Rationaalisen kielen sentralisaattorista ja sen määrittämisestä kiintopistemetodilla

Petri Salmela

Pro gradu -tutkielma

2002

TURUN YLIOPISTO
MATEMATIIKAN LAITOS
FIN-20014 TURKU
FINLAND

TURUN YLIOPISTO
Matemaatiikan laitos/Matemaattis-luonnontieteellinen tiedekunta

SALMELA, PETRI Rationaalisen kielen sentralisaattorista ja sen määrittämisestä kiintopistemetodilla

Pro Gradu -tutkielma, 63 s., 7 liites.
Matematiikka
syyskuu 2002

Työssä tarkastellaan rationaalisten kielten kommutointia. Erityisesti tutkitaan maksimaalista annetun kielen kanssa kommutoivaa kieltä, sentralisaattoria, ja siihen liittyvää Conwayn ongelmaa. Conwayn otaksuma väittää, että rationaalisen kielen sentralisaattori on aina myöskin rationaalinen. J.H. Conway esitti väitteen vuonna 1971 (Regular algebra and Finite Machines, Chapman Hall,1971) ja se on yhä todistamatta. Työssä tutustutaan J. Karhumäen (Challenges of Commutation - An advertisement, in: FCT 2001, R. Freivalds, Ed., Lecture Notes in Computer Science 2138, Springer-Verlag, New York 2001, ss. 15-23) esittämään iteratiivisesti etenevään kiintopistemetodiin, jonka tavoitteena on löytää sentralisaattori erään kuvauksen maksimaalisena kiintopisteenä. Useilla kielillä sentralisaattori löytyy tällä menetelmällä hyvinkin nopeasti. On kuitenkin olemassa myös kieliä, joilla metodi johtaa pysähtymättömään laskentaan.

Työssä esitellään aluksi tarvittavaa välineistöä ja niihin liittyviä tuloksia. Tämän jälkeen esitellään kiintopistemetodia ja sentralisaattorin määrittämistä eri tyyppisten pysähtyvien esimerkkien avulla. Tämän jälkeen osoitetaan sopivalla esimerkillä, että jopa äärellinen kieli voi johtaa pysähtymättömään laskentaan kiintopistemetodia käytettäessä. Esimerkkitapausta tarkastellaan tarkasti ja analysoidaan syitä metodin pysätymättömyyteen. Lisäksi mainitaan joitakin muita esimerkkejä pysähtymättömään laskentaan johtavista tapauksista.

Lopuksi esitetään joitakin päätelmiä ja pohdintoja siitä, miten sentralisaattorin tutkimista voisi jatkaa edelleen.

Työssä käytettiin esimerkkien laskemisessa ja johtopäätösten teon apuna Länsi-Ontarion yliopistossa Kanadassa kehitettyä Grail+ -tietokoneohjelmaa.

Asiasanat: rationaaliset kielet, kielten kommutointi, kielen sentralisaattori, kiintopistemetodi, Conwayn ongelma.

Sisältö

1 Johdanto

Algebrallisia systeemejä tutkittaessa keskeisessä asemassa ovat usein erilaiset yhtälöt ja niiden ratkaiseminen. Usein ratkaisun löydettävyyteen vaikuttaa se, onko yhtälössä käytetty operaatio kommutatiivinen kyseisessä systeemissä. Yhtälö, joka kommutatiivisessa systeemissä on hyvinkin yksinkertainen, saattaa kommutoimattomassa järjestelmässä osoittautua erittäin vaikeaksi. Formaalien kielten yhteydessä käytettävä kertolaskuoperaatio, katenaatio, on eräs esimerkki ei-kommutatiivisesta operaatiosta.

Tässä työssä tarkastellaan erästä formaalien kielten kommutointiin liittyvää ongelmaa. Kahden kielen, X ja Y, kommutointi voidaan esittää hyvin yksinkertaisella yhtälöllä $XY = YX$, mutta kysymys, millaisilla ehdoilla tämä yhtälö on voimassa, on osoittautunut varsin vaikeaksi ongelmaksi. Työssä keskitytään kielen X sentralisaattorin, eli laajimman kielen X kanssa kommutoivan kielen, etsimiseen kiintopistemetodiksi kutsutulla menetelmällä.

Aluksi työssä käydään läpi tarvittavia käsitteitä ja perustuloksia sekä todistetaan joitakin rationaalisten kielten sulkeumatuloksia. Seuraavaksi esitellään sentralisaattorin käsite ja esitetään siihen liittyviä tuloksia ja sentralisaattorin ominaisuuksia. Muotoillaan, yhä avoin, *Conwayn ongelma*, joka väittää, että rationaalisen kielen sentralisaattori on aina rationaalinen. Esitellään lisäksi muutama tämän ongelman muunnos.

Tämän jälkeen muodostetaan iteratiivinen *kiintopistemetodi* kielen X sentralisaattorin löytämiseksi. Conway esitteli kiintopistemetodin yksinkertaisen version kirjassaan [1]. Tässä työssä käytetään kuitenkin Karhumäen [5] esittämää tehokkaampaa versiota kiintopistemetodista. Kiintopistemetodin käyttöä esitellään usealla erilaisella rationaalisella esimerkkitapauksella. Erityisesti kiinnostavat äärelliset kielet. Menetelmää tarkasteltaessa huomataan, että se rationaalisten kielten tapauksessa antaa sentralisaattorin usein jo muutaman iteraatioaskeleen jälkeen. Toisaalta esitetään esimerkkinä myös varsin yksinkertainen, vain muutaman sanan kieli, jolla iteraatio ei pääty lainkaan, vaikka sentralisaattori onkin esitettävissä hyvin yksinkertaisesti.

Työn lopuksi pohditaan joitakin esiin tulleita ajatuksia, miten sentralisaattoriin liittyviä ongelmia, kuten Conwayn ongelmaa, voitaisiin tutkia tarkemmin.

Rationaalisten kielten operointiin ja kiintopistemetodin suoritukseen käytettiin tämän työn yhteydessä tietokoneohjelmistoa nimeltä *Grail+* [8]. Ohjelma käsittelee rationaalisia kieliä esittämällä ne äärellisinä automaatteina, joille suoritetaan halutut operaatiot.

2 Käsitteitä ja määritelmiä

Tässä kappaleessa kerrataan ja määritellään joitakin tarvittavia peruskäsitteitä sekä todistetaan näille muutama sulkeumatulos.

2.1 Kielistä

Aluksi kerrataan formaalien kielten teorian perusvälineistö. Kielten perustana on äärellinen *kirjainten* joukko Σ, eli *aakkosto*. Kirjaimille määritellään binäärioperaationa *katenaatio*, eli kirjainten kirjoittaminen peräkkäin. Esimerkiksi $a \cdot b = ab$. Aakkoston Σ generoimasta vapaasta puoliryhmästä käytetään merkintää Σ^+ ja sen alkioita kutsutaan *sanoiksi*. Kun tähän puoliryhmään lisätään ykkösalkioksi tyhjä sana 1, saadaan vastaava vapaasti generoitu monoidi Σ^*.

Aakkoston Σ *kieliksi* kutsutaan monoidin Σ^* osajoukkoja. Sanojen katenaatio laajennetaan kielten väliseksi binäärioperaatioksi määrittelemällä

$$AB = \{\alpha\beta \quad | \quad \alpha \in A,\ \beta \in B\}, \qquad A, B \subseteq \Sigma^*.$$

Tällä binäärioperaatiolla varustettu potenssijoukko $\wp(\Sigma^*)$ muodostaa nyt monoidin, jonka neutraalialkiona on singletonjoukko $\{1\}$. Kielen X sisältämien sanojen lukumäärästä käytetään tässä työssä merkintää $|X|$ ja merkintä $|w|$ tarkoittaa sanan $w \in \Sigma^*$ pituutta.

Kielten A ja B joukko-opillisesta unionista $A \cup B$ käytetään tämän esityksen säännöllisissä lausekkeissa merkintää $A + B$. Merkintää voidaan perustella sillä, että katenaatiolla ja unionilla varustettu potenssijoukko $\wp(\Sigma^*)$ on puolirengas, jonka nolla-alkiona on tyhjä joukko $\emptyset$. Erityisesti ovat voimassa distributiivilait

$$\begin{aligned}(A+B)C &= \{uv \in \Sigma^* \,|\, u \in A+B, v \in C\} \\ &= \{uv \in \Sigma^* \,|\, u \in A, v \in C\} + \{uv \in \Sigma^* \,|\, u \in B, v \in C\} \\ &= AC + BC \qquad \forall A, B, C \subseteq \Sigma^*\end{aligned}$$

ja

$$C(A+B) = CA + CB \qquad \forall A, B, C \subseteq \Sigma^*.$$

Samalla tavoin distributiivilait ovat voimassa äärettömille summille

$$\begin{aligned}(\sum_{i=0}^{\infty} A_i)C &= \{uv \in \Sigma^* \,|\, u \in \sum_{i=0}^{\infty} A_i, v \in C\} \\ &= \sum_{i=1}^{\infty} \{uv \in \Sigma^* \,|\, u \in A_i, v \in C\} \\ &= \sum_{i=1}^{\infty} (A_i C) \qquad \forall A_i, C \subseteq \Sigma^*\end{aligned}$$

ja

$$C \sum_{i=0}^{\infty} A_1 = \sum_{i=0}^{\infty} (CA_i) \qquad A_i, C \subseteq \Sigma^*.$$

Kielten potenssilla on luonnollinen merkityksensä kertolaskun lyhennysmerkintänä ja se määritellään asettamalla rekursiivisesti

$$A^0 = \{1\}, \quad A^1 = A \quad \text{ja} \quad A^n = AA^{n-1}.$$

Potenssien äärettöminä unioneina saadaan *Kleenen plus- ja tähtioperaatiot*

$$A^+ = A + A^2 + A^3 + \cdots = \sum_{i>0} A^i = \bigcup_{i=1}^{\infty} A^i$$

ja

$$A^* = \{1\} + A + A^2 + A^3 + \cdots = \sum_{i\geq 0} A^i = \bigcup_{i=0}^{\infty} A^i.$$

Kieli A^+ on kielen A generoima puoliryhmä. Vastaavasti A^* on kielen A generoima monoidin Σ^* alimonoidi. A^+ ja A^* eivät ole välttämättä vapaasti generoituja, kuten Σ^+ ja Σ^*.

Joukkoina kielille ovat luonnollisestikin voimassa kaikki tavalliset joukko-opin säännöt, kuten de Morganin kaava

$$(\Sigma^* \setminus A) \cup (\Sigma^* \setminus B) = \Sigma^* \setminus (A \cap B)$$

ja sen duaali kaava. Vastaavat kaavat ovat tietenkin voimassa myös silloin, jos perusjoukkona käytetään monoidin Σ^* sijasta puoliryhmää Σ^+.

Määritelmä 2.1. Sanan $u \in \Sigma^*$ sanotaan olevan sanan $w \in \Sigma^*$ *prefiksi*, eli *vasen tekijä*, jos $w = uv$ jollakin sanalla $v \in \Sigma^*$. Tällöin käytetään myöskin merkintää $u = wv^{-1}$, eli u on sana, joka saadaan, kun sanan w lopusta poistetaan sana v.

Vastaavasti sana $u \in \Sigma^*$ on sanan $w \in \Sigma^*$ *suffiksi*, eli *oikea tekijä*, jos $w = vu$ jollakin sanalla $v \in \Sigma^*$ ja tästä käytetään merkintää $u = v^{-1}w$.

Seuraavaksi laajennetaan sanojen prefiksin ja suffiksin määritelmässä käyttöön otetut merkinnät kielille.

Määritelmä 2.2. Kielen L_1 *vasen osamäärä* kielen L_2 suhteen on kieli

$$L_2^{-1}L_1 = \{u_2^{-1}u_1 \mid u_1 \in L_1, u_2 \in L_2\}.$$

$L_2^{-1}L_1$ on siis kieli, joka koostuu kielen L_2 sanoilla alkavien kielen L_1 sanojen loppuosista.

Aivan vastaavasti kielen L_1 *oikea osamäärä* kielen L_2 suhteen on kieli

$$L_1L_2^{-1} = \{u_1u_2^{-1} \mid u_1 \in L_1, u_2 \in L_2\}.$$

Esimerkki 2.3. Olkoot $X = \{aa, ab, ba, abb, bab\}$, $Y_1 = \{a\}$ ja $Y_2 = \{b, ab\}$. Nyt osamääräkielinä saadaan

$$\begin{aligned} Y_1^{-1}X &= \{a, b, bb\}, & XY_1^{-1} &= \{a, b\}, \\ Y_2^{-1}X &= \{1, a, b, ab\} \quad \text{ja} & XY_2^{-1} &= \{1, a, b, ab\}. \end{aligned}$$

Määritelmä 2.4. Kielen L prefiksi on sen kaikkien sanojen kaikkien prefiksien joukko

$$\mathrm{Pref}(L) = \{u \in \Sigma^* \mid uv \in L,\ v \in \Sigma^*\}$$

ja vastaavasti kielen suffiksi on sen sanojen kaikkien suffiksien joukko

$$\mathrm{Suf}(L) = \{u \in \Sigma^* \mid vu \in L,\ v \in \Sigma^*\}.$$

Esimerkki 2.5. Olkoon $X = \{aa, ab, ba, abb, bab\}$. Nyt

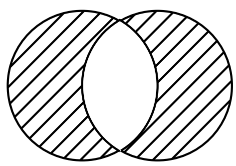
$$\text{Pref}(X) = \{1, a, b, aa, ab, ba, abb, bab\}$$

ja

$$\text{Suf}(X) = \{1, a, b, aa, ab, ba, bb, abb, bab\}.$$

Tässä määritelmässä on prefiksien ja suffiksien perusjoukkona käytetty monoidia Σ^*. Myöhemmin tässä työssä tarkastellaan kieliä pääasiassa puoliryhmän Σ^+ osajoukkoina, jolloin käytetään 1-vapaita prefiksejä ja suffikseja. Ellei sekaannuksen vaaraa ole, näistä käytetään samoja merkintöjä, kuin edellä.

Merkintä 2.6. Otetaan käyttöön lyhennysmerkintä kahden joukon A ja B *symmetriselle erotukselle* $A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$.



Merkintä 2.7. Olkoon sana $w = a_1 a_2 \cdots a_n$, jossa $a_i \in \Sigma$. Nyt sanan w *peilikuva* on $w^R = a_n \cdots a_2 a_1$. Vastaavasti kielen L sanojen peilikuvista muodostuvasta kielestä, *peilikuvakielestä*, käytetään merkintää L^R.

Esimerkki 2.8. Olkoot $w = aab$ ja $L = \{a, aab, aba, bba\}$. Nyt $w^R = baa$ ja $L^R = \{a, baa, aba, abb\}$.

2.2 Rationaalisten kielten sulkeumaominaisuuksia

Rationaalisten kielten perheeksi kutsutaan pienintä kielten joukkoa, joka sisältää kaikki singletonkielet $\{a\}$, jossa $a \in \Sigma$, ja joka on suljettu rationaalisten operaatioiden, eli *katenaation*, *unionin* sekä Kleenen tähtioperaation, eli *iteraation*, suhteen.

Todetaan, että rationaalisten kielten joukko on suljettu kappaleessa 4 käytettävien operaatioiden suhteen. Määritelmänsä mukaan rationaalisten kielten perhe on suljettu rationaalisten operaatioiden suhteen.

Oletetaan tunnetuksi, että rationaaliset kielet voidaan määritellä antamalla kyseisen kielen tunnistava äärellinen automaatti [7]. Tässä esityksessä *deterministisistä* ja *epädeterministisistä* äärellisistä automaateista, *DFA* ja *NFA*, käytetään seuraavia määritelmiä.

Määritelmä 2.9. Deterministiseksi äärelliseksi automaatiksi (*DFA*) kutsutaan viisikkoa

$$\mathcal{A} = (Q, \Sigma, \delta, q_0, F),$$

jossa

- Q on äärellinen *tilojen* joukko,
- Σ on äärellinen aakkosto,
- $\delta : Q \times \Sigma \to Q$ on *siirtymäfunktio*,
- q_0 on *aloitustila* ja
- F on *lopetustilojen* joukko.

Laajennetaan siirtymäfunktio δ funktioksi $\delta^* : Q \times \Sigma^* \to Q$ määrittelemällä

$$\delta^*(q, 1) = q \qquad \text{ja} \qquad \delta^*(q, aw) = \delta^*(\delta(q, a), w)$$

kaikilla sanoilla $w \in \Sigma^*$ ja kirjaimilla $a \in \Sigma$. Nyt DFA:n $\mathcal{A}$ tunnistamaksi kieleksi kutsutaan kieltä

$$\mathrm{L}(\mathcal{A}) = \{w \in \Sigma^* \mid \delta^*(q_0, w) \in F\}.$$

Määritelmä 2.10. Epädeterministiseksi äärelliseksi automaatiksi (*NFA*) kutsutaan viisikkoa

$$\mathcal{A} = (Q, \Sigma, E, Q_0, F),$$

jossa Q, Σ ja F ovat kuten deterministisessä automaatissa, mutta

- $E \subseteq Q \times \Sigma \times Q$ on *siirtymärelaatio* ja
- Q_0 on aloitustilojen joukko.

NFA $\mathcal{A}$ tunnistaa kielen

$$\begin{aligned}\mathrm{L}(\mathcal{A}) = \{w \in \Sigma^* \mid \exists a_1, \ldots, a_n \in \Sigma, q_0, \ldots, q_n \in Q : w = a_1 \cdots a_n, \\ q_0 \in Q_0, q_n \in F \text{ ja } (q_{i-1}, a_i, q_i) \in E \text{ kaikilla } i = 1, \ldots, n\}.\end{aligned}$$

Käyttämällä hyväksi äärellisiä automaatteja todistetaan seuraavat tulokset.

Lause 2.11. *Rationaalisen kielen X prefiksien joukko* $\mathrm{Pref}(X)$ *ja suffiksien joukko* $\mathrm{Suf}(X)$ *ovat myöskin rationaalisia.*

Todistus. Oletetaan, että $\mathcal{A} = (Q, \Sigma, Q_0, E, F)$ on redusoitu NFA, joka määrittelee kielen X, siis $X = \mathrm{L}(\mathcal{A})$. Redusoidulla automaatilla tarkoitetaan tässä yhteydessä automaattia, jonka kaikki tilat ovat saavutettavissa alkutilasta ja jonka kustakin tilasta päästään johonkin lopputilaan. Muodostetaan automaatit

$$\mathcal{A}_1 = (Q, \Sigma, Q_0, E, Q)$$

ja

$$\mathcal{A}_2 = (Q, \Sigma, Q, E, F).$$

Selvästikin $\mathrm{Pref}(X)$ ja $\mathrm{Suf}(X)$ ovat rationaalisia, sillä $\mathrm{Pref}(X) = \mathrm{L}(\mathcal{A}_1)$ ja $\mathrm{Suf}(X) = \mathrm{L}(\mathcal{A}_2)$. □

Lemma 2.12. *Jos kieli L on rationaalinen, on myös sen peilikuvakieli L^R rationaalinen.*

Todistus. Rationaaliselle kielelle L voidaan muodostaa sen tunnistava epädeterministinen äärellinen automaatti

$$\mathcal{A} = \{Q, \Sigma, Q_0, E, F\}.$$

Kun automaatin alkutilat vaihdetaan lopetustiloiksi ja päin vastoin, sekä muodostetaan siirtymärelaatio

$$E' = \{(q_1, a, q_2) \in Q \times \Sigma \times Q \mid (q_2, a, q_1) \in E\}$$

kääntämällä kukin relaation E siirtymä, saadaan automaatti

$$\mathcal{A}' = \{Q, \Sigma, F, E', Q_0\}.$$

Automaatti $\mathcal{A}$ hyväksyy nyt sanan w täsmälleen silloin, kun automaatti $\mathcal{A}'$ hyväksyy sanan w^R. □

Lause 2.13. *Rationaalisten kielten joukko on suljettu vasemman ja oikean osamäärän suhteen. Toisin sanoen, jos kielet L_1 ja L_2 ovat rationaalisia, niin samoin ovat myös kielet $L_1^{-1}L_2$ ja $L_2L_1^{-1}$.*

Todistus. Todistetaan väite konstruoimalla kielen $L_1^{-1}L_2$ tunnistava äärellinen automaatti käyttäen hyväksi kielet L_1 ja L_2 tunnistavia automaatteja. Oikean osamäärän rationaalisuus saadaan todistettua vasemman osamäärän rationaalisuuden ja lemman 2.12 avulla.

Oletetaan, että kielet L_1 ja L_2 ovat rationaaliset ja että $L_1 = \mathrm{L}(\mathcal{A}_1)$ ja $L_2 = \mathrm{L}(A_2)$, kun

$$\begin{aligned}\mathcal{A}_1 &= (Q_1, \Sigma, q_{01}, \delta_1, F_1)\\ \mathcal{A}_2 &= (Q_2, \Sigma, q_{02}, \delta_2, F_2).\end{aligned}$$

Lisäksi voidaan olettaa, että automaattien $\mathcal{A}_1$ ja $\mathcal{A}_2$ kaikki tilat ovat saavutettavissa alkutilasta. Muodostetaan nyt automaatti

$$\mathcal{A}' = (Q_{12}, \Sigma, (q_{01}, q_{02}), \delta_{12}, F_{12}),$$

jossa

- $Q_{12} = (Q_1 \cup \{q\}) \times Q_2$,
- $F_{12} = \{(q_1, q_2) \in Q_{12} \mid q_2 \in F_2\}$ ja
- $\delta_{12} : Q_{12} \times \Sigma \to Q_{12}$ on määritelty kaikilla kirjaimilla $a \in \Sigma$ seuraavasti:
  - jos $\delta_1(q_1, a)$ ja $\delta_2(q_2, a)$ ovat molemmat määriteltyjä, niin $\delta_{12}((q_1, q_2), a) = (\delta_1(q_1, a), \delta_2(q_2, a))$.
  - jos $\delta_2(q_2, a)$ on määritelty, mutta $\delta_1(q_1, a)$ ei, niin $\delta_{12}((q_1, q_2), a) = (q, \delta_2(q_2))$.

Nyt kieli $\mathrm{L}(\mathcal{A}') = \mathrm{L}(\mathcal{A}_2) = L_2$, sillä kullakin sanalla laskenta etenee automaatissa $\mathcal{A}'$ täsmälleen automaatin $\mathcal{A}_2$ määräämällä tavalla. Koska kaikki automaatin $\mathcal{A}_1$ lopputilat ovat saavutettavia, ovat ne kielen L_2 sanat, jotka alkavat kielen L_1 sanoilla, tarkalleen ne, joiden laskenta automaatissa $\mathcal{A}'$ etenee jonkin joukoon

$$Q_0 = \{(q_1, q_2) \in Q_{12} \,|\, q_1 \in F_1\}$$

kuuluvan tilan kautta. Asettamalla nämä joukon Q_0 tilat automaatin alkutiloiksi, saadaan kielen $L_1^{-1}L_2$ tunnistava NFA

$$\mathcal{A} = (Q_{12}, \Sigma, Q_0, \delta_{12}, F_{12}).$$

Kieli $L_1^{-1}L_2$ on siis rationaalinen.

Kielen $L_2L_1^{-1}$ rationaalisuus seuraa edellä olevasta sekä lemmasta 2.12, kun huomataan, että

$$L_2L_1^{-1} = \left((L_1^R)^{-1}L_2^R\right)^R.$$

□

Lause 2.14. *Kahden rationaalisen kielen, L_1 ja L_2, erotus $L_2 \setminus L_1$ on rationaalinen.*

Todistus. Koska L_1 ja L_2 ovat rationaaliset, voidaan olettaa, että $L_1 = \mathrm{L}(\mathcal{A}_1)$ ja $L_2 = \mathrm{L}(\mathcal{A}_2)$. Tässä $\mathcal{A}_1$ ja $\mathcal{A}_2$ ovat lauseen 2.13 merkintöjen mukaiset automaatit. Nyt voidaan muodostaa deterministinen automaatti

$$\mathcal{A} = (Q_{12}, \Sigma, (q_{01}, q_{02}), \delta_{12}, F_{12}),$$

joka määritellään muuten samoin, kuin lauseen 2.13 automaatti $\mathcal{A}'$, mutta lopetustilojen joukoksi F_{12} valitaan joukko

$$F_{12} = \{(q_1, q_2) \,|\, q_1 \notin F_1, q_2 \in F_2\}.$$

Nyt $\mathcal{A}$ hyväksyy täsmälleen ne sanat $w \in \Sigma^*$, jotka kuuluvat kieleen L_2, mutteivät kieleen L_1. □

Seuraus 2.15. Koska rationaalisten kielten perhe on suljettu unionin ja erotuksen suhteen, on se suljettu myöskin leikkauksen ja symmetrisen erotuksen suhteen. Nämä tulokset nähdään suoraan De Morganin kaavasta

$$A \cap B = \Sigma^* \setminus ((\Sigma^* \setminus A) \cup (\Sigma^* \setminus B))$$

ja symmetrisen erotuksen määritelmästä

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

3 Sentralisaattori

Tässä kappaleessa tutkitaan kahden kielen kommutointia, määritellään sentralisaattorin käsite ja tutustutaan eräisiin sentralisaattorin ominaisuuksiin. Lopuksi esitellään Conwayn esittämä ongelma, joka on yhä avoin.

3.1 Kielten kommutointi

Kielten X ja Y kommutointi esitetään hyvin yksinkertaisella yhtälöllä $XY = YX$. Yleisten ehtojen löytäminen kielten kommutoille on kuitenkin osoittautunut varsin vaikeaksi ongelmaksi. Esimerkiksi on yleisesti tunnettua [7], että yksittäiset sanat $u, v \in \Sigma^*$ kommutoivat keskenään, siis $uv = vu$, silloin ja vain silloin, kun $u = t^i$ ja $v = t^j$ jollakin sanalla $t \in \Sigma^*$ ja luvuilla $i, j \geq 0$. Vastaavan kaltaista yleistä ehtoa ei kielten kommutoinnille kuitenkaan ole löydetty.

Jos kielten X ja Y kommutointia tarkastellaan yksittäisten sanojen tasolla, tarkoittaa kommutointiyhtälö $XY = YX$ seuraavaa. Kaikilla sanoilla $x_1 \in X$ ja $y_1 \in Y$ ovat olemassa sellaiset sanat $x_2 \in X$ ja $y_2 \in Y$, joilla

$$x_1 y_1 = y_2 x_2$$

sekä sellaiset sanat $x_3 \in X$ ja $y_3 \in Y$, joilla

$$y_1 x_1 = x_3 y_3.$$

Esimerkki 3.1. Kielet $X = \{a, aaa, b, ba, ab, aba\}$ ja $Y = X \cup \{aa\}$ kommutoivat. Tämä nähdään huomaamalla ensinnäkin, että luonnollisestikin $XX \subseteq XY$.

Toiseksi huomataan, että

$$\begin{array}{ll} aa \cdot a = a \cdot aa & aa \cdot aaa = aaa \cdot aa \\ aa \cdot b = a \cdot ab & aa \cdot ba = a \cdot aba \\ aa \cdot ab = aaa \cdot b & aa \cdot aba = aaa \cdot ba, \end{array}$$

eli siis $\{aa\}X \subseteq XY$. Yhdessä nämä antavat $YX = (X \cup \{aa\})X \subseteq XY$. Koska X ja Y ovat kielinä "symmetrisiä", eli $X = X^R$ ja $Y = Y^R$, saadaan

$$\begin{aligned} YX &\subseteq XY \\ (YX)^R &\subseteq (XY)^R \\ X^R Y^R &\subseteq Y^R X^R \\ XY &\subseteq YX. \end{aligned}$$

Näin ollen $XY = YX$.

3.2 Sentralisaattori ja sen ominaisuuksia

Kiinnitämme nyt toisen kielistä X ja Y sekä tarkastelemme maksimaalista tämän kanssa kommutoivaa kieltä.

Määritelmä 3.2. Aakkoston Σ kielelle X määritellään kaksi *sentralisaattoria* (*centralizer*), $*$-sentralisaattori $\mathcal{C}_*(X)$ ja $+$-sentralisaattori $\mathcal{C}_+(X)$. Sentralisaattori $\mathcal{C}_*(X)$ on maksimaalinen kieli, siis monoidin Σ^* osajoukko, joka kommutoi kielen X kanssa. $\mathcal{C}_+(X)$ on taas puolestaan maksimaalinen kielen X kanssa kommutoiva puoliryhmän Σ^+ osajoukko.

Varmistetaan vielä, että sentralisaattorit ovat aina olemassa ja yksikäsitteiset. Kielen X kanssa kommutoivia kieliä on aina olemassa, sillä kaikilla luvuilla $n \geq 0$ on selvästi voimassa $X^n X = X X^n$. Erityisesti $X^* X = X X^*$. Myöskin puoliryhmän Σ^+ osajoukoista löytyy aina kielen X kanssa kommutoivia kieliä. Jos $1 \notin X$, ainakin X^+ kommutoi kielen X kanssa. Jos taas $1 \in X$, voidaan kommutoivaksi esimerkiksi valita itse Σ^+, sillä $\Sigma^+ X = \Sigma^+ = X \Sigma^+$.

Sentralisaattorin $\mathcal{C}_*(X)$ olemassaolo ja yksikäsitteisyys nähdään, kun huomataan, että kaikkien kielen X kanssa kommutoivien kielten unioni

$$\bigcup\{A \subseteq \Sigma^* \mid AX = XA\}$$

kommutoi, distributiivilain ansiosta, myöskin kielen X kanssa sekä sisältää kaikki muut kommutoivat kielet. Tämä on juuri haluttu maksimaalinen kommutoiva kieli. Sentralisaattorin $\mathcal{C}_+(X)$ olemassaolo ja yksikäsitteisyys perustellaan aivan vastaavalla tavalla korvaamalla Σ^* puoliryhmällä Σ^+.

Toteamme eräitä sentralisaattorin ominaisuuksia.

Lause 3.3. *Kielen X $*$-sentralisaattori $\mathcal{C}_*(X)$ on monoidi ja vastaavasti $+$-sentralisaattori $\mathcal{C}_+(X)$ on puoliryhmä.*

Todistus. Kieli A on monoidi (vastaavasti puoliryhmä) tarkalleen silloin, jos $A = A^*$ (vastaavasti $A = A^+$). Todistetaan $*$-sentralisaattoria koskeva väittämä. $+$-sentralisaattoria koskeva todistus on aivan samanlainen. Osoitetaan ensin induktiolla, että $\mathcal{C}_*(X)^n X = X\mathcal{C}_*(X)^n$ kaikilla luvuilla $n \geq 0$. Ensinnäkin

$$\mathcal{C}_*(X)^0 X = \{1\}X = X = X\{1\} = X\mathcal{C}_*(X)^0.$$

Toisaalta, jos oletetaan, että väite pätee, kun $n \leq k$, saadaan sentralisaattorin määritelmän ja induktio-oletuksen nojalla, että

$$\mathcal{C}_*(X)^{k+1}X = \mathcal{C}_*(X)^k\mathcal{C}_*(X)X = \mathcal{C}_*(X)^k X\mathcal{C}_*(X) = X\mathcal{C}_*(X)^{k+1}.$$

Siispä $\mathcal{C}_*(X)^n X = X\mathcal{C}_*(X)^n$ jokaisella luvulla n ja näin ollen $\mathcal{C}_*(X)^* X = X\mathcal{C}_*(X)^*$. Koska $\mathcal{C}_*(X)$ on maksimaalinen, on oltava $\mathcal{C}_*(X)^* \subseteq \mathcal{C}_*(X)$. Koska taas triviaalisti on voimassa $\mathcal{C}_*(X) \subseteq \mathcal{C}_*(X)^*$, ovat nämä kaksi joukkoa samat

$$\mathcal{C}_*(X)^* = \mathcal{C}_*(X).$$

$\mathcal{C}_*(X)$ on siis monoidi. □

Lause 3.4. *Jos $1 \in X$, niin $\mathcal{C}_*(X) = \Sigma^*$ ja $\mathcal{C}_+(X) = \Sigma^+$.*

Todistus. Jos $1 \in X$, niin $\Sigma^* X = \Sigma^* = X\Sigma^*$. Toinen väite todistuu aivan samoin. □

Ellei sekaannuksen vaaraa ole, käytetään sentralisaattorista vain merkintää $\mathcal{C}(X)$. Jatkossa tulemme tarkastelemaan pääasiassa $+$-sentralisaattoreita. Näin siksi, että $*$-sentralisaattorin tapauksessa tarkastelu on joko triviaali taikka helposti muunnettavissa $+$-sentralisaattorin vastaavasta tarkastelusta. Tilanne nimittäin jakautuu neljään osaan sen mukaan, tutkitaanko $+$- vai $*$-sentralisaattoria ja onko tyhjä sana 1 kielessä X vai ei.

Jos $1 \in X$, on kyseessä lauseen 3.4 mukainen tilanne. Tosin $+$-sentralisaattorin tutkiminen ei tällöin ole välttämättä edes mielekästä.

Jos taas $1 \notin X$, niin tilanne, jossa tutkitaan sentralisaattoria $\mathcal{C}_+(X)$, on tässä tarkasteltava tapaus. Sentralisaattoria $\mathcal{C}_*(X)$ koskevat tulokset taas ovat analogisia sentralisaattorin $\mathcal{C}_+(X)$ vastaavien kanssa sillä erotuksella, että perusjoukkona toimii puoliryhmän Σ^+ sijasta monoidi Σ^*. Toisin sanoen $+$-sentralisaattoria koskevat tulokset voidaan muuntaa vastaaviksi $*$-sentralisaattoria koskeviksi vaihtamalla tarvittaessa merkin $+$ tilalle merkki $*$, eli hyväksymällä tyhjän sanan esiintyminen kielissä. Tämä ei välttämättä tarkoita, että $\mathcal{C}_*(X) = \mathcal{C}_+(X) \cup \{1\}$. Esimerkissä 3.7 ja kappaleessa 5.3 käsitelty kieli $X = \{a, ab, ba, bb\}$ on eräs esimerkki kielestä, jolla $\mathcal{C}_*(X) \neq \mathcal{C}_+(X) \cup \{1\}$.

Lause 3.5. *Kielen X sentralisaattori on myöskin kielen X^+ sentralisaattori, eli*

$$\mathcal{C}(X) = \mathcal{C}(X^+).$$

Todistus. Todistetaan väite osoittamalla, että yhtälön molemmat puolet sisältyvät toisiinsa.

Osoitetaan induktiolla, että $\mathcal{C}(X)X^n = X^n\mathcal{C}(X)$ kaikilla luvuilla $n \geq 1$, jolloin $\mathcal{C}(X)X^+ = X^+\mathcal{C}(X)$. Aluksi, kun $n = 1$, määritelmän mukaisesti

$$\mathcal{C}(X)X = X\mathcal{C}(X).$$

Jos väite pätee, kun $n \leq k$, niin

$$\mathcal{C}(X)X^{k+1} = X\mathcal{C}(X)X^k = X^{k+1}\mathcal{C}(X),$$

eli $\mathcal{C}(X) \subseteq \mathcal{C}(X^+)$.

Väitteen toinen puoli saadaan huomaamalla, että $X \subseteq X^+ \subseteq \mathcal{C}(X^+)$, ja että $\mathcal{C}(X^+)$ on puoliryhmä, jolloin

$$X\mathcal{C}(X^+) \subseteq X^+\mathcal{C}(X^+) = \mathcal{C}(X^+)X^+ = \underbrace{\mathcal{C}(X^+)X^*}_{\subseteq \mathcal{C}(X^+)} \cdot X \subseteq \mathcal{C}(X^+)X.$$

Samoin $\mathcal{C}(X^+)X \subseteq X\mathcal{C}(X^+)$, joten $X\mathcal{C}(X^+) = \mathcal{C}(X^+)X$. Sentralisaattorin $\mathcal{C}(X)$ maksimaalisuuden vuoksi

$$\mathcal{C}(X^+) \subseteq \mathcal{C}(X).$$

□

Sentralisaattorin etsiminen voidaan aloittaa määrittämällä sille ensin ala- ja ylärajat.

Lause 3.6. *Olkoon X annettu aakkoston Σ kieli ja $\mathcal{C}(X)$ sen sentralisaattori. Tällöin sentralisaattorille on voimassa*

$$X^+ \subseteq \mathcal{C}(X) \subseteq \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+). \tag{1}$$

Jos kyseessä on $+$-sentralisaattori, tulkitaan $\mathrm{Pref}(X^+)$ *ja* $\mathrm{Suf}(X^+)$ *1-vapaiksi.*

Todistus. Ensimmäinen sisältyminen on selvä, sillä X^+ kommutoi kielen X kanssa ja näin ollen sisältyy sentralisaattoriin.

Toinen sisältyminen saadaan tarkastelemalla sentralisaattorin sisältymistä kumpaankin kieleen $\mathrm{Pref}(X^+)$ ja $\mathrm{Suf}(X^+)$ erikseen. Osoitetaan, että $\mathcal{C}(X) \subseteq \mathrm{Pref}(X^+)$. Sisältyminen $\mathcal{C}(X) \subseteq \mathrm{Suf}(X^+)$ nähdään samoin. Kullakin kokonaisluvulla $n \geq 1$ ja sanoilla $z_1 \in \mathcal{C}(X)$, $x_1, \ldots, x_n \in X$ on olemassa sanat $z_2, \ldots, z_{n+1} \in \mathcal{C}(X)$ ja $y_1, \ldots, y_n \in X$, joilla on voimassa

$$z_i x_i = y_i z_{i+1},$$

kun $1 \leq i \leq n$. Näin siis on voimassa

$$z_1 x_1 \cdots x_n = y_1 \cdots y_n z_{n+1}.$$

Koska oletamme, että $1 \notin X$, niin $|y_i| \geq 1$ kaikilla indekseillä i ja siis tarpeeksi suurella luvun n arvolla $|y_1 \cdots y_n| \geq |z_1|$. Näin siis $z_1 \in \mathrm{Pref}(X^+)$ kaikilla sentralisaattorin alkioilla z_1, joten $\mathcal{C}(X) \subseteq \mathrm{Pref}(X^+)$. □

Sentralisaattorille voidaan koettaa löytää myös tarkempia rajoja. Esimerkiksi kieli

$$S = \{w \in \Sigma^+ \mid wX \subseteq XX^+ \text{ ja } Xw \subseteq XX^+\} \tag{2}$$

sisältyy sentralisaattoriin. Tämä nähdään seuraavasti. Kielen S määritelmän mukaisesti $X^+ \subseteq S$. Nyt $SX \subseteq XX^+ \subseteq XS$ ja vastaavasti toisin päin $XS \subseteq XX^+ = X^+X \subseteq SX$, eli $XS = SX$.

Näin siis on voimassa

$$X^+ \subseteq S \subseteq \mathcal{C}(X). \tag{3}$$

Varsin usein sentralisaattori onkin joko kieli X^+ tai kieli S. Edellisessä tapauksessa tietenkin tarkemmin sanottuna $X^+ = S = \mathcal{C}(X)$. On kuitenkin olemassa myös kieliä X, joiden sentralisaattori on aidosti suurempi, kuin kieli S.

Esimerkki 3.7. Esimerkkinä tilanteesta, jolloin kaavassa (3) on voimassa aito sisältyminen $S \subset \mathcal{C}(X)$, voidaan käsitellä tapausta, jossa $X = \{a, ab, ba, bb\}$. Kappaleessa 5.3 todistetaan tarkemmin, että $\mathcal{C}(X) = \Sigma^+ \setminus \{b\}$. Erityisesti nyt $\mathcal{C}(X) \neq S$, sillä esimerkiksi $bab \in \mathcal{C}(X)$, mutta $bab \cdot bb = babbb \notin XX^+$, eli $bab \notin S$.

Kieli S on määritelty niiden sanojen joukkona, joilla on tietty ominaisuus. Jotta kieltä S voitaisiin käsitellä helpommin, ja erityisesti, jotta se voitaisiin laskea koneellisesti, on sille etsittävä toisenlainen, laskukaavan muotoinen, esitystapa.

Lause 3.8. *Kieli S voidaan esittää muodossa*

$$S = \Sigma^+ \setminus \left((\Sigma^+ \setminus XX^+)X^{-1} \cup X^{-1}(\Sigma^+ \setminus XX^+)\right). \tag{4}$$

Todistus. Lähdetään liikkeelle kaavassa (2) annetusta kielen S määritelmästä. Koska

$$w \in S \iff (\forall a \in X)(wa \in XX^+ \wedge aw \in XX^+),$$

on siis voimassa

$$\begin{aligned} w \notin S &\iff (\exists a \in X)(wa \notin XX^+ \vee aw \notin XX^+) \\ &\iff (\exists a \in X)(wa \in \Sigma^+ \setminus XX^+ \vee aw \in \Sigma^+ \setminus XX^+). \end{aligned}$$

Nyt kielen S komplementti saadaan muodossa

$$\begin{aligned}\Sigma^+ \setminus S &= \{w \in \Sigma^+ \mid (\exists a \in X)(wa \in \Sigma^+ \setminus XX^+)\} \\ &\quad \cup \{w \in \Sigma^+ \mid (\exists a \in X)(aw \in \Sigma^+ \setminus XX^+)\} \\ &= (\Sigma^+ \setminus XX^+)X^{-1} \cup X^{-1}(\Sigma^+ \setminus XX^+).\end{aligned}$$

Näin ollen kieli S on tämän komplementti

$$S = \Sigma^+ \setminus \left((\Sigma^+ \setminus XX^+)X^{-1} \cup X^{-1}(\Sigma^+ \setminus XX^+)\right).$$

□

Lauseessa 3.8 esitetyssä kaavassa esiintyvät operaatiot ovat rationaalisuuden säilyttäviä ja kohtuullisen helposti toteutettavissa äärellisille automaateille. Näin annetusta rationaalisesta kielestä X riippuva kieli S voidaan laskea esimerkiksi Grail+:aa käyttäen.

Tässä työssä esitetyissä esimerkeissä käytetään aakkostona binääristä aakkostoa $\Sigma = \{a, b\}$. Binäärisen aakkoston käyttö ei kuitenkaan rajoita tulosten yleisyyttä, sillä suuremmalla kirjainmäärällä varustetut aakkostot voidaan koodata sopivalla koodauksella binääriselle aakkostolle.

Lause 3.9. *Olkoon $\Sigma = \{a_1, a_2, \ldots, a_n\}$ äärellinen aakkosto. Sopivasti valitulla koodauksella ψ voidaan aakkoston Σ mielivaltainen kieli X koodata binääriselle aakkostolle $\{0, 1\}$ siten, että*

$$\psi(\mathcal{C}(X)) = \mathcal{C}(\psi(X)).$$

Todistus. Valitaan koodisanojen joukoksi $\Gamma = \{10^i1 | i = 1, 2, \ldots, n\}$ ja suoritetaan koodaus käyttämällä morfismia

$$\psi : \Sigma^* \to \Gamma^*,\ \psi(a_i) = 10^i1.$$

Kielen $X \subseteq \Sigma^+$ koodattu kuva on siis $\psi(X) \subseteq \Gamma^+$. Nyt $\mathcal{C}(\psi(X)) \subseteq \Gamma^+$, sillä $\mathcal{C}(\psi(X)) \subseteq \mathrm{Pref}(\psi(X)^+) \cap \mathrm{Suf}(\psi(X)^+) \subseteq \Gamma^+$. Koska $\mathcal{C}(X)X = X\mathcal{C}(X)$, on myös voimassa

$$\psi(\mathcal{C}(X))\psi(X) = \psi(\mathcal{C}(X)X) = \psi(X\mathcal{C}(X)) = \psi(X)\psi(\mathcal{C}(X))$$

ja siis $\psi(\mathcal{C}(X)) \subseteq \mathcal{C}(\psi(X))$.

Toisaalta, koska morfismi ψ on injektio, on

$$\begin{aligned}\psi^{-1}(\mathcal{C}(\psi(X)))X &= \psi^{-1}(\mathcal{C}(\psi(X)))\psi^{-1}(\psi(X)) = \psi^{-1}(\mathcal{C}(\psi(X))\psi(X)) \\ &= \psi^{-1}(\psi(X)\mathcal{C}(\psi(X))) = X\psi^{-1}(\mathcal{C}(\psi(X))),\end{aligned}$$

eli $\psi^{-1}(\mathcal{C}(\psi(X))) \subseteq \mathcal{C}(X)$ ja siis $\mathcal{C}(\psi(X)) \subseteq \psi(\mathcal{C}(X))$.

Kokonaisuudessaan siis $\mathcal{C}(\psi(X)) = \psi(\mathcal{C}(X))$. □

Äärellisen mielivaltaisen aakkoston kielen X sentralisaattoria voidaan siis tutkia koodaamalla kieli binääriselle aakkostolle ja tutkimalla koodatun kielen sentralisaattoria.

Edellä olevassa ei koodaukseksi voida kuitenkaan valita mitä tahansa koodausta.

Esimerkki 3.10. Olkoon $\Sigma = \{a, b, c, d\}$ aakkosto joka koodataan binääriselle aakkostolle $\{0, 1\}$ kuvauksella

$$\psi : \Sigma^* \to \{0,1\}^*,\ \psi(a) = 0000, \psi(b) = 0011, \psi(c) = 1100, \psi(d) = 1111.$$

Jos tutkittavaksi kieleksi valitaan $X = \Sigma = \{a, b, c, d\}$, saadaan sen sentralisaattoriksi $\mathcal{C}(X) = X^+ = \Sigma^+$ ja koodatuksi kieleksi

$$\psi(X) = \{0000, 0011, 1100, 1111\} = \{00, 11\}^2.$$

Näillä valinnoilla kuitenkin kielen X kuvan sentralisaattori ei ole sama, kuin sentralisaattorin kuva, sillä

$$\mathcal{C}(\psi(X)) = \{00, 11\}^+ \neq (\{00, 11\}^2)^+ = \psi(\mathcal{C}(X)).$$

3.3 Conwayn ongelma

Sentralisaattoriin liittyy useita yhä avoimia ongelmia. Esittelemme tässä erään, jo vuonna 1971 esitetyn, mutta yhä ratkaisemattoman ongelman.

Conwayn ongelma. *Onko rationaalisen kielen sentralisaattori aina rationaalinen?*

John Horton Conway esitti tämän kysymyksen yli 30 vuotta sitten kirjassaan [1]. Conway käytti tosin sentralisaattorin sijasta nimitystä *normalizer*. Vieläkään ongelmaan ei ole löydetty vastausta. Toisin sanoen sitä ei ole pystytty todistamaan oikeaksi, mutta kukaan ole myöskään löytänyt vastaesimerkkiä. Rationaalisen kielen sentralisaattorista ei ole vielä onnistuttu edes selvittämään, onko kyseessä aina rekursiivinen kieli. Samat ongelmat pysyvät yhä voimassa, vaikka rajoituttaisiin tarkastelemaan äärellisiä kieliä.

Myönteinen vastaus Conwayn ongelmaan on todistettu eräissä erikoistapauksissa, esimerkiksi, kun kieli X on rationaalinen prefiksikoodi, kun X on säännöllinen ω-koodi [4], kun $|X| \leq 2$ [3], tai kun $|X| = 3$ [2].

Sentralisaattorista tiedetään kuitenkin, että sen komplementti on rekursiivisesti numeroituva. Toisin sanoen, on olemassa algoritmi, joka äärellisessä suoritusajassa, ennemmin tai myöhemmin, hyväksyy kunkin kielen X sentralisaattorin komplementtiin kuuluvan sanan, mutta ei välttämättä pysähdy sentralisaattoriin kuuluvilla sanoilla. Tällainen algoritmi muodostuu tässä työssä määriteltävän kiintopistemetodin sivutuotteena.

4 Kiintopistemetodi

Tässä kappaleessa esitellään *kiintopistemetodiksi* kutsuttu menetelmä, jolla pyritään löytämään annetun kielen X sentralisaattori $\mathcal{C}(X)$. Erityisesti rationaalisille kielille kiintopistemetodi voidaan toteuttaa helposti jollakin äärellisiä automaatteja käsittelemään kykenevällä tietokoneohjelmalla, kuten *Grail+*:lla [8]. Koska Grail+ käsittelee kieliä ne tunnistavien äärellisten automaattien avulla, viitataan kielen X tunnistavaan automaattiin jatkossa vain kieltä X vastaavana automaattina. Erityisesti nämä automaatit oletetaan deterministisiksi ja minimoiduiksi.

Kiintopistemetodin ideana on muodostaa kuvaus $\varphi : \wp(\Sigma^+) \to \wp(\Sigma^+)$, jonka maksimaalinen kiintopiste on kielen X sentralisaattori $\mathcal{C}(X)$ ja tämän jälkeen sopivasta sentralisaattorin sisältävästä kielestä aloittaen löytää sentralisaattori iteratiivisesti.

Olkoon X annettu 1-vapaa kieli. Määritellään kielten X_i jono rekursiivisesti määräämällä

$$X_0 = \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+) \tag{5}$$

ja

$$X_{i+1} = X_i \setminus \left(X^{-1}(XX_i \Delta X_i X) \cup (XX_i \Delta X_i X)X^{-1}\right), \qquad i \geq 0. \tag{6}$$

Käytetään nyt kaikkien X_i-kielten äärettömästä leikkauksesta merkintää

$$Z_0 = \bigcap_{i \geq 0} X_i. \tag{7}$$

Lause 4.1. *Kieli Z_0 on kielen X sentralisaattori, siis $Z_0 = \mathcal{C}(X)$.*

Todistus. Koska kukin X_{i+1} saadaan vähentämällä jokin kieli edellisestä kielestä X_i, on selvää, että $X_{i+1} \subseteq X_i$ kullakin indeksillä $i \geq 0$. Lauseen 3.6 nojalla tiedetään, että $\mathcal{C}(X) \subseteq X_0$. Jos $\mathcal{C}(X) \subseteq X_i$ jollakin indeksillä, niin

$$\mathcal{C}(X)X = X\mathcal{C}(X) \subseteq XX_i \cap X_iX.$$

Siis

$$\mathcal{C}(X)X \cap (XX_i \Delta X_iX) = \emptyset$$

ja edelleen

$$\mathcal{C}(X) \cap \left(X^{-1}(XX_i \Delta X_iX) \cup (XX_i \Delta X_iX)X^{-1}\right) = \emptyset,$$

joten myös $\mathcal{C}(X) \subseteq X_{i+1}$. Induktiolla siis nähdään, että $\mathcal{C}(X) \subseteq X_i$ kaikilla indekseillä $i \geq 0$. Näin ollen, siis myöskin $\mathcal{C}(X) \subseteq Z_0$.

Toisaalta $Z_0X = XZ_0$ ja näin ollen $Z_0 \subseteq \mathcal{C}(X)$. Jos olisi voimassa $Z_0X \neq XZ_0$, niin silloin olisi olemassa jokin sellainen sana $w \in Z_0$, jolla joko $wX \not\subseteq XZ_0$ tai $Xw \not\subseteq Z_0X$. Nyt voidaan olettaa, että kyseessä on edellinen tapaus. Kielen Z_0 määritelmän johdosta tämä tarkoittaa sitä, että jostakin indeksistä k alkaen $wX \not\subseteq XX_i$, kun $i \geq k$. Nyt kuitenkin, koska w kielen Z_0 sanana on myös jokaisessa kielessä X_i, erityisesti kielessä X_k, huomataan, että $X_kX \neq XX_k$ ja siis $w \in (XX_k \Delta X_kX)X^{-1}$. Nyt siis $w \notin X_{k+1}$, vastoin oletusta $w \in Z_0$.

Koska sisältyminen on voimassa molemmin päin, on siis voimassa $Z_0 = \mathcal{C}(X)$. $\square$

Edellä oleva kaava (6) määrittelee kuvauksen

$$\varphi : \wp(\Sigma^+) \to \wp(\Sigma^+),\ \varphi(Y) = Y \setminus \left(X^{-1}(YX\Delta XY) \cup (YX\Delta XY)X^{-1}\right). \quad (8)$$

Kuvauksen kiintopisteitä ovat tarkalleen ne kielet Y, joilla $YX\Delta XY = \emptyset$, eli joilla $XY = YX$. Maksimaalisena kielen X kanssa kommutoivana kielenä $\mathcal{C}(X)$ on siis kuvauksen φ maksimaalinen kiintopiste.

Kielen X sentralisaattori löydetään suorittamalla metodia rekursiivisesti niin kauan, kunnes $XX_i\Delta X_iX = \emptyset$, eli kunnes $X_{i+1} = X_i$. Useassa tapauksessa metodi antaa sentralisaattorin jo muutaman, yleensä kahdesta viiteen, laskentakierroksen jälkeen. Ikävä kyllä joillakin kielillä sentralisaattoria ei kuitenkaan saavuteta äärellisellä askelmäärällä, vaan ainoastaan lausekkeen (7) mukaisena raja-arvona.

Jos rationaalisen kielen sentralisaattori saavutetaan äärellisellä askelmäärällä, on myös sentralisaattori rationaalinen, sillä kussakin vaiheessa on käytetty vain rationaalisuuden säilyttäviä operaatioita. Jos sentralisaattori kuitenkin saadaan vain raja-arvona, ei sen rationaalisuudesta ole takeita, sillä ääretön leikkaus ei välttämättä säilytä rationaalisuutta. Esimerkiksi kielet

$$A_i = \{a^{2^k} \mid 1 \leq k \leq i\} \cup a^{2^i} a^*, \qquad i = 1, 2, 3, \ldots$$

ovat rationaalisia, muuta niiden leikkaus

$$A = \bigcap_{i=1}^{\infty} A_i = \{a, a^2, a^4, a^8, a^{16}, \ldots\} = \{a^{2^i} \mid i = 1, 2, 3, \ldots\}$$

ei ole rationaalinen.

Kiintopistemetodi ei siis antanut varmaa vastausta siihen, onko rationaalisen kielen sentralisaattori aina rationaalinen. Rationaalisen kielen sentralisaattorin komplementille saadaan kuitenkin seuraava tulos.

Lause 4.2. *Rationaalisen kielen X sentralisaattorin $\mathcal{C}(X)$ komplementti $\Sigma^+ \setminus \mathcal{C}(X)$ on rekursiivisesti numeroituva kieli.*

Todistus. Kiintopistemetodi antaa kielen $\Sigma^+ \setminus \mathcal{C}(X)$ jäsenyysprobleeman selvittämiseen menetelmän, joka pysähtyy kaikilla kieleen $\Sigma^+ \setminus \mathcal{C}(X)$ kuuluvilla syötteillä. Jäsenyysprobleemaa voidaan tutkia suorittamalla kiintopistemetodia. Jos sana w kuuluu sentralisaattorin komplementtiin, huomataan tämä siitä, että $w \notin X_i$ jollakin luvulla i. Jos taas w ei kuulu sentralisaattorin komplementtiin, ei menetelmä välttämättä pysähdy lainkaan. □

Jos kiitopistemetodi pysähtyy jollakin syötteellä X, on tilanne yksinkertainen. Pysähtymättömässä tapauksessa voidaan sentralisaattorista tehdä joitakin päätelmiä seuraavilla kahdella tavalla.

Kappaleessa 3 määritellystä kielestä S tiedetään, että

$$X^+ \subseteq S \subseteq \mathcal{C}(X).$$

Koska S voidaan laskea lauseen 3.8 mukaisella kaavalla, voimme vertailla kieliä X^+ ja S. Jos nämä ovat eri kielet, tiedämme, että $X^+ \subset \mathcal{C}(X)$, eli kieli X^+ ei ole kielen X sentralisaattori.

Toista tapaa varten määrittelemme kaksi kieleen X liittyvää joukkoa. Olkoon Σ aakkosto, jossa on vähintään kaksi kirjainta, ja X tämän aakkoston 1-vapaa kieli. Nyt kielen $\mathrm{Pref}(X^+)$ sanaa w sanotaan *haaroittumispisteeksi (branching point)*, jos $w\,\mathrm{Pref}_1(X) \subseteq \mathrm{Pref}(X^+)$. Tässä merkinnällä $\mathrm{Pref}_1(X)$ tarkoitetaan kielen X kaikkien 1-pituisten prefiksien joukkoa. Haaroittumispistettä w sanotaan *kriittiseksi*, jos $w \notin X^+$. Nyt selvästikin kaikki kielen X^+ sanat ovat haaroittuvia. Käytetään kieleen X liittyvien haaroittumispisteiden kielestä merkintää B ja kriittisten pisteiden kielestä merkintää C. Seuraavasta lauseesta saamme äärellisille kielille toisen keinon, joka kertoo joissakin tapauksissa, onko $\mathcal{C}(X) = X^+$.

Lause 4.3. *Olkoon $X \subseteq \Sigma^+$ äärellinen kieli. Jos $C = \emptyset$, niin $\mathcal{C}(X) = X^+$.*

Todistus. $\mathcal{C}(X)$ on puoliryhmä, joten jos $z \in \mathcal{C}(X)$ ja $x \in X \subseteq \mathcal{C}(X)$, niin $zx \in \mathcal{C}(X)$. Näin siis myös $zx \in \mathrm{Pref}(X^+)$. Erityisesti siis $z\,\mathrm{Pref}_1(X) \subseteq \mathrm{Pref}(X^+)$, joten $z \in B$. Koska oletimme, että $C = \emptyset$, ei z ole kriittinen, eli $z \in X^+$. Siispä $\mathcal{C}(X) = X^+$. □

Lauseen 4.3 mukaan, jos äärellisen kielen X haaroittumispisteiden joukko on tyhjä, on sen sentralisaattori kieli X^+. Väite ei kuitenkaan päde toiseen suuntaan, sillä on olemassa äärellisiä kieliä, joiden sentralisaattori on X^+, mutta joiden haaroittumispisteiden joukko on epätyhjä. Eräs esimerkki tällaisesta kielestä on kappaleessa 6 tarkemmin käsiteltävä kieli $X = \{a, bb, aba, bab, bbb\}$.

Kiintopistemetodin rekursiivista kaavaa voidaan yksinkertaistaa hieman. Itse menetelmä ja sen välivaiheet eivät kuitenkaan muutu millään lailla.

Lause 4.4. *Kiintopistemetodin rekursiivisessa kaavassa voidaan symmetriset erotukset Δ korvata sopivaan suuntaan olevilla tavallisilla erotuksilla. Tällöin*

$$X_{i+1} = X_i \setminus \left(X^{-1}\left(XX_i \setminus X_iX\right) \cup \left(X_iX \setminus XX_i\right)X^{-1}\right).$$

Todistus. Alkuperäisen kaavan symmetriset erotukset voidaan kirjoittaa tavallisten erotusten unioneina

$$\begin{aligned} X_{i+1} &= X_i \setminus \left(X^{-1}(XX_i \Delta X_i X) \cup (XX_i \Delta X_i X) X^{-1} \right) \\ &= X_i \setminus \left(X^{-1} \left((XX_i \setminus X_i X) \cup (X_i X \setminus XX_i) \right) \right. \\ &\qquad \left. \cup \left((XX_i \setminus X_i X) \cup (X_i X \setminus XX_i) \right) X^{-1} \right). \end{aligned}$$

Oikeat ja vasemmat osamäärät voidaan ottaa unionien osasista erikseen, jolloin

$$\begin{aligned} X_{i+1} &= X_i \setminus \left(X^{-1}(XX_i \setminus X_i X) \cup X^{-1}(X_i X \setminus XX_i) \right. \\ &\qquad \left. \cup (XX_i \setminus X_i X) X^{-1} \cup (X_i X \setminus XX_i) X^{-1} \right). \end{aligned}$$

Seuraavaksi huomataan, että koska $X_i X \setminus XX_i$ ei sisällä kielen XX_i sanoja ja vastaavasti $XX_i \setminus X_i X$ ei sisällä kielen $X_i X$ sanoja, on

$$X^{-1}(X_i X \setminus XX_i) \cap X_i = (XX_i \setminus X_i X) X^{-1} \cap X_i = \emptyset.$$

Näin näitä vastaavat osuudet voidaan jättää pois kaavasta ja jäljelle jää väitetty kaava

$$X_{i+1} = X_i \setminus \left(X^{-1} \left(XX_i \setminus X_i X \right) \cup \left(X_i X \setminus XX_i \right) X^{-1} \right).$$

□

Jos vielä vertaamme kiintopistemetodissa käytettävää kuvausta uudessa muodossaan

$$\varphi(Y) = Y \setminus \left(X^{-1}(XY \setminus YX) \cup (YX \setminus XY) X^{-1} \right)$$

kielen S kaavaan

$$S = \Sigma^+ \setminus \left(X^{-1}(\Sigma^+ \setminus XX^+) \cup (\Sigma^+ \setminus XX^+) X^{-1} \right),$$

voimme huomata yhtäläisyyksiä. Kaavojen samankaltaisuus näkyy vielä selvemmin, jos kieli S kirjoitetaan muodossa

$$S = \Sigma^+ \setminus \left(X^{-1}(X\Sigma^+ \setminus X^+ X) \cup (\Sigma^+ X \setminus XX^+) X^{-1} \right).$$

Kiintopistemetodissa käytettävä kuvaus on siis eräänlainen tarkennettu versio kielen S kaavasta. Sentralisaattorin yläraja, josta sanoja poistetaan, on pienennetty kielestä Σ^+ kieleen Y, joka käytännössä on jokin kielistä X_i. Samalla kieli X^+ on korvattu myös kielellä Y. Sanojen poisto tapahtuu siis hieman varovaisemmin, sillä kielen S kaavassa saattaa erotuksessa hävitä myös sentralisaattorin alkioita.

Kieli X_i voidaan antaa myös samanlaisen joukkomerkinnän avulla, kuin kieli S kaavassa (2).

Lause 4.5. *Kieli X_{i+1} voidaan kirjoittaa muodossa*

$$X_{i+1} = \{w \in X_i \,|\, wX \subseteq XX_i \;\; \textit{ja} \;\; Xw \subseteq X_iX\}$$

Todistus. Tämän väitteen todistaminen tapahtuu samoin, kuin lausessa 3.8. Jos merkitään

$$T = \{w \in X_i \,|\, wX \subseteq XX_i \;\; \text{ja} \;\; Xw \subseteq X_iX\},$$

niin

$$\begin{aligned}
w \in T \quad &\iff \quad w \in X_i \quad \text{ja} \quad (\forall a \in X)(wa \in XX_i \;\; \wedge \;\; aw \in X_iX) \\
w \in X_i \setminus T \quad &\iff \quad w \in X_i \quad \text{ja} \quad (\exists a \in X)(wa \notin XX_i \;\; \vee \;\; aw \notin X_iX) \\
&\iff \quad w \in X_i \quad \text{ja} \quad (\exists a \in X)(wa \in X_iX \setminus XX_i \;\; \vee \;\; aw \in XX_i \setminus X_iX).
\end{aligned}$$

Näin ollen

$$X_i \setminus T = \big((X_iX \setminus XX_i)X^{-1} \cup X^{-1}(XX_i \setminus X_iX)\big) \cap X_i$$

ja edelleen

$$T = X_i \setminus \big((X_iX \setminus XX_i)X^{-1} \cup X^{-1}(XX_i \setminus X_iX)\big) = X_{i+1}.$$

□

5 Esimerkkejä

Käymme esimerkkeinä läpi muutamia rationaalisia kieliä ja etsimme niille sentralisaattorit. Osa kielistä on äärellisiä ja osa äärettömiä. Tarkastelemme samalla, onko sentralisaattori X^+, S vai jotakin muuta muotoa. Esimerkkejä tarkastellaan eri tavoin. Osa esimerkeistä käydään läpi tarkasti askel askeleelta, esittäen kukin välivaiheena saatava kieli rationaalisena lausekkeena, ja osan yhteydessä tyydytään esittämään Grail+:n laskemat välivaiheet äärellisinä automaatteina.

Ensiksi todistamme erään ominaisuuden, joka on kaikilla niillä kielillä X, joilla $X^+ \neq S$.

Lause 5.1. *Olkoon X aakkoston Σ kieli. Jos $X^+ \neq S$, niin kullakin kielen X sanalla w on voimassa jompikumpi seuraavista väitteistä.*

a) $v = uw$, joillakin sanoilla $u \in \Sigma^+$, $v \in X$.

b) $w = uv$, joillakin sanoilla $u \in \Sigma^+$, $v \in X$.

Sana w on siis jonkin toisen kielen X sanan v aito suffiksi, taikka jokin toinen kielen X sana v on sanan w aito suffiksi.

Sama tulos pätee myös, jos suffiksien sijasta tarkastellaan prefiksejä.

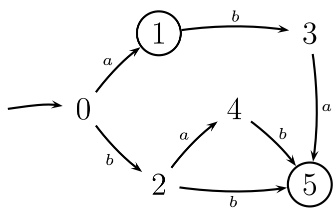
Todistus. Olkoon w mielivaltainen kielen X sana ja $t \in S \setminus X^+$. Silloin $tw \in XX^+$, eli sana tw voidaan jakaa jollain tavalla kieleen X kuuluviin tekijöihin. Koska $t \notin X^+$, ei sana tw jakaudu tekijöihin sanojen t ja w välistä tässä tekijöihin jaossa. Sanan tw oikeanpuoleisin kieleen X kuuluva tekijä tässä tekijöihinjaossa on siis joko sanan w jokin aito suffiksi, taikka jokin sana uw siten, että $t = su$ ja $s \in X^+$.

Prefiksejä koskeva väite todistuu aivan samoin. □

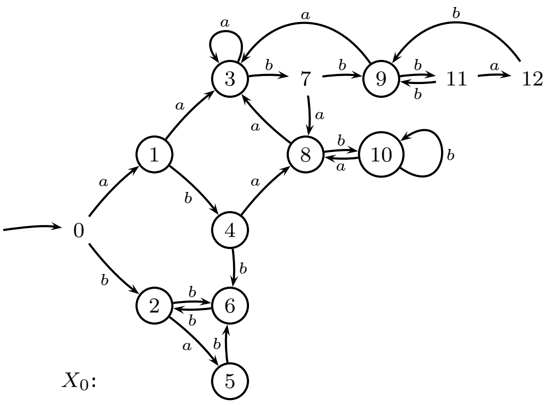
5.1 $X^+ = S = \mathcal{C}(X)$

Tarkastelemme kieltä $X = \{a, bb, aba, bab\}$ esimerkkinä äärellisestä kielestä, jolla $X^+ = S = \mathcal{C}(X)$. Lauseen 5.1 perusteella selvästikin $X^+ = S$, sillä kielestä X ei löydy sanaa, jonka aito suffiksi sana bb olisi, eikä sanaa, joka olisi sanan bb aito suffiksi.

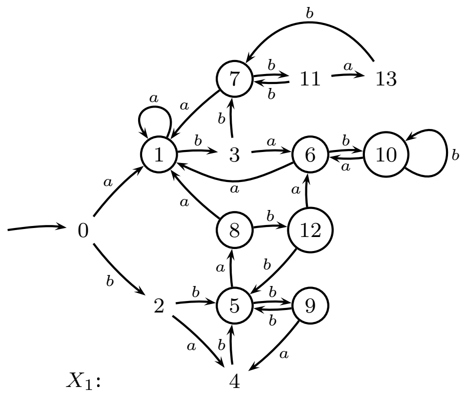
Kielen sentralisaattori etsitään kiintopistemetodilla. Sentralisaattori löytyy jo kolmannella askeleella. Iteroinnin askelia X_i vastaavat, Grail+:lla lasketut, minimoidut DFA:t voidaan esittää seuraavanlaisina graafeina. Graafeissa automaatin tilat on numeroitu alkaen alkutilasta 0 ja lopetustilat on merkitty ympyröimällä ne. Yhtäsuuruus $X_3 = X^+$ saadaan laskemalla samoin Grail+:lla kielen X^+ minimoitu automaatti ja vertaamalla sitä kielen X_3 automaattiin.



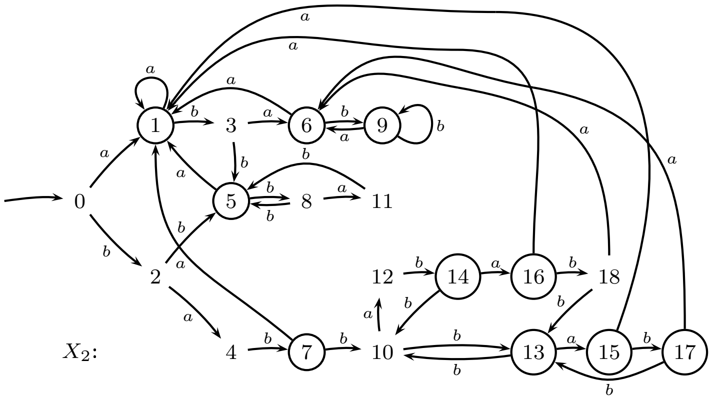
$X = \{a, bb, aba, bab\}$



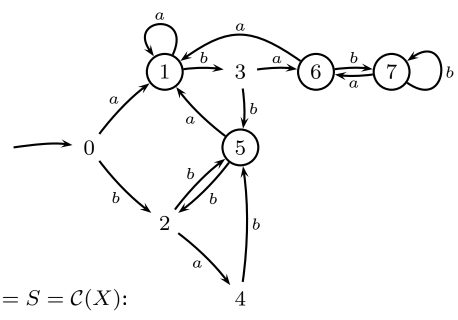
X_0:



X_1:



X_2:



$X_3 = X^+ = S = \mathcal{C}(X)$:

Automaatteja tarkastelemalla voidaan havaita, miten niihin muodostuu iteraation jossain vaiheessa kuvioita, jotka säilyvät seuraavissa askelissa, aina lopulliseen sentralisaattoriin saakka. Esimerkiksi jo kielen X_0 automaatissa oleva tilojen 3,7,8,9,10,11 ja 12 muodostama kuvio säilyy seuraavissakin askelissa. Vain tilojen numerot vaihtuvat. Kielen X_1 automaatissa näitä tiloja vastaavat samassa järjestyksessä tilat 1,3,6,7,10,11 ja 13, kielen X_2 automaatissa tilat 1,3,6,5,9,8 ja 11 sekä lopulta kielen X_3 automaatissa tilat 1,3,6,5,7,2 ja 4. Iteraation askelia vastaavien automaattien tilamäärät yleensä kasvavat kun siirrytään seuraavaan askeleeseen. Tämän voidaan ajatella johtuvan siitä, että uusien tilojen avulla seuraava askel esittää aina edellistä tarkempaa likiarvoa etsitystä sentralisaattorista lisäämällä tarkentavia sääntöjä. Toisaalta, kun automaattiin muodostuu useita keskenään vastaavanlaisia kuvioita, näiden vastintilat usein samaistuvat ja tämä vähentää automaatin tilojen määrää. Jos sentralisaattori saavutetaan jollakin iteraation askeleella, tällöin tilamäärä usein putoaa melko pieneen määrään verrattuna edelliseen askeleeseen, kuten kävi tässä tapauksessa.

5.2 $X^+ \subset S = \mathcal{C}(X)$

Äärellisen esimerkin löytäminen kieleksi X, jolla $X^+ \subset S$, osoittautuu vaikeaksi. Jos tämän tyyppisiä äärellisiä kieliä on olemassa, saattavat ne olla melko harvinaisia. Esimerkiksi valitsemme siis aakkoston $\Sigma = \{a, b\}$ äärettömän rationaalisen kielen $X = \Sigma a \Sigma^* a + b \Sigma^* b \Sigma$. Kun tälle kielelle suoritetaan kiintopistemetodia, löydetään sentralisaattori kolmella askeleella. Kuten taulukosta 1 käy ilmi, kieliä X_i vastaavien automaattien tilamäärä ei viimeisessä askeleessa putoa, kuten edellisessä esimerkissä.

	lopput.	tiloja	siirtymiä
X	5	13	25
X_0	3	5	9
X_1	7	18	35
X_2	9	22	43
X_3	11	26	51
X_4	11	26	51
S	11	26	51
X^+	6	14	27

Taulukko 1: Kieltä $X = \Sigma a \Sigma^* a + b \Sigma^* b \Sigma$, kiintopistemetodin askelia sekä kieliä S ja X^+ vastaavien automaattien lopputilojen, tilojen ja siirtymien määrät.

Etsitään sentralisaattori kiintopistemetodilla. Ensimmäiseksi etsitään iteroinnin lähtökohta, eli kieli $X_0 = \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+)$. Tiedetään, että sisältyminen $X^+ \subseteq X_0$ on aina voimassa, joten muodostetaan kieli $X_0 \setminus X^+$. Kielen X^+ lyhimpien sanojen pituus on 3. Tätä lyhyemmistä sanoista kieleen X_0 kuuluvat sanat a, b, aa, ba ja bb.

Pidemmät sanat voidaan jakaa alkukirjaimen mukaan kahteen joukkoon. Olkoon $w \in X_0 \setminus X^+$. Oletetaan ensin, että $w \in a\Sigma\Sigma^+$. Nyt $w \notin \mathrm{Pref}(b\Sigma b\Sigma^* X^*)$, koska sanan w ensimmäinen kirjain on a. Siispä $w \in \mathrm{Pref}(\Sigma a\Sigma^* aX^*)$ ja siis $w \in aa\Sigma^+$. Toisaalta, koska $w \notin X$, on oltava $w \in aa\Sigma^* b$. Symmetrisesti nyt, koska $w \notin \mathrm{Suf}(X^*\Sigma a\Sigma^* a)$, on oltava voimassa $w \in \mathrm{Suf}(X^* b\Sigma^* b\Sigma)$. Näin siis $w \in aa\Sigma^* bb$. Lisäksi, koska $aa\Sigma^* ab\Sigma^* bb \subseteq X^2$, on sanan w oltava kielestä aab^*a^*bb, joka selvästi sisältyy kokonaisuudessaan kieleen $X_0 \setminus X^+$.

Jos taas oletetaan, että $w \in b\Sigma\Sigma^+$, voidaan sanasta w päätellä seuraavaa. Koska $b\Sigma^* b\Sigma \subseteq X^+$, nähdään, että $w \in b\Sigma^* a\Sigma$. Edelleen, koska $w \notin \mathrm{Suf}(X^* b\Sigma^* b\Sigma)$, on oltava $w \in \mathrm{Suf}(X^*\Sigma a\Sigma^* a)$, joten $w \in b\Sigma^* aa$. Nyt, koska $baa + ba\Sigma^* aa \subseteq X^+$, on sanan w toisen kirjaimen oltava b, eli $w \in bb\Sigma^* aa$. Seuraavaksi nähdään, että kielen $bb\Sigma^* aa$ sana kuuluu kieleen X^+ tarkalleen silloin, jos se kuuluu kieleen $b\Sigma^* b\Sigma\Sigma a\Sigma^* a$, sillä selvästikin sanan alkuosan on oltava muotoa $b\Sigma^* b\Sigma$ ja loppuosan muotoa $\Sigma a\Sigma^* a$. Nyt etsimme kielen $bb\Sigma^* aa \setminus b\Sigma^* b\Sigma\Sigma a\Sigma^* a$. Tutkitaan sanojen bb ja aa välissä olevaa osuutta Σ^* kolmen kirjaimen ryhmissä. Alun bb:n jälkeen tulevan kolmen kirjaimen sanan on oltava jokin sanoista aab, abb tai bab, sillä $bb\Sigma\Sigma a\Sigma^* aa + bbbbb\Sigma^* aa \subseteq b\Sigma^* b\Sigma\Sigma a\Sigma^* a$. Vastaavasti sanan aab jälkeen voi olla mikä tahansa näistä kolmesta sanasta, sanan abb jälkeen vain abb ja sanan bab jälkeen vain sana bab. Toisinsanoen

$$w \in bb(aab)^*(abb)^*(1 + \Sigma + \Sigma^2)aa + bb(aab)^*(bab)^*(1 + \Sigma + \Sigma^2)aa.$$

Käydään näistä sanoista vielä läpi kaikki eri mahdollisuudet. Kieleen X^+ sisältyvistä kielistä on alleviivattu muotoa $b\Sigma\Sigma a$ oleva osuus.

$$
\begin{aligned}
bb(aab)^*(abb)^*aa &\subseteq X_0 \setminus X^+ \\
bb(aab)^*(a\underline{bb)^+\Sigma a}a &\subseteq X^+ \\
bb(aab)^*\Sigma aa &\subseteq X_0 \setminus X^+ \\
bb(aab)^*(ab\underline{b)^+\Sigma\Sigma a}a &\subseteq X^+ \\
bb(aa\underline{b)^+\Sigma\Sigma a}a &\subseteq X^+ \\
b\underline{b\Sigma\Sigma a}a &\subseteq X^+ \\
bb(aab)^*(\underline{bab)^+a}a &\subseteq X^+ \\
bb(aab)^*(\underline{bab)^+a}aa &\subseteq X^+ \\
bb(aab)^*(bab)^+baa &\subseteq X_0 \setminus X^+ \\
bb(aab)^*(ba\underline{b)^+\Sigma\Sigma a}a &\subseteq X^+
\end{aligned}
$$

Näin siis

$$
\begin{aligned}
b\Sigma\Sigma \cap (X_0 \setminus X^+) &= bb(aab)^*(abb)^*aa + bb(aab)^*\Sigma aa + bb(aab)^*(bab)^+baa \\
&= bb(aab)^*(abb)^*aa + bb(aab)^*a(abb)^*aa + bb(aab)^*b(abb)^*aa
\end{aligned}
$$

ja iteroinnin lähtökohdaksi saatu kieli on kokonaisuudessaan

$$
\begin{aligned}
X_0 &= \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+) \\
&= X^+ + aab^*a^*bb \\
&\quad + bb(aab)^*(abb)^*aa \\
&\quad + bb(aab)^*a(abb)^*aa \\
&\quad + bbb(abb)^*aa \\
&\quad + a + b + aa + ba + bb.
\end{aligned}
$$

Iteroinnin askelet etenevät seuraavasti. Kielestä X_0 poistuvat seuraavat sanat annetuilla syillä siirryttäessä kieleen X_1.

$\underline{a}$	$a \cdot bbb \in X_0X$, mutta $abbb \notin XX_0$.
$\underline{b}$	$aaa \cdot b \in XX_0$, mutta $aaab \notin X_0X$.
$\underline{aa}$	$babba \cdot aa \in XX_0$, mutta $babbaaa \notin X_0X$.
$\underline{bb}$	$bb \cdot baaba \in X_0X$, mutta $bbbaaba \notin XX_0$.
$\underline{aaa^n bb}$	$aaa \cdot aaa^n bb \in XX_0$, mutta $aaaaaa^n bb \notin X_0X$.
	$(n \geq 0)$
$\underline{aab^n bb}$	$aab^n bb \cdot bbb \in X_0X$, mutta $aab^n bbbbb \notin XX_0$.
	$(n \geq 0)$
$\underline{bb(abb)^n aa}$	$aabba \cdot bb(abb)^n aa \in XX_0$, mutta $aab(bab)^i \cdot \overbrace{(bab)^j baa}^{\in X} \notin X_0X$.
	$(n \geq 0, i + j = n + 1)$
$\underline{bb(aab)^n aa}$	$bb(aab)^n aa \cdot baabb \in X_0X$, mutta $bba(aba)^{n+1} abb \notin XX_0$.
	$(n \geq 0)$

Kieleksi X_1 saadaan siis

$$
\begin{aligned}
X_1 \;=\; & X^+ + aab^+a^+bb \\
& +bb(aab)^+(abb)^+aa \\
& +bb(aab)^*a(abb)^*aa \\
& +bbb(abb)^*aa \\
& +ba.
\end{aligned}
$$

Vastaavalla tavalla tavalla etsitään sanat, jotka poistuvat kielestä X_1, kun siirrytään seuraavassa iteraation askeleessa kieleen X_2.

$\underline{ba}$	$ba \cdot aaa \in X_1X$, mutta $baaaa \notin XX_1$.
$\underline{bbb(abb)^n aa}$	$aaa \cdot bbb(abb)^n aa \in XX_1$, mutta $aaabbb(abb)^n aa \notin X_1X$.
	$(n \geq 0)$
$\underline{bb(aab)^n aaa}$	$bb(aab)^n aaa \cdot bbb \in X_1X$, mutta $bb(aab)^n aaabbb \notin XX_1$.
	$(n \geq 0)$
$\underline{bbaabb(abb)^n aa}$	$aaa \cdot bbaabb(abb)^n aa \in XX_1$, mutta $aaabbaabb(abb)^n \notin X_1X$.
	$(n \geq 0)$
$\underline{bb(aab)^n aabbaa}$	$bb(aab)^n aabbaa \cdot bbb \in X_1X$, mutta $bb(aab)^n aabbaabbb \notin XX_1$.
	$(n \geq 0)$

Näin iteraation seuraava askel on kieli

$$\begin{aligned} X_2 &= X^+ + aab^+a^+bb \\ &\quad + bb(aab)^+(abb)^+aa \\ &\quad + bb(aab)^+aabb(abb)^+aa. \end{aligned}$$

Lopuksi siirryttäessä kielestä X_2 kieleen X_3, poistuvat loputkin sentralisaattoriin kuulumattomat sanat.

$\underline{aab^n abb}$ $aab^n abb \cdot aaa \in X_2X$, mutta $aab^n bbaaa \notin XX_2$, sillä vasempana, kieleen X kuuluvana tekijänä, olisi jokin a-loppuisista sanoista $aab^n a, aab^n abba$ tai $aab^n abbaa$, mutta näitä vastaavilla oikeilla tekijöillä $bbaaa, aa, a \notin X_2$.
$(n \geq 0)$

$\underline{aab^n aabb}$ $aab^n aabb \cdot aaa \in X_2X$, mutta $aab^n aabbaaa \notin XX_2$ samoin kuin edellä, sillä $abbaaa, bbaaa, aa, a \notin X_2$.
$(n \geq 0)$

$\underline{aaba^n bb}$ $bbb \cdot aaba^n bb \in XX_2$, mutta $bbbaaba^n bb \notin X_2X$ edelleen vastaavasti kuin edellä, sillä nyt oikeana, kieleen X kuuluvana tekijänä, olisi jokin b-alkuisista sanoista $ba^n bb, baaba^n bb$ tai $bbaaba^n bb$, mutta näitä vastaavat vasemmat tekijät $bbbaa, bb$ ja b eivät kuulu kieleen X_2.
$(n \geq 0)$

$\underline{aabba^n bb}$ $bbb \cdot aabba^n bb \in XX_2$, mutta $bbbaabba^n bb \notin X_2X$ täsmälleen samoin perusteluin kuin muissakin kohdissa.
$(n \geq 0)$

Näin kieleksi X_3 muotoutuu kieli

$$\begin{aligned} X_3 &= X^+ + aabbb^+aaa^+bb \\ &\quad + bb(aab)^+(abb)^+aa \\ &\quad + bb(aab)^+aabb(abb)^+aa. \end{aligned}$$

Kiintopistemetodi pysähtyy tähän kieleen, sillä $XX_3 = X_3X$. Tämän näkeminen suoraan rationaalisesta lausekkeesta ei ole kuitenkaan itsestäänselvää. Tämä nähdään kuitenkin esimerkiksi seurauksena siitä, että

$$
\begin{aligned}
XX_3 &= X\left(X^+ + aabbb^+aaa^+bb + bb(aab)^+(abb)^+aa + bb(aab)^+aabb(abb)^+aa\right) \\
&= XX^+ \\
&\quad +\Sigma a\Sigma^* a \cdot aabbb^+aaa^+bb \\
&\quad +b\Sigma^* b\Sigma \cdot aabbb^+aaa^+bb \\
&\quad +\Sigma a\Sigma^* a \cdot bb(aab)^+(abb)^+aa \\
&\quad +b\Sigma^* b\Sigma \cdot bb(aab)^+(abb)^+aa \\
&\quad +\Sigma a\Sigma^* a \cdot bb(aab)^+aabb(abb)^+aa \\
&\quad +b\Sigma^* b\Sigma \cdot bb(aab)^+aabb(abb)^+aa \\
&= XX^+ \\
&\quad +\Sigma a\Sigma^* aaa \cdot bbb^+aaa^+bb \\
&\quad +b\Sigma^* b\Sigma aabb \cdot b^+aaa^+bb \\
&\quad +\Sigma a\Sigma^* abbaa \cdot b(aab)^*(abb)^+aa \\
&\quad +b\Sigma^* b\Sigma bb \cdot (aab)^+(abb)^+aa \\
&\quad +\Sigma a\Sigma^* abbaa \cdot b(aab)^*aabb(abb)^+aa \\
&\quad +b\Sigma^* b\Sigma bb \cdot (aab)^+aabb(abb)^+aa \\
&\subseteq XX^+
\end{aligned}
$$

ja vastaavin perusteluin $X_3X \subseteq XX^+$. Tällöin siis $X_3 \subseteq S$. Koska kuitenkin aina $S \subseteq \mathcal{C}(X) \subseteq X_i$, on nyt oltava voimassa $X_3 = S = \mathcal{C}(X)$. Lisäksi selvästikin $X^+ \neq X_3$, sillä esimerkiksi $aabbbaaabb \in X_3 \setminus X^+$, joten $X^+ \subset S = \mathcal{C}(X)$.

5.3 $X^+ = S \subset \mathcal{C}(X)$

Eräs hyvin yksinkertainen esimerkki äärellisestä kielestä, jolla $X^+ = S$, mutta sentralisaattori on jotakin muuta, on aakkoston $\Sigma = \{a\}$ kieli $X = \{aa\}$. Koko puoliryhmä $\Sigma^+ = a^+$ selvästikin kommutoi kielen X kanssa, joten se on kielen X sentralisaattori. Kieli $X^+ = (aa)^+$ on kaikkien parillisenpituisten a-kirjainten jonojen kieli. Koska parillisen ja parittoman luvun summa on pariton, sisältää kieli S vain parillisen mittaisia sanoja. Siispä $S = X^+$. Näin siis $X^+ = S \subset \mathcal{C}(X)$.

Sanotaan, että kieli $\rho(X)$ on kielen X *juuri*, jos $X = \rho(X)^n$ jollakin luvulla n. Jos kielen X ainoa juuri on X itse, sanotaan kieltä X *primitiiviseksi*. Tällöin on yleisemmin voimassa, että jos kieli X ei ole primitiivinen, kuten edellä olevassa esimerkissä, on aina

$$X^+ \neq \mathcal{C}(X),$$

sillä aina

$$\rho(X)^+ \subseteq \mathcal{C}(X),$$

koska kieli $\rho(X)^+$ kommutoi luonnollisestikin kielen $X = \rho(X)^n$ kanssa.

Toiseksi esimerkiksi kielestä, jolla $X^+ = S \subset \mathcal{C}(X)$, valitsemme jo esimerkissä 3.7 mainitun kielen $X = \{a, ab, ba, bb\}$. Se, että $S = X^+$, nähdään joko lauseesta 5.1 taikka laskemalla molemmat kielet Grail+:lla ja vertaamalla niitä.

Kielen X prefiksien ja suffiksien joukot ovat samat

$$\mathrm{Pref}(X) = \mathrm{Suf}(X) = \{a, b, ab, ba, bb\}.$$

Etsitään kieli X_0. Ensimmäiseksi teemme huomion

$$\begin{aligned}\left(\Sigma^2\right)^+ &= \{a\,a, ab, ba, bb\}^+ \subseteq X^+ \subseteq \mathrm{Pref}(X^+)\\ \left(\Sigma^2\right)^* \Sigma &\subseteq X^*\Sigma \subseteq X^* \,\mathrm{Pref}(X) = \mathrm{Pref}(X^+)\end{aligned}$$

Näin siis $\Sigma^+ = (\Sigma^2)^+ + (\Sigma^2)^*\Sigma \subseteq \mathrm{Pref}(X^+)$. Päinvastainen sisältyminen $\mathrm{Pref}(X^+) \subseteq \Sigma^+$ on itsestäänselvää. Vastaavasti nähdään, että $\mathrm{Suf}(X^+) = \Sigma^+$. Näin siis

$$X_0 = \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+) = \Sigma^+.$$

Sentralisaattori löydetään seuraavasti. Väitetään, että sentralisaattori on kieli $Z = \Sigma^+ \setminus \{b\}$. Tämä kieli voidaan jakaa kahteen osaan sen mukaan, kummalla kirjaimella sanat alkavat. Vastaavasti jako voidaan tehdä viimeisen kirjaimen mukaan. Siis

$$Z = \Sigma^+ \setminus \{b\} = a\Sigma^* + b\Sigma^+ = \Sigma^* a + \Sigma^+ b.$$

Nyt Z kommutoi kielen X kanssa, sillä

$$ZX = a\Sigma^* X + b\Sigma^+ X = a \cdot \Sigma^* X + b\Sigma \cdot \Sigma^* X = (a + b\Sigma)\Sigma^* X \subseteq XZ$$

ja

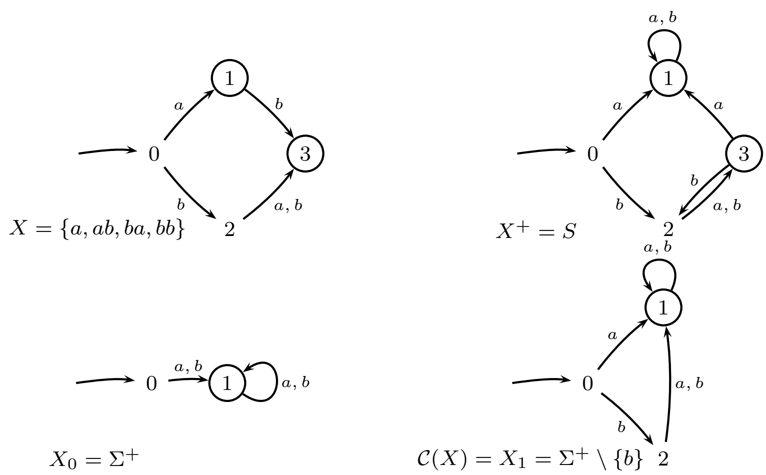
$$XZ = X\Sigma^* a + X\Sigma^+ b = X\Sigma^* \cdot a + X\Sigma^* \cdot \Sigma b = X\Sigma^*(a + \Sigma b) \subseteq ZX.$$

Sentralisaattori ei ole koko $X_0 = \Sigma^+$, sillä sana b ei kuulu sentralisaattoriin, koska $ba \in X_0 X$, mutta $ba \notin XX_0$. Siispä $\mathcal{C}(X) = Z = \Sigma^+ \setminus \{b\}$. Erityisesti $X^+ \neq \mathcal{C}(X)$, sillä esimerkiksi $bbb \in \mathcal{C}(X)$, mutta $bbb \notin X^+$.

Kielellä X siis

$$X^+ = S \subset \mathcal{C}(X).$$

Lisäksi $\mathcal{C}_*(X) \neq \mathcal{C}_+(X) \cup \{1\}$, sillä myös $b \in \mathcal{C}_*(X)$ ja siis $\mathcal{C}_*(X) = \Sigma^*$.



Kuva 1: Minimoidut automaatit kielille X ja X^+ sekä kiintopistemetodin askelille X_0 ja $X_1 = \mathcal{C}(X)$.

5.4 $X^+ \subset S \subset \mathcal{C}(X)$

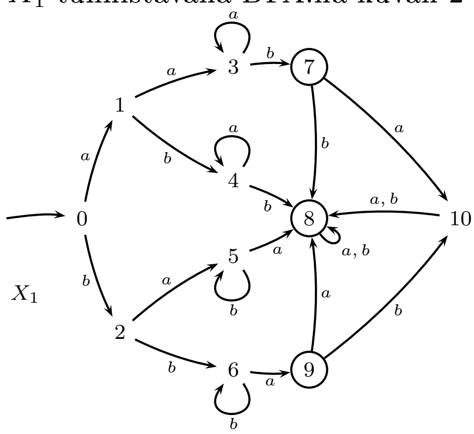
Käsitellään vielä esimerkki, jossa kielellä X ovat voimassa aidot sisältymiset $X^+ \subset S \subset \mathcal{C}(X)$. Valitaan

$$X = a\Sigma^+ b + b\Sigma^+ a.$$

Nyt

$$X_0 = \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+) = (a\Sigma^* + b\Sigma^*) \cap (\Sigma^* a + \Sigma^* b) = \Sigma^+ \cap \Sigma^+ = \Sigma^+.$$

Grail+ antaa kielen X_1 tunnistavana DFA:na kuvan 2 automaatin. Automaatin



Kuva 2: Grail+:lla laskettu kielen X_1 tunnistava automaatti.

antama kieli voidaan helpohkosti esittää rationaalisena lausekkeena

$$\begin{aligned} X_1 &= aaa^*b + aaa^*bb\Sigma^* + aaa^*ba\Sigma^+ + aba^*b\Sigma^* \\ &+ bbb^*a + bbb^*aa\Sigma^* + bbb^*ab\Sigma^+ + bab^*a\Sigma^*. \end{aligned} \tag{9}$$

Sama kieli voidaan esittää myöskin komplementtinsa avulla muodossa

$$X_1 = \Sigma^+ \setminus (a^+ + b^+ + aba^* + bab^* + aa^+ba + bb^+ab). \tag{10}$$

Tämäkin nähdään melko helposti samasta automaatista. Kieli X_1 saadaan kiintopistemetodin välivaiheena, koska

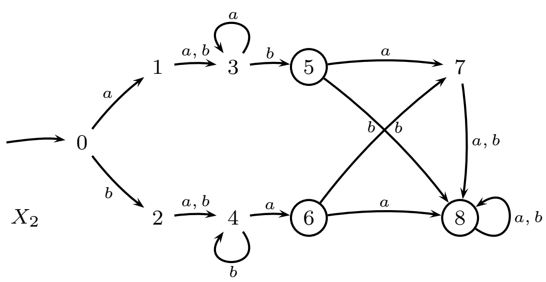
$$\begin{aligned} a^+ \cdot aab \cap XX_0 &= \emptyset \\ b^+ \cdot bba \cap XX_0 &= \emptyset \\ aba^* \cdot aab \cap XX_0 &= \emptyset \\ bab^* \cdot bba \cap XX_0 &= \emptyset \\ baa \cdot aa^+ba \cap X_0X &= \emptyset \\ abb \cdot bb^+ab \cap X_0X &= \emptyset. \end{aligned}$$

Näin siis

$$a^+ + b^+ + aba^* + bab^* + aa^+ba + bb^+ab \subseteq X^{-1}(XX_0 \Delta X_0X) \cup (X_0X \Delta XX_0)X^{-1}.$$

Tästä ei tietenkään nähdä, ovatko nämä kaikki kielestä X_0 poistettavat sanat, mutta koska sentralisaattori saavutetaan jo seuraavalla askeleella, emme tutki tätä vaihetta tarkemmin.

Seuraavana välivaiheena Grail+ antaa kuvan 3 mukaisen automaatin. Rationaalisena lausekkeena kieli X_2 voidaan ilmaista muodossa



Kuva 3: Grail+:lla laskettu kielen X_2 tunnistava automaatti.

$$\begin{aligned} X_2 &= a\Sigma a^*b + a\Sigma a^*ba\Sigma^+ + a\Sigma a^*bb\Sigma^* \\ &+ b\Sigma b^*a + b\Sigma b^*ab\Sigma^+ + b\Sigma b^*aa\Sigma^*. \end{aligned} \qquad (11)$$

Myös tämä kieli voidaan esittää melko yksinkertaisesti komplementtinsa avulla, jolloin käy paremmin ilmi, mitkä osuudet kielestä X_1 poistetaan iteraation tässä askeleessa.

$$\begin{aligned} X_2 &= \Sigma^+ \setminus (a^+ + b^+ + aba^* + bab^* + aa^+ba + bb^+ab + aba^*ba + bab^*ab) \\ &= X_1 \setminus (aba^*ba + bab^*ab). \end{aligned} \tag{12}$$

Perusteluksi kielten aba^*ba ja bab^*ab poistamiselle kielestä X_1 riittää nähdä, että

$$\begin{aligned} aba^*ba \cdot baa \cap XX_1 &= \emptyset \quad \text{ja} \\ bab^*ab \cdot abb \cap XX_1 &= \emptyset. \end{aligned}$$

Nämä pitävät molemmat paikkansa, sillä jos kielistä aba^*babaa ja bab^*ababb erotetaan vasemmalta puolelta kieleen X kuuluvat osuudet aba^*b ja bab^*a, kuuluvat jäljelle jäävät osuudet $abaa$ ja $babb$ kielen X_1 komplementtiin, kuten käy ilmi kaavasta (10).

Näin saavutettu kieli X_2 on sentralisaattori $\mathcal{C}(X)$, koska $XX_2 = X_2X$. Tämä voidaan varmistaa laskemalla X_2X auki käyttämällä kielen X määritelmää ja kielelle X_2 saatua lauseketta (11). Näin saadun lausekkeen

$$\begin{aligned} X_2X &= a\Sigma a^*b \cdot a\Sigma^+b + a\Sigma a^*b \cdot b\Sigma^+a + a\Sigma a^*ba\Sigma^+ \cdot a\Sigma^+b + a\Sigma a^*ba\Sigma^+ \cdot b\Sigma^+a \\ &+ a\Sigma a^*bb\Sigma^* \cdot a\Sigma^+b + a\Sigma a^*bb\Sigma^* \cdot b\Sigma^+a + b\Sigma b^*a \cdot a\Sigma^+b + b\Sigma b^*a \cdot b\Sigma^+a \\ &+ b\Sigma b^*ab\Sigma^+ \cdot a\Sigma^+b + b\Sigma b^*ab\Sigma^+ \cdot b\Sigma^+a + b\Sigma b^*aa\Sigma^* \cdot a\Sigma^+b + b\Sigma b^*aa\Sigma^* \cdot b\Sigma^+a \end{aligned}$$

kunkin termin voidaan osoittaa sisältyvän kieleen XX_2. Suurin osa termeistä, esimerkiksi $a\Sigma a^*b \cdot a\Sigma^+b$, sisältyy selvästi kieleen $XX \subseteq XX_2$. Toisaalta esimerkiksi $a\Sigma a^*bb\Sigma^* \cdot a\Sigma^+b \subseteq XX_2$, sillä $a\Sigma a^*bb \cdot a\Sigma^+b \subseteq XX$ ja $a\Sigma a^*b \cdot b\Sigma^+a\Sigma^+b \subseteq XX_2$, koska kaavan (12) mukaan $b\Sigma^+a\Sigma^+b \cap (\Sigma^+ \setminus X_2) = \emptyset$. Näin $X_2X \subseteq XX_2$. Vastaavasti, koska $X = X^R$ ja $X_2 = X_2^R$, saadaan

$$XX_2 = (X_2^R X^R)^R = (X_2X)^R \subseteq (XX_2)^R = X_2^R X^R = X_2X.$$

Lopulta sisältymiset $X^+ \subset S \subset \mathcal{C}(X)$ saadaan valitsemalla sopivat esimerkkisanat. Esimerkiksi

$$abbbaX = abbba \cdot a\Sigma^+b + abbba \cdot b\Sigma^+a = abbb \cdot aa\Sigma^+b + abb \cdot bab\Sigma^+a \subseteq XX^+$$

ja

$$Xabbba = a\Sigma^+b \cdot abbba + b\Sigma^+a \cdot abbba = a\Sigma^+bab \cdot bba + b\Sigma^+aa \cdot bbba \subseteq XX^+,$$

joten $abbba \in S$, mutta selvästikin $abbba \notin X^+$. Toisaalta $abbaa \in \mathcal{C}(X)$, koska kuvan 3 automaatti tunnistaa tämän sanan, mutta $abbaa \notin S$, sillä esimerkiksi $abbaa \cdot baa \notin XX^+$.

6 Sentralisaattori raja-arvona

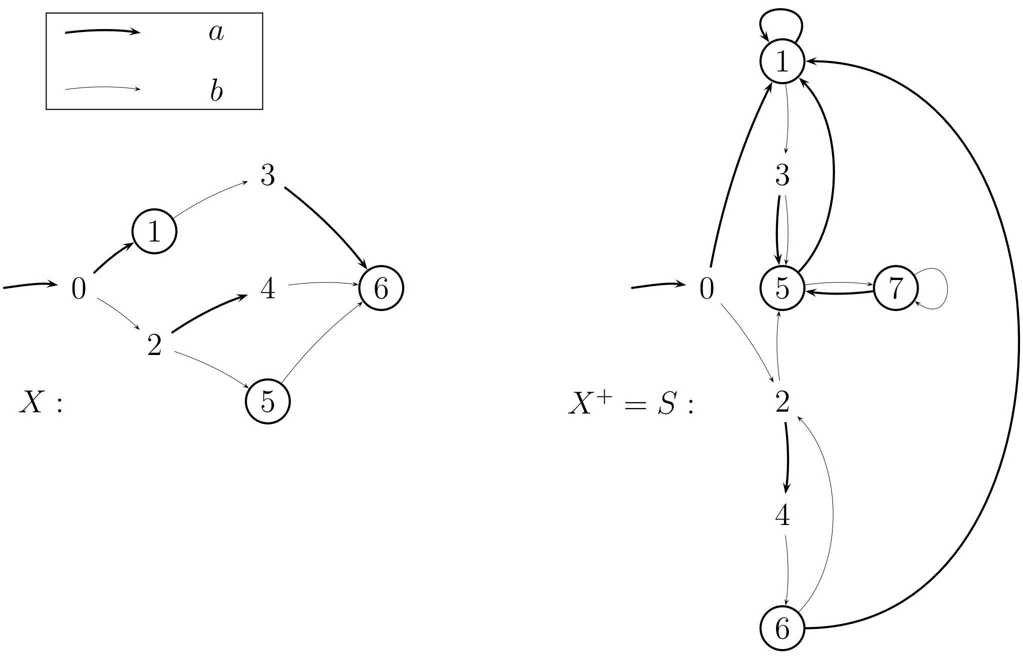
Seuraavaksi esitellään ja analysoidaan eräs esimerkki tapauksesta, jossa kiintopistemetodi johtaa pysähtymättömään laskentaa. Sentralisaattori saadaan siis vain menetelmän raja-arvona. Esimerkki on erityisen mielenkiintoinen siksi, että se osoittaa, että metodi voi johtaa pysähtymättömään laskentaa, vaikka annettu kieli X on äärellinen. Esimerkin kielessä on vain viisi sanaa. Toinen mielenkiintoinen huomio tapauksesta on, että vaikka kiitopistemetodi ei onnistuu tuottamaan ratkaisua äärellisellä askelmäärällä, on kielen sentralisaattori lopulta kuitenkin yksinkertaisesti vain X^+.

6.1 Tapauksen esittely

Tässä tapauksessa tarkasteltavana kielenä on aakkoston $\Sigma = \{a, b\}$ äärellinen kieli $X = \{a, bb, aba, bab, bbb\}$.

Jos tarkastellaan kysymystä, onko $\mathcal{C}(X) = X^+$, voidaan yrittää soveltaa kappaleen 4 lopussa esiteltyä kahta menetelmää. Ensin vertaillaan kieliä X^+ ja S. Jos nämä kielet poikkeavat toisistaan, on välttämättä $\mathcal{C}(X) \neq X^+$. Kuitenkin, jos kieli S lasketaan tietokoneella, käyttäen lauseen 3.8 kaavaa, nähdään, että $S = X^+$. Sama tulos saadaan käyttämällä lausetta 5.1, sillä sana bab ei ole minkään muun kielen X sanan aito suffiksi, eikä toisaalta mikään sanan bab aito suffiksi kuulu kieleen X. Tästä tarkastelusta ei siis ole suurta apua tämän kielen yhteydessä.

Toisena keinona voidaan tutkia kielen X haaroittumispisteitä ja näistä eri-



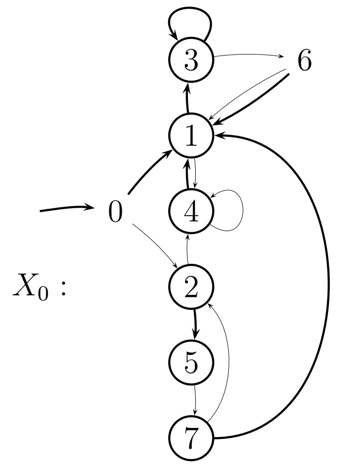
Kuva 4: Kielet X ja X^+ tunnistavat äärelliset automaatit

tyisesti kriittisiä pisteitä. Lauseen 4.3 mukaan, jos äärellisen kielen X kriittisten pisteiden joukko C on tyhjä, on kielen sentralisaattori $\mathcal{C}(X) = X^+$. Tässä tarkasteltavan kielen X kriittisten pisteiden joukko on kuitenkin epätyhjä. Esimerkiksi sana b on kielen X haaroittumispiste, sillä $b \in \mathrm{Pref}(X^+)$ ja $b\,\mathrm{Pref}_1(X) = \{ba, bb\} \subseteq \mathrm{Pref}(X^+)$. Sana b on lisäksi kriittinen, sillä $b \notin X^+$. Näin ollen $C \neq \emptyset$. Siispä kriittisten pisteiden tutkiminen ei paljasta, että $\mathcal{C}(X) = X^+$.

Jos kielen X sentralisaattoria lähdetään etsimään kiintopistemetodilla, huomataan hyvin pian, että iterointi ei näytä tuottavan tulosta, vaan kieliä X_i esittävien minimoitujen determinististen äärellisten automaattien tilamäärä kasvaa tietyllä nopeudella ja tietyn kuvion mukaisesti. Kieli osoittautuu siis näillä menetelmillä varsin hankalasti käsiteltäväksi.

6.2 Kiintopistemetodi Grail+:lla

Muodostetaan Grail+:lla kieltä X edustava automaatti sekä suoritetaan sille kiintopistemetodia usean iteraation verran. Tarkastellaan iteraatioaskelina saatuja minimoituja automaatteja ja tehdään havaintoja niistä löytyvistä säännöllisyyksistä. Iteroinnin lähtökohdaksi saadaan kuvan 5 mukaisen automaatin esittämä kieli X_0.



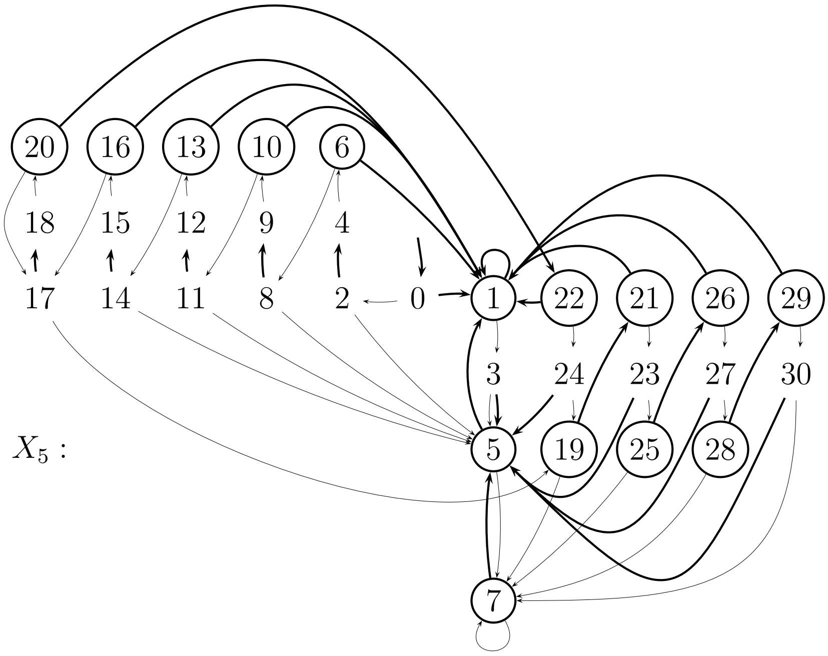
Kuva 5: Iteraation aloituskielen X_0 tunnistava automaatti.

Kielelle X kiintopistemetodilla laskettuja iteraatioaskelia vastaavien automaattien tilamääriä, lopputilojen määriä ja siirtymien määriä on luetteloitu taulukossa 2. Näistä luvuista nähdään, että jo muutaman iteraatioaskeleen jälkeen automaattien kasvu vakiintuu. Kussakin askeleessa tilojen määrä lisääntyy kuudella, lopputilojen kolmella sekä siirtymien määrä yhdellätoista.

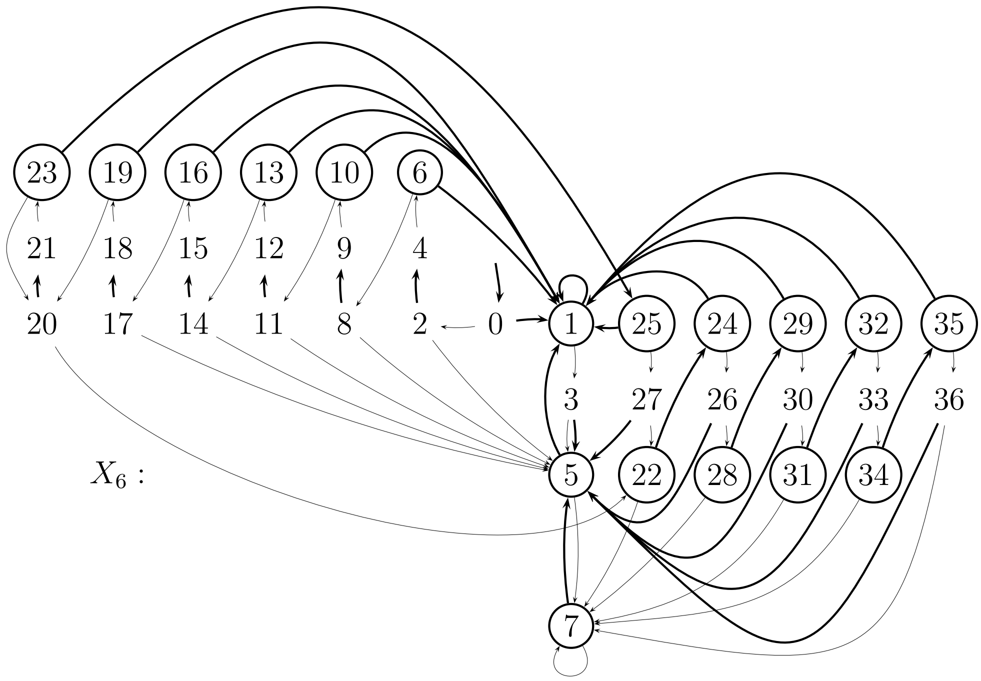
Jos piirrämme kahta peräkkäistä iteraatioaskelta vastaavat automaatit, voimme nähdä, minkälaisen säännöllisyyden mukaan automaatit kasvavat. Kuvista 6 ja 7 huomaamme, miten kieliä X_5 ja X_6 esittävät automaatit koostuvat kahdesta sarjasta kolmen tilan ryppäitä. Kielen X_5 automaatissa tilat 2,4,6,8–18 sekä 20 kuuluvat toiseen sarjaan ja tilat 1,3,5,19 sekä 21–28 kuuluvat toiseen. Kun siirrytään kieltä X_6 esittävään automaattiin, huomattaan, että molemmat sarjat ovat pidentyneet yhdellä kolmen tilan ryppäällä. Yhteensä uusia tiloja on siis kuusi kappaletta, joista lopputiloja kolme. Uusia siirtymiä on 11 kappaletta. Jokaisella iteraatioaskeleella vastaavat automaatit kasvavat saman säännön mukaisesti. Iteraation lähestyessä ääretöntä, molemmat sarjat kasvavat äärettömän pituiseksi. Tällöin samantyyppisten ryppäiden vastintilat voidaan samaistaa ja tuloksena saadaan kielen X^+ tunnistava automaatti.

	lopput.	tiloja	siirtymiä		lopput.	tiloja	siirtymiä
X_0	6	8	15	X_{20}	60	121	222
X_1	5	9	17	X_{21}	63	127	233
X_2	6	13	24	X_{22}	66	133	244
X_3	9	19	35	X_{23}	69	139	255
X_4	12	25	46	X_{24}	72	145	266
X_5	15	31	57	X_{25}	75	151	277
X_6	18	37	68	X_{26}	78	157	288
X_7	21	43	79	X_{27}	81	163	299
X_8	24	49	90	X_{28}	84	169	310
X_9	27	55	101	X_{29}	87	175	321
X_{10}	30	61	112	X_{30}	90	181	332
X_{11}	33	67	123	X_{31}	93	187	343
X_{12}	36	73	134	X_{32}	96	193	354
X_{13}	39	79	145	X_{33}	99	199	365
X_{14}	42	85	156	X_{34}	102	205	376
X_{15}	45	91	167	X_{35}	105	211	387
X_{16}	48	97	178	X_{36}	108	217	398
X_{17}	51	103	189	X_{37}	111	223	409
X_{18}	54	109	200	X_{38}	114	229	420
X_{19}	57	115	211	X_{39}	117	235	431

Taulukko 2: Kielen X ensimmäisiä iteraatioaskelia vastaavien minimoitujen determinististen automaattien tilamäärät, lopputilojen määrät ja siirtymien määrät.



Kuva 6: Kieltä X_5 edustava automaatti.



Kuva 7: Kieltä X_6 edustava automaatti.

6.3 Sentralisaattori

Todistetaan seuraavaksi väite $\mathcal{C}(X) = X^+$ formaalisti. Esitetään kielet todistuksessa rationaalisilla lausekkeilla. Suoritetaan kielelle X edelleen kiintopistemetodia, mutta jotta sentralisaattori saadaan laskettua, korvataan kuitenkin yksittäiset iteraatioaskeleet induktiolla, joka suorittaa ne kaikki kerralla. Aloitetaan laskenta määrittämällä ensin iteroinnin lähtökohta, eli kieli $X_0 = \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+)$.

Lause 6.1. *Iteroinnin aloituskieli X_0 voidaan ilmaista muodossa*

$$X_0 = X^+ + (bab)^*b(bab)^* + (bab)^*ab(bab)^* + (bab)^*ba(bab)^*.$$

Todistus. Merkitään

- $Y_1 = (bab)^*b(bab)^*$,
- $Y_2 = (bab)^*ab(bab)^*$ ja
- $Y_3 = (bab)^*ba(bab)^*$.

Nyt sisältyminen

$$X^+ + Y_1 + Y_2 + Y_3 \subseteq X_0$$

nähdään helposti. Luonnollisestikin $X^+ \subseteq X_0$, sillä X^+ sisältyy sekä prefiksiinsä, että suffiksiinsa. Sisältyminen $Y_1 \subseteq X_0$ saadaan huomaamalla, että

$$(bab)^*b(bab)^* = b(a\,bb)^*(bab)^* \subseteq \mathrm{Suf}(X)X^* = \mathrm{Suf}(X^+) \quad \text{ja} \tag{13}$$
$$(bab)^*b(bab)^* = (bab)^*(bb\,a)^*b \subseteq X^*\,\mathrm{Pref}(X) = \mathrm{Pref}(X^+), \tag{14}$$

eli siis $Y_1 \subseteq \mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+) = X_0$. Samalla huomataan, että

$$(bab)^*b(bab)^* \cap X^+ = \emptyset,$$

sillä kielen $(bab)^*b(bab)^*$ sanat voidaan esittää muodossa $X^*\,\mathrm{Pref}(X)$ vain kaavan (14) mukaisella tavalla, jolloin sanan loppuun jäävä b kuuluu kieleen $\mathrm{Pref}(X)$, muttei kieleen X. Samoin yhtälöistä

$$(bab)^*ab(bab)^* = ab(bab)^* + ba(bb\,a)^*(bab)^+ = (bab)^*(a\,bb)^*ab$$

ja

$$(bab)^*ba(bab)^* = ba(bb\,a)^*(bab)^* = (bab)^*ba + (bab)^+(a\,bb)^*ab$$

nähdään, että $Y_2, Y_3 \subseteq X_0$, sekä $Y_2 \cap X^+ = Y_3 \cap X^+ = \emptyset$.

Seuraavaksi todistetaan sisältyminen toiseen suuntaan. Koska $\mathrm{Pref}(X) = \mathrm{Suf}(X) = \{a, b, ab, ba, bb, aba, bab, bbb\} = X + \{b, ab, ba\}$, on

$$X_0 = X^+ + (\{b, ab, ba\}X^* \cap X^*\{b, ab, ba\})\,.$$

Kieli X^+ sisältyy selvästi väitteen oikeaan puoleen $X^+ + Y_1 + Y_2 + Y_3$, joten tarkastellaan muotoa $\{b, ab, ba\}X^*$ olevia sanoja. Ensinnäkin $\{b, ab, ba\} \in Y_1 + Y_2 + Y_3$. Seuraavaksi tarkastelemme sanoja uvw, joissa $u \in \{b, ab, ba\}$, $v \in X$ sekä $w \in X^*$ ja tutkimme, mitä muotoa olevat sanat uvw kuuluvat joko kieleen X^+ taikka kieleen $X^+\{b, ab, ba\}$. Jos $u = ab$, päädymme sanan v eri arvoilla seuraavanlaiseen tarkasteluun:

- $v = a \implies uvw = ab \cdot a \cdot w = aba \cdot w \in X^+$,
- $v = bb \implies uvw = ab \cdot bb \cdot w = a \cdot bbb \cdot w \in X^+$,
- $v = aba \implies uvw = ab \cdot aba \cdot w = a \cdot bab \cdot a \cdot w \in X^+$,
- $v = bab \implies uvw = ab \cdot bab \cdot w = (abb)abw$,
- $v = bbb \implies uvw = ab \cdot bbb \cdot w = a \cdot bb \cdot bb \cdot w \in X^+$.

Näistä tapauksista siis vain $v = bab$ vaatii tarkempaa tarkastelua. Tarkastelu on kuitenkin helppo, sillä sanan loppuosa abw on edelleen alkuperäistä muotoa abX^*. Sanasta w voidaan siis purkaa kieleen X kuuluvia vasempia tekijöitä, kunnes sana loppuu. Rekursiivisesti nähdään, että

$$abX^* \cap X^*\{b, ab, ba\} \subseteq X^+ + ab(bab)^* \subseteq X^+ + Y_2.$$

Tutkitaan vastaavalla tavalla tilanne $u = ba$.
Jos $u = ba$, niin

- $v = a \implies uvw = ba \cdot a \cdot w$,

- $v = bb \implies uvw = ba \cdot bb \cdot w = (bab)bw$,
- $v = aba \implies uvw = ba \cdot aba \cdot w$,
- $v = bab \implies uvw = ba \cdot bab \cdot w = (bab)abw$,
- $v = bbb \implies uvw = ba \cdot bbb \cdot w = bab \cdot bb \cdot w \in X^+$.

Näistä nähdään, että koska $b, ba, baa \notin X$, ei sanaksi v ei kelpaa kumpikaan sanoista a tai aba. Tällöinhän sana uvw ei kuuluisi kieleen $X^+\{b, ab, ba\}$. Lisäksi tilanne on triviaali, jos $v = bbb$, jolloin $uvw \in X^+$. Lisätarkastelua vaativat siis tilanteet $v = bb$, ja $v = bab$. Näistä jälkimmäinen johtaa edellisen tarkastelun kaltaiseen tilanteeseen, jossa sanan loppu $abw \in abX^*$. Tällöin siis

$$(ba)babX^* \cap X^*\{b, ab, ba\} \subseteq X^+ + (bab)ab(bab)^* \subseteq X^+ + Y_2.$$

Muotoa $ba \cdot bbX^*$ olevien sanojen kohtalo ratkeaa seuraavan tarkastelun myötä.

Jos $u = b$, niin

- $v = a \implies uvw = b \cdot a \cdot w = ba \cdot w$,
- $v = bb \implies uvw = b \cdot bb \cdot w = bbb \cdot w \in X^+$,
- $v = aba \implies uvw = b \cdot aba \cdot w = bab \cdot a \cdot w \in X^+$,
- $v = bab \implies uvw = b \cdot bab \cdot w = (bba) \cdot b \cdot w$,
- $v = bbb \implies uvw = b \cdot bbb \cdot w = bb \cdot bb \cdot w \in X^+$.

Näin siis tapauksissa $v = a$ ja $v = bab$ vaaditaan lisätarkastelua, sillä muissa tapauksissa sana uvw kuuluu selvästi kieleen $X^+ + Y_1 + Y_2 + Y_3$. Jos $v = bab$, on sanan loppu edelleen alkuperäistä muotoa bX^*, joten tämä kohta koskee muotoa bX^* olevien sanojen lisäksi kaikkia muotoa $(bba)^*bX^*$ oleva sanoja. Nyt

$$(bba)^*b = b(bab)^* \subseteq (bab)^*b(bab)^* = Y_1$$

ja

$$b(bab)^* \cdot a = (bba)^*ba = ba + (bba)^*bb \cdot aba \subseteq Y_2 + X^+.$$

Myöskin $b(bab)^+aw = (bb\,a)^+baw = (bb\,a)^*bb\,abaw \subseteq X^+$, joten lisää tutkittavia sanoja löytyy vain sanoista baw. Edellä olevan mukaisesti tällaiset sanat voivat olla vain muotoa $ba\,babX^* = (bab)abX^*$ tai muotoa $ba\,bbX^* = (bab)bX^*$. Näistä edellinen johti väistämättä kielen $X^+ + (bab)ab(bab)^*$ sanaan. Jäkimmäinen taas johtaa rekursiivisesti sanaan, joka kuuluu johonkin kielistä X^+, $(bab)^+b(bab)^*$ tai $(bab)^+(bab)ab(bab)^*$. Näin siis

$$bX^* \cap X^*\{b, ab, ba\} \subseteq X^+ + Y_1 + Y_2 + Y_3,$$

joten myös sisältyminen

$$X_0 = X^+ + (\{b, ab, ba\}X^* \cap X^*\{b, ab, ba\}) \subseteq X^+ + Y_1 + Y_2 + Y_3$$

on voimassa. □

Systemaattisempi tapa etsiä kieli X_0 olisi kielien manipulointi automaatteina. Kielten esittäminen automaattien avulla, prefiksin, suffiksin, unionien ja leikkauksen suorittaminen sekä tulosten minimointi ja lopulta vertailu käsin tehtynä on kuitenkin varsin työlästä ja virhealtista. Edellä esitetty saadaan laskettua helposti antamalla tietokoneen laskea sekä X_0, että $X^+ + Y_1 + Y_2 + Y_3$ ja tarkistamalla, että nämä ovat samat.

Seuraavaksi todistetaan, lauseen 6.1 merkinnöin, että kielet Y_1, Y_2 ja Y_3 eivät sisälly sentralisaattoriin $\mathcal{C}(X)$, jolloin kaavasta

$$X^+ \subseteq \mathcal{C}(X) \subseteq X^+ + Y_1 + Y_2 + Y_3$$

nähdään, että $\mathcal{C}(X) = X^+$.

Lemma 6.2. *Sanat b, ab ja ba eivät kuulu sentralisaattoriin $\mathcal{C}(X)$.*

Todistus. Ensinnäkin väite $b \notin \mathcal{C}(X)$ nähdään seuraavasti. Selvästi

$$b \in (bab)^*b(bab)^* = Y_1 \subseteq X_0$$

ja $a \in X$, joten $ab \in XX_0$. Kuitenkin $ab \notin X_0X$, ja siis $b \notin \mathcal{C}(X)$.
Samoin $ab \in (bab)^*ab(bab)^* = Y_2 \subseteq X_0$, $a \in X$ ja $aab \in XX_0$, mutta $aab \notin X_0X$. Kolmantena vielä $ba \in (bab)^*ba(bab)^* = Y_3 \subseteq X_0$, $a \in X$ ja $baa \in X_0X$, mutta $baa \notin XX_0$.

Siispä $b, ab, ba \notin \mathcal{C}(X)$. □

Lemma 6.3. *Seuraavat identiteetit ovat voimassa:*

$$ab(bab)^* \cap \mathcal{C}(X) = ba(bab)^* \cap \mathcal{C}(X) = b(bab)^* \cap \mathcal{C}(X) = \emptyset.$$

Todistus. Valitaan kielestä $ab(bab)^*$ mielivaltainen sana $ab(bab)^k$ $(k \geq 1)$ ja katenoidaan tämän alkuun kielen X sana a. Näin saatu sana $a \cdot ab(bab)^k$ kuuluu kieleen XX_0. Lopusta lähtien voidaan sanasta erottaa kieleen X kuuluvana tekijänä vain sana bab. Kuitenkin sana $a(abb)^{k-1}ab \cdot bab$ ei kuulu kieleen X_0X, sillä selvästi

$$a(a\,bb)^{k-1}ab \notin X_0 = X^+ + Y_1 + Y_2 + Y_3.$$

Siis $ab(bab)^k \notin \mathcal{C}(X)$, kun $k \geq 0$, eli $ab(bab)^* \cap \mathcal{C}(X) = \emptyset$.

Seuraavaksi valitaan mielivaltainen kielen $ba(bab)^*$ sana $ba(bab)^k$, jossa $k \geq 1$. Nyt $ba(bab)^k \cdot bab \in X_0X$, mutta $ba(bab)^k(bab) = bab \cdot ab(bab)^k \notin X\mathcal{C}(X)$, sillä edellä olleen mukaan $ab(bab)^k \notin \mathcal{C}(X)$. Nyt siis $ba(bab)^k \notin \mathcal{C}(X)$, kun $k \geq 0$ ja edelleen $ba(bab)^* \cap \mathcal{C}(X) = \emptyset$.

Aivan samoin, kun $b(bab)^k$ $(k \geq 1)$ on mielivaltainen kielen $b(bab)^*$ sana, nähdään, että $a \cdot b(bab)^k \in XX_0$, mutta $ab(bab)^{k-1} \cdot bab \notin \mathcal{C}(X)X$. Näin siis $b(bab)^k \notin \mathcal{C}(X)$ $(k \geq 0)$ ja $b(bab)^* \cap \mathcal{C}(X) = \emptyset$. □

Seuraavassa vaiheessa käytämme induktiota.

Lemma 6.4. *Yhtälöt*

$$\begin{aligned}
(bab)^*b(bab)^* \cap \mathcal{C}(X) &= \emptyset, \\
(bab)^*ab(bab)^* \cap \mathcal{C}(X) &= \emptyset \quad \textit{ja} \\
(bab)^*ba(bab)^* \cap \mathcal{C}(X) &= \emptyset
\end{aligned}$$

ovat voimassa.

Todistus. Todistus menee samoin jokaisella kolmesta kielestä. Merkitään $v \in \{b, ab, ba\}$ ja tutkitaan kieltä $(bab)^*v(bab)^*$. Osoitamme, että kaikilla kokonaisluvuilla $n, i \geq 0$ on voimassa $(bab)^n v(bab)^i \notin \mathcal{C}(X)$. Kun $n = 0$ on kaikilla kokonaisluvuilla $i \geq 0$ väite $v(bab)^i \notin \mathcal{C}(X)$ voimassa.

Oletetaan nyt, että kaikilla kokonaisluvuilla $i \geq 0$ väite $(bab)^n v(bab)^i \notin \mathcal{C}(X)$ pätee aina, kun $n \leq k$ jollain kokonaisluvulla k. Jos $n = k + 1$, näemme, että

$$(bab)^{k+1} v(bab)^i \cdot (bab) \in X_0 X,$$

mutta

$$(bab) \cdot (bab)^k v(bab)^{i+1} \notin X\mathcal{C}(X),$$

sillä induktio-oletuksen mukaan $(bab)^k v(bab)^{i+1} \notin \mathcal{C}(X)$.

Siispä $(bab)^{k+1} v(bab)i \notin \mathcal{C}(X)$, joten

$$Y_1 \cap \mathcal{C}(X) = Y_2 \cap \mathcal{C}(X) = Y_3 \cap \mathcal{C}(X) = \emptyset.$$

□

Lause 6.5. *Kielen $X = \{a, bb, aba, bab, bbb\}$ sentralisaattori on*

$$\mathcal{C}(X) = X^+.$$

Todistus. Lauseen väite on suora seuraus edeltävistä lemmoista, sillä

$$\mathcal{C}(X) = X_0 \cap \mathcal{C}(X) = (X^+ + Y_1 + Y_2 + Y_3) \cap \mathcal{C}(X) = X^+.$$

□

Seuraus 6.6. Kielen $Y = \{a, bb, aba, abb, bab, bba, bbb\}$ sentralisaattori on myöskin

$$\mathcal{C}(Y) = \mathcal{C}(X) = X^+.$$

Todistus. Kielten Y ja X generoimat puoliryhmät ovat samat, eli $Y^+ = X^+$, sillä $X \subseteq Y \subseteq X^+$. Lisäksi lauseen 3.5 nojalla $\mathcal{C}(Y) = \mathcal{C}(Y^+) = \mathcal{C}(X^+) = \mathcal{C}(X) = X^+$. □

6.4 Analysointi

Tarkastellaan, miten kiintopistemetodin suoritus etenee edellä olevalla kielellä X ja mietitään, miksi metodi päätyy pysähtymättömään iterointiin.

Vaikka tutkittu kieli X on äärellinen ja sen sentralisaattori on niinkin yksinkertainen kuin X^+, ei sitä saatu laskettua kiintopistemetodilla äärellisellä iteraatiokertojen määrällä. Syynä kiintopistemetodin huonoon menestykseen tässä tapauksessa voidaan pitää sen kyvyttömyyttä induktioon. Jos esimerkiksi kaikki kielen $(bab)^*ab(bab)^*$ sanat kirjoitetaan alareunasta äärettömiin jatkuvan kolmion muotoon, kuten kuvassa 8, voidaan kiintopistemetodin etenemistä havainnollistaa seuraavasti. Kun ensimmäisessä askeleessa siirrytään kielestä X_0

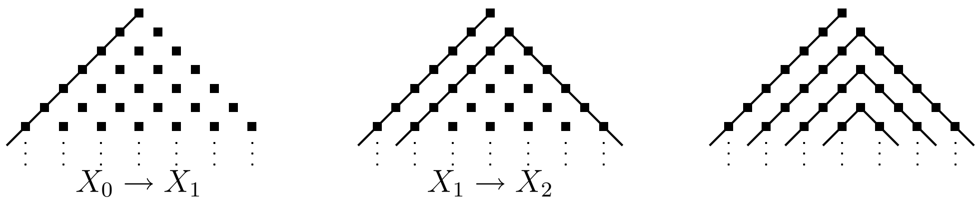
$$
\begin{array}{ccccccccc}
 & & & & ab & & & & \\
 & & & ab(bab) & & (bab)ab & & & \\
 & & ab(bab)^2 & & (bab)ab(bab) & & (bab)^2ab & & \\
 & ab(bab)^3 & & (bab)ab(bab)^2 & & (bab)^2ab(bab) & & (bab)^3ab & \\
ab(bab)^4 & & (bab)ab(bab)^3 & & (bab)^2ab(bab)^2 & & (bab)^3ab(bab) & & (bab)^4ab \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

Kuva 8: Kieli $(bab)^*ab(bab)^*$ kolmion muotoon kirjoitettuna.

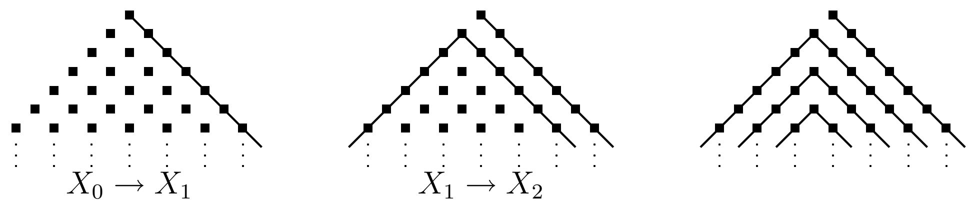
kieleen X_1, karsitaan kolmion esittämästä joukosta pois, lemman 6.3 mukaisesti, kieli $ab(bab)^*$. Kolmiosta poistuu siis vasemmanpuoleinen pintakerros kuvan 9 vasemmanpuoleisimman kuvion tavoin. Seuraavat iteraation askeleet pyyhkivät samoin kukin vuorollaan jäljellä olevasta kolmiosta pintakerroksia, kuten lemman 6.4 induktiotodistuksessa. Koska kolmio on alareunastaan ääretön, iteraatio ei pääty äärellisellä askelmäärällä, vaan jatkuu loputtomiin. Laskenta etenee vastaavasti myös kielillä $(bab)^*ba(bab)^*$ ja $(bab)^*b(bab)^*$, kuten on havainnollistettu kuvissa 10 ja 11. Osa näistä pintakerroksista on kolmioille yhteisiä, esimerkiksi $(bab)ab(bab)^* = ba(bab)^+$.

Kiintopistemetodia voidaan koettaa nopeuttaa valitsemalla iteroinnin aloituskieleksi X_0 jonkin kieltä $\mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+)$ suppeamman, sentralisaattorin sisältävän, kielen. Voimme esimerkiksi merkitä

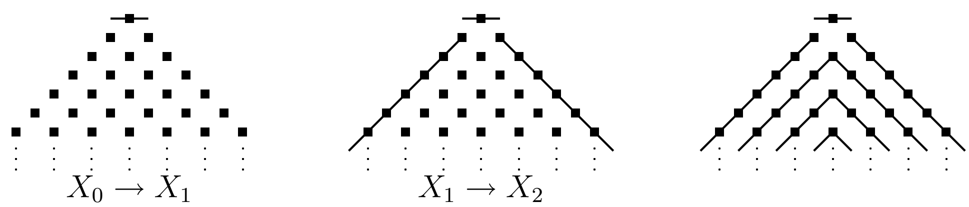
$$
\begin{aligned}
B_p &= \{w \in \mathrm{Pref}(X^+) \mid w \,\mathrm{Pref}_1(X) \subseteq \mathrm{Pref}(X^+)\} = B \\
B_s &= \{w \in \mathrm{Suf}(X^+) \mid \mathrm{Suf}_1(X) w \subseteq \mathrm{Suf}(X^+)\},
\end{aligned}
$$



Kuva 9: Kielen $(bab)^*ab(bab)^*$ osien poistuminen iteraation edetessä.



Kuva 10: Kielen $(bab)^*ba(bab)^*$ osien poistuminen iteraation edetessä.



Kuva 11: Kielen $(bab)^*b(bab)^*$ osien poistuminen iteraation edetessä.

jolloin B_p on siis haaroittumispisteiden joukko ja B_s vastaava suffiksien avulla muodostettu joukko. Nyt voimme valita

$$X_0 = B_p \cap B_s,$$

sillä lauseen 4.3 todistuksen perusteella $\mathcal{C}(X) \subseteq B = B_p$ ja vastaavasti $\mathcal{C}(X) \subseteq B_s$. Kuitenkin esimerkiksi edellä tarkastellun kielen X kohdalla tällä aloituskielen vaihdoksella ei ole suurta merkitystä, sillä

$$\big(\mathrm{Pref}(X^+) \cap \mathrm{Suf}(X^+)\big) \setminus (B_p \cap B_s) = ab(bab)^* + (bab)^*ba,$$

joka poistetaan muutenkin jo askeleen $X_0 \to X_1$ yhteydessä.

6.5 Muita esimerkkejä

On olemassa myöskin rationaalisia kieliä, joilla kiintopistemetodi ei pysähdy ja joilla sentralisaattori ei ole X^+. Eräs tällainen kieli on $X = a\Sigma^+ b + b\Sigma^* ba$, sillä laskemalla kielet X^+ ja S nähdään, että $X^+ \neq S$, joten $X^+ \subset \mathcal{C}(X)$. Tälle kielelle Grail+:lla suoritetun kiintopistemetodin eri vaiheiden automaattien tilamääriä on lueteltu taulukossa 3. Taulukosta käy ilmi, että kielestä X_7 lähtien kullakin

	lopputil.	tiloja	siirtymiä
X	2	7	14
X_0	2	4	8
X_1	4	15	30
X_2	6	16	32
X_3	9	24	48
X_4	13	33	66
X_5	18	44	88
X_6	23	53	106
X_7	29	67	134
X_8	33	75	150
X_9	37	83	166
X_{10}	41	91	182
X_{11}	45	99	198

	lopputil.	tiloja	siirtymiä
X_{12}	49	107	214
X_{13}	53	115	230
X_{14}	57	123	246
X_{15}	61	131	262
X_{16}	65	139	278
X_{17}	69	147	294
X_{18}	73	155	310
X_{19}	77	163	326
X_{20}	81	171	342
X_{21}	85	179	358
X^+	4	12	24
S	8	24	48
Z	10	28	56

Taulukko 3: Kieleen $X = a\Sigma^+ b + b\Sigma^* ba$ liittyviä kieliä vastaavien automaattien tila- ja siirtymämääriä.

askeleella vastaavan automaatin tilamäärä kasvaa kahdeksalla. Tarkastelemalla automaatteja tarkemmin huomataan, että ne säilyttävät muuten saman rakenteen, mutta jotkin osat automaatista laajenevat tietyn säännön mukaisesti. Kuvassa 12 esitetään kielen X_{11} ratkaisevat osat pääpiirteissään. Automaatissa on kaksi erillistä ”ketjua”, jotka molemmat kasvavat neljän tilan jaksoissa. Jokaisella iteraation askeleella kumpaankin ketjuun tulee neljä jakson mukaista tilaa lisää. Vain ketjujen viimeiset tilat poikkeavat hieman muun osan kaavasta. Iteraation edetessä ketjun päät siirtyvät yhä kauemmas ja kauemmas. Tällöin kahden peräkkäisen kielen erotukseen kuuluvat sanat ovat yhä pidempiä ja pidempiä. Kun iteroinnin askelten määrä lähenee ääretöntä, lähenee ketjujen pituus ääretöntä ja ketjut voidaan korvata silmukalla. Näin voidaan esimerkiksi kieltä X_{11} vastaavassa automaatissa siirtymä $98 \xrightarrow{b} 13$ korvata siirtymällä $98 \xrightarrow{b} 95$ ja siirtymä $93 \xrightarrow{b} 13$ korvata siirtymällä $93 \xrightarrow{b} 87$ sekä lopuksi minimoida saatu automaatti. Käytetään merkintää Z tämän automaatin tunnistamasta kielestä. Tämän automaatin tilojen ja siirtymien määrät ovat myöskin taulukossa 3. Kielen Z rationaalinen lauseke on ilmeisestikin jonkin verran monimutkaisempi, kuin esimerkiksi kielen X^+ lauseke. Sen selvittäminen on automaatin tila- ja siirtymämäärän vuoksi kuitenkin melko työläs tehtävä. Oletettavasti kuitenkin $Z = \mathcal{C}(X)$. Grail+ vahvistaa ainakin, että

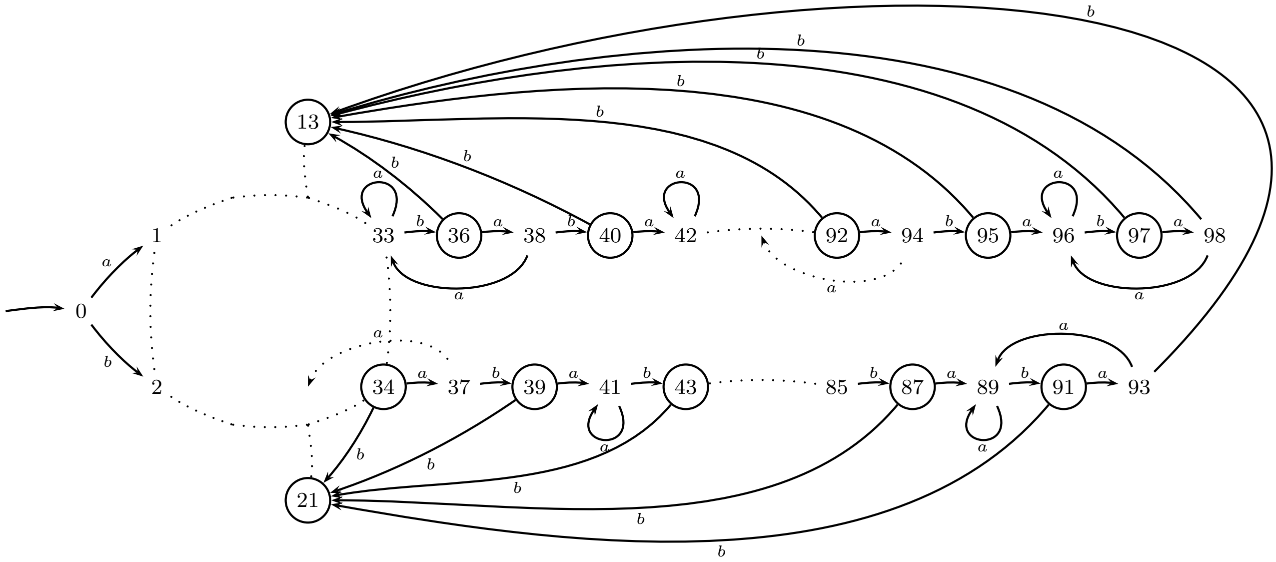
$$X^+ \subset S \subset Z$$

ja että Z kommutoi kielen X kanssa, joten varmasti

$$X^+ \subset S \subset Z \subseteq \mathcal{C}(X).$$

Kaikilla kiintopistemetodissa pysähtymättömään iteraation johtavilla kielillä jaksollisuus ei ole näin selvästi nähtävissä. Esimerkiksi kielellä $X = \Sigma a \Sigma^* a \Sigma + \Sigma b \Sigma^* b \Sigma$ iteraation askelten tilamäärät kasvavat kaksivaiheisesti. Joka toisella askeleella tilamäärä lisääntyy kahdellatoista ja joka toisella lisäys on kahdeksantoista tilaa.

Äärellisellä kielellä $X = \{a, bb, aab, aba, abb, baa, bab, bba, bb\}$ taas tilamäärän kasvu ei näytä vakiintuvan, ainakaan viidelläkymmenellä ensimmäisellä iteraatioaskeleella. Lisäksi joka toisella askeleella tilamäärä kasvaa ja joka toisella vähenee. Lisääntyvien tilojen määrä kavaa vähitellen ja vähenevien tilojen määrä on aina noin puolet seuraavasta lisäyksestä. Automaatteina tarkasteltuna nähdään, että askelia vastaaviin automaatteihin muodostuu kuitenkin selkeää säännönmukaisuutta, joka kiintopistemetodin edetessä vakiintuu edellä olevan esimerkin kaltaiseksi yhä pidemmäksi ketjuksi. (Kuva 13)



Kuva 12: Kieli X_{11} pääpiirteissään.

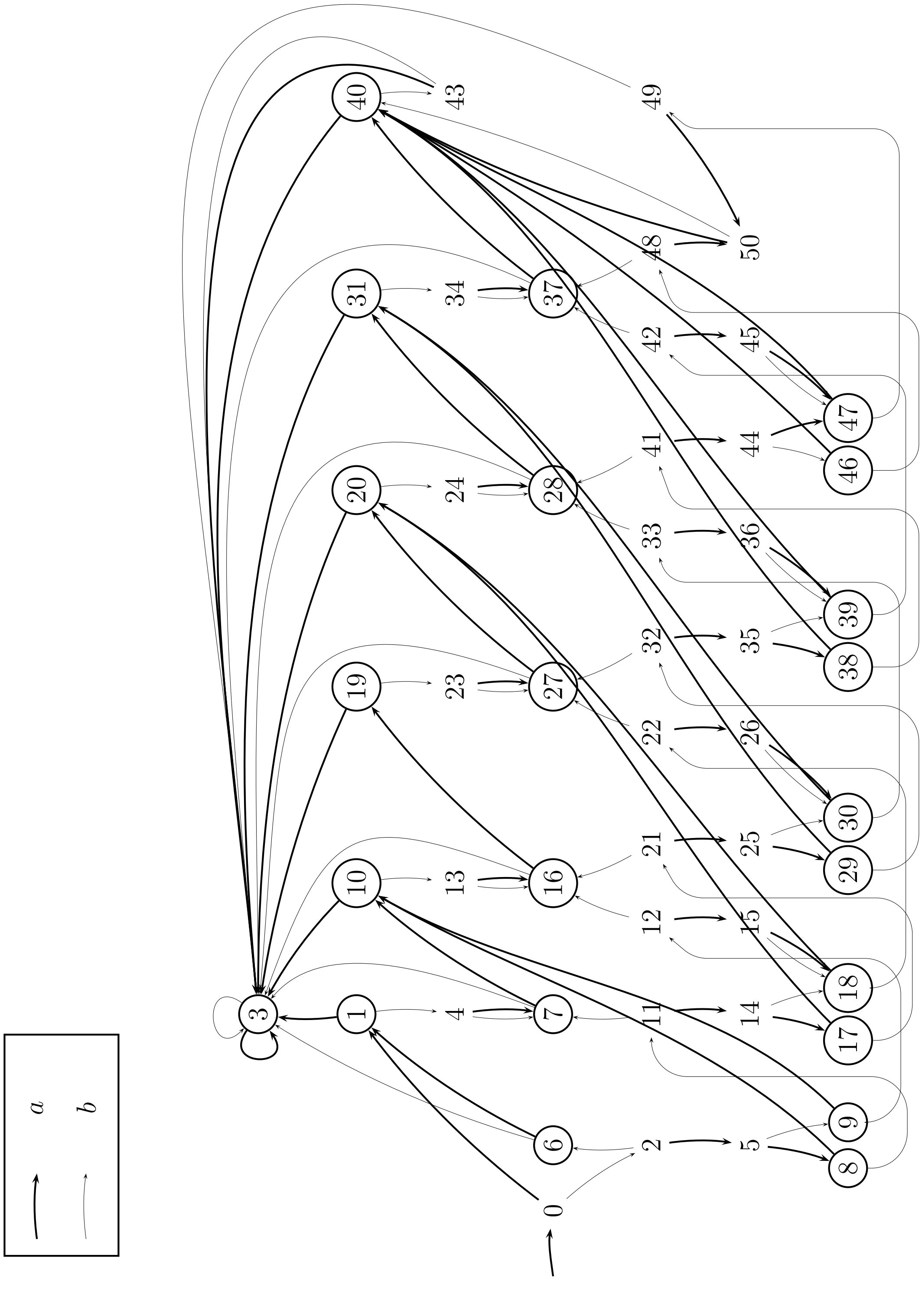
Verrataan edellä olevasta kielen $X = a\Sigma^+ b + b\Sigma^* ba$ kiintopistemetodin suorituksesta saatua kieliä X_i vastaavien automaattien jonoa desimaalilukujen

$$0,12345; 0,123412345; 0,1234123412345; 0,12341234123412345\dots$$

muodostamaan jonoon, joka lähestyy lukua $0,123412341234\dots = \frac{1234}{9999}$. Jokaisessa askeleessa jaksolliseen osuuteen tulee yksi jakso lisää ja kaikki muu pysyy muuttumattomana. Nyt kielen $X = \{a, bb, aab, aba, abb, baa, bab, bba, bb\}$ sentralisaattoria etsittäessä saatavaa automaattien jonoa voidaan verrata lukujonoon

$$0,12798; 0,1218798098; 0,12127098; 0,12121978068798; 0,121212707698\dots,$$

joka lähenee lukua $0,121212\dots = \frac{12}{99}$. Lukujonossa jaksollinen osuus kasvaa kahden askeleen välein, mutta lisäksi lukujen loppuun kerääntyy kasvava määrä jaksotonta materiaalia, joka ei kuitenkaan vaikuta jonon raja-arvoon.



Kuva 13: Kieli $X = \{a, bb, aab, aba, abb, baa, bab, bba, bbb\}$ / Iteraatio X_{12}

7 Johtopäätöksiä

Tehdään muutamia huomioita ja päätelmiä kiintopistemetodin toiminnasta ja pohditaan, miten voitaisiin tunnistaa, kuuluuko annettu sana kielen X sentralisaattoriin.

Muodostavatko kiintopistemetodilla saatujen peräkkäisten kielten erotusten $X_i \setminus X_{x+1}$ lyhimpien sanojen w_i pituudet $|w_i|$ monotonisesti kasvavan jonon? Tällöin voitaisiin laskea kiintopistemetodilla n iteraatiota ja olla varmoja, että kaikki kielen X_n sanat, joiden pituus on pienempi kuin n, kuuluvat sentralisaattoriin. Näin saataisiin tuloksena, että rationaalisen kielen sentralisaattori olisi aina rekursiivinen. Esimerkiksi kappaleen 6 kielen $X = \{a, bb, aba, bab, bbb\}$ kohdalla tämä pitää paikkansa. Kuitenkin esimerkiksi myöhemmin kappaleessa 6 käsitelty esimerkki $X = a\Sigma^+b + b\Sigma^*ba$ on eräs kieli, jolla edellä olevan merkinnän mukaisesti $|w_2| > |w_3|$. Oletettavasti tälläkin kielellä lukujen $|w_i|$ jono on jatkossa monotonisesti kasvava, kieliä X_i vastaaviin automaatteihin ilmestyvän toistuva jakson ansiosta.

Sen sijaan $X = \Sigma^4 - a^4 + a + bb$ on kieli, jolle suoritettu kiintopistemetodi ei ainakaan ensimmäisen 16 suorituskerran jälkeen osoita minkäänlaista merkkiäkään jaksollisuudesta. Metodin peräkkäisten askelten erotusten lyhimpinä sanoina saadaan seuraavat sanat.

$$
\begin{aligned}
w_0 &= b \\
w_1 &= aba \\
w_2 &= baaaab \\
w_3 &= baaaaab \\
w_4 &= bbbbb \\
w_5 &= abaaaab \\
w_6 &= abaaaba \\
w_7 &= bbbbbbb \\
w_8 &= abaaabaaaba \\
w_9 &= bbbbbbbbb \\
w_{10} &= abaaabaaabaaaab \\
w_{11} &= baaabbbbbbbbb \\
w_{12} &= bbbbbbbbbbb \\
w_{13} &= baaabbbbbbbbbbb \\
w_{14} &= bbbbbbbbbbbbb \\
w_{15} &= baaabaaabbbbbbbbbbb
\end{aligned}
$$

Sanojen pituudet näyttävät vaihtelevan varsin säännöttömästi. Sentralisaattorin rekursiivisuuteen riittäisi myös se, että jono $|w_i|$ voitaisiin jakaa äärelliseen määrään ei-väheneviä osajonoja. Tällöin sanan $w \in \mathcal{C}(X)$ tunnistamiseksi riittäisi laskea kiintopistemetodia, kunnes kukin osajono on ylittänyt sanan w pituuden. Edellä olevista sanoista nähdään, että ainakin jonojen $|w_{3n}|, |w_{3n+1}|$ ja $|w_{3n+2}|$ alut ovat ei-vähenevät. Jatkosta on kuitenkin vaikea sanoa mitään varmaa.

Kiintopistemetodin kaltaisten, kielten tai automaattien äärettömään jonoon johtavien, tilanteiden kannalta olisi hyödyllistä löytää jokin sopiva mitta, jolla voitaisiin ilmaista kahden kielen tai automaatin etäisyys toisistaan. Näin raja-arvotarkasteluissa voitaisiin käyttää analyysistä tuttuja menetelmiä. Voidaan kuitenkin kysyä, löytyykö tällaista yleistä mittaa, joka toimii kaikissa tapauksissa, vai pitäisikö mitta valita kussakin tilanteessa tapauskohtaisesti. Eräs tällainen mitta voisi olla esimerkiksi symmetrisen erotuksen $A \Delta B$ lyhimmän sanan pituuden käänteisluku.

A Grail+

Tämän esityksen esimerkkien laskemisessa ja tarkistamisessa sekä päättelyn apuna on käytetty Länsi-Ontarion yliopistossa Kanadassa kehitettyä *Grail+*-ohjelmistoa. Grail+ on äärellisten automaattien, säännöllisten lausekkeiden ja äärellisten kielten käsittelyyn tarkoitettu symbolisen laskennan ohjelma. Ohjelma on kirjoitettu C++-kielellä ja sen lähdekoodi on saatavilla internetosoitteesta *http://www.csd.uwo.ca/research/grail/grail.html.*

Työssä käytettiin esijulkaisua ohjelman versiosta 3.0. Ohjelmaan jouduttiin tekemään pieni korjaus, jotta automaattien katenointi saatiin toimimaan oikein. Lisäksi *Arto Lepistö* lisäsi ohjelmistoon useita uusia operaatioita, joita tarvittiin kiintopistemetodin toteuttamiseen.

Kukin tarkasteltu kieli esitettiin äärellisenä automaattina ja talletettiin Grail+:n ymmärtämänä tekstitiedostona. Itse Grail+ koostuu useasta ajettavasta ohjelmatiedostosta, joista kukin suorittaa syötteenä annetuille automaateille jonkin operaation. Syötteensä nämä ohjelmat saavat joko tiedostosta taikka edelleen ohjattuna edelliseltä ohjelmalta. Tulostus puolestaan voidaan ohjata joko seuraavalle ohjelmalle, näytölle taikka tiedostoon.

Käytettyjä Grail+:n operaatioita olivat:

fmcat Kahden automaatin katenointi.

fmcross Kahden automaatin karteesinen tulo (eli leikkaus).

fmdiff Kahden automaatin erotus.

fmemptyof Poistaa automaatin esittämästä kielestä tyhjän sanan, jos se on siinä.

fmenum Luettelee halutun määrän (Oletuksena 100) automaatin sanoja pituusjärjestyksessä.

fmexec Suorittaa laskennan annetulla automaatilla ja annetulla sanalla. Tulos joko hyväksytty tai hylätty.

fmlq Automaatin esittämän kielen vasen osamäärä toisen automaatin esittämän kielen suhteen.

fmminrev Automaatin minimointi. Toteutettu vuorotellen kääntämällä automaatti esittämään peilikuvakieltä ja muuttamalla deterministiseksi.

fmplus Kleenen plusoperaatio annetulle automaatille.

fmpref Automaatin esittämän kielen prefiksi.

fmrq Automaatin esittämän kielen oikea osamäärä toisen automaatin esittämän kielen suhteen.

fmstar Kleenen tähtioperaatio annetulle automaatille.

fmstats Tilastotietoa annetusta automaatista. Tilamäärä, siirtymiä, alkutiloja, lopputiloja.

fmsuff Automaatin esittämän kielen suffiksi.

fmsymdiff Kahden automaatin symmetrinen erotus.

fmunion Kahden automaatin unioni.

isomorph Tarkistaa, ovatko annetut automaatit tilojen numerointia lukuunottamatta samat. Tulos joko *isomorphic* tai *nonisomorphic*.

retofm Muuttaa syötteeksi saamansa rationaalisen lausekkeen vastaavaksi automaatiksi.

Iteroinnin aloituskielen X_0 ja seuraavien askelten X_i laskemiseen käytettiin seuraavia DOS:in komentojonotiedostoja.

laskex0.bat:

```
@echo off
rem laskex0.bat
rem Laskee aloituskielen x0
if "%2"=="" goto :virhe
echo - %1plus
fmplus %1| fmminrev > %1plus
echo - %1pref
fmpref %1plus| fmemptyof| fmminrev > %1pref
echo - %1suff
fmsuff %1plus| fmemptyof| fmminrev > %1suff
echo - %2
fmcross %1pref %1suff | fmminrev > %2
goto :end
:virhe
echo Käyttö: laskex0 X X0
:end
```

laskexi.bat:

```
@echo off
rem laskexi.bat
rem Laskee seuraavan xi:n
if "%3"=="" goto :virhe
echo - %1%2
fmcat %1 %2| fmminrev > %1%2
echo - %2%1
fmcat %2 %1| fmminrev > %2%1
echo - %1%2-%2%1
fmdiff %1%2 %2%1| fmminrev > %1%2-%2%1
echo - %2%1-%1%2
fmdiff %2%1 %1%2| fmminrev > %2%1-%1%2
echo - %1d%2
fmunion %1%2-%2%1 %2%1-%1%2 | fmminrev > %1d%2
echo - %1d%2lq
fmlq %1d%2 %1| fmminrev > %1d%2lq
echo - %1d%2rq
fmrq %1d%2 %1| fmminrev > %1d%2rq
echo - %2u
fmunion %1d%2lq %1d%2rq| fmminrev > %2u
echo - %3
fmdiff %2 %2u| fmminrev > %3
goto :end
:virhe
echo Käyttö: laskexi X Xi Xi+1
:end
```

Kielten S, B ja C laskemiseen käytetyt komentojonotiedostot rajoittuvat aakkostoon $\{a, b\}$, mutta ovat laajennettavissa. Lisäksi *laskeBC.bat* vaatii kielikohtaisia muutoksia sen mukaan, mikä on kieli $\mathrm{Pref}_1(X)$.

laskeS.bat:

```
@echo off
REM laskeS.bat
REM Laskee kielen S annetulla kielellä X.
REM Aakkostona {a,b}.
echo - S%1%1p
fmplus %1|fmminrev|fmcat %1|fmminrev > S%1%1p
echo - Ssigmaplus
echo a+b|retofm|fmplus|fmminrev > Ssigmaplus
echo - Sc%1%1p
fmdiff Ssigmaplus S%1%1p|fmminrev > Sc%1%1p
echo - Sc%1%1plq
fmlq Sc%1%1p %1 |fmemptyof|fmminrev > Sc%1%1plq
echo - Sc%1%1prq
fmrq Sc%1%1p %1 |fmemptyof|fmminrev > Sc%1%1prq
echo - Sc
fmunion Sc%1%1plq Sc%1%1prq|fmminrev > Sc
echo - S
fmdiff Ssigmaplus Sc|fmminrev > S
```

Kieleen X liittyvä haaroittumispisteiden kieli B ja kriittisten pisteiden kieli C laskettiin tiedostolla:

laskeBC.bat:

```
@echo off
REM laskeBC.bat
REM Lasketaan syötekielen haaroittuvien sanojen joukko B ja
REM kriittisten sanojen joukko C.
REM Skripti rajoittuu aakkostoon {a,b}
echo - Bpref%1p
fmplus %1|fmminrev|fmpref|fmminrev|fmemptyof|fmminrev > Bpref%1p
echo - Bpref%1
fmpref %1|fmemptyof|fmminrev > Bpref%1
echo - Bpref1%1
echo a+b|retofm|fmminrev|fmcross Bpref%1|fmminrev > Bpref1%1
echo - Bpref%1ppref1%1
fmcat Bpref%1p Bpref1%1 |fmminrev > Bpref%1ppref1%1
echo - Binters
fmcross Bpref%1ppref1%1 Bpref%1p|fmminrev > Binters
REM Lasketaan Binters:n oikea osamäärä kullakin Pref_1X:n
REM kirjaimella.
REM Tämä täytyy säätää kielikohtaisesti.
REM Tässä Pref_1X={a,b}
echo - Ba
echo a|retofm|fmminrev|fmrq Binters|fmemptyof |fmminrev > Ba
echo - Bb
echo b|retofm|fmminrev|fmrq Binters|fmemptyof |fmminrev > Bb
echo - B
REM B on näiden leikkaus.
fmcross Ba Bb|fmminrev > B
echo - C
fmplus %1|fmminrev|fmdiff B|fmminrev > C
```

```
echo *** Onko C tyhjä? ***
isempty C
```

Viitteet

[1] J. H. Conway, Regular Algebra and Finite Machines, Chapman Hall, 1971

[2] J. Karhumäki, I. Petre, On the Centralizer of a Finite Set, in: U. Montanari et al. (Eds.): ICALP 2000, Lecture Notes in Computer Science 1853, Springer–Verlag, Berlin Heidelberg 2000, ss. 536–546

[3] C. Choffrut, J. Karhumäki, N. Ollinger, The commutation of finite sets: A challenging problem, Theoretical Computer Science 273, ss. 69–79, 2002

[4] T. Harju, I. Petre, On Commutation and Primitive Roots of Codes, Theoretical Computer Science (ilmestyy)

[5] J. Karhumäki, Challenges of Commutation - An advertisement, in: Fundamentals of Computation Theory 2001, R. Freivalds, Ed., Lecture Notes in Computer Science 2138, Springer–Verlag, New York 2001, ss. 15–23

[6] J. Karhumäki, Some open problems on combinatorics of words and related areas, Proceedings of Algebraic Systems, Formal Languages and Computations, RIMS Proceedings 1166, Research Institute for Mathematical Sciences, 2000, 118–130

[7] J. Karhumäki, Automata and Formal Languages. Luentomoniste, Turun yliopisto, 1997.

[8] Grail+ 3.0 -ohjelmisto, Länsi-Ontarion yliopisto, Kanada, <URL:http://www.csd.uwo.ca/research/grail/grail.html>