Taloustieteen matemaattiset menetelmät: Osan 2 kertausmateriaalia

Topi Hokkanen*

22. marraskuuta 2019

Tiivistelmä

Nämä muistiinpanot ovat keittokirjatyyliset merkinnät kurssin toisen osan asioista. Näitä ei tule käyttää kurssikirjan tai varsinaisten luentomuistiinpanojen substituuttina. Erityisesti kurssin toisen osan kohdalla, kirjoittaja suosittelee vahvasti kurssikirjan (Simon and Blume, 1994) hankkimista, sillä se on eräs selkeimmistä optimointia ja taloustieteen matemaattisia perustyökaluja käsittelevistä kirjoista. Olen koonnut loppuun joitain kirjallisuusviitteitä eri tasoilta, jotka voivat myös olla hyödyllisiä jatkoa ajatellen. Soveltamiemme lauseiden todistuksiin paneutuminen jää myös lukijan harrastuneisuuden varaan.

*Aalto-yliopiston kauppakorkeakoulu, taloustieteen laitos.

1 Rajoitettu optimointi

Käydään läpi intuitiivinen käsittely yhtälö- ja epäyhtälörajoitetulle optimointitehtävälle, sekä rajoitteiden **NDCQ** -ehdolle.

1.1 Optimointi yhtälörajoitteilla

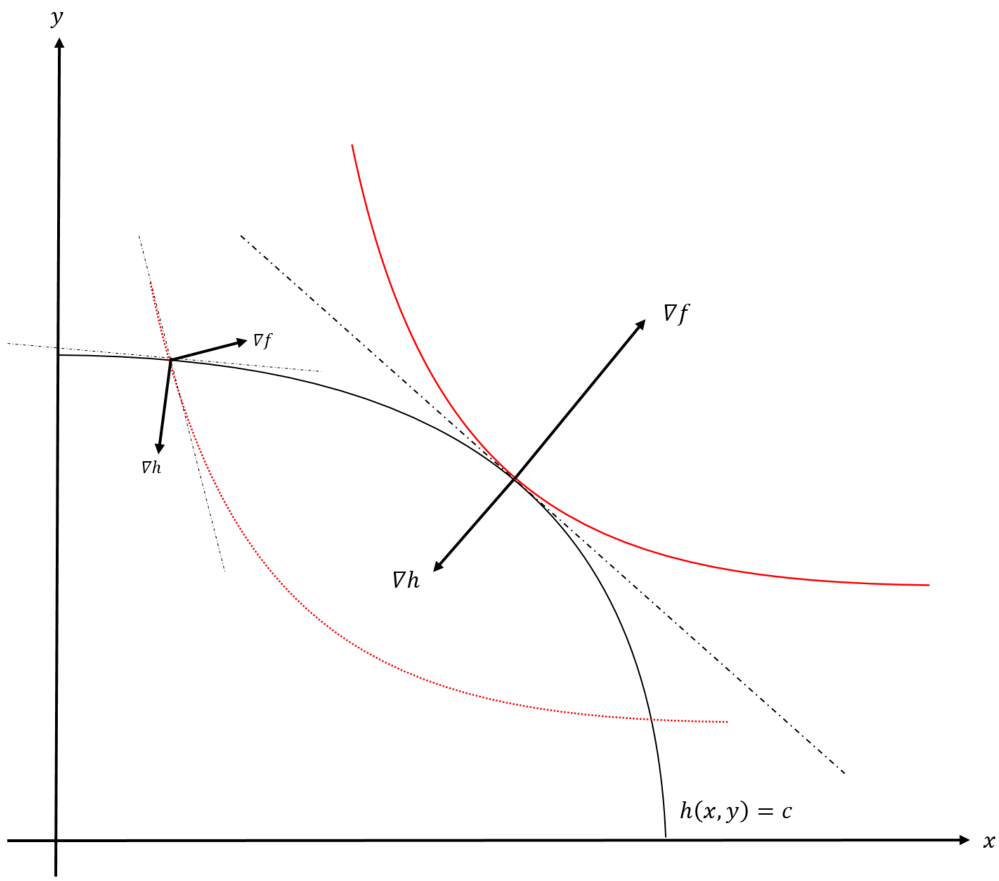
Kehitetään ratkaisumenetelmä yhtälörajoitetulle tehtävälle intuitiivisesti. Olkoon ratkaistavanamme vaikkapa seuraavankaltainen tehtävä:

$$\begin{aligned} \max_{x,y} f(x,y), \quad &\text{s.t.} \\ h(x,y) &= c \end{aligned}$$

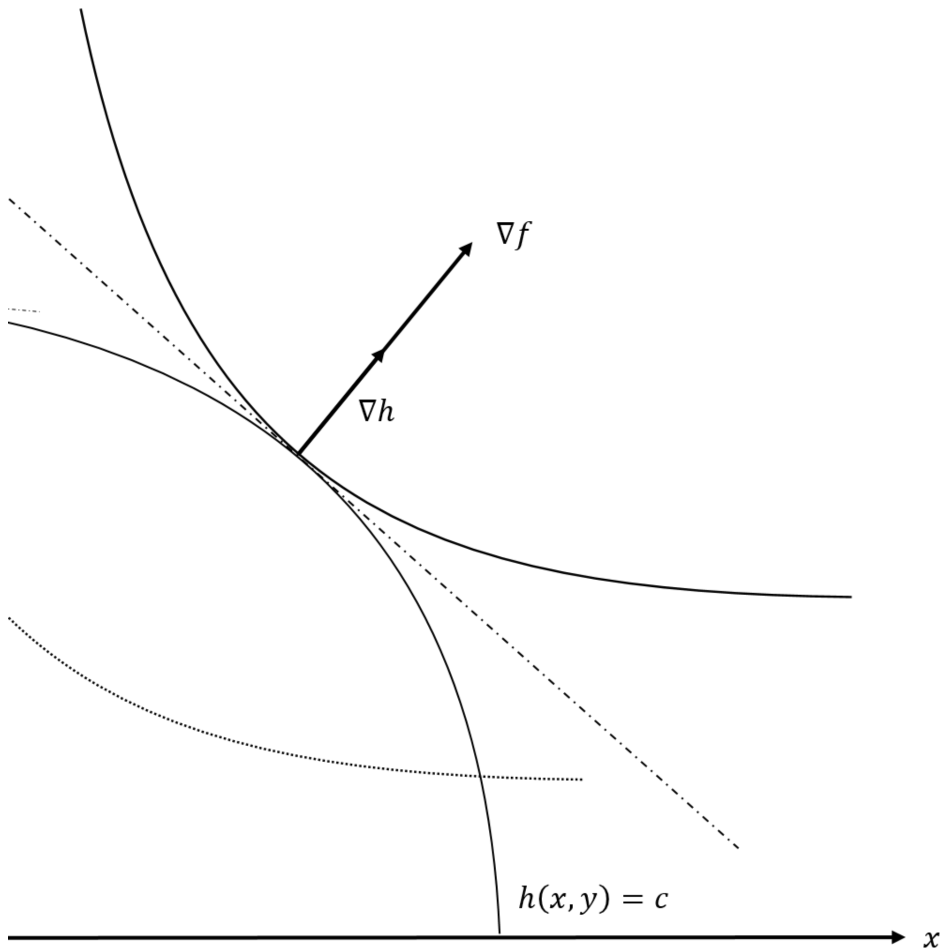
Eli haluamme siis maksimoida jonkun funktion f, mutta meidän on otettava myös huomioon funktio h, joka antaa meille yhtälön, jonka kaikkien ratkaisukandidaattiemme (x^*, y^*) on toteutettava. Kuvaan 1 on piirretty jokin rajoitefunktio $h(x,y) = c$ (x,y)- koordinaatistossa. Jokainen tämän käyrän piste (x,y) toteuttaa siis rajoitteen h. Tämä on nyt optimointiongelmamme *käypä joukko*. Piirretään samaan kuvaajaan myös funktion f eri tasokäyriä (punaiset käyrät). Etsimme korkeinta funktion f tasokäyrää, joka *sivuaa* rajoitetta $h(x,y)$. Tiedämme aiemman osan materiaaleista, että jos meillä on funktiolle f jonkin pisteen $\mathbf{x}^*$ kautta kulkeva tasa-arvokäyrä, silloin tähän pisteeseen piirretty funktion tasa-arvokäyrän tangentti (kuvassa suorat katkoviivat) on kohtisuorassa funktion gradienttiin $\nabla f(\mathbf{x}^*)$ nähden.

Rajoitteen gradientille pätee sama asia. Eli siis rajoitteen gradientti ∇h on myös kohtisuorassa rajoitteelle piirrettyä tangenttia vasten. Kuvat 1 ja 2 havainnollistavat käsillä olevaa tilannetta.

Yhtälörajoite

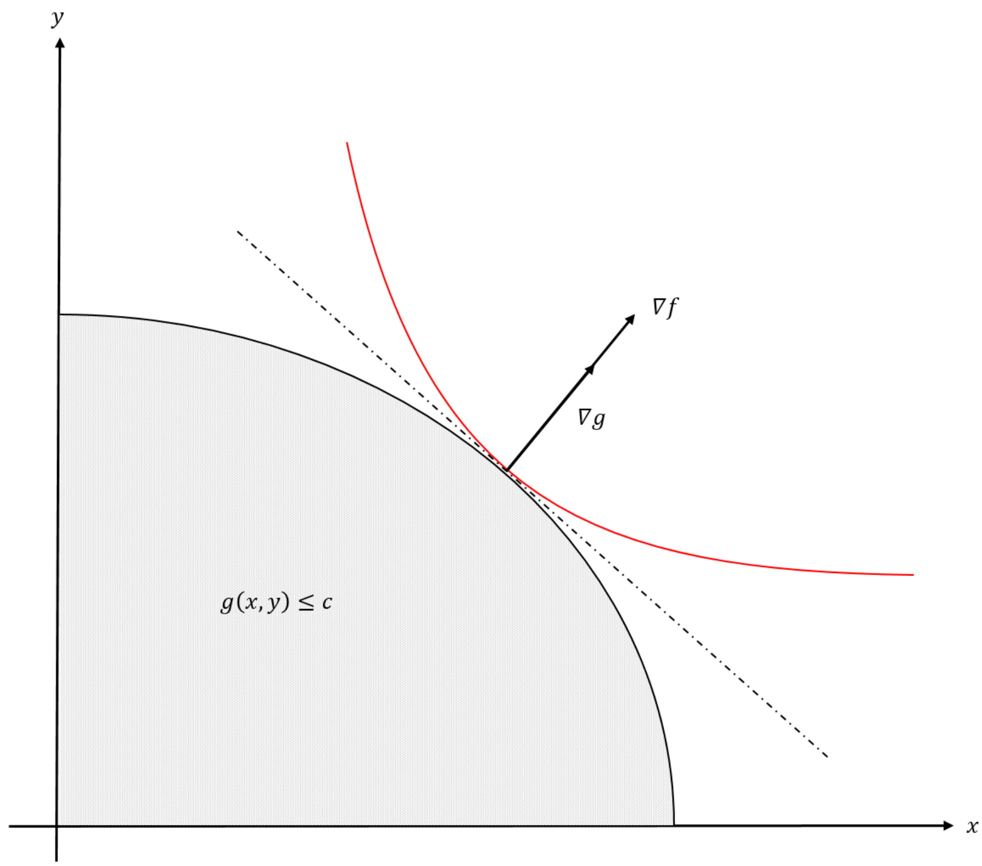


Kuva 1: Optimissa gradientit ovat toistensa lineaariyhdisteitä.



Kuva 2: Rajoitteen ja funktion gradientit voivat olla myös samansuuntaiset.

Epäyhtälörajoite



Kuva 3: Epäyhtälörajoitetussa tehtävässä gradienttien pitää osoittaa optimissa samaan suuntaan.

Kuvissa 1 ja 2 on esitetty optimipisteen sivuamisehto graafisesti. Optimipisteessä gradientit ∇f ja ∇h ovat joko vastakkain tai samansuuntaiset, sillä f:n tasa-arvokäyrä sivuaa rajoitetta optimissa. Toisin sanoen, voimme esittää funktiomme gradientin ∇f jonain rajoitteen gradientin ∇h lineaariyhdisteenä optimipisteessä, joka matemaattisesti on siis seuraava ehto:

$$\nabla f\left(\mathbf{x}^*\right) = \mu \nabla h\left(\mathbf{x}^*\right)$$

jollekin $\mu \neq 0$. Vaihtoehtoisesti voimme esittää tämän muodossa

$$\nabla f\left(\mathbf{x}^*\right) - \mu \nabla h\left(\mathbf{x}^*\right) = \mathbf{0}$$

jonka lisäksi meillä on tietenkin vielä alkuperäinen rajoite: $h(x, y) = c$. Määrittelemällä uuden funktion, jota kutsumme *Lagrangen funktioksi* seuraavalla tavalla:

$$\mathcal{L}(x, y, \mu) = f(x, y) - \mu\left[h(x, y) - c\right]$$

huomaamme, että jos käytämme aiemmin opittuja rajoittamattoman optimoinnin tekniikoita tähän funktioon, saamme tasan tarkkaan intuitiivisesti johtamamme ensimmäisen kertaluvun ehdot, sillä ottamalla gradientin $\nabla \mathcal{L}$ ja asettamalla sen nollaksi saamme:

FOC

$$\mathcal{L}_x = f_x - \mu h_x = 0 \tag{1}$$
$$\mathcal{L}_y = f_y - \mu h_y = 0 \tag{2}$$
$$\mathcal{L}_\mu = h(x,y) - c = 0 \tag{3}$$

Tässä pitää tietenkin huomata, että funktiomme $\mathcal{L}$ on nyt *kolmen* muuttujan funktio. On tärkeää pitää mielessä, että Lagrangen menetelmä, ja Lagrangen funktio on lopun viimein vain yksinkertainen menetelmä muistaa aiemmat ensimmäisen kertaluvun ehdot muuntamalla yhtälörajoitettu tehtävä rajoittamattomaksi optimointitehtäväksi lisäämällä uudeksi muuttujaksi Lagrangen kertoja μ.

1.2 Optimointi epäyhtälörajoitteilla

Jos yhtälörajoitteen sijaan ongelmamme onkin epäyhtälörajoitettu, ts. meillä on jokin ongelma

$$\max_{x,y} f(x,y), \quad \text{s.t.}$$
$$g(x,y) \leq c$$

Joudumme muuttamaan käsittelyämme hieman. Aiemmassa kuvassa etsimme ratkaisukandidaatteja (x,y), jotka ovat rajoitteellamme $h(x,y) = c$. Nyt rajoitteemme on muotoa $g(x,y) \leq c$, ja täten ongelmamme käypä joukko ovat kaikki pisteet (x,y), joille

Joko $g(x,y) < c$, eli olemme rajoitejoukon sisäpisteessä, tai

$g(x,y) = c$, eli rajoite sitoo yhtälörajoitteena.

Kuvassa 3 tätä edustaa varjostettu alue. Etsimme edelleen korkeinta funktion f tasokäyrää, joka on käyvässä joukossamme $X := g(x,y) \leq c$. Nyt kuitenkin rajoite on epäyhtälömuotoinen, joten meillä on kaksi tapausta: joko optimissa rajoite g on *sitova*, eli $g(x,y) = c$, jolloin meillä on sama sivuamisehto kuin aiemmin, **tai sitten** optimi on jossakin käyvän joukon sisäpisteessä, jossa $g(x,y) < c$.

On syytä huomioida, että epäyhtälörajoitteilla optimoidessamme emme välttämättä tiedä etukäteen sitooko epäyhtälörajoite optimissa, ja tämän vuoksi joudumme (jälleen intuitiivisesti) kehittämään hieman lisätyökaluja ongelman ratkaisemiseen[1].

Kysymys: Mihin suuntaan gradientit nyt osoittavat ja mitä väliä sillä on?

Yhtälörajoitetussa tehtävässä meitä ei kiinnostanut se, mihin suuntaan rajoitteen ja funktion gradientit osoittavat, sillä optimiehtomme vaati meitä vain etsimään pisteen, jossa gradientit ovat vastakkain (kuvat 1,2). Nyt meidän pitää pohtia asiaa hieman tarkemmin.

Tarkimmat lukijat lienevät huomanneet, että kuvien 1 ja 2 perusteella yhtälörajoitteen gradientti voi osoittaa kumpaan "suuntaan" tahansa. Epäyhtälörajoitetussa tehtävässä (kuva 3) rajoitteen gradientti on taasen piirretty osoittamaan vain yhteen suuntaan. Tämä johtuu siitä, että gradientti osoittaa funktion nopeinta kasvusuuntaa. Yhtälörajoitteelle tämä suunta ei aina ole selvä, mutta epäyhtälörajoitteelle se on, sillä:

- Rajoitteen gradientti ∇g osoittaa aina "rajoitteesta poispäin", sillä tähän suuntaan rajoite "kasvaa" nopeimmin. Toisin sanoen rajoitteen $g(x,y) \leq c$ gradientti osoittaa aina joukkoa $g(x,y) \geq c$ kohti.

[1]Käsittelemämme ehdot ovat eräs variantti niin kutsutuista Karush-Kuhn-Tucker-ehdoista (KKT), joihin voitte tutustua halutessanne.

- *Entä sitten funktiomme gradientti eli* ∇f*?*

Pohditaan asiaa hiukan. Jos optimipisteemme $\mathbf{x}^* = (x^*, y^*)$ on kuten kuvassa käyvän joukon reunalla, silloin selkeästi pätee $g(x, y) = c$. Tällöin edellisen perusteella $\nabla f(\mathbf{x}^*) = \lambda \nabla g(\mathbf{x}^*)$. Mutta jos $\lambda < 0$, silloin ∇f osoittaa rajoitteeseen päin, ja silloin piste $\mathbf{x}^*$ ei voi olla optimi. Täten meidän tulee nyt, erona aikaisempaan vaatia että $\lambda \geq 0$.

Tällöin gradientit ∇f ja ∇g osoittavat samaan suuntaan.

Syy tälle vaatimukselle on se, että jos ∇f osoittaa rajoitteeseen, silloin voimme kasvattaa tavoitefunktiomme f arvoa siirtymällä rajoitteeseen päin (joukon sisäpisteeseen), ja tällöin selkeästi myös rajoitteemme pätee, koska sisäpisteessä $g(x, y) < c$. Näin ollen tämänkaltaisessa optimissa (x^*, y^*) rajoite g ei ole aktiivinen. Eli, mikäli ∇f osoittaa rajoitteeseen päin, rajoite g ei sido optimissa. Tällöin meillä on käsissämme *rajoittamaton optimipiste* funktiolle f.

Seuraavaksi pohdimme, voisimmeko jotekin ilmaista tämän kompaktisti (vastaus: voimme.). Jos nyt merkitsemme tulomuodossa vaatimuksen:

$$\lambda \left[g(x, y) - c \right] = 0 \tag{4}$$

ja vaadimme myös että

$$\lambda \geq 0 \tag{5}$$

$$g(x, y) \leq c \tag{6}$$

silloin olemme tiivistäneet kaiken yllä olevan varsin lyhyeen esitykseen. Ehtoa (4) nimitämme **Complementary slackness-ehdoksi** . Olemme nimittäin pelkistäneet yllä olevaan tulomuodossa seuraavat vaatimukset:

1. Jos $\lambda > 0$, niin $g(x,y) = c$ ehdon (4) perusteella, eli rajoite sitoo ja olemme reunapisteessä (vrt. kuva 3) tai

2. $g(x,y) < c$, jolloin $\lambda = 0$ ja olemme funktion f rajoittamattomassa optimissa.

Täten epäyhtälörajoitetulle tehtävälle:

$$\max_{x,y} f(x,y), \quad \text{s.t.}$$
$$g(x,y) \leq c$$

kirjoitamme jälleen Lagrangen funktion seuraavasti:

$$\mathcal{L}(x,y,\lambda) = f(x,y) - \lambda\left[g(x,y) - c\right] \tag{7}$$

mutta vaatimamme ensimmäisen kertaluvun ehdot (gradientin nollaehto + CS-ehto) ovat nyt:

FOC

$$\mathcal{L}_x = f_x - \lambda g_x = 0 \tag{8}$$
$$\mathcal{L}_y = f_y - \lambda g_y = 0 \tag{9}$$

sekä lisäksi

$$\lambda\left[g(x,y) - c\right] = 0 \tag{10}$$
$$\lambda \geq 0 \tag{11}$$
$$g(x,y) \leq c \tag{12}$$

Ehdot (8) ja (9) ovat Lagrangen funktion osittaisderivaatat valintamuuttujien suhteen, ehto (10) on complementary slackness -ehto, jonka tulkinnan kävimme jo läpi. Ehto (11) vaatii sen, että rajoitteen ja funktion gradientit osoittavat samaan suuntaan ja viimeinen ehto (12) on vain alkuperäinen epäyhtälörajoitteemme, jonka tietenkin tulee täyttyä optimipisteessä.

Huomoikaa myös se, että kun merkitsemme rajoitteitamme muodossa

$$g(x, y) \leq c$$

Silloin on luontevaa myös kirjoittaa päätösmuuttujien ei-negatiivisuusehdot $(x \geq 0, y \geq 0)$ muotoon:

$$-x \leq 0$$
$$-y \leq 0$$

Tässä esityksessä olemme jatkaneet kahden muuttujan tapaustamme. Useamman muuttujan ja rajoitteen tapaus etenee samalla tavalla, ja jokaiselle epäyhtälörajoitteelle on muodostettava oma CS-ehtonsa jne. Ennen kuin yleistämme ehdot käsittämään kumpiakin rajoitteita, käsitellään lyhyesti nk. NDCQ -ehto.

1.3 Rajoitteiden ei-degeneroituvuusehto (NDCQ)

Tähän asti emme ole puhuneet mitään tästä ehdosta. NDCQ-ehto on kuitenkin tarkastettava aina kun etsimme ratkaisua rajoitettuun optimointiongelmaan. Yhtälörajoitetun tehtävämme yhteydessä puhuimme siitä, että etsimme korkeinta funktion f tasokäyrää, joka *sivuaa* rajoitetta $h(x, y) = c$.

Käytetään samaa esimerkkiä NDCQ -ehdon tarkasteluun[2] .

Implisiittifunktiolause (ks. luentomoniste 1.4) antaa f:n tasokäyrän kulmakertoimeksi optimissa $\mathbf{x}^*$ (assistentti on laiska, seuraavassa merkintä $f_x = \frac{\partial f(\mathbf{x}^*)}{\partial x}$):

$$-\frac{f_x}{f_y}$$

kun taas rajoitteen h kulmakerroin on

$$-\frac{h_x}{h_y}$$

Koska etsimme sivuamispistettä, tällöin pitää päteä

$$\frac{f_x}{f_y} = \frac{h_x}{h_y}$$

merkitään $f_x/f_y = \mu = h_x/h_y$, tällöin saamme yhtälöparin

$$f_x - \mu h_x = 0$$
$$f_y - \mu h_y = 0$$

Näemme, että nämä ovat alkuperäisen yhtälörajoitteen kanssa täsmälleen ensimmäisen kertaluvun ehtomme (1), (2), (3).

[2]Katsokaa myös tarkkaan luentomateriaali II 1-3. Aina kun törmäätte johonkin, jossa puhutaan siitä, mikä on Jacobin matriisin $D(h(\mathbf{x}^*))$ rangi/aste, silloin on kyseessä NDCQ!

Tämä menetelmä ei kuitenkaan toimi, jos optimissa $\mathbf{x}^*$:

$$h_x = 0 \ \text{ ja } \ h_y = 0$$

Tästä saamme tässä tapauksessa rajoitteiden ei-degeneroituvuusehdon, eli NDCQ -ehdon.

Esimerkki 1.

$$\max_{x,y} f(x,y), \quad s.t.$$
$$h(x,y) = c$$

tämän tehtävän NDCQ -ehto on se, että optimissa (x^*, y^*), *rajoitteen osittaisderivaatat* h_x, h_y *eivät saa olla kummatkin nollia.*

Luennolla esiin tullut seikka:

Tämän tehtävän kohdalla (yksi yhtälörajoite) on helppo nähdä, että NDCQ -ehto pelkistyy siihen, että optimissa $\mathbf{x}^*$ rajoitteen gradientti $\nabla h(\mathbf{x}^*) \neq \mathbf{0}$. Toisin sanoen, ratkaisu ei saa olla rajoitteen *kriittinen piste*, käyttäen terminologiaa rajoittamattomasta optimoinnista. Rajoitteen kriittinen piste tarkoittaa kuitenkin **eri asiaa** silloin kun meillä on tehtävä, jossa on monta yhtälö- tai epäyhtälörajoitetta.

1.3.1 NDCQ ja monta yhtälö- tai epäyhtälörajoitetta

Monen yhtälö- tai epäyhtälörajoitteen tapauksessa NDCQ -ehto tarkoittaa sitä, että optimissa $\mathbf{x}^*$ meidän on tarkasteltava sitovien rajoitteiden Jacobin matriisin astetta. **Huom!** Kirja (Simon and Blume, 1994) määrittelee rajoitteiden kriittiseksi pisteeksi nyt sellaisen pisteen, jossa tämän matriisin aste ei ole täysi. Käsitellään ensin NDCQ -ehto kaikissa tapauksissa ja esitetään sitten näiden muistamiseen helpompi keino.

Yhtälörajoitettu tehtävä jossa m rajoitetta:

Rajoitteista muodostetun Jacobin matriisin $D(h(\mathbf{x}^*))$ on oltava **täyttä astetta, ts. sen rangin on oltava m**.

Epäyhtälörajoitettu tehtävä jossa n sitovaa rajoitetta:

Sitovista epäyhtälörajoitteista muodostetun Jacobin matriisin $D(g(\mathbf{x}^*))$ on oltava täyttä astetta, ts. sen rangin on oltava **n**.

m yhtälörajoitetta ja n sitovaa epäyhtälörajoitetta:

Jacobin matriisi, jossa ylemmät rivit ovat sitovien epäyhtälörajoitteiden $D(g(\mathbf{x}^*))$ ja alemmat rivit yhtälörajoitteiden $D(h(\mathbf{x}^*))$

$$\begin{bmatrix} D(g(\mathbf{x}^*)) \\ D(h(\mathbf{x}^*)) \end{bmatrix}$$

on astetta $\mathbf{n} + \mathbf{m}$ eli suurinta mahdollista astetta.

Tästä on helppo nähdä, että on turha sekoittaa mieltään gradienteilla tai kriittisillä pisteillä ym. Tehtävässä, jossa on vain yksi rajoite, silloin rajoitteiden Jacobin matriisi on vain vektori, ja ehto sanoo että tämän ”matriisin” aste pitää olla 1. Toisin sanoen, se ei saa olla nollavektori.

Miten ihmeessä tällaista voi muistaa?

Miten muodostat NDCQ -ehdon

Voit ajatella, että NDCQ -ehto liittyy aina jonkin rajoitteisiin liittyvän Jacobin matriisin asteeseen. Täten:

1. Ratkaise ensin optimipisteet.
2. Yhtälörajoitetussa tehtävässä asia on helppo: yhtälörajoitteiden on oltava aina sitovia, joten jos $\mathbf{m}$ rajoitetta, silloin rajoitteiden Jacobin matriisi pitää olla astetta $\mathbf{m}$.
3. Epäyhtälöiden kanssa, tarkista jokaiselle optimipisteelle sitovien rajoitteiden lukumäärä $\mathbf{n}$, josta jälleen sitovien epäyhtälörajoitteiden Jacobin matriisin pitää olla astetta $\mathbf{n}$ (ei-sitovista rajoitteista ei tarvitse välittää).
4. Jos tehtävässä on kumpaakin rajoitetyyppiä, tutki kummatkin aiemmat kohdat. Muodosta Jacobin matriisi laittamalla kummatkin aiemman kohdan matriisit päällekkäin. Tämän matriisin aste pitää olla $\mathbf{m} + \mathbf{n}$.

1.4 Yhtälö- ja epäyhtälörajoitteet

Kun olemme käyneet läpi kummankin rajoitetyypin erikseen, käydään nopeasti tehtävä jossa on kumpaakin rajoitetyyppiä. Olkoon meillä tuttu tehtävämme (nyt kummallakin ehdolla):

$$\begin{aligned} \max_{x,y} f(x,y), \quad \text{s.t.} \\ h(x,y) = c \\ g(x,y) \leq d \end{aligned}$$

Lagrangen funktio tälle tehtävälle on tietenkin

$$\mathcal{L}(x,y,\mu,\lambda) = f(x,y) - \mu\,[h(x,y) - c] - \lambda\,[g(x,y) - d]$$

ja ensimmäisen kertaluvun ehdot ovat yhdistelmä aiemmista tapauksista tuttuja ehtoja:

$$\begin{aligned} \mathcal{L}_x = f_x - \mu h_x - \lambda g_x = 0 \quad (13) \\ \mathcal{L}_y = f_y - \mu h_y - \lambda g_y = 0 \quad (14) \\ \mathcal{L}_\mu = h(x,y) - c = 0 \quad (15) \\ \lambda\,[g(x,y) - d] = 0 \quad (16) \\ \lambda \geq 0 \quad (17) \\ g(x,y) \leq d \quad (18) \end{aligned}$$

Rajoitteiden NDCQ -ehto käytiin aiemmin jo läpi.

1.5 Optimointitehtävien ratkaiseminen

Kuten harjoituksista on käynyt ilmi, on optimointiongelmien ratkominen joskus työlästä, varsinkin jos tehtävässä on monia epäyhtälörajoitteita. Siksi ratkaisussa kannattaa yleensä aloittaa pienellä viekkaudella, eli yrittää ensin päätellä mitkä rajoitteista voisivat olla sitovia ja mitkä ei-sitovia [3]. Valitettavasti kuitenkin jossain tapauksissa ratkaisemiseen ei ole muuta menetelmää kuin brute force -menetelmä, jossa joudumme yksinkertaisesti käymään läpi kaikki tapaukset ja etsimään mahdolliset ristiriidat.

Rajoitetun optimointitehtävän ratkaisu

1. Kirjoita tehtävä optimointiongelman muotoon, mikäli se ei ole sitä valmiiksi.
2. Tarkastele, olisiko loogista aloittaa jostakin kandidaatista, kuten vaikkapa sisäpisteratkaisusta.
3. Jos kandidaatti löytyy intuitiivisesti, täyttää ensimmäisen kertaluvun ehdot ja NDCQ -ehdon, pohdi ovatko ensimmäisen kertaluvun ehdot riittävät (millainen on käypä joukko, onko meillä konkaavi/konveksi tavoitefunktio jne.).
4. Jos tämä menetelmä ei toimi, silloin joudut tarkastelemaan kaikkia kandidaattipisteitä (saamiasi kriittisiä pisteitä) ja kaikkia sitovien/ei-sitovien rajoitteiden yhdistelmiä.
5. Käy läpi kaikki yhdistelmät, ja etsi Lagrangen kertoimien arvot, jotka eivät johda ristiriitaan (vrt. harjoitus 7)
6. Tarkasta NDCQ -ehto näissä pisteissä, ja mikäli tarpeen
7. Tutki toisen kertaluvun ehdoilla onko kyseessä maksimi/minimi.

[3] esimerkiksi hyödyn maksimointitehtävä konkaavilla hyötyfunkiolla ja ei-negatiivisuusrajoitteilla hyödykkeiden kulutukselle, tässä tapauksessa tarkastelu kannattanee aloittaa pisteestä, jossa kaikki kulutukset ovat positiivisia jne..

1.6 Toisen kertaluvun ehdot, Simon and Blume (1994) kappale 19.3

Rajoittamattomassa optimoinnissa toisen kertaluvun ehdot tarkoittivat Hessen matriisin definiittisyyden tarkastelua. *Rajoitetussa optimoinnissa* työkalu, jota joskus joudumme käyttämään on *reunustettu Hessen matriisi*. Reunustus tulee siitä, että nyt rajoittamattoman optimin sijaan, käypä joukko on rajoitettu, ja meidän tulee ottaa huomioon se, että rajoitteemme määrittävät nyt joukon käypiä pisteitä joista etsimme optimia. Täten meidän tulee tarkistaa Hessen matriisin definiittisyys *tietyssä käyvässä joukossa*.

1.6.1 Miten reunustettu Hessen matriisi muodostetaan?

Reunustettu Hessen matriisin muodostamiseen tarvitsemme kaksi asiaa:

1. Optimointiongelman Lagrangen funktion $\mathcal{L}$ ja

2. Ongelman sitovat yhtälö ja/tai epäyhtälörajoitteet h, g

Reunustettu Hesse muodostetaan ensin ottamalla Lagrangen funktion Hessen matriisi $D_x^2\mathcal{L}(\mathbf{x}^*)$ ja reunustamalla se sitovien yhtälö ja/tai epäyhtälörajoitteiden Jacobin matriisilla. Täten esimerkiksi yhden yhtälörajoitteen tehtävällemme reunustettu Hesse olisi:

$$D^2\mathcal{L} = \begin{bmatrix} \mathbf{0} & D\mathbf{h}(\mathbf{x}^*) \\ [D\mathbf{h}(\mathbf{x}^*)]^T & D_x^2\mathcal{L} \end{bmatrix}$$

eli siis

$$D^2\mathcal{L} = \begin{bmatrix} 0 & h_x & h_y \\ h_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ h_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

Jos taas optimissa sitoo yksi epäyhtälörajoite g, silloin matriisi on:

$$D^2\mathcal{L} = \begin{bmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{bmatrix}$$

Jos taas optimissa sekä yhtälö, että epäyhtälörajoitteet sitovat, silloin matriisi on muotoa:

$$D^2\mathcal{L} = \begin{bmatrix} & \mathbf{0} & & \begin{bmatrix} D\mathbf{g}(\mathbf{x}^*) \\ D\mathbf{h}(\mathbf{x}^*) \end{bmatrix} \\ [D\mathbf{g}(\mathbf{x}^*)]^T & & [D\mathbf{h}(\mathbf{x}^*)]^T & D_x^2\mathcal{L} \end{bmatrix}$$

1.6.2 Mitä tästä matriisista tulee tutkia?

Haluamme määrittää reunustetun Hessen matriisin definiittisyyden tietyssä joukossa, jonka määrittäävät optimissa sitovat rajoitteet. Olkoon meillä tehtävässä n valintamuuttujaa, k sitovaa yhtälörajoitetta ja e sitovaa epäyhtälörajoitetta. Toisen kertaluvun riittävä ehto maksimille on se, että tarkistamme seuraavat asiat:

1. Suurimpien johtavien pääminoreiden merkkien vaihtelu ja
2. $\det D^2\mathcal{L}$ pitää olla samanmerkkinen kuin termi $(-1)^n$.

1.6.3 Monta johtavaa pääminoria pitää tarkastella?

- Yhtälörajoitetussa tehtävässä tarkastamme suurimman $(n-k)$ johtavan pääminorin merkin
- Epäyhtälörajoitetussa tehtävässä $(n-e)$ ja

- Yhtälö- ja epäyhtälörajoitetussa tehtävässä $(n - e - k)$

1.7 Ensimmäisen kertaluvun ehtojen riittävyys

Kuten harjoitukset 6 & 7 ovat näyttäneet, toisen kertaluvun ehtojen tarkistaminen on työlästä. Useimmiten emme halua joutua tekemään sitä, jos voimme välttyä siltä. Tämän takia usein annettua optimointitehtävää voi analysoida hetken ja selvittää ovatko ensimmäisen kertaluvun ehdot sekä *välttämättömät* että *riittävät* ehdot optimille. Tässä monisteessa emme todista seuraavia lauseita, mutta ensimmäisen kertaluvun ehdot ovat myös riittävät jos seuraavat ehdot täyttyvät:

FOC riittävät kun

Optimointitehtävän käypä joukko X on konveksi joukko, ja tavoitefunktio f on konkaavi tai aidosti kvasikonkaavi funktio. Weierstrassin lause antaa ehdot sille, milloin optimointiongelmalla on ylipäätään ratkaisu (ks. luentomoniste tai Simon and Blume (1994) kappale 30.)

2 Differenssiyhtälöt

Monesti taloustieteessä on tarpeellista kirjoittaa malli diskreetissä ajassa jatkuvan ajan sijaan. Esimerkkinä tästä vaikkapa makrotaloustieteessä käytettävät kasvumallit, RCK-kasvumalli ja monet muut mukavat mallit. Tällöin matemaattinen olio, joka kohdataan on nimeltään *differenssiyhtälö*, ja toki tarvitsemme myös työkalut sellaisten taltuttamiseen. Kurssilla keskitymme lineaaristen, ensimmäisen kertaluvun differenssiyhtälöiden ratkaisuun, perehtyminen korkeamman kertaluvun yhtälöjärjestelmiin jätetään oman harrastuneisuuden varaan. Kuten kurssin ensimmäisellä osalla, aloitetaan käymällä läpi lineaarialgebraa ja matriisien diagonalisointia. Vaikka tämä vaikuttaa nyt hieman irralliselta, kuten hyvässä mysteerissä ainakin, salaisuus paljastuu lopussa kun pääsemme differenssiyhtälöjärjestelmien kimppuun.

2.1 Matriisin ominaisarvot ja ominaisvektorit

Ominaisarvot

Sanomme neliömatriisin $\mathbf{A}$ ominaisarvoiksi λ sellaisia lukuja, jotka ovat ratkaisuja seuraavaan yhtälöön:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{19}$$

Eli ominaisarvot[4] ovat siis sellaisia lukuja (reaali- tai kompleksi-), että jos vähennämme ne matriisin $\mathbf{A}$ lävistäjältä, tulee matriisista $\mathbf{A}$ singulaarinen matriisi. Yhtälöä (19) kutsumme *karakteristiseksi yhtälöksi* tai polynomiksi, ja tällä tavalla voimme aina ratkaista jonkun neliömatriisin $\mathbf{A}$ ominaisarvot.

[4] Kurssisivulla on hyvä linkki videoon, jossa selostetaan juurta jaksaen ominaisarvoja ja -vektoreita. Suosittelen tutustumaan tähän tenttivalmistautumisessa.

Ominaisvektorit

Neliömatriisin $\mathbf{A}$ *ominaisvektori* taas on vektori $\mathbf{v}$, $\mathbf{v} \neq \mathbf{0}$, joka toteuttaa seuraavan yhtälön

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \tag{20}$$

Huom! Nollavektori on triviaalisti aina ominaisvektori, mutta se ei ole kovin kiinnostava vektori, joten siksi vaadimme aina että $\mathbf{v} \neq 0$. Voimme ajatella tätä niin, että kun kuvaamme vektorin $\mathbf{v}$ matriisilla $\mathbf{A}$, ainoa tulos tästä on se, että vektori $\mathbf{v}$ skaalautuu vain matriisin $\mathbf{A}$ ominaisarvojen λ suuntaan. Ominaisvektorin yhtälöstä (20) näemme myös heti, että jos $\mathbf{v}$ on matriisin $\mathbf{A}$ ominaisvektori, silloin myös $r\mathbf{v}, r \in \mathbb{R}$ on matriisin $\mathbf{A}$ ominaisvektori. Kurssin ensimmäisen osan asiat muistaville tämä huomio on varmastikin tuttu, sillä ominaisarvon määritelmä sanoo suoraan että ryhmän (20) kerroinmatriisi on singulaarinen.

2.1.1 Muunnosmatriisi ja matriisin diagonalisointi

Jos meillä on neliömatriisi $\mathbf{A}$, jolla on ominaisarvot λ, haluamme etsiä matriisin $\mathbf{P}$ siten että seuraava yhtälö pätee:

$$\mathbf{P^{-1}AP} = \mathbf{D} \tag{21}$$

jossa $\mathbf{D}$ on lävistäjämatriisi, jonka lävistäjällä on matriisin $\mathbf{A}$ ominaisarvot. Toisin sanoen, etsimme jotakin matriisia $\mathbf{P}$, joka muuntaa matriisin $\mathbf{A}$ yhtälön (21) avulla lävistäjämatriisiksi. Millainen matriisin $\mathbf{P}$ pitäisi olla?

Tarkastellaan 2×2 matriisia $\mathbf{A}$ ja aloitetaan yhtälöstä (21). Jos

$$\mathbf{P^{-1}AP = D}$$

silloin selkeästi $\mathbf{AP} = \mathbf{PD}$. Voimme kirjoittaa tämän seuraavalla tavalla:

$$\mathbf{AP} = \begin{bmatrix} \mathbf{p_1} & \mathbf{p_2} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

jossa merkkaamme $\mathbf{p_i}$ matriisin $\mathbf{P}$ sarakkeita. Saamme:

$$\begin{bmatrix} \mathbf{Ap_1} & \mathbf{Ap_2} \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{p_1} & \lambda_2\mathbf{p_2} \end{bmatrix}$$

jolloin $\mathbf{Ap_1} = \lambda_1\mathbf{p_1}$ ja $\mathbf{Ap_2} = \lambda_2\mathbf{p_2}$. Tällöin matriisin $\mathbf{P}$ sarakkeet *ovat matriisin $\boldsymbol{A}$ ominaisvektoreita*, sillä ensimmäinen sarake toteuttaa:

$$\mathbf{p_1} : (\mathbf{A} - \boldsymbol{\lambda_1})\mathbf{p_1} = \mathbf{0}$$

ja toinen sarake toteuttaa:

$$\mathbf{p_2} : (\mathbf{A} - \boldsymbol{\lambda_2})\mathbf{p_2} = \mathbf{0}$$

Näin siis muunnosmatriisi $\mathbf{P}$ muodostuu alkuperäisen matriisin $\mathbf{A}$ ominaisvektoreista, jossa $\mathbf{p_1}$ on ensimmäistä ominaisarvoa λ_1 vastaava ominaisvektori jne.

Muunnosmatriisi P

Olkoon meillä neliömatriisi $\mathbf{A}$ ja etsimme matriisia $\mathbf{P}$, joka toteuttaa

$$\mathbf{P^{-1}AP} = \mathbf{D}$$

Tällöin $\mathbf{P}$ muodostetaan matriisin $\mathbf{A}$ ominaisvektoreista $\mathbf{v_{\lambda_i}}$ seuraavasti:

$$\mathbf{P} = \begin{bmatrix} \mathbf{v_{\lambda_1}} & \mathbf{v_{\lambda_2}} & \cdots & \mathbf{v_{\lambda_n}} \end{bmatrix}$$

eli $\mathbf{P}$:n sarakkeet ovat kyseistä ominaisarvoa vastaava $\mathbf{A}$:n ominaisvektori.

2.2 1. kl differenssiyhtälöt, ratkaisumenetelmät

Jos meillä on differenssiyhtälö

$$x_{k+1} = ax_k$$

on tästä helppo nähdä, että jollekin alkuarvolle x_0:

$$\begin{aligned} x_1 &= ax_0 \\ x_2 &= a^2x_0 \\ &\vdots \\ x_k &= a^kx_0 \end{aligned}$$

ja tällöin yhtälön pitkän aikavälin käytös riippuu ainoastaan vakiosta a.

Jos $\mid a \mid > 1$, silloin yhtälö räjähtää käsiin, jos taas $\mid a \mid < 1$, järjestelmä suppenee kohti nollaa, ja jos $a = 1$, silloin $x_k = x_0$.

Olkoon meillä nyt joku differenssiyhtälö muuttujalle x, joka näyttää seuraavalta:

$$x_{k+1} = ax_k + b, \quad a \neq 1$$

on tällaisen yhtälön ratkaisu muotoa:

$$x_k = x_k^H + x_P$$

jossa x_k^H on homogeenisen differenssiyhtälön ratkaisu, ja x_P on jokin differenssiyhtälön erityisratkaisu. Homogeeninen yhtälö saadaan, kun vakio $b = 0$, jolloin differenssiyhtälön ratkaisu on iteroimalla:

$$x_k^H = a^k C$$

jossa C on jokin tuntematon vakio. Erityisratkaisu saadaan vaikkapa merkitsemällä

$$\overline{x} = a\overline{x} + b \Leftrightarrow \overline{x} = \frac{b}{1-a}$$

nyt siis

$$x_k = a^k C + \frac{b}{1-a}$$

Jos meillä olisi jokin alkuarvo, kuten vaikkapa että $x_0 = 1$, silloin voisimme ratkaista vakion C seuraavalla tavalla:

$$
\begin{aligned}
t = 0 : x_0 = a^0 C + \frac{b}{1-a} \\
1 = C + \frac{b}{1-a} \Leftrightarrow C = \frac{1-a-b}{1-a}
\end{aligned}
$$

tällöin siis differenssiyhtälön ratkaisu on:

$$x_k = a^k \left(\frac{1-a-b}{1-a} \right) + \frac{b}{1-a}$$

2.3 1. kl lineaarinen differenssiyhtälöjärjestelmä

Differenssiyhtälöjärjestelmä on rekursiivinen yhtälöjärjestelmä, jossa muuttujan $\mathbf{z}$ arvo ajan hetkellä $k+1$ riippuu sen aiemmasta arvosta ajanhetkellä k. Esimerkiksi seuraava on differenssiyhtälösysteemi:

Esimerkki 2.

$$\mathbf{z_{k+1}} = \mathbf{A}\mathbf{z_k} \tag{22}$$

Jos $\mathbf{z}$ on 1x1, silloin olemme aiemmassa, yhden muuttujan tapauksessa. Silloin kuin $\mathbf{z}$ on vektori, on meillä käsissämme differenssiyhtälöjärjestelmä. Käsitellään seuraavassa tapausta $\mathbf{z} = \begin{bmatrix} x & y \end{bmatrix}^T$, sillä se riittää havainnollistamaan tarvitsemamme tekniikat.

2.3.1 Diagonaalinen kerroinmatriisi

Palataan systeemiin (22). Aiemman kohdan perusteella on varsin helppo havaita, että jos kerroinmatriisimme $\mathbf{A}$ olisi *lävistäjämatriisi*, eli sen ei-diagonaalialkiot olisivat kaikki nollia, silloin systeemi (22) olisi helppo ratkaista. Itse asiassa osaisimme tehdä sen helposti.

Esimerkki 3. *Olkoon* $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

Systeemi (22) näyttää nyt seuraavalta:

$$\mathbf{z_{k+1}} = \mathbf{A}\mathbf{z_k} \tag{23}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix} \tag{24}$$

eli siis

$$\begin{aligned} x_{k+1} &= x_k \\ y_{k+1} &= 2y_k \end{aligned}$$

Tämän ratkaisu on selkeästi

$$\begin{aligned} x_k &= 1^k x_0 \\ y_k &= 2^k y_0 \end{aligned}$$

eli

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} 1^k & 0 \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{25}$$

Jos tiedätte jo jotain matriisien ominaisarvoista, tässä ratkaisussa matriisi $\begin{bmatrix} 1^k & 0 \\ 0 & 2^k \end{bmatrix}$ saattaa näyttää tutulta, ja oikeastaan voisimme merkitä sitä eri tavalla. Tehdään se kuitenkin vasta myöhemmin. Koska kerroinmatriisi $\mathbf{A}$ oli alunperinkin jo meille mukavassa muodossa, oli ryhmän ratkaisu helppoa.

Entä jos kerroinmatriisi ei olekaan diagonaalimatriisi heti kättelyssä?

2.3.2 Ei-diagonaalinen kerroinmatriisi

Olkoon meillä nyt ryhmä:

$$\mathbf{z_{k+1}} = \mathbf{B}\mathbf{z_k} \tag{26}$$

jossa $\mathbf{B}$ on jokin symmetrinen neliömatriisi $\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, jossa $a, b, c, d \in \mathbb{R}$, mutta $a, b, c, d \neq 0$. Nyt meillä on kytketty järjestelmä, koska jos kirjoitamme systeemin auki seuraavasti:

$$x_{k+1} = ax_k + by_k$$
$$y_{k+1} = cx_k + dy_k$$

näemme, että x riippuu *kummankin* muuttujan aiemmasta arvosta, kuten myös muuttuja y. Haluaisimme kuitenkin ratkaista ryhmän samalla tavalla kuin aiemmassa tapauksessa, joten joudumme etsimään jonkun keinon jolla muuntaa matriisi $\mathbf{B}$ diagonaalimatriisiksi. Teemme tämän tekemällä *muuttujanvaihdoksen*:

$$\mathbf{z} = \mathbf{PZ} \tag{27}$$

jossa käytämme aiemmin johtamaamme muunnosmatriisia $\mathbf{P}$ siten että[5]:

$$\mathbf{P^{-1}BP} = \mathbf{D}$$

[5]Vaadimme tietenkin, että matriisilla $\mathbf{P}$ on sopivat ominaisuudet, kuten kääntyvyys ym.

Koska nyt muuttujanvaihtomme kanssa voimme edetä seuraavalla tavalla: Alkuperäinen systeemi on

$$\mathbf{z_{k+1}} = \mathbf{Bz_k} \tag{28}$$

tehdään muuttujanvaihdos (27), kirjoitetaan systeemi uudestaan:

$$\mathbf{PZ_{k+1}} = \mathbf{BPZ_k}$$

kerrotaan vasemmalta käänteismatriisilla $\mathbf{P^{-1}}$:

$$\mathbf{Z_{k+1}} = \mathbf{P^{-1}BPZ_k}$$

ja koska $\mathbf{P^{-1}BP} = \mathbf{D}$, silloin

$$\mathbf{Z_{k+1}} = \mathbf{DZ_k}$$

ratkaistaan iteroimalla kuten aiemmin:

$$\mathbf{Z_k} = \mathbf{D^k Z_0}, \quad \text{jossa} \quad \mathbf{D^k} = \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} \tag{29}$$

meidän pitää vielä palata alkuperäisiin muuttujiin $\mathbf{z}$, ja tämä onnistuu kun huomaamme että

$$\mathbf{z} = \mathbf{PZ} \Leftrightarrow \mathbf{Z} = \mathbf{P^{-1}z}$$

jolloin sijoittamalla yhtälöön (29) saadaan:

$$\mathbf{z_k} = \mathbf{P}\mathbf{D^k}\mathbf{P^{-1}}\mathbf{z_0}, \quad \text{jossa} \quad \mathbf{D^k} = \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} \tag{30}$$

tämä on järjestelmämme (26) ratkaisu jollekin alkuarvolle $\mathbf{z_0}$, jonka saimme käyttämällä hyväksi matriisin $\mathbf{B}$ ominaisarvoja.

Differenssiyhtälöjärjestelmiä ratkaistaessa haluamme aina tehdä asiat itsellemme mahdollisimman helpoiksi, eli haluaisimme ratkoa "irtikytkettyjä" ryhmiä, joissa kerroinmatriisi on lävistäjämatriisi. Mikäli matriisi ei ole tällainen, otamme jonkun lineaarikuvauksen muuttujista (muuttujanvaihdos (21)), joka muuntaa vanhan ryhmän uudeksi, "irtikytketyksi" ryhmäksi, jonka voimme ratkaista iteroimalla. Kun olemme saaneet tämän ratkaistua, voimme ottaa vain käänteisen lineaarikuvauksen ja siirtyä takaisin alkuperäisiin muuttujiin. Prosessi, jota käytämme on kuvattu harjoituksen 8 mallivastauksissa.

2.4 Markov-prosessit

Stokastinen prosessi on sääntö, joka kertoo meille sen, miten jonkin järjestelmän tila muuttuu sen tilahistorian funktiona. Markov-prosessi on erityistapaus tästä, ja siinä järjestelmän tila hetkellä $k+1$ riippuu ainoastaan tilasta ajanhetkellä k, eli ainoa tärkeä asia on edellisen periodin tila (vrt. harjoitus 9 tehtävä 4). Täten, mikäli $\mathbf{x_{k+1}}$ on järjestelmän tilavektori ajanhetkellä $k+1$, on Markov-prosessi seuraavanlainen differenssiyhtälöjärjestelmä:

$$\mathbf{x_{k+1}} = \mathbf{M}\mathbf{x_k} \tag{31}$$

ja siirtymätodennäköisyyksien $(n \times n)$ neliömatriisia $\mathbf{M}$ sanomme *Markov-matriisiksi.* Koska matriisin alkiot ovat siirtymätodennäköisyyksiä tilojen

välillä, tulee jokaisen alkion m_{ij} toteuttaa $m_{ij} \geq 0$, sillä todennäköisyydet eivät voi olla negatiivisia. Matriisissa rivin i sarakkeen j luku kertoo todennäköisyyden, jolla seuraavan periodin tila on i kun nykyisen periodin tila on j, josta saamme ehdon että matriisin $\mathbf{M}$ sarakesumman pitää olla yksi jokaiselle sarakkeelle.
On helppo näyttää, että mikäli $m_{ij} \geq 0$, $\sum_{i=1}^{n} m_{ij} = 1$ kaikille j, silloin

$\lambda_1 = 1$ on yksi matriisin $\mathbf{M}$ ominaisarvo (voitte tehdä tämän jos haluatte kertausta lineaarialgebrasta).

Yleisesti pätee, mikäli Markov-matriisi $\mathbf{M}$ on nk. tavanomainen, että $\lambda_1 = 1$ on eräs sen ominaisarvoista, ja muut ominaisarvot $\lambda_j \in (-1, 1)$. Tämä on meille kiinnostava tieto siksi, että tällöin prosessi (31) suppenee kohti ominaisarvoa $\lambda_1 = 1$ vastaavaa ominaisvektoria.

Lause 1. *(S&B 23.6): Olkoon* $\boldsymbol{A}$ $n \times n$ *-matriisi, jolla on n erillistä reaalista ominaisarvoa* $\lambda_1, \cdots, \lambda_n$ *ja vastaavat ominaisvektorit* $\mathbf{v_1}, \cdots, \mathbf{v_n}$ *. Tällöin yleinen ratkaisu differenssiyhtälöjärjestelmälle* $\mathbf{z_{k+1}} = \mathbf{A}\mathbf{z_k}$ *on:*

$$\mathbf{z_k} = c_1\lambda_1^k\mathbf{v_1} + c_2\lambda_2^k\mathbf{v_2} + \cdots + c_n\lambda_n^k\mathbf{v_n}$$

Esimerkki 4. *Olkoon meillä prosessi:*

$$\mathbf{x_{k+1}} = \mathbf{M}\mathbf{x_k} \tag{32}$$

jossa $\boldsymbol{M}$ *on edellä mainitut ehdot täyttävä tilansiirtomatriisi (Markov-matriisi). Lauseen 1 perusteella osaamme jo nyt sanoa jotakin prosessin suppenemisesta. Koska matriisin* $\boldsymbol{M}$ *suurin ominaisarvo on* $\lambda_1 = 1$*, ja loput ovat rajattu avoimelle välille* $(-1, 1)$*, silloin yleisessä ratkaisussa, kun annamme* $k \to \infty$*:*

$$\lambda_2^k, \cdots, \lambda_n^k \to 0$$

ja prosessi suppenee kohti ominaisarvoa λ_1 *vastaavan ominaisvektorin määrittämää rajajakaumaa.*

Esimerkit Markov-prosesseista löytyvät harjoituksesta 9

Viitteet

Chiang, A. C. (1984): *Fundamental methods of mathematical economics.* McGraw-Hill.

De la Fuente, A. (2000): *Mathematical methods and models for economists.* Cambridge University Press.

Dixit, A. K., J. J. Sherrerd, et al. (1990): *Optimization in economic theory.* Oxford University Press on Demand.

Luenberger, D. G. (1997): *Optimization by vector space methods.* John Wiley & Sons.

Simon, C. P., and L. Blume (1994): *Mathematics for economists*, vol. 7. Norton New York.