

Epälineaarista ohjelmoinnista

Pekka Vuollo

Pro gradu-tutkielma

Sivulaudatur-opinnot

Matemaattisten tieteiden laitos

Oulun yliopisto

Toukokuu 2005

Sisältö

Alkusanat	iii
Johdanto	1
1 Klassinen rajoittamaton optimointi	3
1.1 Yhden muuttujan funktiot	3
1.2 Useamman muuttujan funktiot	7
1.3 Hessin matriisin definiittisyyden yhteys optimointiin	12
1.3.1 Matriisin definiittisyyden selvittäminen determinant-	
tien avulla	14
1.3.2 Matriisin definiittisyyden selvittäminen ominaisarvo-	
jen avulla	18
1.4 Paikalliset ääriarvot ja satulapiste	21
1.5 Globaalit ääriarvot ja koersiiviset funktiot	26
1.5.1 Lisätietoa globaaleista ääriarvoista	26
1.5.2 Koersiiviset funktiot	27
2 Konvekksi optimointi	31
2.1 Konveksit joukot ja funktiot	31
2.2 Konveksisuuden yhteys optimointiin	33
2.3 Konveksien funktioiden tunnistaminen	38
3 Geometrinen optimointi	44
3.1 Optimointi aritmeettis-geometrisen keskiarvoepäyhtälön avulla	44

3.2	Varsinainen geometrinen optimointi	50
3.2.1	Geometrisen optimoinnin kehittelyä	50
3.2.2	Geometrisen optimoinnin menetelmä	56
Kirjallisuutta		61

Alkusanat

Olen opiskellut Oulun yliopistossa päätoimisesti kemian koulutusohjelmassa, mutta jo pitkään haaveenani on ollut suorittaa laudaturopinnot kokonaisuudessaan myös matemaatiikassa. Tämän tutkielman valmistuttua tuo haave toteutui. Haluankin kiittää lämpimästi kaikkia niitä, jotka ovat olleet apuna ja tukena matematiikan opintojeni eri vaiheissa.

Johdanto

Optimoinnissa pyritään ratkaisemaan ongelmia parhaalla mahdollisella tavalla. Tämä voi tarkoittaa esimerkiksi voiton maksimointia, tappioiden minimointia, riskin minimointia tai vaikkapa työmenetelmien tehokkuuden maksimointia. Luonnontieteellisissä mittauksissa mittalaitteiden parametrit pyritään optimoimaan siten, että saatavat vasteet olisivat mahdollisimman luotettavia. Teollisuudessa taas koko toimintaa pyritään ohjaamaan siten, että saataisiin työkoneet mahdollisimman tehokkaasti käyttöön ja voitot maksimaaliseksi. Ongelmien optimaalista ratkaisua tarvitaan lähes kaikilla sovellutusaloilla.

Optimoinnin historia alkaa jo antiikin Kreikasta, jossa matemaatikot ratkoivat joitakin geometrisia optimointiongelmia. Eukleides esimerkiksi osoitti jo 300 eaa, että neliö on suurin sellaisista suorakulmoista, joiden särmien yhteenlaskettu pituus on kiinnitetty.

Antiikin jälkeen seuraavat edistysaskeleet optimoinnin saralla otettiin vasta 1600-luvulla, kun koko tieteellinen tutkimus kiihtyi huimasti. Tämän tutkielman kannalta mielenkiintoisen havainnon teki Fermat vuonna 1646. Hän ehdotti, että funktion ääriarvoja tulee etsiä gradientin nollakohdista. Newton ja Leibniz puolestaan laativat analyysin, jonka pohjalta onnistuttiin kehittämään variaatiolaskenta. Alkanut kehitys johti siihen, että ensimmäiset varsinaiset optimointialgoritmit julkaisiin viimein 1800-luvulla.

1900-luvun alussa luotiin optimoinnin kannalta keskeiset konveksisuuskäsitteet ja vuonna 1917 julkaistiin ensimmäinen alan varsinainen oppikirja,

Harris Hancockin *Theory of Minima and Maxima*. Toisen maailmasodan jälkeen optimointia alettiin jo soveltaa taloustieteessä ja operaatiotutkimuksessa. Suuri läpimurto oli Simplex-menetelmä, jonka Danzig kehitti lineaaristen optimointitehtävien ratkaisemiseen. Viimeisten 70 vuoden aikana optimointialgoritmien tutkimusta on kiihdyttänyt erityisesti elektronisten laskentamenetelmien nopea kehittyminen. [1]

Optimointitehtäviä luokitellaan nykyään optimoitavien kohdefunktioiden määrän ja laadun sekä rajoitefunktioiden määrän ja laadun mukaan. Lineaarisessa optimoinnissa esimerkiksi kohdefunktioita on yksi ja kohdefunktio sekä rajoitteet ovat kaikki lineaarisia. Epälineaarisessa optimoinnissa taas tällaisia funktioiden lineaarisuuteen liittyviä rajoitteita ei ole. [2]

Tässä tutkielmassa keskitytään epälineaarisen optimoinnin perusteisiin. Ensimmäisessä luvussa tarkastellaan klassista rajoittamatonta optimointia. Teorian kehittäminen aloitetaan yhden muuttujan funktioista, mutta pääpaino on useamman muuttujan funktioiden optimoinnissa. Toisessa luvussa taas käsitellään konveksia optimointia ja kolmannessa luvussa keskitytään vielä konveksin optimoinnin mielenkiintoiseen sovellukseen, geometriseen optimointiin. Optimointimenetelmiä, jotka perustuvat optimointitehtävien iteratiiviseen ratkaisemiseen, ei sen sijaan tämän tutkielman puitteissa käsitellä. Pääasiallisena lähteenä on käytetty teosta [3].

Luku 1

Klassinen rajoittamaton optimointi

Aloitetaan tämä luku tarkastelemalla yhden muuttujan funktioiden optimointia ja laajennetaan sen jälkeen käsittely koskemaan myös useamman muuttujan funktioita. Selvitetään, kuinka optimoinnin kannalta mielenkiintoiset pisteet voidaan löytää ja kuinka niiden laatu saadaan varmistettua.

1.1 Yhden muuttujan funktiot

Esitetään ensin tarpeelliset funktioiden optimointiin liittyvät peruskäsitteet. Tarkastellaan ainoastaan minimiin liittyviä käsitteitä, sillä maksimille ne ovat vastaavanlaiset.

Määritelmä 1.1.1. Oletetaan, että funktio $f(x)$ on reaaliarvoinen ja että se on määritelty välillä I . Väli I voi olla nyt joko ääretön tai äärellinen sekä joko avoin, puoliavoin tai suljettu. Tällöin kyseisellä välillä sijaitseva piste x^* on välin:

- (a) globaali minimi, jos $f(x^*) \leq f(x)$ aina, kun $x \in I$;
- (b) aito globaali minimi, jos $f(x^*) < f(x)$ aina, kun $x \in I$ ja $x \neq x^*$;

- (c) paikallinen minimi, jos on olemassa sellainen positiivinen luku δ , että $f(x^*) \leq f(x)$ aina, kun $x^* - \delta < x < x^* + \delta$;
- (d) aito paikallinen minimi, jos on olemassa sellainen positiivinen luku δ , että $f(x^*) < f(x)$ aina, kun $x^* - \delta < x < x^* + \delta$ ja $x \neq x^*$;
- (e) kriittinen piste, jos derivaattafunktio $f'(x^*) = 0$.

Aletaan nyt tutkimaan yhden muuttujan funktion optimointia. Ensin tulee selvittää, mistä funktion ääriarvokohtia eli minimeitä ja maksimeita kannattaa etsiä.

Lause 1.1.1. *Oletetaan, että funktio $f(x)$ on derivoituva välillä I ja että $x^* \in I$. Jos x^* on funktion paikallinen minimi tai maksimi, niin x^* on joko välin I päätepointe tai $f'(x^*) = 0$.*

Todistus. Oletetaan, että x^* on funktion $f(x)$ paikallinen minimi ja että x^* ei ole välin I päätepointe. Koska funktio on derivoituva välillä I , $f'(x^*)$ on olemassa. Nyt täytyy osoittaa, että $f'(x^*) = 0$.

Derivaatta $f'(x^*)$ on erotusosamäärän raja-arvon avulla ilmaistuna

$$f'(x^*) = \lim_{x \rightarrow x^*} \frac{f(x) - f(x^*)}{x - x^*}. \quad (1.1)$$

Koska x^* on paikallinen minimi, niin määritelmän (1.1.1) nojalla $f(x^*) \leq f(x)$ aina, kun x on riittävän lähellä minimikohtaa x^* . Tällöin kaavassa (1.1) on osoittaja $f(x) - f(x^*) \geq 0$. Jos nyt $x^* < x$, niin $x - x^* > 0$ ja saadaan

$$\frac{f(x) - f(x^*)}{x - x^*} \geq 0.$$

Jos taas $x^* > x$, niin $x - x^* < 0$ ja saadaan

$$\frac{f(x) - f(x^*)}{x - x^*} \leq 0.$$

Kaava (1.1) siis johtaa tulokseen $f'(x^*) \geq 0$ tai $f'(x^*) \leq 0$ riippuen siitä, mistä suunnasta x lähestyy pistettä x^* . Raja-arvoksi saadaan täten välttämättä, että $f'(x^*) = 0$. □

Esimerkki 1.1.1. Tutkitaan funktiota $f(x) = \frac{2}{3}x^3 - 2x^2 + 1$, joka on määritelty aina, kun $x \in \mathbb{R}$. Funktiolla ei ole globaaleja ääriarvoja, sillä

$$\lim_{x \rightarrow +\infty} f(x) = +\infty \quad \text{ja} \quad \lim_{x \rightarrow -\infty} f(x) = -\infty.$$

Kriittiset pisteet eli derivaattafunktion

$$f'(x) = 2x^2 - 4x = x(2x - 4)$$

nollakohdat ovat puolestaan $x = 0$ ja $x = 2$. Jos funktion käyttäytymistä kriittisten pisteiden läheisyydessä tutkittaisiin tarkemmin, nähtäisiin, että pisteen $x = 0$ lähellä funktio $f(x) < f(0) = 1$. Siis piste $x = 0$ on aito paikallinen maksimi. Piste $x = 2$ osoittautuisi puolestaan aidoksi paikalliseksi minimiksi, sillä lähellä sitä funktio $f(x) > f(2) = -\frac{5}{3}$.

Kun yhden muuttujan funktiota aletaan optimoimaan, tulee siis ensin etsiä funktion kriittiset pisteet ja tutkia sitten niiden laatu. On myös tutkittava määrittelyvälin päätepisteiden laatu. Edellisessä esimerkissä olisi voitu kokeilemalla tutkia, ovatko kriittiset pisteet minimeitä tai maksimeita. Kriittisten pisteiden laatu voidaan kuitenkin selvittää myös optimoitavan funktion toiseen derivaattaan perustuvalla menetelmällä, joka esitetään lauseessa (1.1.3). Lauseen todistamisessa tarvitaan apuna Taylorin kehitelmää, joka on todistettu kokonaisuudessaan lähdeoteoksessa [4].

Lause 1.1.2 (Taylorin kehitelmä). *Oletetaan, että $f(x)$, $f'(x)$ ja $f''(x)$ ovat olemassa suljetulla välillä $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$. Oletetaan lisäksi, että x ja x^* ovat kaksi eri pistettä kyseisellä välillä. Tällöin pisteiden x ja x^* välisellä janalla on olemassa sellainen piste z , että*

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{f''(z)}{2}(x - x^*)^2.$$

Esitetään seuraavaksi optimoitavan funktion toiseen derivaattaan perustuva menetelmä, jolla kriittisten pisteiden laatu saadaan selville.

Lause 1.1.3. Oletetaan, että $f(x)$, $f'(x)$ ja $f''(x)$ ovat kaikki jatkuvia välillä I ja että piste $x^* \in I$. Oletetaan lisäksi, että x^* on funktion $f(x)$ kriittinen piste.

- (a) Jos $f''(x) \geq 0$ aina, kun $x \in I$, niin x^* on funktion $f(x)$ globaali minimi;
- (b) Jos $f''(x) > 0$ aina, kun $x \in I$ ja $x \neq x^*$, niin x^* on funktion $f(x)$ aito globaali minimi;
- (c) Jos $f''(x^*) > 0$, niin x^* on funktion $f(x)$ aito lokaali minimi.

Todistus. Oletuksen perusteella $f'(x^*) = 0$. Jos nyt piste $x \in I$ ja $x \neq x^*$, niin Taylorin kehitelmän nojalla

$$f(x) = f(x^*) + \frac{f''(z)}{2}(x - x^*)^2. \quad (1.2)$$

Tässä z on piste, joka sijaitsee pisteiden x ja x^* välisellä janalla.

- (a) Jos $f''(x) \geq 0$ aina, kun $x \in I$, niin myös $f''(z) \geq 0$. Tällöin kaavan (1.2) termi $f''(z)(x - x^*)^2/2 \geq 0$ ja saadaan tulokseksi, että $f(x) \geq f(x^*)$. Siis x^* on globaali minimi.
- (b) Todistus menee kuten edellä kohdassa (a). Nyt vain $f''(x) > 0$, joten tulokseksi saadaan epäyhtälö $f(x) > f(x^*)$. Täten x^* on aito globaali minimi.
- (c) Funktio $f''(x)$ on oletuksen nojalla jatkuva ja nyt $f''(x^*) > 0$. Tällöin on olemassa sellainen positiivinen luku δ , että $f''(x) > 0$ aina, kun tarkastellaan väliä $x^* - \delta < x < x^* + \delta$. Siis, jos piste x on riittävän lähellä pistettä x^* mutta kuitenkin $x \neq x^*$, niin termi $f''(z)(x - x^*)^2/2 > 0$. Näin ollen pisteen x^* välittömässä läheisyydessä $f(x) > f(x^*)$ ja kyseessä on aito lokaali minimi. $\square$

Huomautus. Edellisen lauseen (1.1.3) tulokset pätevät luonnollisesti myös maksimeille. Korvataan vain kohtien (a), (b) ja (c) merkinnät $f''(x) \geq 0$, $f''(x) > 0$ ja $f''(x^*) > 0$ merkinnöillä $f''(x) \leq 0$, $f''(x) < 0$ ja $f''(x^*) < 0$.

Lisäksi on huomattava tilanne, jossa x^* on funktion $f(x)$ kriittinen piste, mutta $f''(x^*) = 0$. Tällöin lause (1.1.3) ei anna tietoa kyseisen kriittisen pis-

teen laadusta, vaan funktion käyttäytymistä tällaisen pisteen läheisyydessä tulee tutkia muulla tavalla.

Esimerkki 1.1.2. Tarkastellaan funktiota $f(x) = \frac{2}{3}x^3 - 2x^2 + 1$, jota tutkittiin jo esimerkissä (1.1.1). Tällöin jo todettiin, että funktiolla ei ole globaaleja ääriarvoja. Sen sijaan piste $x = 0$ on funktion aito lokaali maksimi ja piste $x = 2$ taas aito lokaali minimi. Varmistetaan vielä kriittisten pisteiden $x = 0$ ja $x = 2$ laatu lauseessa (1.1.3) esitetyllä optimoitavan funktion toiseen derivaattaan perustuvalla menetelmällä.

Funktion toinen derivaatta on

$$f''(x) = 4x - 4,$$

joten $f''(0) = -4 < 0$ ja $f''(2) = 4 > 0$. Näin ollen piste $x = 0$ on funktion aito lokaali maksimi ja piste $x = 2$ taas aito lokaali minimi. Päädyttiin siis samaan tulokseen kuin esimerkissä (1.1.1); nyt vain kriittisten pisteiden laadun selvittäminen kävi helpommin.

1.2 Useamman muuttujan funktiot

Aloitetaan tarkastelu määrittelemällä useamman muuttujan funktion $f(\mathbf{x})$ optimointiin liittyvät peruskäsitteet. Merkinnällä $\mathbf{x}$ tarkoitetaan tulevissa tarkasteluissa vektoriavaruuden $\mathbb{R}^n$ vektoria $(x_1, x_2, \dots, x_n)$. Nyt esitetään taas ainoastaan minimiin liittyvät käsitteet, mutta maksimille voitaisiin laatia vastaavat. Määritelmä (1.2.1) on analoginen edellä yhden muuttujan funktiolle esitetyn määritelmän (1.1.1) kanssa.

Määritelmä 1.2.1. Oletetaan, että joukko D on vektoriavaruuden $\mathbb{R}^n$ osajoukko ja että funktio $f(\mathbf{x})$ on reaaliarvoinen ja määritelty joukossa D . Lisäksi ilmaistaan avointa $\mathbf{x}^*$ keskeistä palloa merkinnällä

$$B(\mathbf{x}^*, \delta) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}^*\| < \delta\}, \quad \delta > 0.$$

Tällöin joukossa D sijaitseva piste $\mathbf{x}^*$ on:

- (a) globaali minimi, jos $f(\mathbf{x}^*) \leq f(\mathbf{x})$ aina, kun $\mathbf{x} \in D$;
- (b) aito globaali minimi, jos $f(\mathbf{x}^*) < f(\mathbf{x})$ aina, kun $\mathbf{x} \in D$ ja $\mathbf{x} \neq \mathbf{x}^*$;
- (c) paikallinen minimi, jos on olemassa sellainen positiivinen luku δ , että $f(\mathbf{x}^*) \leq f(\mathbf{x})$ aina, kun $\mathbf{x} \in D$ ja $\mathbf{x} \in B(\mathbf{x}^*, \delta)$;
- (d) aito paikallinen minimi, jos on olemassa sellainen positiivinen luku δ , että $f(\mathbf{x}^*) < f(\mathbf{x})$ aina, kun $\mathbf{x} \in D$, $\mathbf{x} \neq \mathbf{x}^*$ ja $\mathbf{x} \in B(\mathbf{x}^*, \delta)$;
- (e) kriittinen piste, jos gradientti $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Siis, jos

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, n.$$

Lauseessa (1.1.1) osoitettiin, että yhden muuttujan funktioiden ääriarvot ovat joko määrittelyjoukon päätepisteissä tai kriittisissä pisteissä. Osoitetaan seuraavassa lauseessa, että useamman muuttujan funktioiden kohdalla tilanne on vastaavanlainen.

Lause 1.2.1. *Oletetaan, että joukko D on vektoriavaruuden $\mathbb{R}^n$ osajoukko ja että funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat olemassa joukossa D . Oletetaan myös, että piste $\mathbf{x}^*$ on kyseisen joukon sisäpiste. Tällöin, jos $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ paikallinen minimi, niin $\mathbf{x}^*$ on myös funktion $f(\mathbf{x})$ kriittinen piste.*

Todistus. Piste $\mathbf{x}^*$ on sisäpiste, joten on olemassa sellainen positiivinen luku δ , että avoin pallo $B(\mathbf{x}^*, \delta)$ sisältyy joukkoon D . Koska $\mathbf{x}^*$ on lisäksi paikallinen minimi, niin

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \tag{1.3}$$

aina, kun $\mathbf{x} \in B(\mathbf{x}^*, \delta)$.

Määritellään nyt yhden muuttujan funktio

$$g(x_1) = f(x_1, x_2^*, x_3^*, \dots, x_n^*).$$

Koska funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat olemassa, niin funktio $g(x_1)$ on derivoituva. Ehdon (1.3) perusteella taas $g(x_1^*) \leq g(x_1)$

välillä $x_1^* - \delta < x_1 < x_1^* + \delta$. Piste x_1^* ei ole kyseisen välin päätepiste, joten lauseen (1.1.1) nojalla $g'(x_1^*) = 0$. Toisaalta

$$g'(x_1^*) = \frac{\partial f}{\partial x_1}(x_1^*, x_2^*, x_3^*, \dots, x_n^*) = \frac{\partial f}{\partial x_1}(\mathbf{x}^*),$$

joten on osoitettu, että $(\partial f / \partial x_1)(\mathbf{x}^*) = 0$. Vastaavalla tavalla voidaan osoittaa, että $(\partial f / \partial x_i)(\mathbf{x}^*) = 0$ aina, kun $i = 2, 3, \dots, n$. Siis gradientti $\nabla f(\mathbf{x}^*) = \mathbf{0}$ ja $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste. $\square$

Vastaavalla tavalla voidaan osoittaa, että jos piste $\mathbf{x}^*$ on määrittelyjoukon sisäpiste ja funktion $f(\mathbf{x})$ paikallinen maksimi, niin se on myös kyseisen funktion kriittinen piste. Tarkastellaan kriittisten pisteiden määrittämistä vielä esimerkin avulla.

Esimerkki 1.2.1. Tutkitaan funktiota $f(\mathbf{x}) = x_1^2 - 4x_1 + 2x_2^2 + 7$, joka on määritelty aina, kun $\mathbf{x} \in \mathbb{R}^2$. Funktiolla ei ole globaalia maksimia, sillä

$$\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty.$$

Otetaan selville funktion kriittiset pisteet.

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 2x_1 - 4 \quad \text{ja} \quad \frac{\partial f(\mathbf{x})}{\partial x_2} = 4x_2,$$

joten kriittiset pisteet toteuttavat yhtälöt

$$2x_1 - 4 = 0 \quad \text{ja} \quad 4x_2 = 0.$$

Koska tämä yhtälöpari toteutuu ainoastaan, jos $x_1 = 2$ ja $x_2 = 0$, niin piste $(2, 0)$ on funktion ainoa kriittinen piste. Kriittisen pisteen laatu voitaisiin nyt selvittää tutkimalla funktion käyttäytymistä kyseisen pisteen läheisyydessä, mikä voi olla työlästä. Seuraavana tarkastellaankin tehokkaampia menetelmiä kriittisten pisteiden laadun määrittämiseen, joten palataan tähän esimerkkiin vielä myöhemmin uudelleen.

Yhden muuttujan funktion tapauksessa kriittisten pisteiden laadun määrittäminen perustui optimoitavan funktion toisen derivaatan tutkimiseen. Useamman

muuttujan funktioiden kohdalla tilanne on periaatteessa analoginen. Tarkastelussa tarvitaan apuna Taylorin kehitelmää, joka lauseessa (1.1.2) jo esitettiin yhden muuttujan funktiolle. Nyt tarvitaan useamman muuttujan funktiolle laadittua kehitelmää, joka on todistettu lähdeoteoksessa [3]. Ennen kuin kyseinen lause esitetään, on vielä huomautettava seuraavista tarkasteluissa käytettävistä merkinnöistä ja käsitteistä:

- (1) Matriisi $A \in \mathbb{R}_{m \times n}$ on $m \times n$ -matriisi, jonka vaakarivien lukumäärä on m , pystyrivien lukumäärä n ja kaikki alkio (a_{ij}) ovat reaalisia;
- (2) Olkoon matriisi A $m \times n$ -matriisi, jonka (i,j) -alkio on (a_{ij}) . Tällöin sen transpoosi A^T on $n \times m$ -matriisi, jonka (i,j) -alkio on (a_{ji}) ;
- (3) Matriisi A on symmetrinen, jos se on $n \times n$ -matriisi, jonka alkio (a_{ij}) ja (a_{ji}) ovat samat. Tällöin $A = A^T$.

Lause 1.2.2 (Taylorin kehitelmä useamman muuttujan funktiolle).

Oletetaan, että pisteet $\mathbf{x}^$ ja $\mathbf{x}$ ovat vektoriavaruuden $\mathbb{R}^n$ kaksi eri pistettä ja että niiden välinen jana $[\mathbf{x}^*, \mathbf{x}]$ sisältyy vektoriavaruuden $\mathbb{R}^n$ avoimeen osajoukkoon D . Oletetaan lisäksi, että funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat olemassa joukossa D . Tällöin pisteiden $\mathbf{x}^*$ ja $\mathbf{x}$ välisellä janalla $[\mathbf{x}^*, \mathbf{x}]$ on olemassa sellainen piste $\mathbf{z}$, että*

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*).$$

Nyt gradientti ∇f on vektori

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \quad (1.4)$$

ja merkinnällä Hf tarkoitetaan symmetristä $n \times n$ -matriisiä, Hessin matriisiä

$$Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}. \quad (1.5)$$

Taylorin kehitelmän avulla saadaan tietoa kriittisten pisteiden laadusta, kuten seuraava lause osoittaa.

Lause 1.2.3. Oletetaan, että funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat jatkuvia vektoriavaruudessa $\mathbb{R}^n$ ja että piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste.

(a) Piste $\mathbf{x}^*$ on globaali minimi, jos $(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) \geq 0$ aina, kun $\mathbf{x} \in \mathbb{R}^n$ ja $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

(b) Piste $\mathbf{x}^*$ on aito globaali minimi, jos $(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) > 0$ aina, kun $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{x}^*$ ja $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

(c) Piste $\mathbf{x}^*$ on globaali maksimi, jos $(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) \leq 0$ aina, kun $\mathbf{x} \in \mathbb{R}^n$ ja $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

(d) Piste $\mathbf{x}^*$ on aito globaali maksimi, jos $(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) < 0$ aina, kun $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{x}^*$ ja $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$.

Todistus. Piste $\mathbf{x}^*$ on oletuksen mukaan funktion $f(\mathbf{x})$ kriittinen piste, joten ensimmäinen osittaisderivaatta $(\partial f / \partial x_i)(\mathbf{x}^*) = 0$, kun $i = 1, 2, \dots, n$. Siis gradientti $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Jos nyt myös piste $\mathbf{x} \in \mathbb{R}^n$ ja $\mathbf{x} \neq \mathbf{x}^*$, niin Taylorin kehitemmä useamman muuttujan funktiolle saadaan muotoon

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*). \quad (1.6)$$

Tässä piste $\mathbf{z}$ kuulu pisteiden $\mathbf{x}^*$ ja $\mathbf{x}$ väliselle janelle $[\mathbf{x}^*, \mathbf{x}]$.

Tutkitaan lauseen väitettä (a). Väitteen mukaan $(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) \geq 0$. Tällöin kaavan (1.6) perusteella $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ aina, kun $\mathbf{x} \in \mathbb{R}^n$. Täten piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ globaali minimi ja väite on todistettu.

Loput lauseen väitteet (b)-(d) saadaankin nyt todistettua vastaavalla tavalla. $\square$

Lause (1.2.3) antaa siis useita sääntöjä sille, kuinka tekijän

$$(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*)$$

merkistä voidaan päätellä, millainen globaali ääriarvo tarkasteltava kriittinen piste on. Kyseisen merkin selvittäminen kokeilemalla eri arvoyhdistel-

miä pisteille $\mathbf{x}$, $\mathbf{x}^*$ ja $\mathbf{z}$ on kuitenkin työlästä, joten tarvitaan tehokkaampia menetelmiä. Näitä menetelmiä käsitellään tutkielmassa seuraavana.

1.3 Hessin matriisin definiittisyyden yhteys optimointiin

Tarkastellaan ensin matriisiin liittyvää kvadraattista muotoa. Jos matriisi A on symmetrinen $n \times n$ -matriisi, niin siihen liittyy kvadraattinen muoto

$$Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y}, \quad (1.7)$$

missä vektori $\mathbf{y} \in \mathbb{R}^n$. Määritellään seuraavaksi kvadraattisen muodon merkkiin liittyvät käsitteet.

Määritelmä 1.3.1. Olkoon A symmetrinen $n \times n$ -matriisi ja $Q_A(\mathbf{y})$ kyseiseen matriisiin liittyvä kvadraattinen muoto. Tällöin A ja $Q_A(\mathbf{y})$ ovat:

- (a) positiivisesti semidefiniittejä, jos $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} \geq 0$ aina, kun $\mathbf{y} \in \mathbb{R}^n$;
- (b) positiivisesti definiittejä, jos $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} > 0$ aina, kun $\mathbf{y} \in \mathbb{R}^n$ ja $\mathbf{y} \neq \mathbf{0}$;
- (c) negatiivisesti semidefiniittejä, jos $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} \leq 0$ aina, kun $\mathbf{y} \in \mathbb{R}^n$;
- (d) negatiivisesti definiittejä, jos $Q_A(\mathbf{y}) = \mathbf{y}^T A \mathbf{y} < 0$ aina, kun $\mathbf{y} \in \mathbb{R}^n$ ja $\mathbf{y} \neq \mathbf{0}$;
- (e) indefiniittejä, jos on olemassa sellaiset kaksi eri vektoria $\mathbf{y}_1 \in \mathbb{R}^n$ ja $\mathbf{y}_2 \in \mathbb{R}^n$, että $Q_A(\mathbf{y}_1) = \mathbf{y}_1^T A \mathbf{y}_1 > 0$ ja $Q_A(\mathbf{y}_2) = \mathbf{y}_2^T A \mathbf{y}_2 < 0$.

Lauseessa (1.2.2) todettiin, että jos funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat jatkuvia, niin Hessin matriisi $Hf(\mathbf{x})$ on symmetrinen $n \times n$ -matriisi. Kun vielä sijoitetaan kaavaan (1.7) vektorin $\mathbf{y}$ paikalle vektori $\mathbf{x} - \mathbf{x}^*$, voidaan Hessin matriisille kirjoittaa kvadraattinen muoto

$$Q_H(\mathbf{x} - \mathbf{x}^*) = (\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*). \quad (1.8)$$

Verrataan nyt kaavaa (1.8) lauseen (1.2.3) väitteisiin. Nähdään, että optimoitavan funktion $f(\mathbf{x})$ kriittisten pisteiden laatu riippuu itse asiassa Hessin matriisin kvadraattisen muodon $Q_H(\mathbf{x} - \mathbf{x}^*)$ merkistä. Täten lause (1.2.3) voidaan määritelmän (1.3.1) perusteella kirjoittaa seuraavaan muotoon.

Lause 1.3.1. *Oletetaan, että funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat jatkuvia vektoriavuudessa $\mathbb{R}^n$ ja että sen Hessin matriisi on $Hf(\mathbf{x})$. Oletetaan lisäksi, että piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste.*

- (a) *Piste $\mathbf{x}^*$ on globaali minimi, jos $Hf(\mathbf{x})$ on positiivisesti semidefiniitti vektoriavuudessa $\mathbb{R}^n$;*
- (b) *Piste $\mathbf{x}^*$ on aito globaali minimi, jos $Hf(\mathbf{x})$ on positiivisesti definiitti vektoriavuudessa $\mathbb{R}^n$;*
- (c) *Piste $\mathbf{x}^*$ on globaali maksimi, jos $Hf(\mathbf{x})$ on negatiivisesti semidefiniitti vektoriavuudessa $\mathbb{R}^n$;*
- (d) *Piste $\mathbf{x}^*$ on aito globaali maksimi, jos $Hf(\mathbf{x})$ on negatiivisesti definiitti vektoriavuudessa $\mathbb{R}^n$.*

Esimerkki 1.3.1. Tarkastellaan funktiota

$$f(\mathbf{x}) = \frac{1}{2}d_1x_1^2 + \frac{1}{2}d_2x_2^2 + \cdots + \frac{1}{2}d_nx_n^2,$$

missä kertoimet d_i ovat kaikki reaalisia ja nollasta eroavia. Tällöin funktion gradientti

$$\nabla f(\mathbf{x}) = (d_1x_1, d_2x_2, \dots, d_nx_n) = \mathbf{0}$$

ainoastaan, jos $\mathbf{x} = \mathbf{0}$. Siis piste $\mathbf{x} = \mathbf{0}$ on funktion ainoa kriittinen piste. Kriittisen pisteen laatu saadaan nyt pääteltyä lauseen (1.3.1) perusteella, kunhan Hessin matriisin definiittisyys on saatu selville. Funktion $f(\mathbf{x})$ Hessin matriisi on kaavan (1.5) perusteella diagonaalimatriisi

$$Hf(\mathbf{x}) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix},$$

ja kvadraattinen muoto puolestaan

$$Q_H(\mathbf{y}) = \mathbf{y}^T H f(\mathbf{x}) \mathbf{y} = d_1 y_1^2 + d_2 y_2^2 + \cdots + d_n y_n^2.$$

Kvadraattisessa muodossa kaikki neliötermit $y_i^2 \geq 0$. Näin ollen Hessin matriisin $Hf(\mathbf{x})$ definiittisyys riippuu kertoimien d_i arvoista seuraavalla tavalla:

- (1) $Hf(\mathbf{x})$ on positiivisesti semidefiniitti, jos $d_i \geq 0$, kun $i = 1, 2, \dots, n$;
- (2) $Hf(\mathbf{x})$ on positiivisesti definiitti, jos $d_i > 0$, kun $i = 1, 2, \dots, n$;
- (3) $Hf(\mathbf{x})$ on negatiivisesti semidefiniitti, jos $d_i \leq 0$, kun $i = 1, 2, \dots, n$;
- (4) $Hf(\mathbf{x})$ on negatiivisesti definiitti, jos $d_i < 0$, kun $i = 1, 2, \dots, n$;
- (5) $Hf(\mathbf{x})$ on indefiniitti, jos ainakin yksi $d_i < 0$ ja ainakin yksi $d_i > 0$, kun $i = 1, 2, \dots, n$.

Kuten edellisen esimerkin perusteella nähtiin, tarkasteltavan Hessin matriisin definiittisyys voi olla toisinaan helppo määrittää ilman erityisempiä apukeinoja. Yleensä apukeinoja kuitenkin tarvitaan, joten käsitellään nyt niitä.

1.3.1 Matriisin definiittisyyden selvittäminen determinanttien avulla

Määritellään ensin johtavan pääminorin käsite.

Määritelmä 1.3.2. Oletetaan, että A on symmetrinen $n \times n$ -matriisi. Esi-
tetään matriisin A vasemmassa yläkulmassa olevaa $k \times k$ -alimatriisia mer-
kinnällä Δ_k , kun $1 \leq k \leq n$. Tällöin determinantti Δ_k on matriisin A kokoa
 k oleva johtava pääminori. Siis, jos symmetrinen matriisi

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix},$$

niin johtavat pääminorit ovat

$$\Delta_1 = a_{11}, \quad \Delta_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \dots, \quad \Delta_n = \det A.$$

Matriisin johtavat pääminorit ja matriisin definiittisyys liittyvät toisiinsa seuraavassa lauseessa esitetyllä tavalla.

Lause 1.3.2. *Olko A symmetrinen $n \times n$ -matriisi ja Δ_k sen kokoa k oleva johtava pääminori, kun $1 \leq k \leq n$. Tällöin:*

- (a) *A on positiivisesti definiitti matriisi, jos ja vain jos $\Delta_k > 0$ aina, kun $1 \leq k \leq n$;*
- (b) *A on negatiivisesti definiitti matriisi, jos ja vain jos $(-1)^k \Delta_k > 0$, kun $1 \leq k \leq n$.*

Todistus. Lauseen todistaminen voitaisiin tehdä matemaattisen induktion avulla. Ensin todistettaisiin, että väitteet pitävät paikkansa kun $n = 1$. Sitten tehtäisiin oletus, että väitteet ovat totta, kun $n = k$. Lopuksi suoritettaisiin todistuksen varsinainen induktioaskel eli näytettäisiin väitteiden paikkansapitävyys, kun $n = k + 1$. Induktioaskeleen todistus on tällä ketäa kuitenkin niin vaivalloinen, että lauseen todistaminen matemaattisen induktion avulla sivutetaan. Sen sijaan nyt todistetaan esimerkkinä, että väite (b) pitää paikkansa kun $n = 2$.

Olko A symmetrinen 2×2 -matriisi

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Tällöin sen kvadraattinen muoto on

$$Q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2,$$

missä $\mathbf{x} = (x_1, x_2) \neq \mathbf{0}$. On siis kaksi mahdollisuutta:

- (1) $\mathbf{x} = (x_1, 0), \quad x_1 \neq 0;$
- (2) $\mathbf{x} = (x_1, x_2), \quad x_2 \neq 0.$

Tarkastellaan tapauksia erikseen.

- (1) $\mathbf{x} = (x_1, 0), \quad x_1 \neq 0.$

Väitteen (b) mukaisesti matriisi A on negatiivisesti definiitti, joten aina

$Q_A(\mathbf{x}) < 0$. Nyt $Q_A(\mathbf{x}) = a_{11}x_1^2$, joten $Q_A(\mathbf{x}) < 0$ täsmälleen silloin, kun alkio $a_{11} < 0$. Tämän mukaan matriisin A positiiviselle definiittisyydelle on siis olemassa ainakin sellainen ehto, että $a_{11} < 0$.

(2) $\mathbf{x} = (x_1, x_2)$, $x_2 \neq 0$.

Merkitään nyt $x_1 = tx_2$, missä luku $t \in \mathbb{R}$. Tällöin matriisin A kvadraattinen muoto voidaan kirjoittaa muodossa

$$Q_A(\mathbf{x}) = [a_{11}t^2 + 2a_{12}t + a_{22}]x_2^2 = \varphi(t)x_2^2.$$

Nyt $x_2 \neq 0$, joten neliöllinen termi $x_2^2 > 0$. Täten $Q_A(\mathbf{x}) < 0$ täsmälleen silloin, jos funktio $\varphi(t) = [a_{11}t^2 + 2a_{12}t + a_{22}] < 0$ aina, kun $t \in \mathbb{R}$. Tarkastellaan funktiota $\varphi(t)$ seuraavaksi tarkemmin. Sen derivaatta on

$$\varphi'(t) = 2a_{11}t + 2a_{12}.$$

Derivaatafunktiota $\varphi'(t) = 0$ täsmälleen silloin, kun $t = -a_{12}/a_{11}$. Siis funktion $\varphi(t)$ ainoa kriittinen piste $t^* = -a_{12}/a_{11}$. Funktiota $\varphi(t)$ toinen derivaatta puolestaan on

$$\varphi''(t) = 2a_{11}.$$

Lauseen (1.1.3) nojalla tämä piste on siis aito globaali minimi, jos $a_{11} > 0$ ja aito globaali maksimi, jos $a_{11} < 0$. Tapauksen (1) perusteella matriisi A on kuitenkin negatiivisesti definiitti, jos ja vain jos alkio $a_{11} < 0$, joten kyseinen kriittinen piste on nyt aito globaali maksimi. Siis nähdään, että

$$\varphi(t) \leq \varphi(t^*) = \varphi\left(-\frac{a_{12}}{a_{11}}\right) = -\frac{a_{12}^2}{a_{11}} + a_{22} = \frac{1}{a_{11}} \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

aina, kun $t \in \mathbb{R}$. Koska nyt alkio $a_{11} < 0$, niin funktio $\varphi(t) < 0$, jos ja vain jos determinantti

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} = \det A > 0.$$

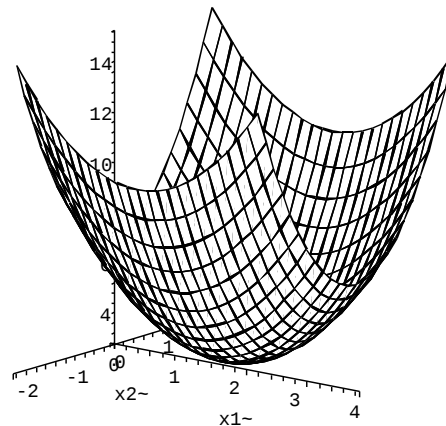
Nyt on siis todistettu, että symmetrinen 2×2 -matriisi A on negatiivisesti definiitti, jos ja vain jos sen johtavat pääminori $\Delta_1 = a_{11} < 0$ ja $\Delta_2 = \det A > 0$.

□

Huomautus. Vaikka matriisi A olisi symmetrinen $n \times n$ -matriisi, niin lauseen (1.3.2) perusteella ei tule tehdä päätelmiä sen semidefiniittisyydestä. Jos esimerkiksi matriisin kaikki johtavat pääminorit Δ_n ovat epänegatiivisia, niin matriisi ei silti välttämättä ole positiivisesti semidefiniitti.

Havainnollistetaan edeltävän lauseen käyttöä optimoinnissa vielä yhden esimerkin avulla.

Esimerkki 1.3.2. Tutkitaan funktiota $f(\mathbf{x}) = x_1^2 - 4x_1 + 2x_2^2 + 7$, joka on määritelty aina, kun $\mathbf{x} \in \mathbb{R}^2$. Jo aiemmin esimerkissä (1.2.1) todettiin, että kyseisen funktion ainoa kriittinen piste $\mathbf{x}^*$ on piste $(2, 0)$. Tällöin ei kuitenkaan vielä osattu päätellä tämän kriittisen pisteen laatua. Funktio on esitetty kuvassa 1.1, josta nähdään kyllä heti, että kriittinen piste on aito globaali minimi.



Kuva 1.1: $f(\mathbf{x}) = x_1^2 - 4x_1 + 2x_2^2 + 7$.

Jatketaan nyt tarkastelua ja osoitetaan vielä matemaattisesti, että kriittinen piste on todella aito globaali minimi. Funktion $f(\mathbf{x})$ Hessin matriisi on symmetrinen 2×2 -matriisi

$$Hf(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix},$$

jonka johtavat pääminorit $\Delta_1 = 2 > 0$ ja $\Delta_2 = \det Hf(\mathbf{x}) = 2 * 4 = 8 > 0$. Lauseen (1.3.2) nojalla Hessin matriisi on siten positiivisesti definiitti ja lauseen (1.3.1) perusteella kriittinen piste $\mathbf{x}^* = (2, 0)$ on funktion $f(\mathbf{x})$ aito globaali minimi.

1.3.2 Matriisin definiittisyyden selvittäminen ominaisarvojen avulla

Edellä esimerkissä (1.3.2) piti selvittää 2×2 -matriisin definiittisyys. Determinanttien avulla tällaisen pienen matriisin definiittisyys saatiin selville pienellä vaivalla. Sen sijaan, jos tarkasteltava matriisi on suuri, tarvitaan tehokkaampi menetelmä. Tässä osiossa käsitelläänkin matriisin definiittisyyden määrittämistä ominaisarvoihin perustuen.

Käsitellään ensin lyhyesti sitä, mitä matriisin ominaisarvoilla tarkoitetaan. Oletetaan, että matriisi A on $\mathbb{C}_{n \times n}$ -matriisi. Luku $\lambda \in \mathbb{C}$ on matriisin A ominaisarvo, jos on olemassa sellainen nollavektorista eroava vektori $\mathbf{x} \in \mathbb{C}^n$, että $A\mathbf{x} = \lambda\mathbf{x}$. Sanotaan, että vektori $\mathbf{x}$ on tällöin ominaisarvoa λ vastaava ominaisvektori.

Helpoin tapa matriisin A ominaisarvojen selvittämiseksi, on tutkia, koska

$$\det(A - \lambda I) = 0.$$

Toisin sanoen ominaisarvot ovat astetta n olevan yhtälön $p_A(\lambda) = \det(A - \lambda I) = 0$ nollakohdat. Yhtälöä $p_A(\lambda)$ kutsutaan matriisin A karakteristiseksi polynomiksi.

Mainittakoon vielä, että jos λ on karakteristisen polynomin k -kertainen nollakohta, niin silloin kyseisen ominaisarvon kertaluku on k . Jos kukin ominaisarvo huomioidaan kertalukunsa verran, on $n \times n$ -matriisin ominaisarvojen kokonaislukumäärä n .

Tarkastellaan seuraavaksi joitakin teorian kannalta tarpeellisia matriiseihin ja ominaisarvoihin liittyviä ominaisuuksia. Apulauseiden väittämät on todistettu lähdeoteoksessa [5].

Apulause 1.3.1. *Olkoon A symmetrinen reaallinen $n \times n$ -matriisi. Tällöin pätevät seuraavat väitteet:*

- (a) *Matriisin A kaikki ominaisarvot ovat reaalisia;*
- (b) *On olemassa n kappaletta sellaisia erilaisia ominaisvektoreita $u_i \in \mathbb{R}^n$, että $Au_i = \lambda_i u_i$, kun $1 \leq i \leq n$, ja että nämä ominaisvektorit ovat ortonormaaleja eli ortogonaalisia yksikkövektoreita. Siis kyseiset ominaisvektorit ovat kohtisuorassa toisiaan vasten ja normi $\|u_i\| = 1$, kun $1 \leq i \leq n$;*
- (c) *Matriisi A voidaan kirjoittaa muodossa $A = CDC^T$, missä matriisi D on matriisin A ominaisarvoista koostuva diagonaalimatriisi*

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

Matriisi C puolestaan koostuu matriisin A ominaisvektoreista, jotka kohdan (b) nojalla ovat reaalisia ja ortogonaalisia ja joita on n kappaletta. Siis

$$C = (u_1, u_2, \dots, u_n),$$

missä u_j on matriisin C pystyrivi $[u_{1j}, u_{2j}, \dots, u_{nj}]^T$.

Apulause 1.3.2. *Olkoon C reaallinen ja ortogonaalinen $n \times n$ -matriisi. Tällöin $C^T C = C C^T = I$ eli matriisin C käänteismatriisi $C^{-1} = C^T$.*

Näiden kahden apulauseen avulla matriisin ominaisarvot saadaan nyt liitettyä matriisin definiittisyyden määrittämiseen. Esitetään seuraavassa vain positiiviseen definiittisyyteen liittyvät väitteet, sillä negatiiviseen definiittisyyteen liittyvät ovat vastaavanlaiset.

Lause 1.3.3. *Olkoon A symmetrinen reaallinen matriisi. Tällöin:*

- (a) *matriisi A on positiivisesti definiitti täsmälleen silloin, jos sen kaikki ominaisarvot $\lambda_i > 0$, kun $1 \leq i \leq n$;*
- (b) *matriisi A on positiivisesti semidefiniitti täsmälleen silloin, jos sen kaikki ominaisarvot $\lambda_i \geq 0$, kun $1 \leq i \leq n$;*
- (c) *matriisi A on indefiniitti täsmälleen silloin, jos ainakin yksi $\lambda_i < 0$ ja ainakin yksi $\lambda_i > 0$, kun $1 \leq i \leq n$.*

Todistus. Matriisi A on symmetrinen ja reaalinen $n \times n$ -matriisi, joten apulauseen (1.3.1) kohdan (c) perusteella

$$A = CDC^T, \quad (1.9)$$

missä D on matriisin A ominaisarvoista koostuva diagonaalimatriisi ja C puolestaan matriisin A ortonormaaleista ominaisvektoreista koostuva matriisi. Apulauseen (1.3.2) perusteella käänteismatriisi $C^{-1} = C^T$, joten kaava (1.9) saadaan mutoon

$$C^T AC = C^T CDC^T C = D. \quad (1.10)$$

Matriisin A kvadraattinen muoto on

$$Q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \quad (1.11)$$

missä vektori $\mathbf{x} \in \mathbb{R}^n$ ja $\mathbf{x} \neq \mathbf{0}$. Valitaan nyt vektori $\mathbf{y} \in \mathbb{R}^n$ sillä tavalla, että vektori $\mathbf{x} = C\mathbf{y}$. Matriisi C on säännöllinen ja kääntyvä, joten yhtälöllä $C\mathbf{y} = \mathbf{0}$ on vain triviaaliratkaisu $\mathbf{y} = \mathbf{0}$. Koska kuitenkin vektori $\mathbf{x} \neq \mathbf{0}$, niin myös vektori $\mathbf{y} \neq \mathbf{0}$. Sijoitetaan kaavaan (1.11) vektoriksi $\mathbf{x} = C\mathbf{y}$, jolloin matriisin A kvadraattinen muoto saa esityksen

$$Q_A(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = (C\mathbf{y})^T A C\mathbf{y} = \mathbf{y}^T C^T A C \mathbf{y} = \mathbf{y}^T D \mathbf{y}. \quad (1.12)$$

Edelleen, kertomalla auki kaavan (1.12) oikea puoli, saadaan

$$Q_A(\mathbf{x}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2. \quad (1.13)$$

Tarkastellaan nyt sitä, millainen on kunkin ominaisarvon vaikutus kaavan (1.12) nojalla kvadraattisen muodon merkkiin. Tätä varten valitaan vektori $\mathbf{y}^{(i)}$ sillä tavalla, että sen komponentti i saa arvon yksi ja muut $n - 1$ kappaletta komponentteja puolestaan arvon nolla. Kuten aiemmin jo todettiin, vektori $\mathbf{y} \neq \mathbf{0}$, joten kaikkien komponenttien arvo ei voi olla nolla. Nyt vektori $\mathbf{x}^{(i)} = C\mathbf{y}^{(i)}$ ja kvadraattiseksi muodoksi tulee

$$Q_A(\mathbf{x}^{(i)}) = \lambda_i y_i^2 = \lambda_i, \quad (1.14)$$

kun $1 \leq i \leq n$. Siten ehto $Q_A(\mathbf{x}) > 0$ pätee aina, jos ja vain jos kaikki ominaisarvot ovat positiivisia. Tämä todistaa lauseen väitteen (a). Ehto $Q_A(\mathbf{x}) \geq 0$ pätee taas aina, jos ja vain jos kaikki ominaisarvot ovat epänegatiivisia. Täten väite (b) on todistettu oikeaksi. Jos on olemassa ainakin yksi positiivinen ja yksi negatiivinen ominaisarvo, saa myös kvadraattinen muoto $Q_A(\mathbf{x})$ sekä positiivisia että negatiivisia arvoja. Tämä todistaa kohdan (c). $\square$

Myöhemmin tässä tutkielmassa on esimerkki (1.4.1), jossa ominaisarvoja käytetään apuna matriisin definiittisyyden määrittämisessä.

1.4 Paikalliset ääriarvot ja satulapiste

Tähän astisessa tarkastelussa on useamman muuttujan funktioista pyritty löytämään vain globaalit ääriarvot. Keskityttään nyt tutkimaan funktion muita huomion arvoisia pisteitä. Aloitetaan paikallisista ääriarvokohdista.

Lause 1.4.1. *Olkoon joukko D vektoriavaruuden $\mathbb{R}^n$ osajoukko ja oletetaan, että funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat jatkuvia joukossa D . Oletetaan lisäksi, että piste $\mathbf{x}^*$ on joukon D sisäpiste ja funktion $f(\mathbf{x})$ kriittinen piste.*

(a) *Piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ aito paikallinen minimi, jos $Hf(\mathbf{x}^*)$ on positiivisesti definiitti.*

(b) *Piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ aito paikallinen maksimi, jos $Hf(\mathbf{x}^*)$ on negatiivisesti definiitti.*

Todistus. Väitteiden (a) ja (b) todistukset ovat samanlaiset, joten todistetaan ainoastaan kohta (b). Käytetään todistuksessa apuna Hessin matriisin $Hf(\mathbf{x})$ johtavia pääminoreita $\Delta_i(\mathbf{x})$.

Väitteen (b) mukaan matriisi $Hf(\mathbf{x}^*)$ on negatiivisesti definiitti, joten lauseen (1.3.2) perusteella $(-1)^i \Delta_i(\mathbf{x}^*) > 0$, kun $i = 1, 2, \dots, n$. Toisaalta funktion

$f(\mathbf{x})$ toiset osittaisderivaatat ovat nyt jatkuvia. Täten kaavan (1.5) perusteella myös Hessin matriisin $Hf(\mathbf{x})$ kaikki johtavat pääminorit $\Delta_i(\mathbf{x})$ ovat jatkuvia funktioita muuttujan $\mathbf{x}$ suhteen.

Edeltävän perusteella siis funktiot $(-1)^i \Delta_i(\mathbf{x})$ ovat aina jatkuvia ja pisteessä $\mathbf{x}^*$ positiivisia. Tällöin jokaista lukua i kohti on olemassa sellainen positiivinen reaaliluku r_i , että

$$(-1)^i \Delta_i(\mathbf{x}) > 0 \quad (1.15)$$

aina, kun $\|\mathbf{x} - \mathbf{x}^*\| < r_i$. Valitaan nyt luku r sillä tavalla, että se on pienin luvuista r_i . Siis luku $r = \min\{r_1, r_2, \dots, r_n\}$. Tämän jälkeen epäyhtälö (1.15) pätee aina, kun $\|\mathbf{x} - \mathbf{x}^*\| < r$. Lauseesta (1.3.2) seuraa tällöin, että Hessin matriisi $Hf(\mathbf{x})$ on negatiivisesti definiitti, kun $\|\mathbf{x} - \mathbf{x}^*\| < r$.

Tarkastellaan seuraavaksi useamman muuttujan funktioille tarkoitettua Taylorin kehitelmää eli lausetta (1.2.2). Koska $\mathbf{x}^*$ on kriittinen piste niin kehitelmän mukaan

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*). \quad (1.16)$$

Tässä piste $\mathbf{z}$ kuuluu pisteiden $\mathbf{x}^*$ ja $\mathbf{x}$ väliselle janelle $[\mathbf{x}^*, \mathbf{x}]$. Koska nyt on valittu, että $\|\mathbf{x} - \mathbf{x}^*\| < r$, niin myös $\|\mathbf{z} - \mathbf{x}^*\| < r$. Siis matriisi $Hf(\mathbf{z})$ on negatiivisesti definiitti ja kaavassa (1.16) on termi

$$\frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*) < 0.$$

Täten $f(\mathbf{x}) < f(\mathbf{x}^*)$, kun $\|\mathbf{x} - \mathbf{x}^*\| < r$ ja $\mathbf{x} \neq \mathbf{x}^*$. $f(\mathbf{x}^*)$ on funktion $f(\mathbf{x})$ aito paikallinen maksimi. $\square$

Huomautus. Jos piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste ja Hessin matriisi $Hf(\mathbf{x}^*)$ on semidefiniitti, ei edeltävää lausetta voida käyttää kriittisen pisteen laadun selvittämiseen. Funktion käyttäytymistä kyseisen pisteen läheisyydessä on tutkittava muulla tavalla.

Paitsi, että kriittinen piste voi olla funktion ääriarvo eli maksimi tai minimi, niin se voi olla myös funktion satulapiste. Määritellään satulapiste.

Määritelmä 1.4.1. Oletetaan, että avoin joukko D on vektoriavaruuden $\mathbb{R}^n$ osajoukko ja että funktio $f(\mathbf{x})$ on reaaliarvoinen ja määritelty joukossa D . Oletetaan lisäksi, että piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ sisäpiste ja kriittinen piste joukossa D . Tällöin piste $\mathbf{x}^*$ on satulapiste, jos sen mielivaltaisen pienessä ympäristössä löytyy sellaiset pisteet $\mathbf{x}$ ja $\mathbf{y}$, että $f(\mathbf{x}) \leq f(\mathbf{x}^*) \leq f(\mathbf{y})$.

Optimoitavan funktion Hessin matriisin definiittisyyttä voidaan käyttää satulapisteiden löytämiseen seuraavan lauseen avulla.

Lause 1.4.2. Oletetaan edelleen, että avoin joukko D on vektoriavaruuden $\mathbb{R}^n$ osajoukko. Oletetaan lisäksi, että funktion $f(\mathbf{x})$ kaikki ensimmäiset ja toiset osittaisderivaatat ovat jatkuvia joukossa D ja että piste $\mathbf{x}^*$ on kyseisen joukon sisäpiste. Jos piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste ja Hessin matriisi $Hf(\mathbf{x}^*)$ on indefiniitti, niin piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ satulapiste.

Todistus. Oletuksen nojalla Hessin matriisi $Hf(\mathbf{x}^*)$ on indefiniitti. Tällöin on siis olemassa sellaiset nollasta eroavat vektorit $\mathbf{y} \in \mathbb{R}^n$ ja $\mathbf{w} \in \mathbb{R}^n$, että

$$\mathbf{y}^T Hf(\mathbf{x}^*) \mathbf{y} > 0 \quad \text{ja} \quad \mathbf{w}^T Hf(\mathbf{x}^*) \mathbf{w} < 0. \quad (1.17)$$

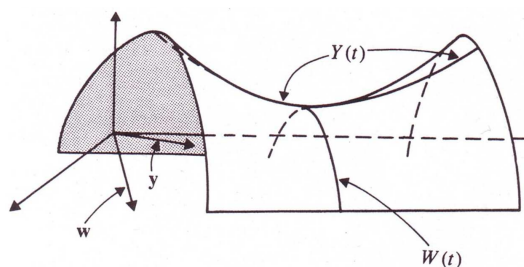
Oletuksena oli myös, että funktion $f(\mathbf{x})$ toiset osittaisderivaatat ovat jatkuvia, joten kaavan (1.17) epäyhtälöt pätevät myös pisteen $\mathbf{x}^*$ välittömässä lähiympäristössä. On siis olemassa sellainen positiivinen reaaliluku ϵ , että

$$\mathbf{y}^T Hf(\mathbf{x}^* + t\mathbf{y}) \mathbf{y} > 0 \quad \text{ja} \quad \mathbf{w}^T Hf(\mathbf{x}^* + t\mathbf{w}) \mathbf{w} < 0 \quad (1.18)$$

aina, kun $|t| < \epsilon$. Määritellään nyt kaksi uutta funktiota

$$Y(t) = f(\mathbf{x}^* + t\mathbf{y}) \quad \text{ja} \quad W(t) = f(\mathbf{x}^* + t\mathbf{w}),$$

jotka on esitetty kuvassa 1.2.



Kuva 1.2: Satulapiste [3].

Piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste. Täten funktioiden $Y(t)$ ja $W(t)$ ensimmäisille derivaatoille

$$Y'(t) = \nabla f(\mathbf{x}^* + t\mathbf{y})\mathbf{y} \quad \text{ja} \quad W'(t) = \nabla f(\mathbf{x}^* + t\mathbf{w})\mathbf{w}$$

pätee, että $Y'(0) = W'(0) = 0$. Siis piste $t = 0$ on funktioiden $Y(t)$ ja $W(t)$ kriittinen piste. Kyseisten funktioiden toiset derivaatat ovat puolestaan

$$Y''(t) = \mathbf{y}^T H f(\mathbf{x}^* + t\mathbf{y})\mathbf{y} \quad \text{ja} \quad W''(t) = \mathbf{w}^T H f(\mathbf{x}^* + t\mathbf{w})\mathbf{w}.$$

Nyt ehtojen (1.18) perusteella $Y''(0) = \mathbf{y}^T H f(\mathbf{x}^*)\mathbf{y} > 0$ ja vastaavasti $W''(0) = \mathbf{w}^T H f(\mathbf{x}^*)\mathbf{w} < 0$. Täten, lauseen (1.1.3) mukaisesti, piste $t = 0$ on funktion $Y(t)$ aito paikallinen minimi ja funktion $W(t)$ aito paikallinen maksimi. Jos siis siirrytään pisteestä $\mathbf{x}^*$ vähänkin vektorin $\mathbf{y}$ suuntaan, funktion $f(\mathbf{x})$ arvot kasvavat. Vastaavasti, jos siirrytään vähänkin vektorin $\mathbf{w}$ suuntaan, funktion $f(\mathbf{x})$ arvot pienenevät. Piste $\mathbf{x}^*$ ei siis ole maksimi eikä minimi, vaan se on satulapiste. $\square$

Esitetään tämän osion loppuun vielä yksi esimerkki funktiosta, jolla on sekä paikallisia ääriarvoja että satulapisteitä. Esimerkissä Hessin matriisien definiittisyys ratkaistaan aiemmin esitetyllä, matriisien ominaisarvoihin perustuvalla menetelmällä.

Esimerkki 1.4.1. Tutkitaan reaaliarvoista funktiota $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1 - 12x_2 + 20$, joka on jatkuva vektoriavuudessa $\mathbb{R}^2$. Esimerkiksi raja-arvo

$$\lim_{x_1 \rightarrow +\infty} f(x_1, 0) = \lim_{x_1 \rightarrow +\infty} (x_1^3 - 3x_1) = +\infty,$$

koska termi x_1^3 on dominoiva termiin $3x_1$ verrattuna. Siis funktiolla $f(x_1, x_2)$ ei ole globaalia maksimia. Vastaavasti raja-arvo

$$\lim_{x_1 \rightarrow -\infty} f(x_1, 0) = \lim_{x_1 \rightarrow -\infty} (x_1^3 - 3x_1) = -\infty,$$

joten funktiolla $f(x_1, x_2)$ ei ole myöskään globaalia minimiä.

Otetaan sitten selville funktion kriittiset pisteet.

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 3x_1^2 - 3 \quad \text{ja} \quad \frac{\partial f(x_1, x_2)}{\partial x_2} = 3x_2^2 - 12,$$

joten kriittiset pisteet toteuttavat yhtälöt

$$3x_1^2 - 3 = 0 \quad \text{ja} \quad 3x_2^2 - 12 = 0.$$

Ensimmäinen yhtälö toteutuu, jos $x_1 = \pm 1$ ja toinen, jos $x_2 = \pm 2$. Kriittisiä pisteitä on siis neljä kappaletta: $(1, 2)$, $(1, -2)$, $(-1, 2)$ ja $(-1, -2)$. Tutkitaan seuraavaksi, ovatko nämä kriittiset pisteet paikallisia ääriarvoja vai satulapisteitä. Funktion $f(x_1, x_2)$ Hessin matriisi on symmetrinen 2×2 -matriisi

$$Hf(x_1, x_2) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{pmatrix}. \quad (1.19)$$

Tutkitaan ominaisarvomenetelmällä, mikä on kriittisen pisteen $(1, 2)$ laatu. Nyt

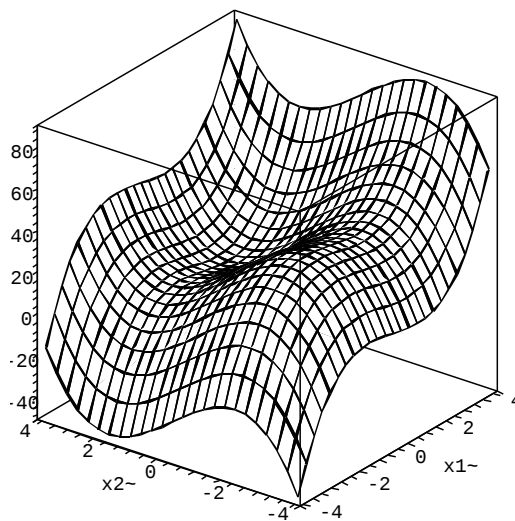
$$Hf(1, 2) = \begin{pmatrix} 6 & 0 \\ 0 & 12 \end{pmatrix}, \quad (1.20)$$

joten determinantti

$$\det(Hf(1, 2) - \lambda I) = \det \begin{pmatrix} 6 - \lambda & 0 \\ 0 & 12 - \lambda \end{pmatrix}.$$

Karakteristiseksi polynomiksi saadaan $p(\lambda) = (6 - \lambda)(12 - \lambda) = \lambda^2 - 18\lambda + 72$, ja sen nollakohdat ovat ominaisarvot $\lambda_1 = 6$ ja $\lambda_2 = 12$. Koska molemmat ominaisarvot ovat positiivisia, on matriisi (1.20) lauseen (1.3.3) nojalla positiivisesti definiitti. Lauseesta (1.4.1) seuraa nyt, että piste $(1, 2)$ on funktion $f(x_1, x_2)$ aito paikallinen minimi.

Vastaavalla tavalla saadaan ratkaistua, että pisteiden $(1, -2)$ ja $(-1, 2)$ tapauksessa ominaisarvot ovat 6 ja -12 ja pisteen $(-1, -2)$ tapauksessa -6 ja -12 . Samanlainen päättely kuin pisteelle $(1, 2)$ antaa nyt ratkaisuksi, että piste $(-1, -2)$ on funktion aito paikallinen maksimi. Pisteet $(1, -2)$ ja $(-1, 2)$ ovat puolestaan lauseen (1.4.2) nojalla funktion satulapisteitä. Optimoitava funktio on vielä esitetty kuvassa 1.3.



Kuva 1.3: $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1 - 12x_2 + 20$.

1.5 Globaalit ääriarvot ja koersiiviset funktiot

1.5.1 Lisätietoa globaaleista ääriarvoista

Yleensä optimointitehtävissä ei olla niinkään kiinnostuneita funktion paikallisista ääriarvoista, vaan ennemminkin globaaleista ääriarvoista. Tällöin halutaan löytää juuri ne optimaaliset olosuhteet, jotka esimerkiksi takaavat tehtaalle mahdollisimman pienet tuotantokustannukset. Edellisten osioiden perusteella optimoitavan funktion globaalit minimi ja maksimi löytyvät joko määrittelyalueen reunapisteistä tai funktion kriittisistä pisteistä.

Lauseessa (1.3.1) tarkasteltiin optimoitavan funktion kriittisen pisteen laatua siinä tapauksessa, että määrittelyalueena on koko vektoriavaruus $\mathbb{R}^n$. Lauseen perusteella kriittinen piste on funktion aito globaali maksimi, jos funktion Hessin matriisi on aina negatiivisesti definiitti. Aito globaali minimi kriittinen piste on puolestaan silloin, kun Hessin matriisi on aina positiivisesti definiitti.

Yleensä todellisissa optimointitehtävissä määrittelyalueelle on asetettu suuria rajoituksia. Tällöin funktion arvo määrittelyalueella olevissa kriittisissä pisteissä sekä alueen reunapisteissä on selvitettävä tarkemmin, jotta globaalit ääriarvot saataisiin selville. Kriittisten pisteiden tutkimisessa voidaan käyttää tällöin apuna lauseen (1.4.1) antamaa tietoa paikallisten ääriarvojen laadusta. Jos määrittelyalue erityisesti on kompakti eli suljettu ja rajoitettu, kannattaa globaaleja ääriarvoja aina etsiä, kuten seuraava lause osoittaa.

Lause 1.5.1. *Olkoon joukko D vektoriavaruuden $\mathbb{R}^n$ suljettu ja rajoitettu epätyhjä osajoukko. Oletetaan lisäksi, että funktio $f(\mathbf{x})$ on jatkuva määrittelyalueessa D . Tällöin $f(\mathbf{x})$ saavuttaa pienimmän ja suurimman arvonsa joukossa D . Toisin sanoen joukossa D on olemassa sellaiset pisteet $\mathbf{x}_{\min}$ ja $\mathbf{x}_{\max}$, että $f(\mathbf{x}_{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}_{\max})$ aina, kun $\mathbf{x} \in \mathbb{R}_n$.*

Tämä lause on yksi analyysin perustuloksista ja nyt sen todistus sivutetaan. Todistus on esitetty kokonaisuudessaan lähdeoteoksessa [6].

1.5.2 Koersiiviset funktiot

Tähän asti luvussa 1 esitetyt menetelmät funktion ääriarvojen laadun tutkimiseksi ovat perustuneet pitkälti optimoitavan funktion Hessin matriisin definiittisyyden selvittämiseen. Aina definiittisyyttä ei kuitenkaan saada selville tai funktiolla voi olla globaali minimi, vaikka Hessin matriisi ei olisikaan positiivisesti definiitti. Optimointia varten tarvitaan siis muita keinoja. Tarkastellaan nyt koersiivisuutta, yhtä tärkeää funktioihin liittyvää ominaisuutta, jota voidaan käyttää apuna optimoinnissa.

Määritelmä 1.5.1. Oletetaan, että funktio $f(\mathbf{x})$ on jatkuva koko avaruudessa $\mathbb{R}^n$. Tällöin funktio on koersiivinen, jos

$$\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty.$$

Siis kaikkia reaalisia vakioita $M > 0$ kohti on olemassa sellainen positiivinen luku r_M , että $f(\mathbf{x}) \geq M$, kun $\|\mathbf{x}\| \geq r_M$.

Koersiivisuuden yhdistää optimointiin seuraava lause.

Lause 1.5.2. *Olko funktio $f(\mathbf{x})$ jatkuva koko vektoriavaruudessa $\mathbb{R}^n$. Jos funktio $f(\mathbf{x})$ on koersiivinen, niin tällöin sillä on ainakin yksi globaali minimi. Jos lisäksi funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat olemassa avaruudessa $\mathbb{R}^n$, niin globaalit minimit kuuluvat funktion kriittisiin pisteisiin.*

Todistus. Funktio $f(\mathbf{x})$ on koersiivinen, joten raja-arvo $\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$. Koska funktion arvo kasvaa nyt rajatta liikuttaessa pisteestä $\mathbf{x} = \mathbf{0}$ pois päin, niin funktio $f(\mathbf{x})$ saa suuria arvoja normin $\|\mathbf{x}\|$ ollessa suuri. Täten on olemassa sellainen positiivinen luku r , että

$$f(\mathbf{x}) > f(\mathbf{0}) \tag{1.21}$$

aina, kun $\|\mathbf{x}\| > r$.

Tutkitaan seuraavaksi origokeskeistä ja r -säteistä joukkoa

$$\bar{B}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq r\},$$

joka on suljettu ja rajoitettu. Muistetaan myös oletus, että funktio $f(\mathbf{x})$ on jatkuva koko avaruudessa $\mathbb{R}^n$. Tällöin funktiolla lauseen (1.5.1) nojalla on minimi joukossa $\bar{B}(\mathbf{0}, r)$. Merkitään tätä minimipistettä $\mathbf{x}^*$. Koska tämä piste on joukon minimi, niin

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \tag{1.22}$$

aina, kun $\mathbf{x}$ kuuluu joukkoon $\bar{B}(\mathbf{0}, r)$. Piste $\mathbf{x} = \mathbf{0}$ kuuluu joukkoon $\bar{B}(\mathbf{0}, r)$, joten nyt pätee myös ehto

$$f(\mathbf{x}^*) \leq f(\mathbf{0}). \quad (1.23)$$

Yhdistämällä ehdot (1.21) ja (1.23) nähdään siten, että

$$f(\mathbf{x}^*) \leq f(\mathbf{0}) < f(\mathbf{x}) \quad (1.24)$$

aina, kun piste $\mathbf{x}$ ei kuulu joukkoon $\bar{B}(\mathbf{0}, r)$. Edelleen ehtojen (1.22) ja (1.24) perusteella nähdään nyt, että $f(\mathbf{x}^*) \leq f(\mathbf{x})$ aina, kun $\mathbf{x} \in \mathbb{R}^n$. Näin ollen on osoitettu, että piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ globaali minimi.

Jos vielä lisäksi funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat jatkuvia avaruudessa $\mathbb{R}^n$, niin gradientti $\nabla f(\mathbf{x}^*)$ on olemassa kaikkialla. Koska piste $\mathbf{x}^*$ on minimi ja määrittelyalueen $\mathbb{R}^n$ sisäpiste, niin lauseen (1.2.1) nojalla $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Siis kyseessä on funktion $f(\mathbf{x})$ kriittinen piste. $\square$

Esimerkki 1.5.1. Tehtävänä on nyt selvittää funktion

$$f(x_1, x_2) = x_1^4 + x_2^4 - 32x_2^2$$

globaali minimi. Funktio on määritelty koko avaruudessa $\mathbb{R}^2$. Otetaan ensin selville funktion kriittiset pisteet.

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 4x_1^3 \quad \text{ja} \quad \frac{\partial f(x_1, x_2)}{\partial x_2} = 4x_2^3 - 64x_2,$$

joten kriittiset pisteet toteuttavat yhtälöt

$$4x_1^3 = 0 \quad \text{ja} \quad 4x_2(x_2^2 - 16) = 0.$$

Molemmat yhtälöt toteuttavia pisteitä eli funktion $f(x_1, x_2)$ kriittisiä pisteitä on kolme kappaletta: $(0, 0)$, $(0, 4)$ ja $(0, -4)$. Funktion $f(x_1, x_2)$ Hessin matriisi on puolestaan symmetrinen 2×2 -matriisi

$$Hf(x_1, x_2) = \begin{pmatrix} 12x_1^2 & 0 \\ 0 & 12x_2^2 - 64 \end{pmatrix}.$$

Lauseen (1.3.1) avulla ei voida nyt tehdä johtopäätöksiä globaalin minimin olemassaolosta, sillä Hessin matriisi ei ole aina positiivisesti definiitti. Esimerkiksi pisteessä $(1, 2)$ matriisin johtava pääminori $\Delta_1 = 12 > 0$, mutta sitä vastoin $\Delta_2 = 12 * (12 * 4 - 64) = -192 < 0$.

Tutkitaan seuraavaksi, onko funktio $f(x_1, x_2)$ koersiivinen. Funktio voidaan kirjoittaa muodossa

$$f(x_1, x_2) = (x_1^4 + x_2^4) \left(1 - \frac{32x_2^2}{x_1^4 + x_2^4} \right).$$

Nyt termi $32x_2^2/(x_1^4 + x_2^4) \rightarrow 0$, kun normi $\|x_1, x_2\| = \sqrt{x_1^2 + x_2^2} \rightarrow +\infty$. Täten funktio $f(x_1, x_2)$ on koersiivinen, sillä

$$\lim_{\|x_1, x_2\| \rightarrow +\infty} f(x_1, x_2) = (x_1^4 + x_2^4)(1 - 0) = +\infty.$$

Lauseen (1.5.2) nojalla, joku koersiivisen funktion kriittisistä pisteistä on globaali minimi. Kokeilemalla nähdään, että $f(0, 0) = 0$ ja $f(0, 4) = f(0, -4) = -256$, joten pisteet $(0, 4)$ ja $(0, -4)$ ovat molemmat funktion globaaleja minimeitä.

Luku 2

Konvekssi optimointi

Luku 1 lopetettiin tarkastelemalla koersiivisten funktioiden optimointia. Tämä koko luku käsittelee toista erikoistapausta, konvekseja funktioita, jotka ovat tärkeitä esimerkiksi monissa taloustieteen optimointiongelmissa. Myöhemmin luvussa 3 tullaan tutustumaan yhteen konvekseihin funktioihin liittyvistä mielenkiintoisista sovelluksista, geometriseen optimointiin. Aloitetaan aiheen käsittely konveksisuuteen liittyvistä peruskäsitteistä.

2.1 Konveksit joukot ja funktiot

Määritelmä 2.1.1. Olkoon joukko C vektoriavaruuden $\mathbb{R}^n$ osajoukko. Tällöin C on konvekssi joukko, jos pisteitä $\mathbf{x}_1$ ja $\mathbf{x}_2$ yhdistävä jana

$$[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1) : 0 \leq \lambda \leq 1\}$$

sisältyy joukkoon C aina, kun pisteet $\mathbf{x}_1$ ja $\mathbf{x}_2$ ovat joukon C pisteitä.

Pisteitä $\mathbf{x}_1$ ja $\mathbf{x}_2$ yhdistävä jana voidaan esittää myös muodossa

$$[\mathbf{x}_1, \mathbf{x}_2] = \{\lambda\mathbf{x}_2 + (1 - \lambda)\mathbf{x}_1 : 0 \leq \lambda \leq 1\}.$$

Tätä esitystä käyttäen määritellään seuraavaksi konvekssi funktio.

Määritelmä 2.1.2. Olkoon C vektoriavaruuden $\mathbb{R}^n$ konvekksi osajoukko ja $f(\mathbf{x})$ reaaliarvoinen funktio, joka on määritelty joukossa C . Tällöin:

(a) funktio $f(\mathbf{x})$ on konvekksi joukossa C , jos

$$f(\lambda \mathbf{x}_1 + [1 - \lambda] \mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2)$$

aina, kun $\mathbf{x}_1$ ja $\mathbf{x}_2$ ovat joukon C pisteitä ja $0 \leq \lambda \leq 1$;

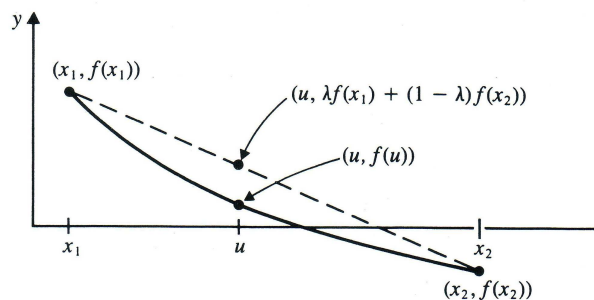
(b) funktio $f(\mathbf{x})$ on aidosti konvekksi joukossa C , jos

$$f(\lambda \mathbf{x}_1 + [1 - \lambda] \mathbf{x}_2) < \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2)$$

aina, kun $\mathbf{x}_1$ ja $\mathbf{x}_2$ ovat joukon C kaksi eri pistettä ja $0 < \lambda < 1$.

Huomautus. Jos konveksin funktion määritelmässä merkinnät $\leq$ ja $<$ korvataan merkinnöillä $\geq$ ja $>$, saadaan konkaavin funktion määritelmä. Tulevissa tarkasteluissa konkaaveita funktioita ei silti tarkastella, koska niitä koskevat tulokset saadaan johdettua helposti konvekseja funktioita koskevista tuloksista. Jos näet funktio $f(\mathbf{x})$ on konkaavi, niin $-f(\mathbf{x})$ on silloin konvekksi.

Tarkastellaan vielä lyhyesti, millaisia konveksit funktiot oikeastaan ovat. Tehdään tarkastelut selkeyden vuoksi yhden muuttujan funktiolle. Kuva 2.1 havainnollistaa selitystä.



Kuva 2.1: Aidosti konvekksi funktio [3].

Olkoon $f(x)$ reaaliarvoinen funktio ja määritelty välillä I . Olkoot lisäksi pisteet x_1 ja x_2 välin I kaksi satunnaisesti valittua pistettä. Tällöin määritel-

män (2.1.2) perusteella $f(x)$ on konvekksi funktio välillä I , jos kyseisellä välillä funktion kuvaaja ei ole koskaan funktion pisteitä $(x_1, f(x_1))$ ja $(x_2, f(x_2))$ yhdistävän janan yläpuolella. Jos lisäksi x_1 ja x_2 ovat välin I eri pisteitä ja funktion $f(x)$ kuvaaja on koko jana funktion pisteitä $(x_1, f(x_1))$ ja $(x_2, f(x_2))$ yhdistävän janan alapuolella, funktio on aidosti konvekksi.

Oletetaan vielä, että konvekksi funktio on derivoituva välillä I . Jos tällöin tarkasteltaisiin pisteeseen $(u, f(u))$ piirretyn tangentin kuvaajaa, nähtäisiin, että saatu tangenttisuora ei ole koskaan funktion kuvaajan yläpuolella. Tangenttisuora on puolestaan koko ajan funktion kuvaajan alapuolella, jos funktio on aidosti konvekksi.

2.2 Konveksisuuden yhteys optimointiin

Seuraava lause osoittaa, että funktioiden konveksisuutta voidaan käyttää apuna optimoinnissa.

Lause 2.2.1. *Olkoon C avaruuden $\mathbb{R}^n$ konvekksi osajoukko ja $f(\mathbf{x})$ määritelty ja konvekksi funktio joukossa C . Olkoon lisäksi $\mathbf{x}^*$ joukon C piste. Tällöin:*

- (a) *jos piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ paikallinen minimi, niin se on myös globaali minimi joukossa C ;*
- (b) *jos piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ aito paikallinen minimi, niin se on myös aito globaali minimi joukossa C .*

Todistus. Todistetaan ainoastaan kohta (a), koska kohdan (b) todistus menee samalla tavalla.

Nyt konveksin joukon C piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ paikallinen minimikohta kyseisessä joukossa. Täten on olemassa sellainen positiivinen luku r , että

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \tag{2.1}$$

aina, kun $\|\mathbf{x} - \mathbf{x}^*\| < r$ ja $\mathbf{x} \in C$.

Tarkastellaan seuraavaksi joukosta C satunnaisesti valittua pistettä $\mathbf{y}$. Yhdistetään pisteet $\mathbf{x}^*$ ja $\mathbf{y}$ janalla

$$[\mathbf{x}^*, \mathbf{y}] = \{\lambda \mathbf{y} + (1 - \lambda) \mathbf{x}^* : 0 \leq \lambda \leq 1\}.$$

Koska pisteet $\mathbf{x}^*$ ja $\mathbf{y}$ ovat molemmat konveksin joukon C pisteitä, niin myös janan pisteet kuuluvat joukkoon C . Huomataan myös, että mitä pienempi arvo luvulle λ valitaan, sitä lähempänä janalla ollaan pistettä $\mathbf{x}^*$. Kun luvulle λ siten valitaan tapeeksi pieni arvo λ_z , päädytään joukon C pisteeseen $\mathbf{z}$, jolle pätee $\|\mathbf{z} - \mathbf{x}^*\| < r$. Siis ehdon (2.1) nojalla

$$f(\mathbf{x}^*) \leq f(\mathbf{z}) = f(\lambda_z \mathbf{y} + [1 - \lambda_z] \mathbf{x}^*). \quad (2.2)$$

Koska oletuksen nojalla funktio $f(\mathbf{x})$ on konvekssi, niin epäyhtälöstä (2.2) seuraa

$$f(\mathbf{x}^*) \leq \lambda_z f(\mathbf{y}) + [1 - \lambda_z] f(\mathbf{x}^*) = \lambda_z f(\mathbf{y}) + f(\mathbf{x}^*) - \lambda_z f(\mathbf{x}^*). \quad (2.3)$$

Nyt, kun vähennetään tekijä $f(\mathbf{x}^*)$ epäyhtälön (2.3) molemmilta puolilta ja jaetaan epäyhtälö vielä positiivisella vakiolla λ_z , saadaan tulokseksi

$$f(\mathbf{x}^*) \leq f(\mathbf{y}).$$

Piste $\mathbf{y}$ oli nyt valittu satunnaisesti joukosta C , joten saatu tulos pätee kaikille joukon C pisteille. Siis piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ globaali minimi joukossa C . $\square$

Ennen kuin voidaan esittää lisää konveksien funktioiden optimointia helpottavia apukeinoja, täytyy tarkastella seuraavaa lausetta. Lauseen todistus perustuu lähdeeteokseen [7].

Lause 2.2.2. *Oletetaan, että joukko C on avaruuden $\mathbb{R}^n$ konvekssi osajoukko. Oletetaan lisäksi, että funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat jatkuvia joukossa C . Tällöin:*

(a) *funktio $f(\mathbf{x})$ on konvekssi joukossa C , jos ja vain jos*

$$f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y})$$

aina, kun pisteet $\mathbf{x}$ ja $\mathbf{y}$ kuuluvat joukkoon C ;

(b) funktio $f(\mathbf{x})$ on aidosti konvekksi joukossa C , jos ja vain jos

$$f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) < f(\mathbf{y})$$

aina, kun pisteet $\mathbf{x}$ ja $\mathbf{y}$ kuuluvat joukkoon C ja $\mathbf{x} \neq \mathbf{y}$.

Todistus. Todistetaan lauseesta kohta (a).

Oletetaan ensin, että funktio $f(\mathbf{x})$ on konvekksi joukossa C ja että pisteet $\mathbf{x}$ ja $\mathbf{y}$ kuuluvat joukkoon C . Tällöin piste $\lambda\mathbf{y} + (1 - \lambda)\mathbf{x}$ kuuluu myös joukkoon C , kun $0 < \lambda \leq 1$. Koska funktio on konvekksi, niin

$$f(\lambda\mathbf{y} + [1 - \lambda]\mathbf{x}) \leq \lambda f(\mathbf{y}) + (1 - \lambda)f(\mathbf{x}),$$

mikä saadaan esitettyä termejä uudelleen järjestelemällä muodossa

$$f(\mathbf{x} + \lambda[\mathbf{y} - \mathbf{x}]) - f(\mathbf{x}) \leq \lambda[f(\mathbf{y}) - f(\mathbf{x})]. \quad (2.4)$$

Tarkastellaan sitten epäyhtälön (2.4) vasenta puolta. Funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat lauseen oletusten nojalla olemassa joukossa C . Täten analyysin perustuloksiin kuuluvan väliarvolauseen nojalla on välillä $[\mathbf{x}, \mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})]$ olemassa sellainen piste $\mathbf{z}$, että

$$\begin{aligned} f(\mathbf{x} + \lambda[\mathbf{y} - \mathbf{x}]) - f(\mathbf{x}) &= \nabla f(\mathbf{z})(\mathbf{x} + \lambda[\mathbf{y} - \mathbf{x}] - \mathbf{x}) \\ &= \nabla f(\mathbf{z})\lambda(\mathbf{y} - \mathbf{x}). \end{aligned} \quad (2.5)$$

Sijoitetaan nyt kaava (2.5) epäyhtälöön (2.4). Tällöin saadaan, että

$$\nabla f(\mathbf{z})\lambda(\mathbf{y} - \mathbf{x}) \leq \lambda[f(\mathbf{y}) - f(\mathbf{x})].$$

Jaetaan tämä epäyhtälö seuraavaksi positiivisella luvulla λ ja siirretään termi $f(\mathbf{x})$ vasemmalle puolelle. Näin on saatu, että

$$f(\mathbf{x}) + \nabla f(\mathbf{z})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}).$$

Huomioidaan vielä, että kun luku $\lambda \rightarrow 0$ niin piste $\mathbf{z} \rightarrow \mathbf{x}$. Tulokseksi saadaan lauseen kohdan (a) väite eli

$$f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}).$$

Puolet kohdan (a) todistuksesta on nyt tehty. Vielä pitää tutkia, pitääkö kohta (a) paikkansa myös silloin, kun tarkastelu tehdään toiseen suuntaan. Oletetaan siis nyt, että $f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y})$ aina, kun pisteet $\mathbf{x}$ ja $\mathbf{y}$ kuuluvat konvekssiin joukkoon C .

Joukko C on konvekksi joten piste $\mathbf{z} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ kuuluu myös joukkoon C , kun $0 < \lambda < 1$. Piste $\mathbf{y}$ voidaan täten esittää muodossa

$$\mathbf{y} = \frac{\mathbf{z} - \lambda\mathbf{x}}{1 - \lambda},$$

josta edelleen seuraa, että

$$\begin{aligned} \mathbf{y} - \mathbf{z} &= \frac{\mathbf{z} - \lambda\mathbf{x}}{1 - \lambda} - \mathbf{z} \\ &= \frac{-\lambda(\mathbf{x} - \mathbf{z})}{1 - \lambda}. \end{aligned} \tag{2.6}$$

Tarkastellaan nyt pisteparia $(\mathbf{y}, \mathbf{z})$. Oletuksen nojalla pätee epäyhtälö

$$f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{y} - \mathbf{z}) \leq f(\mathbf{y}).$$

Sijoitetaan tähän epäyhtälöön kaava (2.6), jolloin

$$f(\mathbf{z}) + \nabla f(\mathbf{z})\left(\frac{-\lambda}{1 - \lambda}\right)(\mathbf{x} - \mathbf{z}) \leq f(\mathbf{y}).$$

Kerrotaan tämä epäyhtälö puolittain positiivisella luvulla $1 - \lambda$ ja järjestetään sitten epäyhtälön vasemman puolen termejä hieman toisin. Saadaan, että

$$f(\mathbf{z}) - \lambda[f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{x} - \mathbf{z})] \leq (1 - \lambda)f(\mathbf{y}).$$

Tästä seuraa edelleen, että

$$f(\mathbf{z}) \leq \lambda[f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{x} - \mathbf{z})] + (1 - \lambda)f(\mathbf{y}). \tag{2.7}$$

Tarkastellaan vielä pisteparia $(\mathbf{x}, \mathbf{z})$. Oletuksen nojalla pätee epäyhtälö

$$f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{x} - \mathbf{z}) \leq f(\mathbf{x}). \tag{2.8}$$

Sijoitetaan tämä epäyhtälö (2.8) edellä johdettuun epäyhtälöön (2.7). Tällöin

$$f(\mathbf{z}) = f(\lambda\mathbf{x} + [1 - \lambda]\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

Näin on osoitettu, että funktio $f(\mathbf{x})$ on konvekksi ja lauseen väite (a) on todistettu myös toiseen suuntaan.

Lauseen kohdan (b) todistus menee periaatteessa samalla tavalla kuin kohdan (a). $\square$

Tarkastellaan vielä lyhyesti edellisen lauseen (2.2.2) ideaa. Funktion $f(\mathbf{x})$ pistettä $(\mathbf{x}, f(\mathbf{x}))$ sivuava tangenttitaso on $P_{\mathbf{x}} = \{(\mathbf{y}, y) \in \mathbb{R}^{n+1} : y = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})\}$. Lauseen kohdan (a) mukaan tämän tangenttitason pisteet $(\mathbf{y}, y)$ eivät ole koskaan funktion pisteiden $(\mathbf{y}, f(\mathbf{y}))$ yläpuolella, jos funktio on konvekksi. Jos funktio taas on aidosti konvekksi, niin lauseen kohdan (b) nojalla pisteet $(\mathbf{y}, y)$ ovat ehdottomasti funktion pisteiden $(\mathbf{y}, f(\mathbf{y}))$ alapuolella, kunhan $\mathbf{y} \neq \mathbf{x}$.

Lauseesta (2.2.2) saadaan nyt yksinkertaisesti johdettua seuraava, optimoinnin kannalta mielenkiintoinen tulos.

Lause 2.2.3. *Oletetaan, että joukko C on avaruuden $\mathbb{R}^n$ konvekksi osajoukko ja että funktion $f(\mathbf{x})$ kaikki ensimmäiset osittaisderivaatat ovat jatkuvia joukossa C . Oletetaan lisäksi, että joukon C piste $\mathbf{x}^*$ on funktion $f(\mathbf{x})$ kriittinen piste. Tällöin:*

- (a) *piste $\mathbf{x}^*$ on globaali minimi, jos funktio $f(\mathbf{x})$ on konvekksi joukossa C ;*
- (b) *piste $\mathbf{x}^*$ on aito globaali minimi, jos funktio $f(\mathbf{x})$ on aidosti konvekksi joukossa C .*

Todistus. Olkoon piste $\mathbf{x}$ satunnaisesti valittu joukon C piste ja lisäksi $\mathbf{x}^*$ funktion $f(\mathbf{x})$ kriittinen piste joukossa C . Luvun 1 perusteella tällöin gradientti $\nabla f(\mathbf{x}^*) = 0$ ja lauseen (2.2.2) nojalla

$$f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = f(\mathbf{x}^*) \leq f(\mathbf{x})$$

Koska piste $\mathbf{x}$ oli nyt satunnaisesti valittu, niin piste $\mathbf{x}^*$ on globaali minimi. Vastaavalla tavalla saataisiin osoitettua, että aidosti konveksin funktion kriittinen piste on aito globaali minimi. $\square$

Tämän osion alussa lauseessa (2.2.1) osoitettiin siis, että konveksin funktion paikalliset minimiit ovat myös globaaleja minimeitä. Nyt lause (2.2.3) osoitti, että itse asiassa konveksin funktion kaikki kriittiset pisteet ovat globaaleja minimeitä.

Jotta funktion konveksisuutta voitaisiin käyttää apuna optimoinnissa, pitää funktion konveksisuus tietysti pystyä osoittamaan. Konveksisuuden osoittaminen määritelmiin (2.1.2) perustuen voi kuitenkin olla työlästä. Seuraavana aletaan siksi tutkimaan konveksien funktioiden tunnistamiseen liittyviä apukeinoja.

2.3 Konveksien funktioiden tunnistaminen

Luvun 1 lauseessa (1.3.1) osoitettiin yleisesti, että funktion $f(\mathbf{x})$ kriittiset pisteet ovat globaaleja minimeitä, jos funktion Hessin matriisi on positiivisesti semidefiniitti. Edellisen osion lause (2.2.3) taas osoitti, että jos funktio on konvekksi, niin sen kaikki kriittiset pisteet ovat globaaleja minimeitä. On siis odotettua, että funktion konveksisuuden määrittäminen saadaan liitettyä funktion Hessin matriisin definiittisyyteen.

Lause 2.3.1. *Olkoon avoin joukko C taas vektoriavaruuden $\mathbb{R}^n$ konvekksi osajoukko ja funktio $f(\mathbf{x})$ konvekksi joukossa C . Olkoon lisäksi funktion kaikki toiset osittaisderivaatat jatkuvia kyseisessä joukossa. Tällöin:*

- (a) funktio $f(\mathbf{x})$ on konvekksi, jos sen Hessin matriisi $Hf(\mathbf{x})$ on positiivisesti semidefiniitti aina, kun $x \in C$;*
- (b) funktio $f(\mathbf{x})$ on aidosti konvekksi, jos sen Hessin matriisi $Hf(\mathbf{x})$ on positiivisesti definiitti aina, kun $x \in C$.*

Todistus. Olkoot $\mathbf{x}$ ja $\mathbf{y}$ kaksi joukon C satunnaisesti valittua pistettä. Koska joukko C on konvekksi, niin se sisältää myös pisteiden $\mathbf{x}$ ja $\mathbf{y}$ välisen janan $[\mathbf{x}, \mathbf{y}]$. Lauseessa (1.2.2) taas esitettiin Taylorin kehitelmä useamman muut-

tujan funktiolle, jonka mukaan

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T Hf(\mathbf{z})(\mathbf{y} - \mathbf{x}).$$

Koska tässä kehitelmässä piste $\mathbf{z}$ kuuluu välille $[\mathbf{x}, \mathbf{y}]$, niin se kuuluu nyt myös joukkoon C . Jos funktio $f(\mathbf{x})$ on positiivisesti semidefiniitti, niin luvun 1 perusteella termi $(\mathbf{y} - \mathbf{x})^T Hf(\mathbf{z})(\mathbf{y} - \mathbf{x})/2 \geq 0$. Tällöin saadaan, että

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}).$$

Lauseen (2.2.2) nojalla funktio on näin ollen konvekksi.

Vastaavasti, jos termi $(\mathbf{y} - \mathbf{x})^T Hf(\mathbf{z})(\mathbf{y} - \mathbf{x})/2 > 0$ ja $\mathbf{y} \neq \mathbf{x}$, niin

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

ja kyseessä on aidosti konvekssi funktio. □

Huomautus. Tarkastellaan erikoistapauksena yhden muuttujan funktiota $f(x)$. Funktio on lauseen (2.3.1) perusteella konvekssi, jos sen toinen derivaatta $f''(x) \geq 0$. Jos taas $f''(x) > 0$, niin funktio on aidosti konvekssi. Näiden tulosten todistaminen tapahtuu juuri kuten edelläkin, mutta nyt apuna kannattaa käyttää lauseessa (1.1.2) esitettyä yhden muuttujan funktion Taylorin kehitelmää.

Esimerkki 2.3.1. Jo aiemmin esimerkeissä (1.2.1) ja (1.3.2) tutkittiin funktiota $f(\mathbf{x}) = x_1^2 - 4x_1 + 2x_2^2 + 7$. Tällöin funktion Hessin matriisin definitisyys selvitettiin sen determinanttien avulla. Matriisi on positiivisesti definiti. Näin ollen kyseinen funktio käy lauseen (2.3.1) perusteella esimerkkinä myös aidosti konveksista funktiosta. Lisäksi aiemmissa esimerkeissä todettiin, että kyseisen funktion ainoa kriittinen piste on aito globaali minimi. Koska funktio on konvekssi, niin lause (2.2.3) antaa nyt saman tuloksen.

Edellisessä esimerkissä tarkateltiin funktiota, jonka definitiisyyden ja siten myös konveksisuuden määrittäminen on helppoa. Usein matriisin definitiisyyden määrittäminen voi olla kuitenkin vaikeaa, vaikka apuna käytettäisiin

luvussa 1 esiteltyjä matriisin determinantteihin ja ominaisarvoihin perustuvia menetelmiä. Seuraavassa lauseessa esitellään muutamia hyviä apukeinoja, jotka helpottavat funktioiden konveksisuuden toteamista ja siten myös niiden optimointia.

Lause 2.3.2. *Olkoon joukko C taas avaruuden $\mathbb{R}^n$ konvekksi osajoukko.*

(a) *Jos funktiot $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})$ ovat kaikki konvekseja joukossa C , niin myös summafunktio $f(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x}) + \dots + f_n(\mathbf{x})$ on konvekksi. Jos taas yksikin funktioista $f_i(\mathbf{x})$ on aidosti konvekksi, niin sitä on myös summafunktio.*

(b) *Jos funktio $f(\mathbf{x})$ on konvekksi joukossa C ja α on positiivinen luku, niin myös funktio $\alpha f(\mathbf{x})$ on konvekksi. Vastaava tulos pätee aidosti konveksille funktiolle.*

(c) *Olkoon funktio $f(\mathbf{x})$ konvekksi joukossa C ja olkoon sen arvojoukko $D \in \mathbb{R}$. Oletetaan lisäksi, että funktio $g(y)$ on konvekksi ja kasvava funktion $f(\mathbf{x})$ arvojoukossa D . Tällöin yhdistetty funktio $g(f(\mathbf{x}))$ on konvekksi. Jos taas funktio $f(\mathbf{x})$ on aidosti konvekksi ja funktio $g(y)$ konvekksi sekä aidosti kasvava, niin yhdistetty funktio on aidosti konvekksi.*

Todistus. Oletetaan todistuksessa, että pisteet $\mathbf{x}$ ja $\mathbf{y}$ kuuluvat konvekssiin joukkoon C . Täten myös janan $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$ pisteet sisältyvät joukkoon C aina, kun $0 \leq \lambda \leq 1$.

(a) Tarkastellaan nyt joukossa C konvekseja funktioita $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})$. Tällöin summafunktio

$$\begin{aligned} f(\lambda\mathbf{x} + [1 - \lambda]\mathbf{y}) &= (f_1 + \dots + f_n)(\lambda\mathbf{x} + [1 - \lambda]\mathbf{y}) \\ &= f_1(\lambda\mathbf{x} + [1 - \lambda]\mathbf{y}) + \dots + f_n(\lambda\mathbf{x} + [1 - \lambda]\mathbf{y}) \\ &\leq \lambda f_1(\mathbf{x}) + [1 - \lambda]f_1(\mathbf{y}) + \dots + \lambda f_n(\mathbf{x}) + [1 - \lambda]f_n(\mathbf{y}) \\ &= \lambda(f_1 + \dots + f_n)(\mathbf{x}) + [1 - \lambda](f_1 + \dots + f_n)(\mathbf{y}). \end{aligned}$$

Summafunktio $f(\mathbf{x})$ on siis osoitettu konveksiksi. Jos sen sijaan yksikin funktioista $f_i(\mathbf{x})$ on aidosti konvekksi, niin edellisestä päättelystä tulee aito epäyhtälö. Tämä tekee summafunktioista aidosti konveksin.

(b) Funktio $f(\mathbf{x})$ on nyt konvekksi ja α on positiivinen luku. Tällöin

$$\begin{aligned}\alpha f(\lambda \mathbf{x} + [1 - \lambda] \mathbf{y}) &\leq \alpha(\lambda f(\mathbf{x}) + [1 - \lambda] f(\mathbf{y})) \\ &= \lambda \alpha f(\mathbf{x}) + [1 - \lambda] \alpha f(\mathbf{y})\end{aligned}$$

ja funktio $\alpha f(\mathbf{x})$ on konvekksi. Jos taas funktio $f(\mathbf{x})$ on aidosti konvekksi, niin edellisestä päättelystä tulee aito epäyhtälö ja funktiosta $\alpha f(\mathbf{x})$ näin ollen aidosti konvekksi.

(c) Olkoon funktio $f(\mathbf{x})$ konvekksi, joten

$$f(\lambda \mathbf{x} + [1 - \lambda] \mathbf{y}) \leq \lambda f(\mathbf{x}) + [1 - \lambda] f(\mathbf{y}). \quad (2.9)$$

Koska oletettiin, että funktio $g(y)$ on kasvava funktion $f(\mathbf{x})$ arvojoukossa, niin epäyhtälön (2.9) nojalla

$$g(f(\lambda \mathbf{x} + [1 - \lambda] \mathbf{y})) \leq g(\lambda f(\mathbf{x}) + [1 - \lambda] f(\mathbf{y})). \quad (2.10)$$

Koska oletuksen nojalla funktio $g(y)$ on myös konvekssi funktion $f(\mathbf{x})$ arvojoukossa, niin

$$g(\lambda f(\mathbf{x}) + [1 - \lambda] f(\mathbf{y})) \leq \lambda g(f(\mathbf{x})) + [1 - \lambda] g(f(\mathbf{y})). \quad (2.11)$$

Epäyhtälöistä (2.10) ja (2.11) seuraa siis, että

$$g(f(\lambda \mathbf{x} + [1 - \lambda] \mathbf{y})) \leq \lambda g(f(\mathbf{x})) + [1 - \lambda] g(f(\mathbf{y})).$$

Näin ollen yhdistetty funktio $g(f(\mathbf{x}))$ on konvekksi. Jos taas edellä funktio $f(\mathbf{x})$ on aidosti konvekksi ja funktio $g(y)$ puolestaan aidosti kasvava, tulee epäyhtälöistä (2.9) ja (2.10) aitoja. Tällöin yhdistetty funktiokin on aidosti konvekksi. $\square$

Päätetään tämä luku yhteen laajempaan esimerkkiin konveksin funktion optimoinnista.

Esimerkki 2.3.2. Etsitään minimi funktiolle

$$f(\mathbf{x}) = x_1^2 + x_2^2 + e^{x_1^2 + x_2^2},$$

joka on määritelty avaruudessa $\mathbb{R}^2$. Otetaan ensin selville funktion kriittiset pisteet.

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 2x_1 + 2x_1e^{x_1^2+x_2^2} \quad \text{ja} \quad \frac{\partial f(\mathbf{x})}{\partial x_2} = 2x_2 + 2x_2e^{x_1^2+x_2^2},$$

joten kriittiset pisteet toteuttavat yhtälöt

$$2x_1(1 + e^{x_1^2+x_2^2}) = 0 \quad \text{ja} \quad 2x_2(1 + e^{x_1^2+x_2^2}) = 0.$$

Termi $e^{x_1^2+x_2^2} > 0$, joten ainoa edelliset yhtälöt toteuttava piste eli kriittinen piste on $(0, 0)$. Funktion $f(\mathbf{x})$ Hessin matriisi olisi erittäin monimutkainen, joten on tutkittava jollain muulla tavalla, onko piste $(0, 0)$ funktion globaali minimi. Tutkitaan nyt, onko funktio $f(\mathbf{x})$ konvekksi. Jos se on, niin lauseen (2.2.3) perusteella kriittinen piste on globaali minimi.

Tarkasteltava funktio voidaan kirjoittaa myös muodossa

$$f(\mathbf{x}) = h(\mathbf{x}) + g(h(\mathbf{x})),$$

missä

$$h(\mathbf{x}) = x_1^2 + x_2^2 \quad \text{ja} \quad g(y) = e^y.$$

Funktio $g(y)$ on aidosti kasvava, sillä derivaatta $g'(y) = e^y > 0$ aina, kun $y \in \mathbb{R}$. Siis luonnollisesti $g(y)$ on aidosti kasvava myös funktion $h(\mathbf{x})$ arvojoukossa. Vastaavasti toinen derivaatta $g''(y) = e^y > 0$, joten lauseen (2.3.1) huomautuksen perusteella funktio $g(y)$ on aina aidosti konvekksi.

Funktion $h(\mathbf{x})$ Hessin matriisi on puolestaan

$$Hf(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

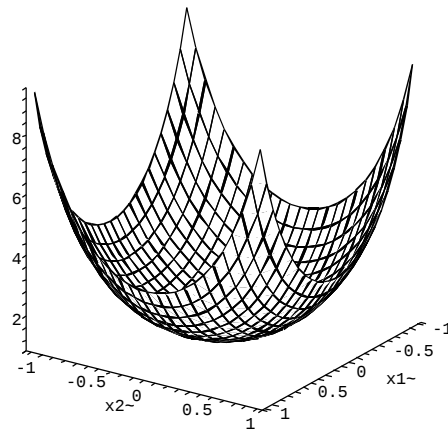
Esimerkiksi luvussa 1 esitetyn determinanttikriteerin avulla, voidaan nopeasti todeta, että matriisi on positiivisesti definiitti. Lauseen (2.3.1) nojalla funktio $h(\mathbf{x})$ on siis aidosti konvekksi.

Käytetään seuraavaksi apuna lauseen (2.3.2) tuloksia. Kohdan (c) perusteella yhdistetty funktio $g(h(\mathbf{x}))$ on aidosti konvekksi. Näin on siksi, että funktio

$h(\mathbf{x})$ todettiin aidosti konveksiksi ja funktio $g(y)$ puolestaan aidosti konveksiksi sekä aidosti kasvavaksi funktion $h(\mathbf{x})$ arvojoukossa. Itse asiassa yhdistetyn funktion aito konveksisuus olisi toteutunut, vaikka $g(y)$ ei olisikaan ollu aidosti konvekksi vaan ainoastaan konvekksi.

Lauseen kohdan (a) perusteella funktio $f(\mathbf{x})$ on siis nyt aidosti konvekksi, sillä summan $h(\mathbf{x}) + g(h(\mathbf{x}))$ molemmat termit ovat konvekseja ja vähintään toinen termeistä on aidosti konvekksi. Nythän peräti molemmat termit ovat aidosti konvekseja.

Näin ollen kriittinen piste $(0, 0)$ on osoittautunut funktion $f(\mathbf{x})$ aidoksi globaaliksi minimiksi. Mainittakoon vielä, että tarkasteltu funktio $f(\mathbf{x})$ on koersiivinen, joten kriittinen piste on myös lauseen (1.5.2) perusteella aito globaali minimi. Optimoitu funktio on vielä esitetty kuvassa 2.2. Funktion arvo kasvaa jyrkästi, kun $\|\mathbf{x}\| \rightarrow +\infty$.



Kuva 2.2: $f(\mathbf{x}) = x_1^2 + x_2^2 + e^{x_1^2 + x_2^2}$.

Luku 3

Geometrinen optimointi

Tässä luvussa keskitytään geometriseen optimointiin, joka on tärkeä menetelmä monien käytännön optimointiongelmien ratkaisemisessa. Periaatteessa geometrinen optimointi on edellisessä luvussa esitetyn konveksin optimoinnin sovellus, sillä menetelmän matemaattinen johtaminen perustuu pohjimmiltaan juuri tiettyihin konveksien funktioiden ominaisuuksiin.

3.1 Optimointi aritmeettis-geometrisen keskiarvoepäyhtälön avulla

Ennen kuin siirrytään varsinaiseen geometriseen optimointiin, käsitellään aritmeettis-geometrista keskiarvoepäyhtälöä. Sen avulla kyetään ratkaisemaan joitakin epälineaarisia optimointiongelmia, niin rajoitettuja kuin rajoittamattomiakin. Ennen tämän tärkeän epäyhtälön esittämistä tarkastellaan kuitenkin sen todistamisessa tarvittavaa apulausetta (3.1.1). Apulause liittyy konvekseihin funktioihin ja sen todistaminen perustuu osittain lähdeeteokseen [7].

Apulause 3.1.1. (a) Oletetaan, että joukko C on avaruuden $\mathbb{R}^n$ konvekksi osajoukko ja että funktio $f(\mathbf{x})$ on konvekksi joukossa C . Oletetaan lisäksi,

että $\delta_1, \delta_2, \dots, \delta_n$ ovat epänegatiivisia lukuja, joiden summa on 1. Tällöin, jos pisteet $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ kuuluvat kaikki joukkon C , niin

$$f\left(\sum_{i=1}^n \delta_i \mathbf{x}_i\right) \leq \sum_{i=1}^n \delta_i f(\mathbf{x}_i). \quad (3.1)$$

(b) Tarkennetaan vielä (a) kohdan oletuksia siten, että funktio $f(\mathbf{x})$ on nyt aidosti konvekssi joukossa C ja kaikki luvut δ_i ovat positiivisia. Tällöin epäyhtälössä (3.1) pätee yhtäsuuruus, jos ja vain jos kaikki pisteet $\mathbf{x}_i$ ovat samoja.

Todistus. Tehdään todistus induktiolla.

Perusaskel.

Tarkastellaan ensin tilannetta $n = 2$. Tutkitaan siis funktion käyttäytymistä pisteessä $\delta_1 \mathbf{x}_1 + \delta_2 \mathbf{x}_2$, joka on kahden pisteen konvekssi kombinaatio. Tällöin lauseen (a)-kohta yksinkertaistuu suoraan konveksin funktion määritelmäksi (2.1.2)(a) ja on siten totta. Lauseen kohta (b) seuraa puolestaan aidosti konveksin funktion määritelmästä (2.1.2)(b). Oletetaan, että lauseen (b)-kohdan oletukset ovat voimassa. Tällöin määritelmän nojalla välttämättä

$$f(\delta_1 \mathbf{x}_1 + \delta_2 \mathbf{x}_2) < \delta_1 f(\mathbf{x}_1) + \delta_2 f(\mathbf{x}_2),$$

jos $\mathbf{x}_1$ ja $\mathbf{x}_2$ ovat eri pisteitä. Yhtäsuuruus on siis epäyhtälössä (3.1) voimassa vain silloin, kun $\mathbf{x}_1 = \mathbf{x}_2$.

Induktioaskel.

Tehdään nyt induktio-oletus eli oletetaan, että lause on konaisuudessaan totta, kun $n = k - 1$. Lauseen todistamiseksi riittää osoittaa, että lause on totta myös arvolla $n = k$.

Olkoon $n = k$. Jos luku $\delta_1 = 0$, niin tarkastelu yksinkertaistuu vastaamaan tilannetta $n = k - 1$. Tällöin näet konveksissa kombinaatiossa $\delta_1 \mathbf{x}_1 + \delta_2 \mathbf{x}_2 + \dots + \delta_k \mathbf{x}_k$ termi $\delta_1 \mathbf{x}_1 = 0$. Induktio-oletuksen nojalla lause on totta, kun $n = k - 1$. Siten se on kokonaisuudessaan totta myös silloin, kun $n = k$ ja $\delta_1 = 0$.

Olkoon nyt luku $\delta_1 \neq 0$ ja lauseen (a)-kohdan oletukset voimassa. Siis tarkastelussa oletetaan ensin, että luvut $\delta_1, \delta_2, \dots, \delta_k$ ovat epänegatiivisia ja niiden

summa 1 ja että lisäksi funktio $f(\mathbf{x})$ on konvekksi. Määritellään konveksin joukon C piste $\mathbf{z}$ siten, että

$$\mathbf{z} = \frac{\delta_1}{\delta_z} \mathbf{x}_1 + \cdots + \frac{\delta_{k-1}}{\delta_z} \mathbf{x}_{k-1},$$

missä $\delta_z = \delta_1 + \cdots + \delta_{k-1}$. Siis summa $\delta_z + \delta_k = 1$ ja koska lauseen kohdan (a) todettiin edellä pitävän paikkansa arvolla $k = 2$, niin

$$\begin{aligned} f\left(\sum_{i=1}^k \delta_i \mathbf{x}_i\right) &= f\left(\sum_{i=1}^{k-1} \delta_i \mathbf{x}_i + \delta_k \mathbf{x}_k\right) \\ &= f(\delta_z \mathbf{z} + \delta_k \mathbf{x}_k) \\ &\leq \delta_z f(\mathbf{z}) + \delta_k f(\mathbf{x}_k). \end{aligned} \quad (3.2)$$

Toisaalta summa $\frac{\delta_1}{\delta_z} + \cdots + \frac{\delta_{k-1}}{\delta_z} = (\delta_1 + \cdots + \delta_{k-1})/\delta_z = \delta_z/\delta_z = 1$, joten induktio-oletuksen nojalla

$$\begin{aligned} \delta_z f(\mathbf{z}) &= \delta_z f\left(\sum_{i=1}^{k-1} \frac{\delta_i}{\delta_z} \mathbf{x}_i\right) \\ &\leq \delta_z \sum_{i=1}^{k-1} \frac{\delta_i}{\delta_z} f(\mathbf{x}_i) \\ &= \sum_{i=1}^{k-1} \delta_i f(\mathbf{x}_i). \end{aligned} \quad (3.3)$$

Sijoitetaan seuraavaksi tämä epäyhtälö (3.3) edellä johdettuun epäyhtälöön (3.2), jolloin saadaan, että

$$f\left(\sum_{i=1}^k \delta_i \mathbf{x}_i\right) \leq \sum_{i=1}^{k-1} \delta_i f(\mathbf{x}_i) + \delta_k f(\mathbf{x}_k) = \sum_{i=1}^k \delta_i f(\mathbf{x}_i). \quad (3.4)$$

Siis lauseen kohta (a) on osoitettu nyt todeksi, myös silloin, kun $n = k$ ja $\delta_1 \neq 0$.

Vielä pitää osoittaa todeksi lauseesta kohta (b) siinä tapauksessa, että $n = k$ ja $\delta_1 \neq 0$. Oletetaan siis nyt, että funktio $f(\mathbf{x})$ on aidosti konvekssi joukossa C ja että kaikki luvut δ_i ovat positiivisia. Olkoon piste $\mathbf{z}$ määritelty, kuten edellä kohtaa (a) todistettaessa.

Kohdan (b) todettiin edellä pitävän paikkansa, kun $k = 2$. Täten

$$f\left(\sum_{i=1}^k \delta_i \mathbf{x}_i\right) = f(\delta_z \mathbf{z} + \delta_k \mathbf{x}_k) = \delta_z f(\mathbf{z}) + \delta_k f(\mathbf{x}_k), \quad (3.5)$$

jos ja vain jos $\mathbf{z} = \mathbf{x}_k$. Induktio-oletuksen nojalla taas

$$\delta_z f(\mathbf{z}) = \delta_z f\left(\sum_{i=1}^{k-1} \frac{\delta_i}{\delta_z} \mathbf{x}_i\right) = \sum_{i=1}^{k-1} \delta_i f(\mathbf{x}_i), \quad (3.6)$$

jos ja vain jos pisteet $\mathbf{x}_1, \dots, \mathbf{x}_{k-1}$ ovat samoja. Sijoitetaan yhtälö (3.6) yhtälöön (3.5). Saadaan ehdoksi, että

$$f\left(\sum_{i=1}^k \delta_i \mathbf{x}_i\right) = \sum_{i=1}^k \delta_i f(\mathbf{x}_i),$$

jos ja vain jos pisteet $\mathbf{x}_1, \dots, \mathbf{x}_k$ ovat kaikki samoja. Näin myös lauseen kohta (b) tuli todistetuksi. $\square$

Esitetään nyt viimein aritmeettis-geometrisen keskiarvoepäyhtälö.

Lause 3.1.1 (A-G-keskiarvoepäyhtälö). *Oletetaan, että luvut $x_1, \dots, x_n$ ovat positiivisia ja että $\delta_1, \dots, \delta_n$ ovat positiivisia lukuja, joiden summa on yksi. Tällöin*

$$\prod_{i=1}^n (x_i)^{\delta_i} \leq \sum_{i=1}^n \delta_i x_i. \quad (3.7)$$

Epäyhtälössä (3.7) pätee yhtäsuuruus, jos ja vain jos $x_1 = \dots = x_n$.

Todistus. Käytetään todistuksessa apuna funktiota $f(x) = -\ln x$, joka on määritelty, kun $x > 0$. Lauseen (2.3.1) huomautuksen perusteella funktio $f(x)$ on aidosti konvekso määrittelyjoukossaan, sillä $f''(x) = 1/x^2 > 0$.

Käytetään seuraavaksi apulauseetta (3.1.1) ja muistetaan tämän todistettavan lauseen oletukset. Nythän luvut $x_1, \dots, x_n$ ja $\delta_1, \dots, \delta_n$ ovat positiivisia ja summa $\delta_1 + \dots + \delta_n = 1$. Koska lisäksi funktio $-\ln x$ on aidosti konvekso, niin apulauseen nojalla

$$-\ln\left(\sum_{i=1}^n \delta_i x_i\right) = f\left(\sum_{i=1}^n \delta_i x_i\right) \leq \sum_{i=1}^n \delta_i f(x_i) = -\sum_{i=1}^n \delta_i \ln x_i. \quad (3.8)$$

Jaetaan seuraavaksi epäyhtälö (3.8) puolittain negatiivisella luvulla -1 ja muokataan sitä sitten perinteisten logaritmin laskusääntöjen avulla. Epäyhtälö saadaan näin muotoon

$$\ln\left(\sum_{i=1}^n \delta_i x_i\right) \geq \sum_{i=1}^n \delta_i \ln x_i = \sum_{i=1}^n \ln(x_i^{\delta_i}) = \ln\left(\prod_{i=1}^n x_i^{\delta_i}\right). \quad (3.9)$$

Logaritmifunktio $g(x) = \ln x$ on määritelty, kun $x > 0$, joten sen derivaatta $g'(x) = 1/x > 0$ koko määrittelyjoukossa. $g(x)$ on siis aidosti kasvava funktio ja täten epäyhtälö (3.9) saadaan kirjoitettua tavoitellussa muodossa

$$\sum_{i=1}^n \delta_i x_i \geq \prod_{i=1}^n x_i^{\delta_i}. \quad (3.10)$$

Apulauseen (3.1.1) perusteella epäyhtälössä (3.8) ja siten myös epäyhtälössä (3.10) pätee yhtäsuuruus, jos ja vain jos luvut $x_1, \dots, x_n$ ovat kaikki samoja. $\square$

Seuraavana yksi esimerkki A-G-keskiarvoepäyhtälön käytöstä optimoinnissa.

Esimerkki 3.1.1. Tarkastellaan rajoitettua optimointiongelmaa. Tehtävänä on tällä kertaa maksimoida funktio $f(x_1, x_2, x_3) = x_1 x_2^2 x_3^3$ sillä tavalla, että rajoitefunktio $g(x_1, x_2, x_3) = x_1^3 + x_2^2 + x_3 = 39$ toteutuu. Ehtona on myös, että luvut x_1, x_2, x_3 ovat positiivisia reaalilukuja.

Ratkaistaan ongelma soveltamalla rajoitefunktioon A-G-keskiarvoepäyhtälöä. Tätä varten rajoitefunktion termejä x_1^3 , x_2^2 ja x_3 on ensin muotoiltava sopivasti. Nyt nämä kolme termiä jaetaan 13 osaan niin, että

$$39 = x_1^3 + x_2^2 + x_3 = 13 \left(\frac{x_1^3 + 3(\frac{1}{3}x_2^2) + 9(\frac{1}{9}x_3)}{13} \right).$$

Tällaiseen termien pilkkomiseen päädyttiin nyt intuitiivisesti ja erilaisia keiloja tekemällä. Mitään erityistä kaavaa termien pilkkomista varten ei ole

käytössä. A-G-keskiarvoepäyhtälön nojalla

$$\begin{aligned}
 39 &= 13 \left(\frac{x_1^3 + 3(\frac{1}{3}x_2^2) + 9(\frac{1}{9}x_3)}{13} \right) \\
 &\geq 13x_1^{\frac{3}{13}} \left(\frac{1}{3} \right)^{\frac{3}{13}} x_2^{\frac{6}{13}} \left(\frac{1}{9} \right)^{\frac{9}{13}} x_3^{\frac{9}{13}} \\
 &= 13 \left(\frac{1}{3} \right)^{\frac{3}{13}} \left(\frac{1}{9} \right)^{\frac{9}{13}} (x_1 x_2^2 x_3^3)^{\frac{3}{13}}. \tag{3.11}
 \end{aligned}$$

Tehtävänä oli maksimoida funktio $f(x_1, x_2, x_3) = x_1 x_2^2 x_3^3$, joten epäyhtälö (3.11) auttaa ratkaisun löytämisessä. Optimoitava funktio maksimoituu, kun epäyhtälössä pätee yhtäsuuruus. A-G-epäyhtälön nojalla taas yhtäsuuruus toteutuu, jos ja vain jos kaikki summan pilkotut 13 termiä ovat yhtä suuria. Siis tällöin

$$x_1^3 = \frac{1}{3}x_2^2 = \frac{1}{9}x_3 = \frac{39}{13} = 3. \tag{3.12}$$

Koska kaikkien lukujen x_i piti olla positiivisia, niin yhtälön (3.12) ainoa ratkaisu ja siten myös funktion $f(x_1, x_2, x_3)$ maksimi on piste $(x_1, x_2, x_3) = (3^{\frac{1}{3}}, 3, 27)$.

Mainittakoon vielä, että tämän tyyppisiin optimointiongelmiin liittyy aina niin sanottu duaaliongelma. Nyt duaaliongelmassa pyrittäisiin minimoimaan funktiota $g(x_1, x_2, x_3) = x_1^3 + x_2^2 + x_3$ ehdolla $x_1 x_2^2 x_3^3 = S_{vakio} > 0$. Lisäksi kaikkien lukujen x_i tulisi edelleen olla positiivisia.

Edellisessä esimerkissä funktion $g(x_1, x_2, x_3) = x_1^3 + x_2^2 + x_3$ termit pilkottiin pienemmiksi, jotta A-G-keskiarvoepäyhtälöä pystyttiin käyttämään tehtävän ratkaisemisessa. Pilkkominen perustui tällöin puhtaasti intuitioon ja kokeilemiseen. Jos tarkasteltava funktio sen sijaan olisi ollut monimutkaisempi, niin termien sopiva pilkkominen olisi varmasti tuottanut vaikeuksia. Tarvittaisiin siis rutiinimenetelmä pilkkomista varten.

Toinen huomion arvoinen asia liittyy tarkasteltavan funktion optimipisteen olemassaoloon. Oletetaan, että funktion termejä ei kovastakaan yrityksestä huolimatta saada pilkottua sopivasti eli A-G-keskiarvoepäyhtälöä ei voida

käyttää. Tällöin on kaksi mahdollisuutta. Joko optimia ei edes ole olemassa tai sitten optimi on olemassa, vaikka sitä ei nyt saatukaan selville. Tarvittaisiin siis myös sellainen menetelmä, joka paljastaisi, onko funktion optimoiminen ylipäättään mahdollista. Seuraavana aletaan käsittelemään varsinaista geometrista optimointia, joka ratkaisee nämä A-G-keskiarvoepäyhtälön käyttöön liittyvät ongelmat.

3.2 Varsinainen geometrisen optimointi

Tässä osiossa johdetaan yksi tärkeistä ja käytännöllisistä epälineaarisen optimoinnin algoritmeista, Geometrisen optimoinnin menetelmä.

3.2.1 Geometrisen optimoinnin kehittelyä

Geometrisen optimoinnin perustehtävänä on minimoida funktio

$$g(\mathbf{x}) = \sum_{i=1}^n c_i \prod_{j=1}^m x_j^{\alpha_{ij}}. \quad (3.13)$$

Tässä luvut c_i ovat positiivisia vakioita ja eksponentit α_{ij} puolestaan reaalilukuja. Funktio $g(\mathbf{x})$ on määritelty aina, kun vektorin $\mathbf{x}^* = (x_1, \dots, x_m)$ kaikki komponentit x_j ovat positiivisia. Perustehtävässä pyritään siis löytämään se vektori $\mathbf{x}^*$, jonka kaikki komponentit x_j^* ovat positiivisia ja joka siis minimoi funktion $g(\mathbf{x})$. Muun muassa edellä esimerkissä (3.12) esiintynyt funktio $g(x_1, x_2, x_3) = x_1^3 + x_2^2 + x_3$ täyttää muotonsa puolesta perustehtävälle asetetut kriteerit.

Määritellään seuraavaksi geometrisen optimoinnin perustehtävään eli funktion (3.13) minimointiin liittyvä duaalitehtävä. Duaalitehtävässä tulee etsiä funktion

$$v(\boldsymbol{\delta}) = \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \quad (3.14)$$

arvon maksimoiva vektori δ^* , joka toteuttaa vektorille δ asetettavat ehdot:

$$\delta_i > 0, \text{ kun } i = 1, \dots, n, \quad (3.15)$$

$$\sum_{i=1}^n \delta_i = 1, \quad (3.16)$$

$$\sum_{i=1}^n \alpha_{ij} \delta_i = 0, \text{ kun } j = 1, \dots, m. \quad (3.17)$$

Nämä kolme ehtoa ovat järjestyksessä ylhäältä alaspäin edettäessä positiivisuusehto, normitusehto ja ortogonaalisuusehto. Vektori δ , joka toteuttaa nämä ehdot, on duaalitehtävän käypä ratkaisu. Jos duaalitehtävällä ei ole ainuttakaan käypää ratkaisua, ei sillä silloin ole myöskään optimiratkaisua. Duaalitehtävää sanotaan konsistentiksi, jos käypiä ratkaisuja on ainakin yksi. Funktioiden $g(\mathbf{x})$ ja $v(\delta)$ määrittelyjoukoissaan saamia arvoja yhdistää seuraavassa lauseessa esitettävä tärkeä tulos.

Lause 3.2.1. *Olkoon funktion $g(\mathbf{x})$ minimointi geometrisen optimoinnin perustehtävä ja funktion $v(\delta)$ maksimointi sitä vastaava duaalitehtävä. Tällöin pätee aina, että*

$$g(\mathbf{x}) \geq v(\delta). \quad (3.18)$$

Todistus. Kirjoitetaan funktio $g(\mathbf{x})$ muodossa

$$g(\mathbf{x}) = \sum_{i=1}^n \delta_i \left(\frac{c_i \prod_{j=1}^m x_j^{\alpha_{ij}}}{\delta_i} \right). \quad (3.19)$$

Positiivisuusehdon (3.15) nojalla kaikkien komponenttien δ_i on oltava positiivisia ja normitusehdon (3.16) perusteella taas summa $\delta_1 + \dots + \delta_n = 1$. Yhtälöön (3.19) voidaan nyt soveltaa lauseessa (3.1.1) esitettyä aritmeettis-

geometrista keskiarvoepäyhtälöä. Tällöin

$$\begin{aligned}
 \sum_{i=1}^n \delta_i \left(\frac{c_i \prod_{j=1}^m x_j^{\alpha_{ij}}}{\delta_i} \right) &\geq \prod_{i=1}^n \left(\frac{c_i \prod_{j=1}^m x_j^{\alpha_{ij}}}{\delta_i} \right)^{\delta_i} \\
 &= \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \left(\prod_{i=1}^n \prod_{j=1}^m x_j^{\alpha_{ij} \delta_i} \right) \\
 &= \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \prod_{j=1}^m x_j^{\alpha_{1j} \delta_1 + \dots + \alpha_{nj} \delta_n} \\
 &= \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \prod_{j=1}^m x_j^{\sum_{i=1}^n \alpha_{ij} \delta_i}. \tag{3.20}
 \end{aligned}$$

Otetaan seuraavaksi huomioon ortogonaalisuusehto (3.17). Sen mukaanhan $\sum_{i=1}^n \alpha_{ij} \delta_i = 0$ aina, kun $j = 1, \dots, m$. Siis epäyhtälössä (3.20)

$$\prod_{j=1}^m x_j^{\sum_{i=1}^n \alpha_{ij} \delta_i} = \prod_{j=1}^m x_j^0 = 1 \tag{3.21}$$

ja näin on osoitettu, että

$$g(\mathbf{x}) = \sum_{i=1}^n \delta_i \left(\frac{c_i \prod_{j=1}^m x_j^{\alpha_{ij}}}{\delta_i} \right) \geq \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i}. \tag{3.22}$$

Verrataan epäyhtälön (3.22) oikeaa puolta yhtälöön (3.14). Nähdään, että kyseessä on duaalitehtävän funktio $v(\boldsymbol{\delta})$ ja siten $g(\mathbf{x}) \geq v(\boldsymbol{\delta})$. $\square$

Tarkastellaan vielä lausetta (3.2.1). Oletetaan, että perustehtävän optimiratkaisu on $\mathbf{x}^*$ ja duaalitehtävän $\boldsymbol{\delta}^*$. Tällöin kyseisen lauseen perusteella pätee erityisesti, että $g(\mathbf{x}^*) \geq v(\boldsymbol{\delta}^*)$. Myöhemmin tullaan itse asiassa havaitsemaan, että $g(\mathbf{x}^*) = v(\boldsymbol{\delta}^*)$. Osoitetaan kuitenkin seuraavana, että jos perustehtävällä on optimiratkaisu $\mathbf{x}^*$, niin tällä ratkaisulla on tällöin vastineensa duaalitehtävän käypien ratkaisujen joukossa.

Lause 3.2.2. *Oletetaan, että vektori $\mathbf{x}^* = (x_1^*, \dots, x_m^*)$ on geometrisen optimoinnin perustehtävän ratkaisu. Tällöin vastaava duaalitehtävä on konsistentti eli sillä on jokin käypä ratkaisu, vektori $\boldsymbol{\delta}^{(t)} = (\delta_1^{(t)}, \dots, \delta_n^{(t)})$. Tämän*

vektorin kaikille komponenteille $\delta_i^{(t)}$ pätee, että

$$\delta_i^{(t)} = \frac{u_i(\mathbf{x}^*)}{g(\mathbf{x}^*)}, \quad \text{missä} \quad u_i(\mathbf{x}^*) = c_i \prod_{j=1}^m (x_j^*)^{\alpha_{ij}}.$$

Todistus. Tulee osoittaa, että duaalitehtävä on konsistentti. Siis toisin sanoen pitää osoittaa, että on olemassa ainakin yksi sellainen vektori $\boldsymbol{\delta}$, joka täyttää asetetut positiivisuus-, normitus- ja ortogonaalisuusehdot. Aletaan kokeilemaan, josko vektori

$$\boldsymbol{\delta}^{(t)} = (\delta_1^{(t)}, \dots, \delta_n^{(t)}) = \left(\frac{u_1(\mathbf{x}^*)}{g(\mathbf{x}^*)}, \dots, \frac{u_n(\mathbf{x}^*)}{g(\mathbf{x}^*)} \right)$$

todella toteuttaisi nämä ehdot.

Tutkitaan ensin positiivisuusehtoa. Luvut c_i ovat kaikki positiivisia ja myös perustehtävän ratkaisun $\mathbf{x}^*$ kaikki komponentit x_i^* ovat positiivisia. Tämä seuraa jo suoraan geometrisen optimoinnin perustehtävän oletuksista. Täten positiivisten lukujen tuloina kaikki funktiot

$$u_i(\mathbf{x}^*) = c_i \prod_{j=1}^m (x_j^*)^{\alpha_{ij}} > 0, \quad (3.23)$$

kun $i = 1, \dots, n$. Siten myös

$$g(\mathbf{x}^*) = \sum_{i=1}^n c_i \prod_{j=1}^m (x_j^*)^{\alpha_{ij}} = \sum_{i=1}^n u_i(\mathbf{x}^*) > 0 \quad (3.24)$$

ja positiivisten lukujen osamääränä

$$\delta_i^{(t)} = \frac{u_i(\mathbf{x}^*)}{g(\mathbf{x}^*)} > 0 \quad (3.25)$$

aina, kun $i = 1, \dots, n$. Siis positiivisuusehto toteutuu.

Tutkitaan seuraavaksi normitusehtoa. Summa

$$\sum_{i=1}^n \delta_i^{(t)} = \frac{\sum_{i=1}^n u_i(\mathbf{x}^*)}{g(\mathbf{x}^*)} = \frac{g(\mathbf{x}^*)}{g(\mathbf{x}^*)} = 1,$$

joten myös normitusehto toteutuu.

Viimeisenä ehtona tarkastellaan vielä ortogonaalisuusehtoa. Vektori $\mathbf{x}^*$ on perustehtävän ratkaisu ja siten funktion $g(\mathbf{x})$ maksimi, joten gradientti $\nabla g(\mathbf{x}^*) = 0$. Näin ollen

$$\frac{\partial g(\mathbf{x}^*)}{\partial x_j} = \frac{\partial \sum_{i=1}^n u_i(\mathbf{x}^*)}{\partial x_j} = 0 \quad (3.26)$$

aina, kun $j = 1, \dots, m$. Lauseesta (3.26) saadaan edelleen, että

$$0 = x_j^* \frac{\partial g(\mathbf{x}^*)}{\partial x_j} = \sum_{i=1}^n x_j^* \frac{\partial u_i(\mathbf{x}^*)}{\partial x_j}. \quad (3.27)$$

Toisaalta

$$\frac{\partial u_i(\mathbf{x})}{\partial x_j} = \frac{\partial (c_i x_1^{\alpha_{i1}} \cdots x_m^{\alpha_{im}})}{\partial x_j} = c_i \alpha_{ij} x_1^{\alpha_{i1}} \cdots x_{j-1}^{\alpha_{i,j-1}} \cdot x_j^{\alpha_{ij}-1} \cdot x_{j+1}^{\alpha_{i,j+1}} \cdots x_m^{\alpha_{im}},$$

joten

$$x_j \frac{\partial u_i(\mathbf{x})}{\partial x_j} = \alpha_{ij} u_i(\mathbf{x}). \quad (3.28)$$

Sijoitetaan nyt tämä lause (3.28) lauseeseen (3.27) jolloin nähdään, että

$$0 = \sum_{i=1}^n \alpha_{ij} u_i(\mathbf{x}^*) = g(\mathbf{x}^*) \sum_{i=1}^n \alpha_{ij} \delta_i^{(t)}.$$

Koska lauseen (3.24) mukaan $g(\mathbf{x}^*) > 0$, niin välttämättä $\sum_{i=1}^n \alpha_{ij} \delta_i^{(t)} = 0$. Ortogonaalisuusehtokin siis täyttyy ja vektori $\boldsymbol{\delta}^{(t)}$ on näin todella duaalitehtävän käypä ratkaisu. Duaalitehtävä on siten konsistentti. $\square$

Yhdistetään seuraavaksi edellisten lauseiden (3.2.1) ja (3.2.2) tulokset mielenkiintoisella tavalla.

Lause 3.2.3. *Oletetaan, että vektori $\mathbf{x}^* = (x_1^*, \dots, x_m^*)$ on geometrisen optimoinnin perustehtävän optimaalinen ratkaisu ja vektori*

$$\boldsymbol{\delta}^{(t)} = (\delta_1^{(t)}, \dots, \delta_n^{(t)}) = \left(\frac{u_1(\mathbf{x}^*)}{g(\mathbf{x}^*)}, \dots, \frac{u_n(\mathbf{x}^*)}{g(\mathbf{x}^*)} \right)$$

puolestaan sitä vastaava duaalitehtävän käypä ratkaisu. Tällöin vektori $\boldsymbol{\delta}^{(t)}$ on duaalitehtävän optimaalinen ratkaisu eli voidaan merkitä $\boldsymbol{\delta}^{(t)} = \boldsymbol{\delta}^$. Lisäksi pätee, että $g(\mathbf{x}^*) = v(\boldsymbol{\delta}^*)$.*

Todistus. Vektori $\boldsymbol{\delta}^{(t)} = (\delta_1^{(t)}, \dots, \delta_n^{(t)})$ on duaalitehtävän käypä ratkaisu, joten se toteuttaa asetetut positiivisuus-, normitus- ja ortogonaalisuusehdot. Normitusehdon nojalla summa $\delta_1^{(t)} + \dots + \delta_n^{(t)} = 1$. Siten funktio $g(\mathbf{x}^*)$ voidaan kirjoittaa muodossa

$$\begin{aligned} g(\mathbf{x}^*) &= g(\mathbf{x}^*)^{\delta_1^{(t)} + \dots + \delta_n^{(t)}} \\ &= g(\mathbf{x}^*)^{\delta_1^{(t)}} \dots g(\mathbf{x}^*)^{\delta_n^{(t)}} \\ &= \left(\frac{u_1(\mathbf{x}^*)}{\delta_1^{(t)}} \right)^{\delta_1^{(t)}} \dots \left(\frac{u_n(\mathbf{x}^*)}{\delta_n^{(t)}} \right)^{\delta_n^{(t)}} \\ &= \frac{\prod_{i=1}^n (u_i(\mathbf{x}^*))^{\delta_i^{(t)}}}{\prod_{i=1}^n (\delta_i^{(t)})^{\delta_i^{(t)}}} \end{aligned} \quad (3.29)$$

Tutkitaan seuraavaksi tarkemmin yhtälön (3.29) osoittajaa. Koska

$$u_i(\mathbf{x}^*) = c_i \prod_{j=1}^m (x_j^*)^{\alpha_{ij}},$$

niin pätee, että osoittaja

$$\begin{aligned} \prod_{i=1}^n (u_i(\mathbf{x}^*))^{\delta_i^{(t)}} &= \prod_{i=1}^n c_i^{\delta_i^{(t)}} \left(\prod_{i=1}^n \prod_{j=1}^m (x_j^*)^{\alpha_{ij} \delta_i^{(t)}} \right) \\ &= \prod_{i=1}^n c_i^{\delta_i^{(t)}} \prod_{j=1}^m (x_j^*)^{\alpha_{1j} \delta_1^{(t)} + \dots + \alpha_{nj} \delta_n^{(t)}} \\ &= \prod_{i=1}^n c_i^{\delta_i^{(t)}} \prod_{j=1}^m (x_j^*)^{\sum_{i=1}^n \alpha_{ij} \delta_i^{(t)}}. \end{aligned} \quad (3.30)$$

Vektori $\boldsymbol{\delta}^{(t)}$ toteuttaa ortogonaalisuusehdon, joten yhtälössä (3.30) potenssi $\sum_{i=1}^n \alpha_{ij} \delta_i = 0$ aina, kun $j = 1, \dots, m$. Siispä

$$\prod_{i=1}^n (u_i(\mathbf{x}^*))^{\delta_i^{(t)}} = \prod_{i=1}^n c_i^{\delta_i^{(t)}} \prod_{j=1}^m (x_j^*)^0 = \prod_{i=1}^n c_i^{\delta_i^{(t)}}.$$

Kun tämä muoto osoittajasta nyt sijoitetaan yhtälöön (3.29), nähdään, että

$$g(\mathbf{x}^*) = \frac{\prod_{i=1}^n c_i^{\delta_i^{(t)}}}{\prod_{i=1}^n (\delta_i^{(t)})^{\delta_i^{(t)}}} = \prod_{i=1}^n \left(\frac{c_i}{\delta_i^{(t)}} \right)^{\delta_i^{(t)}} = v(\boldsymbol{\delta}^{(t)}).$$

Lauseen (3.2.1) perusteella kaikille duaalitehtävän käyville ratkaisuille δ pätee, että $g(\mathbf{x}^*) \geq v(\delta)$. Vektori $\delta^{(t)}$ on yksi näistä käyvistä vektoreista ja nyt on osoitettu, että $g(\mathbf{x}^*) = v(\delta^{(t)})$. Vektori $\delta^{(t)}$ on siis käypien vektoreiden joukossa funktion $v(\delta)$ maksimi. Siten se on myös duaalitehtävän optimiratkaisu ja voidaan merkitä, että $\delta^{(t)} = \delta^*$ ja että täten $g(\mathbf{x}^*) = v(\delta^*)$. $\square$

Tutkitaan vielä edellisen lauseen merkitystä optimoinnin kannalta. Lauseen perusteella geometrisen optimoinnin perustehtävää ratkaistaessa voidaan käyttää apuna vastaavaa duaalitehtävää. Oletetaan, että duaalitehtävä saadaan ratkaistua eli löydetään vektori δ^* , joka on duaalitehtävän maksimi. Lauseen mukaan perustehtävän minimiarvo $g(\mathbf{x}^*) = v(\delta^*)$, joten minimiarvon selvittämiseksi ei näin ollen edes vaadita minimipisteen $\mathbf{x}^*$ ratkaisemista. Tarvittaessa myös vektori $\mathbf{x}^*$ voidaan kyllä selvittää, sillä $\delta_i^* = u_i(\mathbf{x}^*)/v(\delta^*)$ aina, kun $i = 1, \dots, n$. Näin saadaan n kappaletta yhtälöitä, joiden avulla päästään käsiksi komponentteihin x_j^* .

3.2.2 Geometrisen optimoinnin menetelmä

Edellä, osiossa (3.2.1) tehtyjen tarkastelujen yhteenvedoksi voidaan nyt muotoilla käyttökelpoinen algoritmi geometristen optimointiongelmien ratkaisuun.

Algoritmi 3.2.1 (Geometrisen optimoinnin menetelmä). Geometrisen optimoinnin perustehtävänä on minimoida funktio

$$g(\mathbf{x}) = \sum_{i=1}^n u_i(\mathbf{x}),$$

missä $u_i(\mathbf{x}) = c_i \prod_{j=1}^m x_j^{\alpha_{ij}}$. Luvut $c_i > 0$ aina, kun $i = 1, \dots, n$. Vastaavasti myös luvut $x_j > 0$ aina, kun $j = 1, \dots, m$. Tehtävä ratkaistaan seuraavalla tavalla:

Vaihe 1.

Määritetään duaalitehtävän käypä joukko. Etsitään siis tällöin kaikki ne ava-

ruuden $\mathbb{R}^n$ vektorit $\boldsymbol{\delta}$, jotka toteuttavat positiivisuus-, normitus- ja ortogonaalisuusehdot. Tällöin

$$\begin{aligned}\delta_i &> 0, \text{ kun } i = 1, \dots, n, \\ \sum_{i=1}^n \delta_i &= 1, \\ \sum_{i=1}^n \alpha_{ij} \delta_i &= 0, \text{ kun } j = 1, \dots, m.\end{aligned}$$

Vaihe 2.

- (a) Jos duaalitehtävän käypä joukko on tyhjä, ei perustehtävällä ole ratkaisua.
- (b) Jos duaalitehtävällä on täsmälleen yksi käypä ratkaisu, niin se on myös duaalitehtävän optimiratkaisu $\boldsymbol{\delta}^*$. Siirrytään vaiheeseen 4.
- (c) Jos duaalitehtävällä on enemmän kuin yksi käypä ratkaisu, siirrytään vaiheeseen 3.

Vaihe 3.

Etsitään duaalitehtävän käypistä vektoreista se, joka maksimoi funktion

$$v(\boldsymbol{\delta}) = \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i}$$

arvon. Tämä vektori on duaalitehtävän optimiratkaisu $\boldsymbol{\delta}^*$. Siirrytään vaiheeseen 4.

Vaihe 4.

Vektori $\boldsymbol{\delta}^*$ on nyt duaalitehtävän ratkaisu. Perustehtävän minimiarvo $g(\mathbf{x}^*) = v(\boldsymbol{\delta}^*)$. Perustehtävän minimoiva vektori $\mathbf{x}^*$ saadaan puolestaan selville ratkaisemalla yhtälöryhmä

$$\delta_i^* = \frac{u_i(\mathbf{x}^*)}{v(\boldsymbol{\delta}^*)}, \quad i = 1, \dots, n.$$

Havainnollistetaan tämän algoritmin käyttöä esimerkin avulla.

Esimerkki 3.2.1. Geometrisen optimoinnin perustehtävänä on nyt minimoida funktio

$$g(\mathbf{x}) = \frac{1}{x_1 x_2} + x_1 + x_2,$$

missä luvut x_1 ja x_2 ovat positiivisia.

Aloitetaan ongelman ratkaiseminen algoritmin (3.2.1) avulla laatimalla perustehtävää vastaava duaalitehtävä. Duaalitehtävänä on maksimoida funktio

$$v(\boldsymbol{\delta}) = \left(\frac{1}{\delta_1}\right)^{\delta_1} \left(\frac{1}{\delta_2}\right)^{\delta_2} \left(\frac{1}{\delta_3}\right)^{\delta_3}.$$

Suoritetaan algoritmin vaihe 1 eli etsitään nyt kaikki ne vektorit $\boldsymbol{\delta}$, jotka toteuttavat positiivisuus-, normitus- ja ortogonaalisuusehdot:

$$\delta_1 > 0, \quad \delta_2 > 0, \quad \delta_3 > 0, \quad (3.31)$$

$$\delta_1 + \delta_2 + \delta_3 = 1, \quad (3.32)$$

$$-\delta_1 + \delta_2 = 0, \quad (3.33)$$

$$-\delta_1 + \delta_3 = 0. \quad (3.34)$$

Ehdoista (3.33) ja (3.34) seuraa, että $\delta_1 = \delta_2 = \delta_3$. Täten ehdon (3.32) perusteella $\delta_1 = \delta_2 = \delta_3 = \frac{1}{3}$. Kaikki luvut δ_i ovat siis nyt positiivisia ja vektori $\boldsymbol{\delta} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ on duaalitehtävän ainoa käypä ratkaisu.

Siirrytään algoritmin vaiheeseen 2. Nyt duaalitehtävällä on täsmälleen yksi käypä ratkaisu, joten vaiheen 2 kohdan (b) nojalla tämä ratkaisu on myös duaalitehtävän optimiratkaisu $\boldsymbol{\delta}^*$.

Algoritmissa voidaan siirtyä suoraan vaiheeseen 4. Olkoon vektori $\mathbf{x}^*$ perustehtävän optimiratkaisu. Tällöin

$$\begin{aligned} g(\mathbf{x}^*) = v(\boldsymbol{\delta}^*) &= \left(\frac{1}{1/3}\right)^{\frac{1}{3}} \left(\frac{1}{1/3}\right)^{\frac{1}{3}} \left(\frac{1}{1/3}\right)^{\frac{1}{3}} \\ &= 3^{\frac{1}{3}} 3^{\frac{1}{3}} 3^{\frac{1}{3}} = 3. \end{aligned}$$

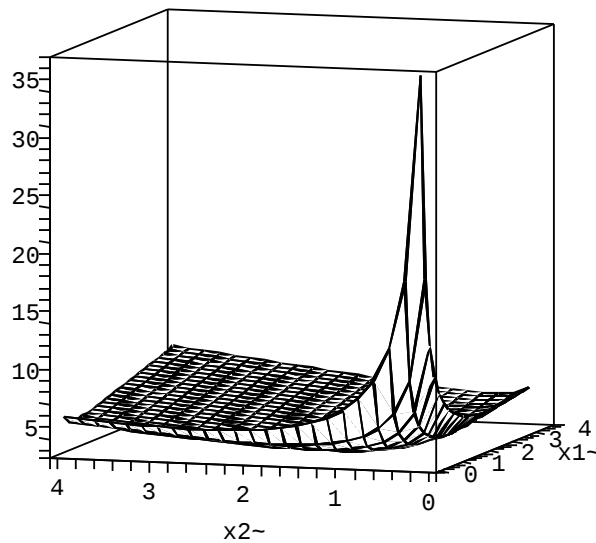
Siis perustehtävän minimiarvo on $g(\mathbf{x}^*) = 3$. Vielä pitää selvittää, mikä on optimipiste $\mathbf{x}^*$. Se saadaan selville seuraavien kolmen yhtälön avulla:

$$\delta_1^* = \frac{(x_1^* x_2^*)^{-1}}{v(\boldsymbol{\delta}^*)} \quad \text{eli} \quad \frac{1}{3} = \frac{(x_1^* x_2^*)^{-1}}{3}, \quad (3.35)$$

$$\delta_2^* = \frac{x_1^*}{v(\boldsymbol{\delta}^*)} \quad \text{eli} \quad \frac{1}{3} = \frac{x_1^*}{3}, \quad (3.36)$$

$$\delta_3^* = \frac{x_2^*}{v(\boldsymbol{\delta}^*)} \quad \text{eli} \quad \frac{1}{3} = \frac{x_2^*}{3}. \quad (3.37)$$

Yhtälö (3.36) toteutuu ainoastaan silloin, kun $x_1^* = 1$ ja vastaavasti yhtälö (3.37) toteutuu vain silloin, kun $x_2^* = 1$. Myös yhtälö (3.35) toteutuu näillä komponenttien x_j^* arvoilla. Siispä funktion $g(\mathbf{x})$ minimi ja siten perustehtävän optimiratkaisu on piste $\mathbf{x}^* = (1, 1)$. Funktion minimiarvo $g(1, 1) = 3$. Funktion $g(\mathbf{x})$ käyttäytymistä origon läheisyydessä on vielä havainnollistettu kuvassa 3.1.



Kuva 3.1: $g(\mathbf{x}) = \frac{1}{x_1 x_2} + x_1 + x_2$.

Edellisessä esimerkissä tarkasteltiin funktiota, jonka optimoiminen luvussa 1 esitetyin menetelmin olisi erittäin työlästä. Kriittisten pisteiden paikantaminen eli gradientin nollakohtien löytäminen olisi jo sinänsä vaikeaa. Mahdolliseksi tehtävä menisi kuitenkin viimeistään silloin, kun kriittisten pisteiden laatua alettaisiin selvittämään funktion Hessin matriisiin perustuen.

Myös käytännön sovellutuksissa epälineaariset optimointiongelmat ovat usein monimutkaisia. Nyt esitelty Geometrisen optimoinnin menetelmä onkin siksi vain yksi monista kehitetyistä optimointimenetelmistä, joilla vältetään gradientin nollakohtien ja Hessin matriisien tutkiminen. Erityisen tärkeitä ovat iteratiiviset optimointimenetelmät. Niissä tehdään ensin arvio todellisen op-

timipisteen sijainnista ja kehitetyn algoritmin avulla aletaan sitten numeerisesti askel askeleelta parantamaan tehtyä arviota. Optimointi lopetetaan, kun paranneltu arvio vastaa riittävällä tarkkuudella todellista optimipistettä. Tämän tutkielman puitteissa ei kuitenkaan puututa näihin iteratiivisiin optimointimenetelmiin.

Kirjallisuutta

- [1] Berg, Kimmo: *Optimointioppi*, Systeemianalyysin laboratorio, Teknillinen korkeakoulu, 2004. Verkkojulkaisu saatavissa: <http://www.sal.tkk.fi/Opinnot/Mat-2.139/>
- [2] Kaleva, Osmo: *Matemaattinen Optimointi 1*, Tampereen teknillinen yliopisto, 2002.
- [3] Peressini, Anthony L.; Sullivan, Francis E.; Uhl, Jerry J.: *The Mathematics of Nonlinear Programming*, Springer-Verlag, New York, 1988.
- [4] Adams, Robert A.: *Calculus - A Complete Course*, 5. painos, Addison-Wesley-Longman, Toronto, 2003.
- [5] Turakainen, Paavo: *Matriisiteoria-luentomateriaali*, Oulun yliopisto.
- [6] Mustonen, Vesa: *Analyysi II-luentomateriaali*, Oulun yliopisto.
- [7] Saranen, Jukka: *Optimointiteoria-luentomateriaali*, Oulun yliopisto.