienin mahdollinen etäisyys kahden mahdollisen koodisanan välillä, toisin sanoen

$$\delta_{\min} = \min\{\, \delta(e(x), e(y)) \mid x, y \in B^m \,\}.$$

[2, s. 379].

Lause 2.17. *Olkoon C koodi, jonka pienin Hamming-etäisyys on $\delta_{\min} = 2k + 1$. Silloin koodi C pystyy korjaamaan k tai vähemmän virheitä. Lisäksi koodilla C voidaan paikantaa $2k$ tai vähemmän virheitä.*

Todistus. [1, s. 123]. Oletetaan, että lähetetään koodisana x ja vastaanotetaan koodisana x_t, jossa on enintään k virhettä. Silloin $\delta(x, x_t) \leq k$. Jos z on jokin muu koodisana kuin x, niin

$$2k + 1 \leq \delta(x, z) \leq \delta(x, x_t) + \delta(x_t, z) \leq k + \delta(x_t, z).$$

Näin ollen $\delta(x_t, z) \geq k + 1$ ja x_t dekoodaataan oikein koodisanaksi x.

Nyt oletetaan, että koodisana x lähetettiin ja koodisana x_t vastaanotettiin ja tapahtui vähintään yksi virhe mutta ei enempää kuin $2k$ virhettä. Silloin $1 \leq \delta(x, x_t) \leq 2k$. Koska minimietäisyys koodisanojen välillä on $2k + 1$, x_t ei voi olla koodisana. Niinpä koodisanasta voidaan löytää virheitä, kun niiden lukumäärä on väliltä $[1, 2k]$. □

Esimerkki 2.18. Tarkastellaan $(2, 5)$-koodausfunktiota $e\colon \mathbb{Z}_2^2 \to \mathbb{Z}_2^5$, jonka määräämät koodisanat ovat 00000, 10011, 01101 ja 11110. Silloin Hamming-etäisyydet ovat

	00000	01101	10011	11110
00000	0	3	3	4
01101	3	0	4	3
10011	3	4	0	3
11110	4	3	3	0

Voimme nyt huomata, että pienin kahden eri koodisanan välinen Hamming-etäisyys on 3, joten koodausfunktio löytää kaksi virhettä ja pystyy korjaamaan yhden virheen.

3 Virheen todennäköisyys

Lause 3.1. *Jos koodisana*

$$x_t = (x_1 x_2 \cdots x_n) \in B^n$$

on vastaanotettu ja todennäköisyydellä p virhettä ei ole syntynyt kirjaimessa x_i $(i = 1, 2, \ldots, n)$*, niin silloin todennäköisyys, että on syntynyt tasan k virhettä on*

$$\binom{n}{k} q^k p^{n-k}.$$

Todistus. [1, s. 120], Todennäköisyys, että koodisanan $x_t \in B^n$ kirjaimessa x_i $(i = 1, \ldots, n)$ on syntynyt virhe on q ja todennäköisyys, että virhettä ei ole syntynyt, on $1 - q = p$. Nyt todennäköisyys, että k virhettä on syntynyt, on näin ollen

$$q^k p^{n-k}.$$

Erilaisia mahdollisia k virhettä käsittäviä koodisanoja eli kombinaatioita on olemassa

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Koska erilaisia kombinaatioita virheiden muodostumiselle on $\binom{n}{k}$ ja todennäköisyys k virheelle on $q^k p^{n-k}$, niin todennäköisyys, että vastaanotetussa koodisanassa x_t on k virhettä, on

$$\binom{n}{k} q^k p^{n-k}.$$

□

Esimerkki 3.2. Vastaanotetaan viesti $x_t \in B^{10}$, jolloin $n = 10$. Oletetaan että todennäköisyys virheeseen on $q = 0{,}004$. Näin ollen todennäköisyys, että virhettä ei ole syntynyt, on $1 - q = 0{,}996 = p$. Todennäköisyys, että virheitä ei ole syntynyt, on

$$p^n = (0{,}996)^{10} \approx 0{,}96.$$

Todennäköisyys, että yksi virhe on syntynyt, on

$$\binom{n}{1} q p^{n-1} = \binom{10}{1} (0{,}004)(0{,}996)^{10-1} = 10 \cdot (0{,}004) \cdot (0{,}996)^9 \approx 0{,}039.$$

Todennäköisyys, että kaksi virhettä on syntynyt, on

$$\binom{n}{2} q^2 p^{n-2} = \binom{10}{2} (0{,}004)^2 (0{,}996)^{10-2} = 45 \cdot (0{,}004)^2 \cdot (0{,}996)^8 \approx 0{,}001.$$

Todennäköisyys, että virheitä on syntynyt enemmän kuin kaksi, on alle 0,001.

4 Dekoodaus

4.1 Lineaariset koodit

Määritelmä 4.1. Tarkastellaan ryhmää $(B^n, \oplus)$ ryhmä. Koodausfunktiota $e\colon B^m \to B^n$ kutsutaan **ryhmäkoodiksi** jos

$$e(B^m) = \{\, e(b) \mid b \in B^m \,\} = Ran(e)$$

on ryhmän B^n aliryhmä. [2, s. 380].

Esimerkki 4.2. Tarkastellaan koodia, joka sisältää koodisanat 00000, 00111, 11100 ja 11011. Huomataan, että koodista löytyy neutraalialkio 00000, ja tiedetään, että jokainen koodisanan on itsensä käänteisluku

$$00000 \oplus 00000 = 0,$$
$$00111 \oplus 00111 = 0,$$
$$11100 \oplus 11100 = 0,$$
$$11011 \oplus 11011 = 0.$$

Näiden lisäksi tarkastellaan, löytyykö kaikkien koodisanojen yhteenlaskun tulokset koodista:

$$00000 \oplus 00111 = 00111,$$
$$00000 \oplus 11100 = 11100,$$
$$00000 \oplus 11011 = 11011,$$
$$00111 \oplus 11100 = 11011,$$
$$00111 \oplus 11011 = 11100,$$
$$11100 \oplus 11011 = 00111.$$

Koska kaikki mahdolliset koodisanojen yhteenlaskut löytyvät koodista, niin silloin koodi on ryhmän $\mathbb{Z}_2^n$ aliryhmä, jolloin se on myös ryhmäkoodi.

Lause 4.3. *Olkoon $\delta_{\min}$ minimietäisyys (m,n)-ryhmäkoodille. Silloin $\delta_{\min}$ on kaikkien (m,n) nollasta eroavien koodisanojen nollasta eroavien painojen minimi. Toisin sanoen*

$$\delta_{\min} = \min\{\, |x| \mid x \neq 0 \,\}.$$

Todistus. [1, s. 125] Havaitaan, että

$$\begin{aligned}\delta_{\min} &= \min\{\, \delta(x,y) \mid x \neq y \,\} \\ &= \min\{\, \delta(x,y) \mid x + y \neq 0 \,\} \\ &= \min\{\, |x \oplus y| \mid x + y \neq 0 \,\} \\ &= \min\{\, |z| \mid z \neq 0 \,\}.\end{aligned}$$

□

Määritelmä 4.4. Olkoon $D = \mathbb{M}_{r\times n}(\mathbb{Z}_2) = [d_{ij}]$ sekä $E = \mathbb{M}_{n\times r}(\mathbb{Z}_2) = [e_{ji}]$ matriisit siten, että $1 \leq i \leq r$ ja $1 \leq j \leq n$. Nyt matriisin operaatio $\oplus$ on esimerkin 2.2 taulukon mukainen laskutoimitus alkioittain ja matriisin operaatio $*$ vastaavasti on

$$\left[(D * E)_{ij}\right] = \left[\sum_{k=1} d_{ik} e_{kj}\right],$$

jossa kertolasku on

$\cdot$	0	1
0	0	0
1	0	1

[2, s. 382]

Esimerkki 4.5.

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 1 \oplus 0\cdot 0 \oplus 1\cdot 0 & 1\cdot 0 \oplus 0\cdot 1 \oplus 1\cdot 0 \\ 0\cdot 1 \oplus 1\cdot 0 \oplus 1\cdot 0 & 0\cdot 0 \oplus 1\cdot 1 \oplus 1\cdot 0 \end{bmatrix} = \begin{bmatrix} 1 \oplus 0 \oplus 0 & 0 \oplus 0 \oplus 0 \\ 0 \oplus 0 \oplus 0 & 0 \oplus 1 \oplus 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Määritelmä 4.6. Olkoon $H \in \mathbb{M}_{m\times n}(\mathbb{Z})$. Matriisin H **nolla-avaruudeksi** $Null(H)$ kutsutaan joukkoa, joka koostuu kaikista koodisanoista x, joilla pätee $Hx = 0$. [1, s. 127].

Lause 4.7. *Olkoon $H \subset \mathbb{M}_{m\times n}(\mathbb{Z}_2)$. Silloin matriisin H nolla-avaruus on ryhmäkoodi.*

Todistus. [1, s. 127]. Koska jokainen joukon $\mathbb{Z}_2^n$ alkio on itsensä käänteisarvo, riittää tarkistaa, että kaikkien matriisin H koodisanojen yhteenlasku kuuluu matriisin nolla-avaruuteen $Null(H)$. Olkoot $x, y \in Null(H)$ koodisanoja matriisin $H \subset \mathbb{M}_{m\times n}(\mathbb{Z}_2)$ nolla-avaruudessa. Nyt siis $Hx = 0$ ja $Hy = 0$, joten $H(x+y) = Hx+Hy = 0+0 = 0$. Koska $x + y$ kuuluu joukon H nolla-avaruuteen, niin silloin se on koodisana. □

Määritelmä 4.8. Koodi C on **lineaarikoodi**, jos se määräytyy jonkin matriisin $H \subset \mathbb{M}_{m\times n}(\mathbb{Z}_2)$ nolla-avaruuden mukaan. [1, s. 127].

Esimerkki 4.9. Olkoon matriisi

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

ja koodi C, joka sisältää koodisanat $x = 0000$, $y = 0001$, $z = 1110$ ja $r = 1111$. Nyt koska $Hx = 0$, $Hy = 0$, $Hz = 0$ ja $Hr = 0$, niin silloin kaikki koodin C koodisanat on määritelty matriisin H nolla-avaruuden mukaan, jolloin koodi on lineaarikoodi.

4.2 Tarkistusmatriisi

Määritelmä 4.10. Olkoon H $r \times n$ -matriisi siten, että $r = n - m$ ja $n > m$ ja sen alkiot kuuluvat joukkoon $\mathbb{Z}_2$. Matriisia H kutsutaan **pariteetin tarkistusmatriisiksi**, jos $H = (A \mid I_r)$, missä A on $r \times m$ -matriisi ja I_r on $r \times r$ -identiteettimatriisi. Toisin sanoen

$$H = \left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1(n-m)} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2(n-m)} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{m(n-m)} & 0 & 0 & \dots & 1 \end{array}\right]$$
$$\underbrace{\qquad\qquad\qquad\qquad}_{m} \quad \underbrace{\qquad\qquad\qquad}_{r}$$

[1, s. 128].

Määritelmä 4.11. Jos $H \in \mathbb{M}_{r\times n}(\mathbb{Z}_2)$ on tarkistusmatriisi, niin silloin nolla-avaruus $Null(H)$ koostuu koodisanoista $x \in \mathbb{Z}_2^n$ joiden m ensimmäistä bittiä ovat satunnaisia informaatiobittejä ja viimeiset r tarkistusbittiä toteuttaa ehdon $Hx = 0$. Lisäksi viimeiset r bittiä toimivat parillisuustarkistusbitteinä m biteille. [1, s. 130].

Esimerkki 4.12. Matriisi

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

on tarkistusmatriisi, koska se sisältää 2×3 -matriisin ja 3×3 -identiteettimatriisin.

Lause 4.13. *Olkoon e_i koodisana, joka sisältää yhden bitin 1 koordinaatissa i ja loput bitit ovat nollia, sekä olkoon matriisi $H \in \mathbb{M}_{m\times n}(\mathbb{Z}_2)$. Tällöin He_i on identtinen matriisin sarakkeen i kanssa. [1, s. 133].*

Todistus. Selvästi H on $r \times n$ -tarkistusmatriisi koodissa C ja vastaanotettu koodisana $e = e_1 \cdots e_i \cdots e_n$, jossa $e_i = 1$ ($i \in \{1, \ldots, n\}$), ja loput bitit ovat nollia. Nyt kun tarkastelemme matriisin tuloa He, niin huomaamme, että se on matriisin H i:s sarake.

$$He_i = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rn} \end{bmatrix} * \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ri} \end{bmatrix}.$$

□

Lause 4.14. *Olkoon H $r \times n$ binäärimatriisi. Tällöin matriisin H nolla-avaruus paikantaa yksittäisen virheen, jos ja vain jos mikää matriisin H sarakkeista ei koostu täysin nollista.*

Todistus. [1, s. 133]. ($\Leftarrow$) Oletetaan että $Null(H)$ pystyy paikantamaan yhden virheen. Tällöin koodin minimietäisyys on oltava vähintää 2. Nyt koodisanojen paino ei voi olla pienempi kuin 2 lukuunottamatta koodisanaa 0. Nyt siis koodisana e_i ei voi kuulua nolla-avaruuteen. Jotta e_i kuuluisi nolla-avaruuteen, niin täytyisi pätеä $He_i = 0$, jolloin siis matriisin H täytyisi sisältää sarake, joka koostuu pelkästää nollista, mikä ei siis ole mahdollista.

($\Rightarrow$) Oletetaan, että mikään sarakkeista ei koostu pelkästään nollista. Nyt siis lauseen 4.13 perusteella voidaan todeta, että He_i (missä $i \in \{1, \ldots, r\}$ on matriisin H sarake i, mistä seuraa, että $He_i \neq 0$. □

Lause 4.15. *Olkoon H $r \times n$ binäärimatriisi. Tällöin matriisin H nolla-avaruus paikantaa ja korjaa yksittäisen virheen, jos ja vain jos mikää matriisin H sarakkeista ei koostu täysin nollista ja mikään matriisin H sarakkeista eivät ole identtisiä.*

Todistus. [1, s. 134]. Olkoon e_i ja e_j koodisanoja, jotka sisältävät bitit 1 indekseissä i ja j ja muutoin koostuvat nollista. Kun $i \neq j$, niin silloin Hamming-paino on $|e_i + e_j| = 2$. Koska $0 = H(e_i + e_j) = He_i + He_j$ pätee vain, kun i ja j ovat identtiset, niin silloin matriisin H nolla-avaruus pystyy korjaamaan yhden virheen. □

4.3 Sivuluokkadekoodaus

Määritelmä 4.16. Olkoon H $r \times n$ -matriisi ja $x \in \mathbb{Z}_2^n$. Koodisanan x **syndroma** on Hx. [1, s. 135].

Lause 4.17. *Olkoon $r \times n$ lineaarikoodin määrittelevä matriisi H ja x_t vastaanotettu n-pituinen koodisana. Avataan koodisana x_t siten, että $x_t = x + e$, jossa x on lähetetty koodisana ja e on tiedonsiirrossa tapahtunut virhe. Silloin vastaanotetun koodisanan x_t syndroma Hx_t on myös virheen e syndroma.*

Todistus. [1, s. 135]. $Hx_t = H(x + e) = Hx + He = 0 + He = He$. □

Lause 4.18. *Olkoon $H \in \mathbb{M}_{r\times n}(\mathbb{Z}_2)$, ja oletetaan, että matriisin H määrittämä lineaarikoodi korjaa yhden virheen. Olkoon x vastaanotettu n-pituinen koodisana, joka on välitetty enintään yhden virheen kanssa. Jos koodisanan x syndroma on* 0*, niin silloin yhtäkään virhettä ei ole tapahtunut. Jos taas koodisanan x syndroma vastaa jotakin matriisin H sarakkeista, niin silloin virhe on tapahtunut ja se voidaan paikantaa ja korjata. Lisäksi jos koodisanan x syndroma ei ole* 0 *tai identtinen matriisin H jonkin sarakkeen kanssa, niin silloin virheitä on tapahtunut enemmän kuin yksi. [1, s. 136].*

Todistus. Koska matriisin H nolla-avaruus $Null(H)$ määrittää lineaarikoodin, niin siis kaikki koodisanat x_t, joilla pätee $Hx_t = 0$, kuuluu silloin lineaarikoodiin. Siis kaikki koodisanat syndromalla 0 kuuluvat lineaarikoodiin. Lauseista 4.13 ja 4.17 seuraa, että jos syndroma vastaa jotakin matriisin H sarakkeista, niin yksi virhe on paikannettu ja korjattavissa. Lauseesta 4.13 voidaan myös päätellä, että jos syndroma ei vastaa mitään matriisin H sarakkeista, niin virheen on täytynyt tapahtua useammassa kuin yhdessä indeksissä i ($i \in \{1, \ldots, n\}$). □

Esimerkki 4.19. Olkoon matriisi

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

ja olkoot vastaanotetut koodisanat $x = 111111$, $y = 101010$ ja $z = 100101$. Silloin

syndromat ovat

$$Hx^T = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 1 + 1\cdot 1 + 0\cdot 1 + 1\cdot 1 + 0\cdot 1 + 0\cdot 1 \\ 0\cdot 1 + 1\cdot 1 + 1\cdot 1 + 0\cdot 1 + 1\cdot 1 + 0\cdot 1 \\ 1\cdot 1 + 0\cdot 1 + 1\cdot 1 + 0\cdot 1 + 0\cdot 1 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 1+1+0+1+0+0 \\ 0+1+1+0+1+0 \\ 1+0+1+0+0+1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$Hy^T = \cdots = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad ja \quad Hz^T = \cdots = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Näin ollen voidaan todeta, että koodisana x sisältää useamman kuin yhden virheen, koodisana y sisältää yhden virheen neljännessä bitissä ja koodisana z ei sisällä yhtäkään virhettä.

Määritelmä 4.20. Olkoon ryhmä H ryhmän G aliryhmä ja $a \in G$. Silloin ryhmän H vasen **sivuluokka** on joukko $aH = \{ ah \mid h \in H \}$ ja oikea sivuluokka on $Ha = \{ ha \mid h \in H \}$. Lisäksi aliryhmä H on normaali aliryhmä, jos kaikilla $a \in G$ pätee $aH = Ha$.

Huomautus. Jos ryhmän G laskuoperaatio on yhteenlasku, niin merkintä on $a+H$ tai $H + a$. [2, s. 340].

Esimerkki 4.21. Tarkastellaan $(2, 5)$ -lineaarikoodia C, joka on määritelty matriisilla

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

Selvitetään koodin C sivuluokat. Sivuluokkia on yhteensä $2^{5-2} = 2^3 = 8$ ja koodisanoja $2^2 = 4$. Olkoon koodisana $x = (x_1x_2x_3x_4x_5)$. Jotta koodisanan x kuuluisi sivuluokkaan, sen täytyy toteuttaa ehto $Hx = 0$. Ehdon toteuttamiseksi katsotaan matriisista H rivi kerrallaan ykkösten sijainnit indeksittäin ja muodostetaan siten

yhtälöryhmät:

$$x_2 + x_3 = 0$$
$$x_1 + x_4 = 0$$
$$x_1 + x_2 + x_5 = 0$$

Yhtälö pitää paikkansa, kun ykkösiä on parillinen määrä. Tästä voidaan päätellä seuraavat koodisanat: (00000), (10011), (01101) ja (11110). Muodostetaan seuraavaksi taulukko sivuluokista jossa jokainen rivi on oma sivuluokkansa.

c	00000	01101	10011	11110
$00001 + c$	00001	01100	10010	11111
$00010 + c$	00010	01111	10001	11100
$00100 + c$	00100	01001	10111	11010
$01000 + c$	01000	00101	11011	10110
$10000 + c$	10000	11101	00011	01110
$00111 + c$	00111	01010	10100	11001
$00110 + c$	00110	01011	10101	11000

Esimerkiksi $00011 + c$ ja $00101 + c$ jätettiin välistä, koska niiden luomat arvot esiintyvät jo taulukossa.

Esimerkki 4.22. Tarkastellaan esimerkin 4.21 koodia ja sivuluokkataulukkoa. Vastaanotetaan koodisanat $x_{t1} = 01001$ ja $x_{t2} = 10110$. Löydämme koodisanan x_{t1} taulukon riviltä 4 jolloin se korjataan koodisanaksi 01101 ja koodisanan x_{t2} riviltä 5 jolloin se korjataan sanaksi 11110.

Koska haluamme saada selville todennäköisimmän koodisanan, niin otamme avuksi Hamming-painon.

Määritelmä 4.23. Binäärisen koodin C sivuluokan koodisanaa, jolla on pienin Hamming-paino (määr. 2.11), kutsutaan **sivuluokan johtajaksi**. [1, s. 137].

Esimerkki 4.24. Tarkastellaan esimerkin 4.21 koodia ja taulukon sivuluokkia. Vali-

taan taulukosta sivuluokan johtajat x_j ja muutetaan taulukko sen mukaiseksi.

$x_j + c$	*sivuluokka*			
00000 + c	00000	01101	10011	11110
00001 + c	00001	01100	10010	11111
00010 + c	00010	01111	10001	11100
00100 + c	00100	01001	10111	11010
01000 + c	01000	00101	11011	10110
10000 + c	10000	11101	00011	01110
10100 + c	10100	11001	00111	01010
00110 + c	00110	01011	10101	11000

Nyt jos vastaanotamme virheellisen koodisanan $x_t = 10010$, niin pystymme laskemaan alkuperäisen koodisanan x sivuluokan johtajan x_j avulla. Huomataan, että koodisana x_t löytyy taulukon riviltä 2, jolloin sen johtaja x_j on 00001. Nyt siis $x_j + x_t = 00001 + 10010 = 10011 = x$.

Lause 4.25. *Olkoon C matriisin H määrittelemä (m,n)-lineaarikoodi, ja oletetaan, että koodisanat $x, y \in B^n$. Koodisanat x ja y ovat samassa koodin C sivuluokassa, jos ja vain jos $Hx = Hy$. Tästä johtuen koodisanat x ja y ovat samassa sivuluokassa, jos ja vain jos niiden syndroma on sama.*

Todistus. [1, s. 138]. Koodisanat x ja y ovat samassa koodin C sivuluokassa täsmälleen silloin kun $x - y \in C$. Toisin sanoen $H(x - y) = Hx - Hy = 0 \iff Hx = Hy$. □

Esimerkki 4.26. Tarkastellaan esimerkin 4.24 mukaista koodia C ja sen sivuluokan johtajia. Jokaista sivuluokkaa kohden löytyy nyt yksilöllinen syndroma:

syndroma	*johtaja*
000	00000
001	00001
010	00010
100	00100
101	01000
011	10000
111	10100
110	00110

Jos vastaanotamme virheellisen koodisanan $x_t = 10110$, niin nyt voimme selvittää alkuperäisen koodisanan x laskemalla syndroman

$$Hx_t = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Silloin meidän on helppo katsoa taulukosta oikea sivuluokan johtaja ja laskea oikea koodisana $x = 01000 + 10110 = 11110$.

Lähteet

[1] Judson, T.W. *Abstract Algebra: Theory and Applications*. Orthogonal Publishing L3c, 2014.

[2] Kolman, B., Busby, R.C. *Discrete Mathematical Structures in Computer Science*. Prentice-Hall International, 1987.