Abelin 2-vapaiden sanojen muodostamisesta 4-kirjaimisessa aakkostossa Seminaariesitelmä

Matti K. Sinisalo

1991

Tarkastelemme seuraavassa ns. Abelin k-vapaiden sanojen (sanojen, joissa ei esiinny permutaatiotoistoja) muodostamiseen liittyviä kysymyksiä lähinnä Veikko Keräsen työn "Abelian squares are avoidable on 4 letters" pohjalta.

Työssään Keränen osoittaa, että 4-kirjaimisessa aakkostossa voidaan muodostaa mielivaltaisen pitkiä abelin 2-vapaita sanoja, ts. sanoja, jotka eivät sisällä kahta peräkkäistä osasanaa, joissa esiintyy täsmälleen sama määrä kutakin kirjainta (esim. 'abacdbadac').

Erdös nosti tämän kysymyksen esille vuonna 1961, mistä lähtien se on ollut avoin ongelma.

Kolmekirjaimisessa aakkostossa a-2-vapaan sanan maksimipituus on 7. Tämä on vanhastaan tunnettu tulos, joka voidaan helposti todistaa ilman tietokonetta.

Pleasants osoitti vuonna 1970, että vähintään viisi kirjainta sisältävässä aakkostossa mielivaltaisen pitkien a-2-vapaiden sanojen muodostaminen on mahdollista.

Thue osoitti tämän vuosisadan alussa, että yli kolme kirjaimisessa aakkostossa voidaan aina muodostaa mielivaltaisen pitkiä 2-vapaita sanoja (ts. sanoja, joissa ei esiinny saman osasanan peräkkäisiä toistoja).

Thue osoitti myös, että binäärisessä aakkostossa voidaan muodostaa mielivaltaisen pitkiä 3-vapaita sanoja (ts. sanoja, joissa mikään osasana ei esiinny kolmea kertaa peräkkäin).

Myöhemmin on todistettu useita sanojen k-vapautta koskevia tuloksia. Useimmissa tapauksissa k-vapaita sanoja on muodostettu ns. *morfismien iteroinnin* avulla.

Myös Keräsen työssä käytetään tätä menetelmää.

Työn keskeinen tulos todistetaan muodostamalla nelikirjaimisten sanojen jou-

kolta itselleen sellainen kuvaus, joka antaa yhdestä sanasta lähtien vaihe vaiheelta yhä pitempiä abelin 2-vapaita sanoja.

Mainittu työ perustuu osittain runsaaseen tietokoneajan käyttöön.

Pyrimme tarkastamaan tästä työstä tietokonetta apuna käyttäen ne tulokset, joiden vaatima laskenta-aika on kohtuullinen.

Keräsen työ sisältää useita teoreettisia tuloksia.

Näiden tulosten tarkastamisessa emme pyri noudattamaan yksityiskohtaisesti Keräsen päättelyä, vaan asetamme tavoitteeksi itse tuloksen tarkastamisen.

Määritelmiä

Aakkosto Σ on äärellinen ei-tyhjä joukko symboleja, joita sanomme *kirjaimiksi*.

Aakkoston Σ *sana* on sen äärellinen kirjainjono.

Tyhjää sanaa merkitsemme λ.

Aakkoston Σ kaikkien sanojen joukkoa merkitsemme Σ^*. Tätä joukkoa sanomme aakkoston Σ generoimaksi *vapaaksi monoidiksi*.

Joukko $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$ on aakkoston Σ generoima *vapaa puoliryhmä*.

Vapaan monoidin Σ^* mielivaltainen osajoukko L on aakkoston Σ *kieli*.

Olkoon $w \in \Sigma^*$, $w = x_1x_2...x_m$, missä $x_1, ..., x_m \in \Sigma$. Sanan w *pituus* (merk. $|w|$) on sanassa w esiintyvien kirjainten lukumäärä ts. $|w| = m$.

Joukko Σ on isomorfinen joukon Σ^* kaikkien 1-pituisten sanojen joukon kanssa. Yleensä nämä joukot samaistetaan ja siten $\Sigma \subseteq \Sigma^+ \subseteq \Sigma^*$.

Sanan $w = x_1x_2...x_m \in \Sigma^*$ *peilikuva* on sana $x_m...x_2x_1$.

Sana $u \in \Sigma^*$ on sanan $w \in \Sigma^*$ *osasana*, jos on olemassa sanat $p, s \in \Sigma^*$, joilla $w = pus$. Sanan w kaikkien osasanojen joukkoa merkitsemme $SW(w)$.

Jos $p = \lambda$, sana u on sanan w *alkuosa* l. *prefiksi*. Jos $s = \lambda$, sana u on sanan w *loppuosa* l. *suffiksi*.

Sanan $w \in \Sigma^*$ kaikkien alkuosien joukkoa merkitsemme $PREF(w)$ ja kaikkien loppuosien joukkoa $SUFF(w)$.

Tietyn kirjaimen $x \in \Sigma$ esiintymien lukumäärää sanassa $w \in \Sigma^*$ merkitsemme $\sharp_x(w)$.

Formaalisti voimme määritellä funktion $\sharp_x : \Sigma^* \to \mathbb{N}$ asettamalla:
1) jokaisella $y \in \Sigma$ on $\sharp_x(y) = 1$, kun $y = x$ ja $\sharp_x(y) = 0$ muulloin ja
2) kaikilla $y \in \Sigma$ ja $w \in \Sigma^*$ on $\sharp_x(yw) = \sharp_x(y) + \sharp_x(w)$.

Helposti todetaan, että $\sharp_x(\lambda) = 0$ ja $\sharp_x(uv) = \sharp_x(u) + \sharp_x(v)$ kaikilla $x \in \Sigma$ ja $u, v \in \Sigma^*$.

Olkoon seuraavassa tarkasteltava aakkosto $\Sigma = \{a_1, a_2, ..., a_n\}$.

Merkintää $\psi(w)$ käytämme sanan w *Parikh-vektorille*

$$\psi(w) = (\sharp_{a_1}(w), \sharp_{a_2}(w), ..., \sharp_{a_n}(w)).$$

Olkoot Σ_1 ja Σ_2 aakkostoja ja h kuvaus $\Sigma_1^* \to \Sigma_2^*$. Kuvaus h on *morfismi*, jos $h(uv) = h(u)h(v)$ aina, kun $u, v \in \Sigma_1^*$.

Erityisesti morfismilla h pätee $h(\lambda) = \lambda$.

Morfismi $h : \Sigma_1^* \to \Sigma_2^*$ tulee yksikäsitteisesti määrättyä, kun annetaan $h(x)$ jokaiselle aakkoston Σ_1 kirjaimelle x.

Jos $\Sigma_1 = \Sigma_2$ (= Σ merk.), morfismi h on *endomorfismi*.

Endomorfismille $h : \Sigma^* \to \Sigma^*$ pätee tulos $h^m(uv) = h^m(u)h^m(v)$ kaikilla $m \in \mathbb{N}$.

Siis h^m on myös endomorfismi kaikilla $m = \mathbb{N}$.

Yleisemmin morfismien yhdistetty kuvaus on aina morfismi.

Morfismille $h : \Sigma_1^* \to \Sigma_2^*$ ja kielelle $L \subseteq \Sigma_1^*$ merkitsemme $h(L) = \{h(w) \mid w \in L\}$.

Morfismi $h : \Sigma_1^* \to \Sigma_2^*$ on *tasaisesti kasvava*, jos $|h(x)| = |h(y)| \geq 2$ kaikilla $x, y \in \Sigma_1$.

Aakkoston Σ *(kirjain)permutaatiolla* tarkoitamme bijektiivistä (surjektiivinen ja injektiivinen, käänteiskuvaus on olemassa) kuvausta $s : \Sigma \to \Sigma$.

Toisin sanoen kaikilla $x, y \in \Sigma$ on $s(x) = s(y)$ silloin ja vain silloin, kun $x = y$.

Identtinen kuvaus muodostaa triviaalin permutaation.

Olkoon Σ jokin aakkosto, $|\Sigma| = n$ ja $s : \Sigma \to \Sigma$ permutaatio. Permutaatio s on *syklinen*, jos se toteuttaa jokaisella $x \in \Sigma$ ehdon $s^i(x) = s^j(x)$ jos ja vain jos $i \equiv j \pmod{n}$.

Esimerkki syklisestä permutaatiosta saadaan indeksoimalla aakkoston Σ kirjaimet $x_1, x_2, ...x_n$. Tämän jälkeen asetetaan $s(x_i) = x_{i+1}$, kun $i = 1, ..., n-1$ ja $s(x_n) = x_1$. Näin määritelty funktio on aakkoston Σ syklinen permutaatio.

Jos s on aakkoston Σ syklinen permutaatio ja $n = |\Sigma|$, niin jokaisella $x \in \Sigma$ on $\Sigma = \{s^j(x) \mid j = 0, 1, ..., n-1\}$.

Lemma 1: Olkoon Σ jokin aakkosto, $|\Sigma| = n$ ja s permutaatio $\Sigma \to \Sigma$. Tällöin on olemassa sellainen $m \in \mathbb{N}$, $m > 0$, että $s^m(x) = x$ kaikilla $x \in \Sigma$.

Todistus. (Periaate) Esitetään s äärellisten syklien tulona. Olkoon $s = s_1 s_2 ... s_r$ ja näiden syklien pituudet $m_1, m_2, ..., m_r$. Valitaan $m = \text{lcm}(m_1, m_2, ..., m_r)$ (tai $m = m_1 m_2 ... m_r$). Mielivaltaisella $x \in \Sigma$ on nyt $s^m(x) = (s_1 s_2 ... s_r)^m(x) =$

$s_1^m s_2^m ... s_r^m(x) = ... = x$. **MOT**

Permutaation $s : \Sigma \to \Sigma$ käänteiskuvaus toteuttaa siis ehdon $s^{-1} = s^{m-1}$.

Permutaatio $s : \Sigma \to \Sigma$ määrittää yksikäsitteisesti endomorfismin $\sigma : \Sigma^* \to \Sigma^*$, kun asetetaan $\sigma(\lambda) = \lambda$ ja kaikilla $x \in \Sigma$, $v \in \Sigma^*$ on $\sigma(xv) = s(x)\sigma(v)$.

Jatkossa tarkoitamme kirjainpermutaatiolla yleisemmin tällaista kirjainpermutaation laajennusta sanoihin.

Syklisellä kirjainpermutaatiolla $\sigma : \Sigma^* \to \Sigma^*$ on käänteiskuvaus $\sigma^{-1} = \sigma^{n-1}$.

Kirjainpermutaatio σ toteuttaa yhtälön $|\sigma(w)| = |w|$ jokaisella $w \in \Sigma^*$. Tämä todetaan helposti induktiolla sanan w pituuden suhteen.

Lemma 2: Olkoon Σ jokin aakkosto ja σ kirjainpermutaatio $\Sigma^* \to \Sigma^*$. Tällöin kaikilla $x \in \Sigma$, $u \in \Sigma^*$ on $\sharp_x(u) = \sharp_{\sigma x}(\sigma u)$.

Todistus. Induktiolla sanan pituuden suhteen.

1) $\sharp_x(\lambda) = 0$. Toisaalta $\sharp_{\sigma x}(\sigma\lambda) = \sharp_{\sigma x}(\lambda) = 0$. Siis väite toteutuu sananpituudella 0.

Olkoon $y \in \Sigma$. Nyt joko $y = x$ tai $y \neq x$.

Olet. $y = x$. Tällöin $\sharp_x(y) = \sharp_x(x) = 1$ ja $\sharp_{\sigma x}(\sigma x) = 1$.

Olet. $y \neq x$. Tällöin $\sigma y \neq \sigma x$. Siis $\sharp_x(y) = \sharp_{\sigma x}(\sigma y) = 0$. Siis väite toteutuu sananpituudella 1.

2) Oletetaan, että väite toteutuu kun sanan u pituus on $\leq n$, $n \geq 0$.

Olkoon $v \in \Sigma$, $|v| = n + 1$. Nyt $v = yu$, $y \in \Sigma$, $u \in \Sigma^*$ ja $|u| = n$. Edelleen $\sharp_x(v) = \sharp_x(yu) = \sharp_x(y) + \sharp_x(u) = \sharp_{\sigma x}(\sigma y) + \sharp_{\sigma x}(\sigma u) = \sharp_{\sigma x}(\sigma y \sigma u) = \sharp_{\sigma x}(\sigma(yu))$.

Induktiolla seuraa väite. **MOT**

Lemma 3: Olkoon Σ jokin aakkosto ja σ kirjainpermutaatio $\Sigma^* \to \Sigma^*$. Olkoot $u, v \in \Sigma^*$ ja $\psi(u) = \psi(v)$. Tällöin $\psi(\sigma u) = \psi(\sigma v)$.

Todistus. Olkoot $a_1, a_2, ..., a_n$ aakkoston Σ kirjaimet. Oletuksen mukaan

$$\psi(u) = (\sharp_{a_1}(u), \sharp_{a_2}(u), ..., \sharp_{a_n}(u)) = \psi(v) = (\sharp_{a_1}(v), \sharp_{a_2}(v), ..., \sharp_{a_n}(v)).$$

Nyt

$$\begin{aligned}\psi(\sigma u) &= (\sharp_{a_1}(\sigma u), \sharp_{a_2}(\sigma u), ..., \sharp_{a_n}(\sigma u)) \\ &= (\sharp_{\sigma^{-1}a_1}(u), \sharp_{\sigma^{-1}a_2}(u), ..., \sharp_{\sigma^{-1}a_n}(u)) \\ &= (\sharp_{\sigma^{-1}a_1}(v), \sharp_{\sigma^{-1}a_2}(v), ..., \sharp_{\sigma^{-1}a_n}(v)) \\ &= (\sharp_{a_1}(\sigma v), \sharp_{a_2}(\sigma v), ..., \sharp_{a_n}(\sigma v)) = \psi(\sigma v). \quad \textbf{MOT}\end{aligned}$$

Lemma 4: Olkoon Σ aakkosto ja $\sigma : \Sigma^* \to \Sigma^*$ jokin kirjainpermutaatio. Sanomme, että vapaan monoidin Σ^* sanat u ja v ovat relaatiossa $\approx$ keskenään,

merk. $u \approx v$, jos on olemassa sellainen $k \in \mathbb{N}$, että $u = \sigma^k(v)$. Tällöin relaatio $\approx$ on ekvivalenssirelaatio.

Todistus. 1) Refleksiivisyys: Valitaan $k = 0$.

2) Symmetrisyys: Olkoon $u \approx v$. On siis olemassa $k \in \mathbb{N}$, jolla $u = \sigma^k(v)$. Lemman 1 mukaan on olemassa sellainen $m \in \mathbb{N}$, $m > 0$, että $\sigma^m(w) = w$ jokaisella $w \in \Sigma^*$. Olkoon $k = pm + q$, $p, q \in \mathbb{N}$, $q < m$. Nyt $u = \sigma^k(v) = \sigma^{pm+q}(v) = \sigma^q(v)$. Siten $\sigma^{m-q}(u) = \sigma^{(m-q)+q}(v) = \sigma^m(v) = v$ ja $v \approx u$.

3) Transitiivisuus: Olkoon $u \approx v$ ja $v \approx w$. Nyt on olemassa k_1 ja k_2 siten, että $u = \sigma^{k_1}(v)$ ja $v = \sigma^{k_2}(w)$. Siis $u = \sigma^{k_1}(v) = \sigma^{k_1}(\sigma^{k_2}(w)) = \sigma^{k_1+k_2}(w)$ ja siten $u \approx w$. **MOT**

Tämä ekvivalenssirelaatio jakaa vapaan monoidin Σ^* yhteispisteettömiin ekvivalenssiluokkiin.

Merkitsemme näiden ekvivalenssiluokkien joukkoa seuraavassa $\Sigma^*/\approx$.

Olkoon $a \in \Sigma$ jokin kirjain ja $\sigma : \Sigma^* \to \Sigma^*$ syklinen kirjainpermutaatio. Tällöin jokainen edellä määritelty ekvivalenssiluokka ($\neq \{\lambda\}$) sisältää täsmälleen yhden a-alkuisen sanan, joka siten voidaan valita luokkansa edustajaksi.

Määritelmä 1. Sanomme, että aakkoston Σ sana on *Abelin k-vapaa* (merk. a-k-vapaa), jos siinä ei esiinny osasanaa $w_1w_2...w_k$, jolla $\psi(w_1) = \psi(w_2) = ... = \psi(w_k)$.

Määritelmästä seuraa välittömästi, että jokaisen a-k-vapaan sanan jokainen osasana on myös a-k-vapaa.

Tällä ominaisuudella on erittäin keskeinen merkitys pyrittäessä konstruoimaan a-k-vapaita sanoja, sillä se antaa mahdollisuuden suorittaa karsintaa mahdollisimman aikaisessa vaiheessa.

Selvästi, jos sana u on a-k_1-vapaa ja $k_2 \geq k_1$, niin sana u on myös a-k_2-vapaa.

Morfismi $h : \Sigma_1^* \to \Sigma_2^*$ on a-k-vapaa, jos $h(w)$ on a-k-vapaa jokaisella vapaan monoidin Σ_1^* a-k-vapaalla sanalla w.

Lemma 5: Olkoon $\sigma : \Sigma^* \to \Sigma^*$ kirjainpermutaatio ja $w \in \Sigma^*$ a-k-vapaa sana, missä $k \in \mathbb{N}$, $k > 1$. Tällöin $\sigma(w)$ on myös a-k-vapaa.

Todistus. Olkoon $w \in \Sigma^*$ k-vapaa. Tehdään vastaoletus, että σw ei ole k-vapaa. Tällöin $\sigma w = pv_1v_2...v_ks$ ja $\psi(v_1) = \psi(v_2) = ... = \psi(v_k)$. Nyt $w = \sigma^{-1}p\sigma^{-1}v_1\sigma^{-1}v_2...\sigma^{-1}v_k\sigma^{-1}s$ ja $\psi(\sigma^{-1}v_1) = \psi(\sigma^{-1}v_2) = ...\psi(\sigma^{-1}v_k)$. Siis w ei voi olla k-vapaa, mikä on ristiriita. **MOT**

Lemman mukaan a-k-vapaus on siis edellä määriteltyjen ekvivalenssiluokkien ominaisuus.

Lemma 6: Olkoot Σ_1 ja Σ_2 aakkostoja, $h : \Sigma_1^* \to \Sigma_2^*$ a-k-vapaa morfismi ja $\sigma_1 : \Sigma_1^* \to \Sigma_1^*$, $\sigma_2 : \Sigma_2^* \to \Sigma_2^*$ kirjainpermutaatioita. Tällöin $h\sigma_1$ ja $\sigma_2 h$ ovat

myös a-k-vapaita morfismeja.

Todistus. Seuraa suoraan edellisestä lemmasta. **MOT**

Endomorfismi $g : \Sigma^* \to \Sigma^*$ *kommutoi* kirjainpermutaation $\sigma : \Sigma^* \to \Sigma^*$ kanssa, jos jokaisella $x \in \Sigma$ on $g(\sigma(x)) = \sigma(g(x))$.

Tällöin jokaisella $w \in \Sigma^*$ on $g(\sigma(w)) = \sigma(g(w))$.

Jokaisella $w \in \Sigma^*$ on myös $g(w) = g(\sigma(\sigma^{-1}w)) = \sigma g(\sigma^{-1}w)$. Siten $\sigma^{-1}g(w) = g(\sigma^{-1}w)$.

Edelleen $g(\sigma^j(w)) = \sigma^j(g(w))$ kaikilla kokonaisluvuilla j ja kaikilla sanoilla $w \in \Sigma^*$.

Endomorfismin $g : \Sigma^* \to \Sigma^*$, joka kommutoi syklisen kirjainpermutaation $\sigma : \Sigma^* \to \Sigma^*$ kanssa, määrittämiseksi riittää tuntea yhden kirjaimen kuva $g(x)$, $x \in \Sigma$.

Tällöin nimittäin jokaisella $y \in \Sigma$ on olemassa $j \in \{0, 1, ..., n-1\}$, jolla $y = \sigma^j x$ ja $g(y) = g(\sigma^j x) = \sigma^j(g(x))$.

Siis kaikki kirjainkuvat tunnetaan ja endomorfismi g määräytyy yksikäsitteisesti.

Tällaisia tietyn kirjainpermutaation $\sigma : \Sigma^* \to \Sigma^*$ kanssa kommutoivia endomorfismeja voidaan siis muodostaa valitsemalla mielivaltainen kirjain $a \in \Sigma$ ja kirjainkuva $g(a) \in \Sigma^+$. Tämän jälkeen valitaan muut kirjainkuvat ehdolla $g(\sigma^j(a)) = \sigma^j(g(a))$ kaikilla $j \in \{0, 1, ..., n-1\}$, missä $n = |\Sigma|$.

Lemma 7: Olkoon Σ jokin aakkosto, $g : \Sigma^* \to \Sigma^*$ endomorfismi ja $\sigma : \Sigma^* \to \Sigma^*$ syklinen kirjainpermutaatio, jonka kanssa g kommutoi. Tällöin $|g(x)| = |g(y)|$ kaikilla $x, y \in \Sigma$.

Todistus. Olkoon n aakkoston Σ kirjainten lukumäärä. Tällöin on olemassa sellainen $j \in \{0, 1, ..., n-1\}$, että $y = \sigma^j(x)$. Nyt

$$|g(y)| = |g(\sigma^j(x))| = |\sigma^j(g(x))| = |g(x)|. \quad \textbf{MOT}$$

Jos luku $|g(x)|$ on riippumaton kirjaimen $x \in \Sigma$ valinnasta, sanomme sitä endomorfismin g *pituudeksi*.

$D0L$-systeemi on kolmikko $G = (\Sigma, g, \alpha_0)$, missä Σ on aakkosto, $g : \Sigma^* \to \Sigma^*$ on endomorfismi ja α_0, ns. *aksiooma*, on jokin aakkoston Σ sana.

$D0L$-systeemi G generoi kielen

$$L(G) = \{g^k(\alpha_0) \mid k \in \mathbb{N}\},$$

missä $g^0(\alpha_0) = \alpha_0$ ja $g^k(\alpha_0) = g(g^{k-1}(\alpha_0))$ kaikilla $k > 0$.

$D0L$-systeemit tarjoavat kätevän menettelytavan kielien ja äärettömien sanojen määrittelyyn.

Jos endomorfismi g ja aksiooma α_0 ovat a-k-vapaita, niin myös järjestelmän generoima kieli on a-k-vapaa.

Huom. Päinvastainen tulos ei välttämättä päde, ts. vaikka kieli $L(G)$ olisi a-k-vapaa, ei endomorfismi g ole välttämättä k-vapaa.

Jotta $D0L$-systeemin $G = (\Sigma, g, \alpha_0)$ generoima kieli $L(G)$ olisi a-2-vapaa riittää, että α_0 on a-2-vapaa ja jokaisella kielessä $L(G)$ esiintyvällä osasanalla w $g(w)$ on a-2-vapaa.

Ääretöntä aakkoston Σ kirjainjonoa sanomme ω-sanaksi. Tällainen ω-sana voidaan määritellä liittämällä jokaiseen joukon $\mathbb{N}^+$ lukuun jokin aakkoston Σ kirjain.

Abelin k-vapaus määritellään ω-sanalle samalla tavalla, kuin sanoille.

ω-sana voidaan konstruoida iteroimalla endomorfismia $g : \Sigma^* \to \Sigma^*$, joka toteuttaa ehdot
1) $\lambda \notin g(\Sigma)$ ja
2) $g(a) = aw$, missä $w \in \Sigma^+$, jollakin kirjaimella $a \in \Sigma$.
Tällaisella morfismilla g, sana $g^i(a)$ on sanan $g^{i+1}(a)$ aito prefiksi aina, kun $i \geq 0$. Jonon $g^i(a)$, $i = 0, 1, 2, ...$ 'raja-arvona' saadaan siis ω-sana.

Lemma 8: Olkoon Σ jokin aakkosto, g aidosti kasvava endomorfismi $\Sigma^* \to \Sigma^*$, joka kommutoi syklisen kirjainpermutaation $\sigma : \Sigma^* \to \Sigma^*$ kanssa, ja $G = (\Sigma, g, \alpha_0)$ $D0L$-systeemi, jonka generoima kieli $L(G)$ on a-k-vapaa. Tällöin on olemassa endomorfismi $g_2 : \Sigma^* \to \Sigma^*$, joka toteuttaa ehdot
1) g_2 ja σ kommutoivat
2) $|g_2(x)| = |g(x)|$ kaikilla $x \in \Sigma$,
3) $D0L$-systeemin $G_2 = (\Sigma, g_2, \alpha_0)$ generoima kieli $L(G_2)$on a-k-vapaa ja
4) kaikilla $x \in \Sigma$ on olemassa sana $v \in \Sigma^*$, jolla $g_2(x) = xv$.

Todistus. Olkoon $a \in \Sigma$ mielivaltainen. Nyt $g(a) = bw$, $b \in \Sigma$, $w \in \Sigma^*$. Valitaan $j \in \{0, 1, ..., n-1\}$ siten, että $\sigma^j(b) = a$. Valitaan $g_2 = \sigma^j g$. Endomorfismien yhdisteenä g_2 on endomorfismi.

1) $g_2(\sigma(x)) = \sigma^j(g(\sigma(x))) = \sigma(\sigma^j(g(x))) = \sigma(g_2(x))$ jokaisella $x \in \Sigma$.

2) $|g_2(x)| = |\sigma(g(x))| = |g(x)|$ jokaisella $x \in \Sigma$ (Kirjainpermutaatio säilyttää sanan pituuden).

3) Seuraa siitä, että kirjainpermutaatio säilyttää a-k-vapauden.

4) Olkoon $x = \sigma^i(a)$, $i \in \{0, 1, ..., n-1\}$. Nyt

$$g_2(x) = \sigma^j g \sigma^i(a) = \sigma^{j+i}(g(a)) = \sigma^{j+i}(bw)$$

$$= \sigma^i(\sigma^j(b))\sigma^{j+i}(w) = \sigma^i(a)\sigma^{j+i}(w) = x\sigma^{j+i}(w).$$

Siten lemma on todistettu. **MOT**

Etsiessämme minimipituista a-2-vapaan ω-sanan generoivaa kirjainpermutaation $\sigma : \Sigma^* \to \Sigma^*$ kanssa kommutoivaa endomorfismia $g : \Sigma^* \to \Sigma^*$ voimme siis rajoittua tarkastelemaan sellaisia endomorfismeja, joilla kirjainkuvan alkukirjain on sama kuin kuvattava kirjain.

Laajamittainen haku ja takaisinpaluu

NP-täydellisten ja eksponentiaalisten tehtävien ratkaisemisessa on yleensä käytössä 2 menetelmää.

Laajamittainen haku (exhaustive search) on menetelmä jossa tietyn hyvin laajan perusjoukon alkioista todetaan yksitellen läpikäyden, onko näillä alkioilla tietty ominaisuus vai ei ole.

Takaisinpaluumenettelyssä (backtracking) ratkaisut muodostetaan osa osalta kokoamalla ja samalla testaamalla, onko ratkaisun löytäminen edelleen mahdollinen.

Laajamittaista hakua käytettäessä etsittyä ominaisuutta testataan jälkikäteen, takaisinpaluumenettelyssä testit suoritetaan mahdollisimman aikaisessa vaiheessa.

Takaisinpaluumenettely on yleensä laajamittaista hakua huomattavasti tehokkaampi menetelmä, koska se karsii huomattavan määrän (yleensä suurimman osan) tutkittavia ratkaisuehdokkaita testien ulkopuolelle ja nopeuttaa siten ratkaisun hakua huomattavasti.

Aina ei kuitenkaan ole mahdollista käyttää takaisinpaluumenetelmää. Tällainen on tilanne silloin, kun tutkittava ominaisuus ei ole rekursiivinen.

Joissakin tapauksissa laajamittainen haku voi olla tehokkaampi teknisistä syistä. Esimerkiksi se voi soveltua paremmin rinnakkaiseen ja vektoroituvaan laskentaan.

Esimerkki: Kahdeksan kuningattaren probleema. Tehtävänä on sijoittaa shakkilaudalle kahdeksan kuningatarta siten, että ne eivät pareittain uhkaa toisiaan. Toisin sanoen kaksi kuningatarta ei koskaan ole samalla vaaka-, pysty- tai vinorivillä.

Laajamittaista hakua käytettäessä numeroidaan ruudut 1-64 ja käydään läpi kaikki mahdolliset 8 kuningattaren sijoittelut shakkilaudalle, joita on $64^8 \approx 2.8 \cdot 10^{14}$ mahdollisuutta. Kussakin tapauksessa erikseen suoritetaan tarvittavat testit.

Testejä voidaan aikaistaa soveltamalla laajamittaista hakua kaikkiin kahdeksan kuningattaren permutaatioihin testaamalla konstruktion yhteydessä vaaka- ja pystyrivit ja jättämällä vinorivitestit lopussa suoritettaviksi. Tällöin selvitään $8! \approx 40320$ testattavalla alkeistapauksella.

Takaisinpaluumenettelyä käytettäessä sijoitetaan aluksi yksi kuningatar ensimmäiselle riville. Tämän jälkeen lisätään muut kuningattaret yksitellen seu-

raaville riveille siten, että kunkin kuningattaren sijoittelun yhteydessä todetaan, että vaaditut ehdot toteutuvat suhteessa aikaisemmin sijoitettuihin kuningattariin. Tarvittaessa palataan siirtämään jo aikaisemmin sijoitettuja kuningattaria uusiin asemiin.

Abelin 2-vapaiden sanojen muodostamisesta

Abelin k-vapaiden sanojen etsiminen on tyypillinen takaisinpaluumenettelyllä ratkaistava tehtävä.

Etsitään m-pituisia abelin k-vapaita sanoja n-kirjaimisessa aakkostossa Σ.

Laajamittaista hakua sovellettaessa muodostetaan kaikki m-pituiset sanat aakkostossa Σ ja testataan jokainen näistä sanoista erikseen. Sanojen lukumäärä on n^m, joka kasvaa hyvin nopeasti lukujen n ja m kasvaessa.

Takaisinpaluumenettelyä käytettäessä muodostetaan sanat lisäämällä kirjaimia yksitellen sanan loppuun siten, että kunkin kirjaimen lisäämisen yhteydessä tarkistetaan, onko uusi sana edelleen abelin k-vapaa.

Abelin k-vapauden lisäksi voidaan sanojen muodostamisen yhteydessä testata myös muita sanoilta vaadittavia ehtoja.

Seuraavassa $\lfloor\alpha\rfloor$ tarkoittaa kaikilla $\alpha \in \mathbb{R}$ suurinta kokonaislukua, joka on pienempi tai yhtäsuuri kuin α.

Lemma 9: Olkoon $w = x_1x_2...x_m$ abelin 2-vapaa sana n-kirjaimisessa aakkostossa σ ja $x_{m+1} \in \Sigma$. Merkitään $\psi_0 = 0$ ja $\psi_i = \psi(x_1x_2...x_i)$ jokaisella $i = 1, 2, ..., m$. Sana $wx_{m+1} = x_1x_2...x_mx_{m+1}$ on abelin 2-vapaa, jos ja vain jos jokaisella $j \in \mathbb{N}$, $0 < j \leq \lfloor\frac{m+1}{2}\rfloor$ on

$$\psi_{m+1} - \psi_{m+1-j} \neq \psi_{m+1-j} - \psi_{m+1-2j}.$$

Todistus. 1) Oletetaan, että edellämainitun testin ehdot toteutuvat, mutta sana wx_{m+1} ei ole abelin 2-vapaa. Tällöin sillä on osasana uv, jolla $\psi(u) = \psi(v)$. Koska w on a-2-vapaa, osasanan uv täytyy olla sanan wx_{m+1} suffiksi, siis $wx_{m+1} = puv$. Valitaan $j = |u| = |v|$. Nyt $\psi_{m+1} - \psi_{m+1-j} = \psi_{m+1-j} - \psi_{m+1-2j}$, eli päädytään ristiriitaan.

2) Oletetaan, että wx_{m+1} on a-2-vapaa, mutta testi ei toteudu. Siis on olemassa $j \in \mathbb{N}$, $0 < j \leq \lfloor\frac{m+1}{2}\rfloor$, jolla $\psi_{m+1} - \psi_{m+1-j} = \psi_{m+1-j} - \psi_{m+1-2j}$. Nyt

$$\psi(x_{m+2-j}...x_mx_{m+1}) = \psi(x_{m+2-2j}...x_{m+1-j}),$$

joten x_{m+1} ei ole a-2-vapaa, mikä on ristiriita. **MOT**

Lemmaa voidaan käyttää myös sanan todistamiseen a-2-vapaaksi.

Lemma 10: m-pituisen sanan todistaminen a-2-vapaaksi lemman 9 testillä vaatii $\lfloor\frac{m^2}{4}\rfloor$ Parikh-vertailun suorittamisen.

Todistus. Helposti todetaan, että m-pituinen sana sisältää kahden peräkkäisen j-pituisen sanan muodostamia pareja, kun $0 < j \leq \lfloor m/2 \rfloor$. Nämä sanaparit alkavat kirjaimista, joiden järjestysluvut tutkittavassa sanassa ovat 1, 2,..., $m + 1 - 2j$. Sanapareja on siis $m+1-2j$ kappaletta ja tarvittavia Parikh-vertailuja yhteensä (merk. $r = \lfloor m/2 \rfloor$)

$$\sum_{j=1}^{r}(m+1-2j) = (m+1)r - 2\frac{r(r+1)}{2} = r(m+1-r-1)$$

$$= r(m-r) = \lfloor \frac{m}{2} \rfloor \left(m - \lfloor \frac{m}{2} \rfloor \right).$$

Jos m on parillinen, saamme vertailujen lukumääräksi $m^2/4$ ja muulloin $(m^2 - 1)/4$. Siis Parikh-vertailuja tarvitaan aina $\lfloor \frac{m^2}{4} \rfloor$ kappaletta. **MOT**

Esim. 5-pituinen sana kuvattuna 85-pituisella endomorfismilla muodostaa $5 \cdot 85 = 425 = m$ -pituisen sanan, jonka testaamiseen tarvitaan edelläolevan perusteella 45156 Parikh-vertailun suorittamisen.

Lemma 11: Kolmekirjaimisessa aakkostossa a-2-vapaan sanan maksimipituus on 7.

Todistus. Merkitään $\Sigma = \{a, b, c\}$. Sovelletaan suoraan takaisinpaluumenettelyä ja lemman 9 testiä. Ensimmäinen 7 merkin mittainen sana, joka löydetään, on *abacaba*. Todetaan, että hakupuun mihinkään haaraan ei voida lisätä 7 useampaa kirjainta. Yhteensä löydetään 3 sellaista 7 merkin mittaista ratkaisua, joita ei saada millään kirjainpermutaatiolla toisistaan (edellisen lisäksi sanat *abacbab* ja *abcbabc*). **MOT**

A-2-vapaus sanoja rajoittavana ehtona

4-kirjaimisen aakkoston n-pituisten sanojen lukumäärä on 4^n.

Seuraavassa taulukossa on esitetty vastaavien a-2-vapaiden sanojen lukumäärät sanan pituuden funktiona.

1	4	-
2	12	3.0000
3	36	3.0000
4	96	2.6667
5	264	2.7500
6	648	2.4545
7	1584	2.4444
8	3576	2.2576
9	7872	2.2013
10	15360	1.9512
11	29184	1.9000
12	51120	1.7516
13	90384	1.7681
14	158448	1.7531
15	286296	1.8069
16	509808	1.7807
17	904296	1.7738
18	1556304	1.7210
19	2638368	1.6953
20	4273512	1.6198
21	6783888	1.5874
22	10308576	1.5196
23	15419640	1.4958
24	22305840	1.4466
25	32204568	1.4438
26	45812088	1.4225
27	65912784	1.4388
28	94069080	1.4272
29	135407400	1.4394
30	194182560	1.4341
31	281332536	1.4488
32	406445496	1.4447
33	590726904	1.4534
34	855581472	1.4484

Abelin kaksi-vapaiden n-pituisten sanojen lukumäärä 4-kirjaimisessa aakkostossa.

Tulokset on saatu yksinkertaisen FORTRAN-ohjelman avulla.

Laskennassa yksi Parikh-vektori sijoitettiin aina yhteen 32-bitin kokonaislukusanaan. Tällainen menettely nopeuttaa käsittelyä vähintään 4-kertaisesti (ehkä jopa 10-kertaisesti, sillä indeksien määrä vähenee, rinnakkaisuus lisääntyy, muistisiirtoja tarvitaan vähemmän, tarkastelussa voidaan käyttää bittioperaatioita, koodi optimoituu paremmin jne.).

Ohjelma muodosti kaikki ab-alkuiset n-pituiset sanat. Saatu lukumäärä kerrot-

tiin luvulla 12 (Kaksi ensimmäistä kirjainta voidaan valita neljästä kirjaimesta $4 \cdot 3$ eri tavalla).

Ohjelmointi suoritettiin FORTRAN-ohjelmointikielellä ja ohjelmakoodin käännös suoritettiin korkeimmalla optimointitasolla (taso 3). Ajot suoritettiin yliopiston IBM9121-keskustietokoneessa.

Vektorointimahdollisuutta ei käytetty, sillä sen ei voida katsoa nopeuttavan käytettyjä ohjelmia (huonosti vektoroituvia).

Tapauksen $n = 34$ laskemiseen kului 51 minuuttia keskusyksikköaikaa. Käytetty keskusyksikköaika oli likimain suoraan verrannollinen ratkaisujen lukumäärään (ts. loppuvaiheessa suoritusaika kasvoi likimain tekijään 1.46^n verrannollisena, kun taas ratkaisujen lukumäärä tekijään 1.44^n verrannollisena).

Välttämättä esiintyvät osasanat

Pyrittäessä määrittämään mielivaltaisen pitkiä a-2-vapaita sanoja syklisen endomorfismin avulla on tärkeää tietää, millaisia n-pituisia sanoja tai niiden syklisiä permutaatioita pitkissä a-2-vapaissa sanoissa välttämättä esiintyy.

Pyrimme nyt tarkastelemaan lähemmin tätä kysymystä.

Toteamme aluksi, että koska kolmekirjaimisissa aakkostoissa pisin mahdollinen a-2-vapaa sana on 7 kirjaimen mittainen, jokainen nelikirjaimisen aakkoston Σ kirjain esiintyy aina viimeistään seitsemän kirjaimen jälkeen.

Jokaisen kirjaimen esiintymistiheys on siis vähintään $1/8$.

Lemma 12: Olkoon Σ 4-kirjaiminen aakkosto, σ jokin syklinen kirjainpermutaatio $\Sigma^* \to \Sigma^*$, $w \in \Sigma^*$ a-2-vapaa sana ja $|w| > 44$. Tällöin w sisältää osasanan, joka on muotoa xy, missä $x, y \in \Sigma$ ja $y = \sigma(x)$.

Lisäksi tulos on paras mahdollinen ts. lyhyemmillä sanoilla tällaista osasanaa ei välttämättä esiinny.

Todistus. Merkitään $\Sigma = a, b, c, d$, missä $\sigma(a) = b$, $\sigma(b) = c$, $\sigma(c) = d$ ja $\sigma(d) = a$.

Liitetään a-2-vapaita sanoja generoivaan tietokoneohjelmaan lisäehto, joka hylkää sanan, jonka loppuosana esiintyy jokin kirjainpareista ab, bc, cd tai da.

Etsimme siis sellaisia sanoja, jossa edellämainittua muotoa olevia osasanoja ei esiinny.

Todetaan, että 45-pituisten sanojen lukumäärä on 0 ja 44-pituisia sanoja on 4 kappaletta, nimittäin sanat

$$\sigma^i(adcadbdcbdbacadcacbacadbdcbdbadbdcacbacadbad),$$

missä $i = 0, 1, 2, 3$. **MOT**

Soveltamalla lemmaa sanojen peilikuviin toteamme, että lemman ehdot toteut-

tavan a-2-vapaan vähintään 45-kirjaimen mittaisen sanan osasanana esiintyy aina myös sana xy, missä $x, y \in \Sigma$ ja $x = \sigma(y)$.

Lemma 13: Olkoon Σ 4-kirjaiminen aakkosto, σ jokin syklinen kirjainpermutaatio $\Sigma^* \to \Sigma^*$, $w \in \Sigma^*$ a-2-vapaa sana ja $|w| > 41$. Tällöin w sisältää osasanan, joka on muotoa xy, missä $x, y \in \Sigma$ ja $y = \sigma^2(x)$.

Lisäksi tulos on paras mahdollinen ts. lyhyemmillä sanoilla tällaista osasanaa ei välttämättä esiinny.

Todistus. Merkitään, kuten edellisen lemman todistuksessa, $\Sigma = a, b, c, d$, missä $\sigma(a) = b$, $\sigma(b) = c$, $\sigma(c) = d$ ja $\sigma(d) = a$.

Liitetään a-2-vapaita sanoja generoivaan tietokoneohjelmaan lisäehto, joka hylkää sanan, jonka loppuosana esiintyy jokin kirjainpareista ac, bd, ca tai db.

Todetaan, että 42-pituisten sanojen lukumäärä on 0 ja 41-pituisia sanoja on 4 kappaletta, nimittäin sanat

$$\sigma^i(abadabcbadabadcdadcbcdcbabcbadabadcdabada),$$

missä $i = 0, 1, 2, 3$. **MOT**

Todettakoon, että tuloksesta seuraa Keräsen työn lemma 2.

Lemma 14: Olkoon Σ 4-kirjaiminen aakkosto, $G = (\Sigma, g, \alpha_0)$ $D0L$-systeemi, missä g on syklinen endomorfismi ja järjestelmän G generoima kieli $L(G)$ a-2-vapaa. Jos $L(G)$ sisältää yli 44-kirjaimisen sanan, niin g toteuttaa 2-*testin*: Jokaisella aakkoston Σ 2-kirjaimisella a-2-vapaalla sanalla w myös $g(w)$ on a-2-vapaa.

Todistus. Endomorfismin g syklisyydestä seuraa, että jos $g(w)$ on a-2-vapaa jollakin sanalla w, niin $g(\sigma^i(w))$ on a-2-vapaa kaikilla $i = 0, 1, 2, 3$.

Lemman 12 mukaan yli 44-kirjaiminen a-2-vapaa sana sisältää aina jonkin sanoista $\sigma^i(ab)$, $i = 0, 1, 2, 3$. Käyttämällä lemmaa peilikuvanaan todetaan, että yli 44-kirjaiminen a-2-vapaa sana sisältää aina jonkin sanoista $\sigma^i(ba)$, $i = 0, 1, 2, 3$. Lemman 13 mukaan yli 41-kirjaiminen aina jonkin sanoista $\sigma^i(ac)$, $i = 0, 1, 2, 3$.

Nämä ovat ainoat a-2-vapaat 2-kirjaimiset sanat ja siten 2-testi on välttämättä voimassa. **MOT**

Keränen on työssään todennut, että 2-testin läpäisevät sykliset endomorfismit ovat vähintään 64 kirjaimen mittaisia.

2-testin avulla saadaan seuraava tulos:

Lemma 15: Olkoon Σ, G, g kuten lemmassa 5. Tällöin välttämättä $\text{first}(g(a)) = \text{last}(g(a))$.

Todistus. Tämä todetaan helposti soveltamalla 2-testiä endomorfismiin g. **MOT**

Huom. Tämä on Keräsen työn lemma 3.

Edellä esitetty 2-testi antaa tehokkaan karsintamenettelyn etsityn tyyppisen endomorfismin g mahdollisille prefix-suffix-pareille.

Lemma 16: Suffix-prefix-pareille, joiden pituudet ovat m_1 (suffix) ja m_2 (prefix), saadaan 2-testin avulla yhteensä

$$\lfloor \frac{m_1^2}{4} \rfloor + \lfloor \frac{m_2^2}{4} \rfloor + 3\lfloor \frac{m_1 m_2}{2} \rfloor$$

Parikh-ehtoa.

Todistus. Kaksi ensimmäistä termiä ilmoittavat prefixin ja suffixin a-2-vapauden tarkistamiseen tarvittavien Parikh-vertailujen lukumäärät.

Viimeinen termi ilmoittaa 2-testin avulla saatavien Parikh-vertailujen lukumäärän, joka on kolme kertaa niiden osasanaparien lukumäärä, jotka muodostuvat kahdesta peräkkäisestä osasanasta, jotka ovat samanpituiset, ja joista ensimmäinen alkaa suffixista ja toinen loppuu prefixiin.

Numeroidaan suffixin kirjaimet $1, 2, ..., m_1$ lopusta lukien ja prefixin kirjaimet $1, 2, ..., m_2$ alusta lähtien.

Merkitään sanaparin alkukohtaa suffixissa i ($1 \leq i \leq m_1$) ja sanaparin pituutta $2j$.

Nyt, jotta sanaparin viimeinen kirjain kuuluisi prefixiin, tulee olla $0 < 2j - i \leq m_2$. Toisin sanoen

$$\frac{i}{2} < j \leq \frac{m_2 + i}{2}.$$

Koska j on kokonaisluku, saadaan

$$\lfloor \frac{i}{2} \rfloor + 1 \leq j \leq \lfloor \frac{m_2 + i}{2} \rfloor.$$

Tällaisten sanaparien kokonaislukumääräksi S saadaan

$$\begin{aligned} S &= \sum_{i=1}^{m_1} \left(\lfloor \frac{m_2 + i}{2} \rfloor - (\lfloor \frac{i}{2} \rfloor + 1) + 1 \right) \\ &= \sum_{i=1}^{m_1} \left(\lfloor \frac{m_2 + i}{2} \rfloor - \lfloor \frac{i}{2} \rfloor \right). \end{aligned}$$

Kun m_2 on parillinen, tästä saadaan

$$S = \sum_{i=1}^{m_1} \left(\lfloor \frac{m_2}{2} \rfloor + \lfloor \frac{i}{2} \rfloor - \lfloor \frac{i}{2} \rfloor \right) = \sum_{i=1}^{m_1} \frac{m_2}{2} = \frac{m_1 m_2}{2}.$$

Kun m_2 on pariton, saadaan vastaavasti

$$S = \sum_{i=1}^{m_1} \left(\frac{m_2 - 1}{2} + \lfloor \frac{i+1}{2} \rfloor - \lfloor \frac{i}{2} \rfloor \right)$$

$$= \frac{m_1(m_2-1)}{2} + \sum_{i=1}^{m_1} \left(\lfloor \frac{i+1}{2} \rfloor - \lfloor \frac{i}{2} \rfloor \right).$$

Summalausekkeen termi = 1, kun i on pariton, ja = 0, muulloin. Summalauseke ilmoittaa siis joukossa $\{1, 2, 3, ..., m_1\}$ olevien parittomien lukujen lukumäärän $(= \lfloor (m_1+1)/2 \rfloor)$. Siten

$$S = \frac{m_1(m_2-1)}{2} + \lfloor \frac{m_1+1}{2} \rfloor.$$

Jos m_1 on parillinen, tästä saadaan

$$S = \frac{m_1(m_2-1)}{2} + \frac{m_1}{2} + \lfloor \frac{1}{2} \rfloor = \frac{m_1 m_2}{2}.$$

Jos m_1 on pariton, saadaan

$$S = \frac{m_1(m_2-1)}{2} + \frac{m_1+1}{2} = \frac{m_1 m_2 + 1}{2}.$$

Siis $S = \frac{m_1 m_2}{2}$, kun m_1 tai m_2 on parillinen ja $S = \frac{m_1 m_2 + 1}{2}$ muulloin. Siten aina $S = \lfloor \frac{m_1 m_2}{2} \rfloor$ ja väite on todistettu. **MOT**

Voimme nyt arvioida, milloin karsinta 2-testin avulla antaa eniten Parikh-ehtoja prefix-suffix-parille.

Merkitään $m_1 + m_2 = m$ (kokonaispituus) ja $m_1 = \alpha m$. Nyt $m_2 = (1-\alpha)m$.

Parikh-vertailujen lukumäärälle saadaan eo. lemman avulla arvio $(\lfloor x \rfloor \approx x)$

$$\frac{m^2}{4} \left(\alpha^2 - (1-\alpha)^2 + 6\alpha(1-\alpha) \right)$$

$$= \frac{m}{4}(1 + 4\alpha - 4\alpha^2) = m^2 \left(\frac{1}{2} - (\frac{1}{2} - \alpha)^2 \right).$$

Todetaan, että eniten Parikh-ehtoja $(m^2/2)$ vaaditaan silloin, kun prefix ja suffix ovat saman pituisia $(\alpha = 1/2)$.

PS. *Tämä kirjoitelma on syntynyt lokakuussa 1991 Veikko Keräsen työn ollessa tarkastettavana julkaisemista varten. Pidin tuolloin aiheesta tämän seminaariesitelmän Oulun yliopiston matematiikan laitoksella. Tarkoituksena oli tutustua Keräsen työhön yhdessä laitoksen henkilökunnan kanssa ja läpikäydä sen apulauseet tarkistusmielessä tietokonetta apuna käyttäen.*

Lähteet

Veikko Keränen: Abelian squares are avoidable on 4 letters, In W. Kuich, editor, *Proc. ICALP '92, Lecture Notes in Comp. Sci.*, volume 623, pages 41-52. Springer-Verlag, Berlin, 1992.