Variaatiolaskenta

Jukka Tuomela

28. helmikuuta 2014

Sisältö

Luku 1

Johdantoa

Nature does nothing in vain and
more is in vain when less will serve;
for Nature is pleased with simplicity and
affects not the pomp of superfluous causes.
Isaac Newton

Because the shape of the whole universe is most perfect
and, in fact, designed by wisest creator,
nothing in all of the world will occur in which
no maximum or minimum rule is somehow shining forth.
Leonhard Euler

Actually the calculus of variations is
simply part of the functional analysis and
plays in it the same role as the theory
of maxima and minima in elementary calculus.
L. C. Young [11]

Variaatiolaskenta on eräänlainen yleistys optimointitehtävistä. Kun tavanomaisissa optimointitehtävissä etsitään jotain pistettä joka minimoisi tai maksimoisi jonkin annetun funktion arvon niin varaatiolaskennan tehtävissä etsitään jotain *funktiota (tai käyrää)* jolla olisi jokin ääriarvo-ominaisuus. Jo heti differentiaalilaskennan alkutaipaleella 1700-luvun alussa ryhdyttiin tutkimaan variaatiolaskennan ongelmia. 1700-luvun puolivälissä Euler ja Lagrange ymmäsivätkin jo perusasiat ja heidän ideoistaan lähdetään tällä kurssilla liikkeelle.

Vaikka variaatiolaskennan historia on siis pitkä niin sitä tutkitaan edelleen aktiivisesti. Kuten ylläolevista lainauksista näkyy, niin laajasti ajateltiin, että jonkinlaiset optimaalisuusperiaatteet liittyisivät läheisesti luonnonlakeihin. Variaatiolaskennalla onkin ollut varsin merkittävä rooli monissa fysikaan teorioissa. Esimerkiksi *stationaarisen vaikutuksen periaate* on suora variaatiolaskun sovellus.[1]

Lisäksi variaatiolaskennalla ja siihen läheisesti liittyvällä *optimisäädöllä* on runsaasti sovelluksia. Esimerkiksi prosessiteollisuudessa haluttaisiin prosessin olosuhteita säätää siten, että ne olisivat parhaat mahdolliset prosessin kannalta. Voitaisiin myös kysyä miten raketti saadaan kuuhun käyttämällä mahdollisimman vähän polttoainetta. Tämäntyyppiset tehtävät johtavat optimisäätöön ja sitä kautta variaatiolaskentaan.

Kirjat [4, 8, 9, 10] ovat hyviä johdatuksia variaatiolaskennan perusteisiin. Modernia teoriaa yksiulotteisessa tapauksessa käsitellään kirjassa [3] ja erittäin perusteellinen ja laaja yleisen teorian esittely löytyy kirjoista [5, 6].

[1] Tämä on englanniksi principle of stationary action. Ennen käytettiin myös termiä pienimmän vaikutuksen periaate (principle of least action).

Luku 2

Eulerin yhtälö

2.1 Perusesimerkkejä

Olkoon $c : [a, b] \to \mathbb{R}^2$ jokin tasokäyrä. Käyrän pituus määritellään tunnetusti kaavalla

$$J(c) = \int_a^b |c'(s)| \, ds$$

Jos käyrä voidaan esittää funktion $y : [a, b] \to \mathbb{R}$ graafina niin $c(s) = \big(s, y(s)\big)$ ja voidaan kirjoittaa

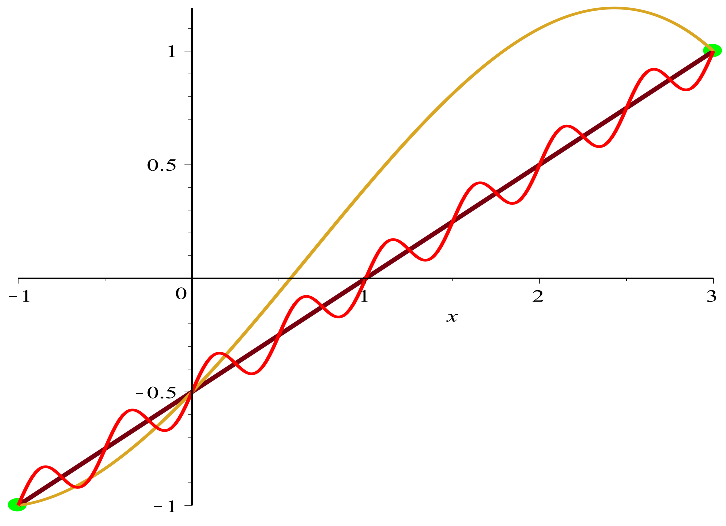
$$J(y) = \int_a^b \sqrt{1 + (y'(x))^2} \, dx \tag{2.1}$$

Jos sitten on annettu kaksi tason pistettä $p = (a, c)$ ja $q = (b, d)$, niin voidaan kysyä mikä on lyhin käyrä näitten pisteitten välillä. Tunnetusti vastaus on suora. Mutta miten tämän voisi osoittaa? Selvästi on monenmoisia polkuja kahden pisteen välillä, kuva 2.1, ja pienen pohdinnan jälkeen on varsin helppo vakuuttua siitä, että sallittujen polkujen avaruus on ääretönulotteinen. Variaatiolaskennan tehtävät ovat siis välttämättä paljon hankalampia kuin klassiset optimoititehtävät joissa etsitään jonkin kuvauksen ääriarvoja jossain sopivassa $\mathbb{R}^n$:n osajoukossa (liite C).

Historiallisesti kenties ensimmäinen variaatiolaskennan ongelma tunnetaan nimellä *Didon ongelma.*[1] Siinä tarkastellaan umpinaista tasokäyrää jonka pituus on kiinnitetty ja kysytään miten käyrä pitäisi valita jotta käyrän sisään jäävän alueen pinta-ala olisi mahdollisimman suuri. ”Duaalinen” muotoilu olisi kiinnittää ensin pinta-ala ja kysyä minkämuotoisella alueella on lyhin reuna. Jo antiikin aikana oltiin vakuuttuneita, että vastaus on ympyrä, mutta tarkka todistus tälle on yllättävän hankala.

Jo differentiaalilaskennan alkuvuosina 1700-luvun alussa tutkittiin variaatiolaskennan tehtäviä, ja kuuluisa esimerkki on *brakistokroniongelma* eli lyhimmän ajan ongelma. Tässä siis pistemassa lähetetään liukumaan ilman kitkaa jotain käyrää pitkin ja haluttaisiin tietää miten käyrä pitää valita jotta aika olisi mahdollisimman pieni. Valitaan $p = (0, 1)$, $q = (1, 0)$ ja olkoon $y : [0, 1] \to \mathbb{R}$ jokin funktio jolle pätee $y(0) = 1$ ja $y(1) = 0$. Ajatellaan, että painovoima vaikuttaa negatiivisen y-akselin suuntaan ja

[1] Dido (tai Alissa) oli Kartagon kuningatar.



Kuva 2.1: Eräitä käyriä kahden pisteen välillä.

että pistemassa lähtee levosta liikkeelle pisteestä p ja liukuu sitten painovoiman vaikutuksesta y:n graafia pitkin pisteeseen q. Mikä y minimoi ajan?

Pistemassa liikkuu siis käyrää $c(s) = \big(s, y(s)\big)$ pitkin. Sen nopeus saadaan selville energian säilymisen laista: potentiaalienergian ja kineettisen energian summa on vakio.

$$E = \tfrac{1}{2}mv^2 + m\mathbf{g}y = \text{vakio}$$

Tässä $\mathbf{g}$ on maan vetovoiman kiihtyvyys ja m on massa. Koska alussa $v = 0$, niin $E = m\mathbf{g}$, joten $v = \sqrt{2\mathbf{g}(1-y)}$. Liukumisajaksi saadaan siten

$$J(y) = \int_c dt = \int_c \frac{ds}{v} = \frac{1}{\sqrt{2\mathbf{g}}}\int_0^1 \sqrt{\frac{1+(y')^2}{1-y}}dx \tag{2.2}$$

2.2 Tarkempi muotoilu

Jotta päästäisiin eteenpäin niin otetaan käyttöön muutama merkintä (liite B). Olkoon $I = [a, b] \subset \mathbb{R}$ ja olkoon $L : I \times \mathbb{R}^2 \to \mathbb{R}$ jokin funktio. Olkoon sitten $V \subset C^1(I)$ jokin sopiva $C^1(I)$:n osajoukko ja olkoon edelleen

$$V_0 = \big\{f \in C^1(I) \,\big|\, f(a) = f(b) = 0\big\}$$

Tarkastellaan kuvausta

$$J : V \to \mathbb{R} \quad , \quad J(y) = \int_a^b L(x, y, y')dx$$

Funktiota L on tapana sanoa *Lagrangen funktioksi.* Selvästi integraali on hyvin määritelty jos L on jatkuva, mutta oletetaan näin ainakin aluksi, että $L \in C^2(I \times \mathbb{R}^2)$.

Määritelmä 2.1.

1. y_* *on* globaali minimi*, jos* $J(y_*) \le J(y)$ *kaikilla* $y \in V$.
2. y_* *on* (heikko) paikallinen minimi*, jos on olemassa jokin* ε *siten, että* $J(y_*) \le J(y)$ *kaikilla* $\|y - y_*\|_{1,\infty} < \varepsilon$.
3. y_* *on* vahva paikallinen minimi*, jos on olemassa jokin* ε *siten, että* $J(y_*) \le J(y)$ *kaikilla* $\|y - y_*\|_{\infty} < \varepsilon$.

✰

Tarkastellaan aluksi paikallista minimiä ja palataan vahvaan paikalliseen minimiin myöhemmin.

Tarkoituksena on johtaa jokin välttämätön ehto minimille (tai maksimille). Olkoon $V_0 \subset V$, valitaan jokin $\varphi \in V_0$, $y \in V$ ja tarkastellaan käyrää $\alpha : \mathbb{R} \to V$ missä $\alpha(s) = y + s\varphi$. Olkoon edelleen $g(s) = (J \circ \alpha)(s)$. Jos y on J:n paikallinen minimi, niin tällöinhän g:llä on paikallinen minimi origossa, joten välttämättä $g'(0) = 0$. Lasketaan siis g:n derivaatta:

$$\begin{aligned}\frac{d}{ds}g(s) =& \frac{d}{ds}\int_a^b L(x, y + s\varphi, y' + s\varphi')dx \\ =& \int_a^b \big(L_y\varphi + L_{y'}\varphi'\big)dx = \Big|_a^b L_{y'}\varphi + \int_a^b \big(L_y - \frac{d}{ds}L_{y'}\big)\varphi\, dx \\ =& \int_a^b \big(L_y - \frac{d}{dx}L_{y'}\big)\varphi\, dx\end{aligned}$$

Siis kaikilla $\varphi \in V_0$ pitää päteä, että

$$g'(0) = \int_a^b \big(L_y - \frac{d}{dx}L_{y'}\big)\varphi\, dx = 0 \tag{2.3}$$

Seuraavaa tulosta on tapana sanoa *variaatiolaskun peruslemmaksi.*

Lemma 2.1. *Olkoon* $f \in C(I)$ *ja oletetaan, että*

$$\int_a^b f(x)\varphi(x)dx = 0$$

kaikilla $\varphi \in V_0$*. Tällöin* $f(x) = 0$ *kaikilla* $x \in I$*.*

Todistus. Harjoitustehtävä. ✧

Mutta tästähän saadaan heti

Lemma 2.2. *Jos* y *on* J*:n minimi, niin* y *on differentiaaliyhtälön*

$$L_y - \frac{d}{dx}L_{y'} = 0 \tag{2.4}$$

ratkaisu. ❖

Määritelmä 2.2. *Yhtälö* (2.4) Eulerin yhtälö *ja sen ratkaisuja sanotaan* ekstremaaleiksi. ✫

Luonnollisesti mikään ei vielä takaa, että Eulerin yhtälön ratkaisu minimoisi J:n, samaan yhtälöön päädyttäisiin jos etsittäisiin maksimia. On myös mahdollista, että ekstremaali ei ole minimi eikä maksimi vaan satulapisteen tyyppinen ratkaisu.

Merkitään

$$\mathcal{E}_L(y) = \frac{d}{dx} L_{y'} - L_y = L_{y'y'} y'' + L_{yy'} y' + L_{xy'} - L_y$$

Eulerin yhtälön muodosta voidaan heti päätellä, että $L_{y'y'}$ on jollain tavalla tärkeä. Differentiaaliyhtälöitten ratkaisujen olemassaololausehan on muotoiltu tapauksessa jossa yhtälö on normaalimuotoinen, joka tässä tarkoittaa, että $L_{y'y'} \neq 0$ välillä $[a, b]$. Toisaalta $L_{y'y'}$ yleisesti ottaen riippuu ratkaisusta joten tilanteen analyysi ei välttämättä ole helppoa.

Tarkastellaan sitten eräitä erikoistapauksia. Jos L ei riipu y':stä niin Eulerin yhtälö on $L_y = 0$. Tämähän on yksinkertaisesti algebrallinen yhtälö josta y pitäisi ratkaista x:n funktiona. Tällöin y:ltä ei voi enää vaatia mitään erillisiä reunaehtoja. Tämä tapaus ei ole kovin kiinnostava.

Esimerkki 2.1. Jos L ei riipu y:stä niin $L_y = 0$ ja Eulerin yhtälöstä saadaan

$$L_{y'} = \text{vakio}$$

Ratkaistaan tämän avulla lyhimmän polun tehtävä (2.1). Tällöin siis

$$L_{y'} = \frac{y'}{\sqrt{1 + (y')^2}} = \text{vakio} = \sqrt{c}$$

Mutta tällöin $y' = \sqrt{c/(1-c)} = \text{vakio} = c_1$ joten $y(x) = c_1 x + c_0$ ja Eulerin yhtälön ratkaisut ovat siis suoria niin kuin pitikin. ✯

Esimerkki 2.2. Jos L ei riipu x:stä niin

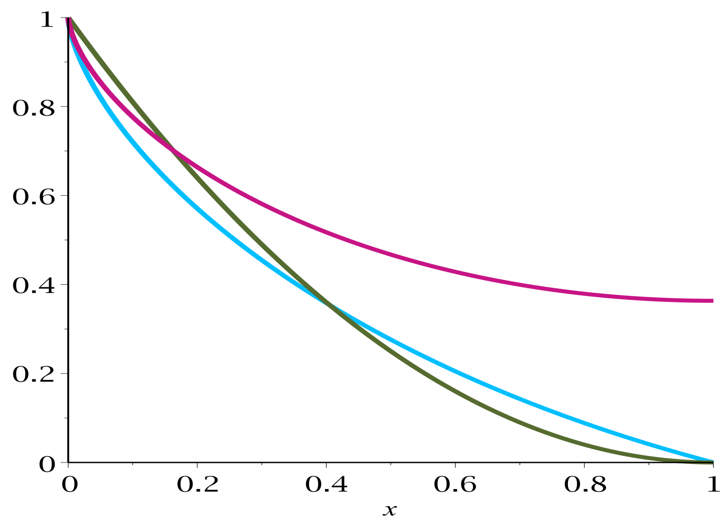
$$\begin{aligned} y'\mathcal{E}_L(y) =& L_{y'y'} y' y'' + L_{yy'} (y')^2 - L_y y' = \frac{d}{dx}\Big(L_{y'} y' - L\Big) = 0 \quad \Rightarrow \\ & L_{y'} y' - L = \text{vakio} \end{aligned}$$

Ratkaistaan tämän avulla lyhimmän ajan tehtävä (2.2). Tällöin

$$\begin{aligned} &\sqrt{2\mathrm{g}}\Big(L_{y'} y' - L\Big) = -\frac{1}{\sqrt{(1-y)(1+(y')^2)}} = \text{vakio} = -\frac{1}{\sqrt{c}} \quad \Rightarrow \\ &(1-y)(1+(y')^2) = c \end{aligned}$$

Tätä ei saada ratkaistua eksplisiittisesti mutta parametrimuotoinen ratkaisu saadaan kun ”arvataan”, että kannattaa asettaa $y'(x(\theta)) = -\cot(\theta)$. Pienen pyörittelyn jälkeen päädytään seuraavaan alkuehdon toteuttavaan ratkaisuun:

$$\begin{aligned} x(\theta) =& c\big(\theta - \tfrac{1}{2}\sin(2\theta)\big) \\ y(\theta) =& 1 - c\sin^2(\theta) \end{aligned}$$



Kuva 2.2: Esimerkin 2.2 ratkaiseva sykloidi on sinisellä. Vertailun vuoksi on piirretty myös parabeli $y_p = (1-x)^2$ jolle ajaksi saadaan $J(y_p) \approx 2.64$. Violetilla on puolestaan esimerkin 2.4 ratkaisu.

Tämä on *sykloidi*. Jotta loppuehto saataisiin voimaan, pitäisi löytää θ_1 ja c siten, että $x(\theta_1) = 1$ ja $y(\theta_1) = 0$. Tämä onnistuu vain numeerisesti: $c \approx 1.15$ ja $\theta_1 \approx 1.2$. Kun merkitään $\alpha(\theta) = \big(x(\theta), y(\theta)\big)$ niin optimaaliseksi ajaksi saadaan

$$J(\alpha) = \int_0^{\theta_1} \frac{|\alpha'(\theta)|}{\sqrt{c\,\sin^2(\theta)}} d\theta = 2\sqrt{c} \int_0^{\theta_1} d\theta = 2\sqrt{c}\,\theta_1 \approx 2.58$$

Ratkaisulla on seuraava mielenkiintoinen ominaisuus. Jos valitaan loppupiste siten, että päädytään sykloidin alimpaan pisteeseen, niin tällöin luonnollisesti $\theta_1 = \pi/2$ ja ajaksi tulee $\sqrt{c}\,\pi$. Mutta nyt jos pistemassa laskettaisiin liukumaan sykloidin mielivaltaisesta pisteestä niin aika alimpaan pisteeseen olisi aina sama. Sykloidi ratkaisee siis myös *isokroniongelman* (saman ajan ongelma). Tästä oltiin kiinnostuneita kun haluttiin kehittää tarkempia kelloja ja pohdittiin miten heilurin taajuus saataisiin riippumattomaksi heilahduksen amplitudista. ✯

Esimerkki 2.3. Muotoillaan Didon ongelma tarkemmin hieman toisessa muodossa. Etsitään funktiota y siten, että y:n graafin ja x-akselin välinen pinta-ala maksimoituu, kun käyrän pituus on annettu. Siis halutaan

$$\begin{aligned} \max \quad : \quad J(y) &= \int_0^1 y\,dx \quad \text{kun} \\ K(y) &= \int_0^1 \sqrt{1+(y')^2}dx = \ell > 1 \end{aligned}$$

Tässä ℓ on jokin annettu vakio ja reunaehtoina on $y(0) = y(1) = 0$. Selvästi tätä ei suoraan voi ratkaista Eulerin yhtälön avulla vaan tarvitaan muita ideoita.

On myös toinen luonnollinen tapa tulkita tehtävä. Jos halutaan maksimoida pinta-ala, niin silloinhan voisi ajatella, että integroimisväli ei ole annettu vaan sekin voidaan valita optimaalisesti. Tällöin saadaan tehtävä

$$\begin{aligned} \max \quad : \quad J(y) &= \int_0^b y\,dx \quad \text{kun} \\ K(y) &= \int_0^b \sqrt{1+(y')^2}dx = \ell > 0 \end{aligned}$$

Reunaehtoina on edelleen $y(0) = y(b) = 0$. Koska integroitavat funktiot eivät eksplisiittisesti riipu x:stä niin on selvää, että tässä yleisyyttä rajoittamatta voidaan valita $a = 0$. ✯

Palataan vielä yhtälöön (2.3) ja kirjoitetaan se hiukan toisin:

$$\delta J_y\,\varphi = \int_a^b \big(L_y\varphi + L_{y'}\varphi'\big)dx$$

Jos nyt ajatellaan, että y on jokin kiinteä funktio, niin δJ_y on lineaarikuvaus $\delta J_y : V \to \mathbb{R}$ missä $V \subset C^1(I)$.

Määritelmä 2.3. *δJ_y on J:n* (ensimmäinen) *variaatio.*

Sanotaan myös, että δJ_y on J:n Gâteaux'n derivaatta tai yksinkertaisesti J:n (ensimmäinen) differentiaali. Näin ajateltuna välttämätön ehto voidaan muotoilla: $\delta J_y = 0$ aivan analogisesti äärellisulotteisen tapauksen kanssa.

Edellä siis tarkasteltiin tilannetta jossa ekstremaalin arvot oli kiinnitetty välin päissä. Miten sitten tilanne muuttuu jos nämä arvot voi valita vapaasti? Olkoon edelleen

$$J(y) = \int_a^b L(x,y,y')dx \quad , \quad J : C^1(I) \to \mathbb{R} \tag{2.5}$$

Nyt siis y voi olla mielivaltainen funktio avaruudessa $C^1(I)$, ja toisaalta kun varioidaan y:tä kuvauksella $\alpha(s) = y + s\varphi$, niin myös φ voi saada mielivaltaisia arvoja välin päissä. Olkoon taas $g(s) = J(y + s\varphi)$ jolloin samoilla laskuilla kuin yllä saadaan välttämätön ehto

$$\begin{aligned} g'(0) =&\delta J_y\,\varphi = \int_a^b \big(L_y\varphi + L_{y'}\varphi'\big)dx \\ =&\Big|_a^b L_{y'}\varphi - \int_a^b \mathcal{E}_L(y)\,\varphi\,dx = 0 \quad \text{kaikilla} \quad \varphi \in C^1(I) \end{aligned}$$

Koska edelleenkin voidaan valita sellaisia φ että $\varphi(a) = \varphi(b) = 0$, niin tästä seuraa että tässäkin tapauksessa $\mathcal{E}_L(y) = 0$. Mutta kun tämän jälkeen valitaan sellaisia funktioita φ jotka eivät ole nollia välin päissä niin tästä sitten seuraa, että $L_{y'}(a) = L_{y'}(b) = 0$. Saatiin siis seuraava tulos.

Lemma 2.3. *Jos y minimoi yhtälössä* (2.5) *annetun kuvauksen, niin y on seuraavan tehtävän ratkaisu:*

$$\begin{cases} \mathcal{E}_L(y) = 0 \\ L_{y'}(a) = L_{y'}(b) = 0 \end{cases}$$

❖

Tässäkin tapauksessa saatiin 2 reunaehtoa, joten voidaan toivoa, että tehtävällä olisi yksikäsitteinen ratkaisu. Näitä reunaehtoja sanotaan joskus *luonnollisiksi reunaehdoiksi.*

Esimerkki 2.4. Pidetään lyhimmän ajan ongelmassa vasen reuna kiinnitettynä mutta valitaan y vapaasti kohdassa $x = 1$. Tässä tapauksessa

$$L_{y'} = \frac{y'}{\sqrt{(1-y)(1+(y')^2)}} = 0$$

Siis parametreiksi saadaan

$$\begin{cases} x(\theta_1) = 1 \\ y'(1) = y'(x(\theta_1)) = -\cot(\theta_1) = 0 \end{cases} \Rightarrow \quad \begin{cases} \theta_1 = \pi/2 \\ c = 2/\pi \end{cases}$$

Ratkaisu on kuvassa 2.2.Tätä vastaavaksi ajaksi saadaan $J(\alpha) = 2\sqrt{c}\,\theta_1 = \sqrt{2\pi} \approx 2.51.$

✪

2.3 Yleistyksiä

Edellä siis johdettiin välttämätön ehto minimille. Perustehtävää voi yleistää muun muassa seuraavilla tavoilla:

1. y on vektori $\Rightarrow$ Eulerin yhtälö on toisen kertaluvun difyhtälösysteemi.
2. x on vektori $\Rightarrow$ Eulerin yhtälö on osittaisdifyhtälö.
3. L riippuu myös y:n korkeammista derivaatoista $\Rightarrow$ Eulerin yhtälö on korkeamman kertaluvun difyhtälö.

Lisäksi herää seuraavanlaisia kysymyksiä:

(i) miten käsitellä yleisemmät reunaehdot?

(ii) miten voisi löytää riittäviä ehtoja minimille/maksimille?

(iii) edellä oletettiin, että ratkaisu on riittävän monta kertaa derivoituva, ja että se ylipäätään on olemassa. Lähempi tarkastelu kuitenkin paljastaa, että tilanne ei ole näin yksinkertainen.

(iv) usein variaatiotehtäviin liittyy erilaisia rajoitusehtoja, kuten Didon ongelmassa. Myös muunlaisia rajoitusehtoja esiintyy. Miten nämä pitäisi käsitellä?

Luonnollisesti on mahdollista tarkastella useita tapauksia yhdessä. Koska korkeamman kertaluvun tehtävät aina voidaan palauttaa ensimmäisen kertaluvun systeemiksi niin tapaus 3. palautuu tapaukseen 1. Kohdan (iii) esimerkit osoittavat, että ei ole oikeastaan selvää mistä funktioluokasta ratkaisua pitäisi etsiä, mikä puolestaan vaikeuttaa entisestään ongelmaa (ii). Kohtien (ii), (iii) ja 2. perusteellisempi käsittely edellyttäisikin siis varsin raskaita funktionaalianalyysin työkaluja [3, 12], ja niihin ei tällä kurssilla voida puuttua.

On kuitenkin eräs geometrisesti luonnollinen välttämätön ehto jonka avulla päästään sekä kiinni eräisiin yleistettyihin ratkaisuihin ja samalla vaivalla voidaan analysoida yleisempiä reunaehtoja. Katsotaan näitä seuraavassa luvussa.

Monissa klassisen mekaniikan sovelluksissa y on vektori. Muodollisesti Eulerin yhtälö voidaan helposti johtaa samaan tapaan kuin skalaaritapauksessa. Näihin tehtäviin liittyy myös usein rajoitusehtoja.

2.4 Harjoitustehtäviä

1. Näytä, että jos Lagrangen funktio riippuu vain y':sta niin suorat ovat ekstremaaleja. Voiko olla muita ekstremaaleja?

2. Todista Lemma 2.1.

Luku 3

Yleistettyjä ratkaisuja ja reunaehtoja

3.1 Esimerkkejä

Ehkäpä on parasta lähteä liikkeelle muutamasta esimerkistä mistä käy ilmi miksi yleistettyjä ratkaisuja ylipäätään tarvitaan. Hieman yllättäen näitä yleistettyjä ratkaisuja esiintyy vaikka tehtävän Lagrangen funktio L olisi sileä.

Esimerkki 3.1. Olkoon

$$J(y) = \tfrac{1}{2}\int_{-1}^{1} y^2\big(2x - y'\big)^2 dx$$

ja asetetaan reunaehdoiksi $y(-1) = 0$ ja $y(1) = 1$. Selvästi $J(y) \geq 0$ kaikilla y. Toisaalta jos määritellään

$$y_*(x) = \begin{cases} 0 & , \ -1 \leq x \leq 0 \\ x^2 & , \ 0 \leq x \leq 1 \end{cases}$$

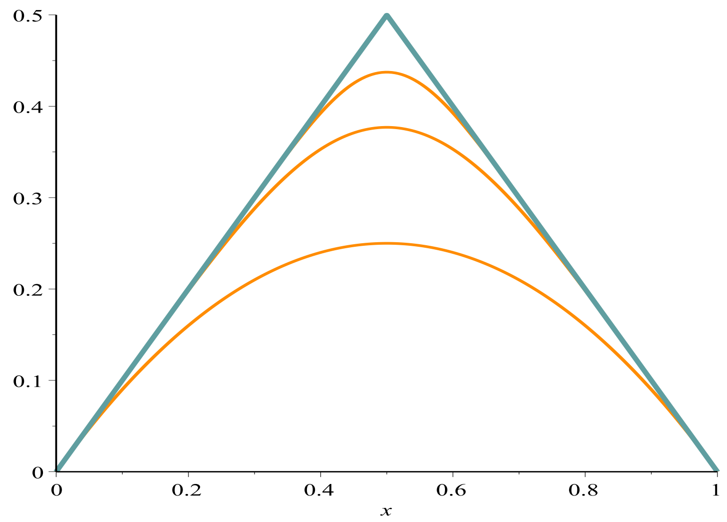
niin havaitaan, että $J(y_*) = 0$ ja y_* toteuttaa reunaehdot. Toisaalta nähdään, että $y_* \in C^1(I)$, mutta $y_* \notin C^2(I)$. Eulerin yhtälö on

$$\mathcal{E}_L(y) = y\big(yy'' + (y')^2 - 2y - 4x^2\big) = 0$$

Yhtälö jakaantuu siis tekijöihin ja nolla on ensimmäisen tekijän ja x^2 toisen tekijän ratkaisu. Tässä mielessä siis y_* toteuttaa Eulerin yhtälön, paitsi kenties kohdassa $x = 0$. Toisaalta jos variaatio kirjoitetaan muodossa

$$\delta J_y\, \varphi = \int_{-1}^{1} \big(L_y \varphi + L_{y'}\varphi'\big)dx = \int_{-1}^{1} y(2x - y')\big((2x - y')\varphi - y\varphi'\big)dx$$

niin havaitaan että $\delta J_{y_*}\, \varphi = 0$ kaikilla φ ilman mitään ongelmia koska tässä ei tarvita toisen derivaatan jatkuvuutta tai olemassaoloa. ✫



Kuva 3.1: Esimerkin 3.2 eräs ratkaisu ja sen joitain sileitä approksimaatioita.

Esimerkki 3.2. Tämä esimerkki on peräisin Eulerilta. Olkoon

$$J(y) = \tfrac{1}{4}\int_0^1 \big((y')^2 - 1\big)^2 dx$$

ja asetetaan reunaehdoiksi $y(0) = y(1) = 0$. Selvästi $J(y) \geq 0$ kaikilla y. Nyt on helppo tarkistaa, että ekstremaalit ovat suoria, joten reunaehdot huomioon ottaen nollafunktio on ratkaisuehdokas. Vastaava arvo $J(0) = 1/4$ on kuitenkin kaukana minimistä. Jos puolestaan valitaan

$$y_*(x) = \begin{cases} x & , \ 0 \leq x \leq 1/2 \\ 1 - x & , \ 1/2 \leq x \leq 1 \end{cases}$$

niin selvästi $J(y_*) = 0$. Mutta nyt y_* on vain jatkuva eikä jatkuvasti derivoituva joten ei ole ihan selvää voidaanko sitä pitää ”hyväksyttävänä” ratkaisuna. Lisäksi tällöin on selvää, että ratkaisuja on ääretön määrä: voidaanhan valita mikä tahansa z siten, että z toteuttaa reunaehdot, $z \in C(I)$ ja $z' = \pm 1$ paitsi äärellisessä määrässä pisteitä.

Lisäksi on selvää, että ratkaisua y_* voidaan approksimoida mielivaltaisen tarkasti sileillä funktioilla vaikkapa kuten kuvassa 3.1, joten $\inf J(y) = 0$. Tehtävällä ei siis ole ratkaisua jatkuvasti derivoituvien funktioitten joukossa. ✮

Esimerkki 3.3. Katsotaan vielä Weierstrassin esimerkki. Olkoon

$$J(y) = \tfrac{1}{2}\int_{-1}^1 x^2 (y')^2 dx$$

ja asetetaan reunaehdoiksi $y(-1) = -1$ ja $y(1) = 1$. Tässäkin $J(y) \geq 0$ kaikilla y. Toisaalta jos $y \in C^1(I)$ niin selvästi $J(y) > 0$ jos y toteuttaa reunaehdot. Helposti nähdään myös että $J(y) > 0$ vaikka sallittaisiin paloittain jatkuvasti derivoituvat ratkaisut kuten edellisessä esimerkissä. Jos kuitenkin voitaisiin osoittaa, että

$\inf J(y) = 0$ niin johtopäätös olisi, että tehtävällä ei ole ratkaisua, ainakaan missään tavallisessa funktioavaruudessa. Juuri tämän Weierstrass halusikin esimerkillään osoittaa.

Tarkistetaan kuitenkin ensin ekstremaalit. Koska L ei riipu y:stä niin

$$L_{y'} = x^2 y' = \text{vakio} = -c_1 \quad \Rightarrow \quad y = c_0 + \frac{c_1}{x}$$

Jos siis $c_1 \neq 0$ niin ekstremaalit eivät ole määritelty halutulla alueella ja toisaalta jos $c_1 = 0$ niin vakiofunktiot eivät toteuta reunaehtoja. Tässä siis ekstremaalit eivät antaneet mitään hyödyllistä informaatiota.

Olkoon sitten

$$y_n(x) = \frac{\arctan(nx)}{\arctan(n)}$$

Tällöin y_n toteuttaa reunaehdot ja suoraviivainen lasku osoittaa, että

$$J(y_n) < \frac{2}{n \arctan(n)}$$

joten $\inf J(y) = 0$. Jos sitten asetetaan

$$y_*(x) = \lim_{n\to\infty} y_n(x) = \begin{cases} -1 & , \ -1 \leq x < 0 \\ 0 & , \ x = 0 \\ 1 & , \ 0 < x \leq 1 \end{cases}$$

niin ei ole selvää voitaisiinko tätä pitää jollain tavalla mielekkäänä ratkaisuna. Jos sallitaan epäjatkuvia ratkaisuja, niin sitten voitaisiin yhtä hyvin asettaa vaikkapa

$$\tilde{y}(x) = \begin{cases} -1 & , \ -1 \leq x < -1/3 \\ 7 & , \ -1/3 \leq x \leq 1/2 \\ 1 & , \ 1/2 < x \leq 1 \end{cases}$$

✯

Osoittautuu, että esimerkin 3.2 tapauksessa päästään eteenpäin perinteisin menetelmin. Itse asiassa tämän ongelman ymmärtämisessä auttaa vapaan reunan tapaus: tässä siis ajatellaan, että väli $I = [a, b]$ ei olekaan kiinteä vaan voi muuttua. Katsotaan siis ensin vapaan reunan tapausta minkä jälkeen tietyntyyppiset yleistetyt ratkaisut tulevat "itsestään".

3.2 Vapaa reuna

Kun ajatellaan, että välin $I = [a, b]$ päät eivät olisikaan kiinteitä vaan voisivat muuttua niin miten sitten variaatio pitäisi tässä tapauksessa määritellä? Tätä kysymystä voi lähestyä eri tavoin mutta edetään nyt tavalla joka ei ole se klassinen tapa vaan geometrisesti luonnollinen tapa. Tämän lähestymistavan etu on myös se, että se soveltuu sellaisenaan myös osittaisdifyhtälöihin, siis tilanteeseen jossa y on monen muuttujan funktio.

Tarkastellaan siis vapaan reunan tehtävää

$$J(y) = \int_a^b L(x, y, y')dx$$

missä nyt ajatellaan, että a ja b voivat muuttua. Perusidea on sama kuin ennen: pitäisi muodostaa jokin perhe variaatioita ja sitten derivoida ”sopivasti”. Lähdetään liikkeelle siitä että on annettu perhe diffeomorfismeja (määritelmä A.3). Olkoon $I = [a, b]$ ja olkoon sitten annettu kuvaukset h_ε joilla on seuraavat ominaisuudet:

(i) $h_\varepsilon : I \to I_\varepsilon$ on diffeomorfismi kun $|\varepsilon| < \varepsilon_0$

(ii) $h_0(x) = x$

(iii) jos määritellään $h(x, \varepsilon) = h_\varepsilon(x)$ niin h on jatkuvasti derivoituva

Määritellään tämän avulla funktioperhe $y_\varepsilon = y \circ h_\varepsilon^{-1}$ ja tätä vastaava variaatio

$$g(\varepsilon) = J(y_\varepsilon) = \int_{I_\varepsilon} L(s, y_\varepsilon, y_\varepsilon')ds$$

Nyt voidaan tulkita, että h_ε on ”vain” koordinaattimuunnos joten y_ε on tavallaan y esitettynä eri koordinaatin avulla. Erityisesti siis ei tarvitse jatkaa funktiota y alkuperäisen määrittelyalueen ulkopuolelle. Siis variaatio on tosiaankin vain itse alueeseen liittyvä joten tässä tavallaan pitäisi derivoida ”alueen suhteen”.

Yleensä kun jokin tehtävä on annettu muuttuvassa alueessa, niin kannattaa muodostaa ekvivalentti tehtävä joka on annettu kiinteässä alueessa. Koska h_ε on diffeomorfismi voidaan kirjoittaa

$$g(\varepsilon) = \int_a^b L\big(h_\varepsilon(x), y(x), y_\varepsilon'(h_\varepsilon(x))\big)h_\varepsilon'(x)\,dx$$

Edelleen

$$y_\varepsilon'(s) = y'\big(h_\varepsilon^{-1}(s)\big)\tfrac{d}{ds}h_\varepsilon^{-1}(s) = \frac{y'\big(h_\varepsilon^{-1}(s)\big)}{h_\varepsilon'\big(h_\varepsilon^{-1}(s)\big)}$$

mistä saadaan

$$g(\varepsilon) = \int_a^b L\big(h_\varepsilon(x), y(x), \frac{y'(x)}{h_\varepsilon'(x)}\big)h_\varepsilon'(x)\,dx$$

Valitaan nyt $h_\varepsilon(x) = x + \varepsilon\,\psi(x)$ missä ψ on jokin mielivaltainen funktio. Selvästi tämä toteuttaa kaikki vaaditut ehdot. Siispä

$$g(\varepsilon) = \int_a^b L\big(x + \varepsilon\,\psi(x), y(x), \frac{y'(x)}{1 + \varepsilon\psi'(x)}\big)\big(1 + \varepsilon\psi'(x)\big)\,dx$$

Tästä sitten derivoimalla saadaan

$$g'(0) = \int_a^b \big(L_x\psi + \big(L - y'L_{y'}\big)\psi'\big)\,dx$$

Määritelmä 3.1. *Olkoon*

$$\partial J_y\,\psi = \int_a^b \big(L_x\psi + \big(L - y' L_{y'}\big)\psi'\big)\,dx$$

∂J_y *on* sisäinen variaatio. ✫

Kuten δJ_y, myös ∂J_y on lineaarikuvaus. Osittaisintegroimalla saadaan

$$\partial J_y\,\psi = \Big|_a^b \big(L - y'\,L_{y'}\big)\psi + \int_a^b y'\,\mathcal{E}_L(y)\psi\,dx$$

Lause 3.1. *Jos y minimoi vapaan reunan tehtävän, niin*

$$\begin{cases} \mathcal{E}_L(y) = 0 \\ L_{y'}(a) = L_{y'}(b) = 0 \\ L(a) = L(b) = 0 \end{cases}$$

Todistus. Jos y minimoi niin myös $\delta J_y = 0$ joten $\mathcal{E}_L(y) = 0$. Koska y:n arvoja ei ole kiinnitetty välin päissä niin lemman 2.3 perusteella $L_{y'}(a) = L_{y'}(b) = 0$. Lopuksi myös sisäisen variaation pitää olla nolla joten $L(a) = L(b) = 0$. ✧

Esimerkki 3.4. Olkoon

$$J(y) = \tfrac{1}{2}\int_0^b \big((y')^2 - y^2 + 1 + 10(y'-1)(y-x+4)\big)dx$$

Otetaan toiseksi reunaehdoksi $y(0) = 0$ mutta toinen reuna on vapaa. Eulerin yhtälö on $y'' + y = 0$ joten ekstremaalit ovat muotoa $y(x) = c\sin(x)$. Vapaan reunan ehdoista saadaan

$$\begin{aligned} &c\cos(b) + 5c\sin(b) - 5b + 20 = 0 \\ &c^2\cos(2b) + 1 + 10(c\cos(b) - 1)(c\sin(b) - b + 4) = 0 \end{aligned}$$

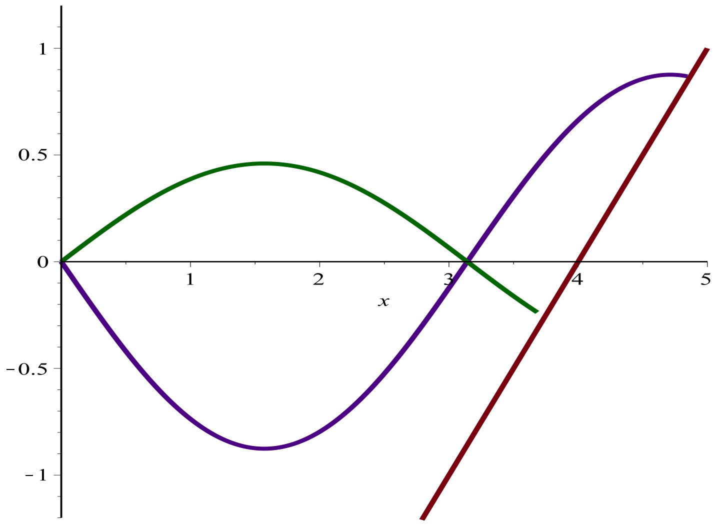
Numeerisesti löydetään ainakin seuraavat ratkaisut:

$$\begin{cases} b \approx 3.7 \\ c \approx 0.46 \end{cases}, \quad \begin{cases} b \approx 4.8 \\ c \approx -0.88 \end{cases}$$

Näitä vastaavat ratkaisut on kuvassa 3.2. Huomaa, että kohdefunktio voidaan esittää muodossa

$$J(y) = \tfrac{1}{2}\int_0^b \big((y')^2 - y^2\big)dx + b + \tfrac{5}{2}\,(y(b) - b + 4)^2 - 40$$

Tämä voidaan tulkita optimisäädön kannalta seuraavasti. Haluttaisiin, että käyrä on loppupisteessä mahdollisimman lähellä suoraa $y = x - 4$. Toisaalta systeemin siirtäminen suoralle maksaa joten ollaan valmiita jonkin verran tinkimään tavoitteista jos ”kustannukset” saadaan pieniksi. ✭



Kuva 3.2: Esimerkin 3.4 kaksi ekstremaalia ja suora $y = x - 4$.

3.3 Kulmaehdot

Olkoon $I = [a, b]$, $a = x_0 < x_1 < \cdots < x_n = b$ ja $I_\ell = [x_{\ell-1}, x_\ell]$. Tarkastellaan seuraavaa funktioavaruutta:

$$C_p^1(I) = \left\{ y : I \to \mathbb{R} \,\middle|\, y_{|I_\ell} \in C^1(I_\ell) \right\}$$

Siis y on jatkuvasti derivoituva, paitsi mahdollisesti äärellisessä määrässä pisteitä. Huomaa, että derivaattojen vasemman- ja oikeanpuoleiset raja-arvot ovat kaikkialla määritelty. Siis jos $I = [-1, 1]$, niin $|x| \in C_p^1(I)$ mutta $\sqrt{|x|} \notin C_p^1(I)$.

Helposti voidaan tarkistaa, että C_p^1 on vektoriavaruus. Pisteitä joissa y ei ole derivoituva sanotaan *kulmapisteiksi*. Kulmapisteitten paikka ei ole etukäteen tiedossa mistä tulee yhteys vapaan reunan tehtävään jossa välin päät voivat muuttua. Pitäisi siis löytää ehtoja sille, että kulmapisteestä olisi jotain etua minimoinnin kannalta. Koska kaikki analyysi on lokaalia, niin riittää tarkastella yhden kulmapisteen mahdollisuutta. Olkoon siis $I = [a, b] = [a, \hat{x}] \cup [\hat{x}, b]$ missä $\hat{x}$ on jokin potentiaalinen kulmapiste. Koska $\hat{x}$ on etukäteen tuntematon niin tässä joudutaan samoihin tarkasteluihin kuin vapaan reunan tapauksessa.

Oletetaan siis, että a ja b ovat kiinteitä ja etsitään ekstremaalia $y \in C_p^1(I)$ joka toteuttaa reunaehdot $y(a) = y(b) = 0$. Minimoitava integraali voidaan nyt kirjoittaa muodossa

$$J(y) = \int_a^{\hat{x}} L(x, y, y')dx + \int_{\hat{x}}^b L(x, y, y')dx$$

Merkitään $y'_+ = \lim_{x \searrow \hat{x}} y'$, $L^- = \lim_{x \nearrow \hat{x}} L(x, y(x), y'(x))$ ja vastaavasti muut toispuoleiset raja-arvot.

$$\begin{aligned}\delta J_y \varphi &= \Big|_a^{\hat{x}} L_{y'} \varphi - \int_a^{\hat{x}} \mathcal{E}_L(y)\, \varphi \, dx + \Big|_{\hat{x}}^b L_{y'} \varphi - \int_{\hat{x}}^b \mathcal{E}_L(y)\, \varphi \, dx \\ &= (L_{y'}^- - L_{y'}^+)\varphi(\hat{x}) = 0\end{aligned}$$

Koska $\varphi(\hat{x})$ on mielivaltainen, niin $L_{y'}$ on jatkuva pisteessä $\hat{x}$ vaikka siis y' on (mahdollisesti) epäjatkuva. Sisäisestä variaatiosta saadaan puolestaan

$$\begin{aligned}\partial J_y\psi &= \Big|_a^{\hat{x}}\big(L - y'\,L_{y'}\big)\psi + \int_a^{\hat{x}} y'\,\mathcal{E}_L(y)\psi\,dx + \Big|_{\hat{x}}^{b}\big(L - y'\,L_{y'}\big)\psi + \int_{\hat{x}}^{b} y'\,\mathcal{E}_L(y)\psi\,dx \\ &= (L^- - y'_- L^-_{y'} - L^+ + y'_+ L^+_{y'})\psi(\hat{x}) = 0\end{aligned}$$

Koska nyt $\psi(\hat{x})$ mielivaltainen niin $L - y' L_{y'}$ on jatkuva $\hat{x}$:ssä. Luonnollisesti jos kulmapisteitä on enemmän, niin näitten ehtojen on oltava voimassa jokaisessa kulmapisteessä. Saatiin siis seuraava tulos.

Lause 3.2. *Jos $y \in C_p^1(I)$ on variaatiotehtävän minimi niin*

(i) $\mathcal{E}_L(y) = 0$ *jokaisella osavälillä*

(ii) $L_{y'}$ *ja* $L - y' L_{y'}$ *ovat jatkuvia kulmapisteissä.*

❖

Määritelmä 3.2. *Edellisen lauseen ehtoa (ii) sanotaan* Weierstrassin ja Erdmanin kulmaehdoksi. *Lauseen ehdot toteuttavaa ratkaisua sanotaan* kulmaekstremaaliksi. ✰

Esimerkki 3.5. Tarkastellaan tehtävää

$$J(y) = \tfrac{1}{2}\int_{-1}^{1} y^2\big(y' - 1\big)^2 dx$$

ja valitaan reunaehdoiksi $y(-1) = 0$ ja $y(1) = 1$. Eulerin yhtälö on

$$\mathcal{E}_L(y) = y\big(yy'' + (y')^2 - 1\big) = 0$$

Tehtävällä on siis kolmenlaisia ekstremaaleja:

(i) suora $y = 0$

(ii) suorat $y = \pm x + c$

(iii) hyperbelit $y^2 - (x - c)^2 = c_0$

Edelleen

$$\begin{aligned} L_{y'} &= y^2(y' - 1) \\ L - y' L_{y'} &= -\tfrac{1}{2}\,y^2(y' - 1)(y' + 1)\end{aligned}$$

Nyt nähdään, että

$$y_*(x) = \begin{cases} 0 & , \ -1 \le x \le 0 \\ x & , \ 0 \le x \le 1 \end{cases}$$

on haluttu kulmaekstremaali. Lisätarkastelut osoittavat, ettei tehtävällä ole muita kulmaekstremaaleja. ✮

3.4 Harjoitustehtäviä

1. Näytä, että esimerkissä 3.2 saatu nollaekstremaali on itse asiassa heikko paikallinen maksimi.

2. Osoita, että Weierstrassin esimerkissä 3.3 ei voi olla kulmaekstremaaleja.

3. Osoita, että esimerkissä 3.5 ei voi olla muita kulmaekstremaaleja. Onko tehtävällä sileitä ekstremaaleja, jotka toteuttavat reunaehdot?

Luku 4

Yleisempiä variaatiotehtäviä

4.1 Korkeampi kertaluku

Tarkastellaan tilannetta, jossa Lagrangen funktio riippuu myös toisesta derivaatasta:

$$J(y) = \int_a^b L(x, y, y', y'')dx \tag{4.1}$$

Oletetaan jälleen, että L on riittävän monta kertaa jatkuvasti derivoituva. Jos taas tarkastellaan variaatioita $\alpha(\varepsilon) = y + \varepsilon\varphi$ ja määritellään $g(\varepsilon) = J(y + \varepsilon\varphi)$, niin samoilla laskuilla kun aiemmin saadaan

$$g'(0) = \delta J_y \varphi = \int_a^b \big(L_y \varphi + L_{y'}\varphi' + L_{y''}\varphi''\big)dx$$

Oletetaan siis tässä, että $\varphi \in C^2(I)$. Tästä osittaisintegroimalla saadaan

$$\delta J_y \varphi = \int_a^b \big(L_y - \frac{d}{dx}L_{y'} + \frac{d^2}{dx^2}L_{y''}\big)\varphi dx + \Big|_a^b L_{y''}\varphi' + \big(L_{y'} - \frac{d}{dx}L_{y''}\big)\varphi$$

Määritelmä 4.1. *Tehtävän* (4.1) Eulerin yhtälö *on*

$$\mathcal{E}_L(y) = L_y - \frac{d}{dx}L_{y'} + \frac{d^2}{dx^2}L_{y''} = 0$$

☆

Eulerin yhtälön ratkaisuja sanotaan edelleen ekstremaaleiksi. Eulerin yhtälö on tässä tapauksessa neljännen kertaluvun difyhtälö joten nyt tarvitaan neljä reunaehtoa, ja klassisen ratkaisun pitäisi siis olla neljä kertaa jatkuvasti derivoituva. Jos siis y on ekstremaali, niin

$$\delta J_y \varphi = \Big|_a^b L_{y''}\varphi' + \big(L_{y'} - \frac{d}{dx}L_{y''}\big)\varphi$$

Valitsemalla sitten sopivia rajoitteita halutulle ekstremaalille, niin tästä saadaan sitten vaadittavat neljä reunaehtoa. Luonnollisesti jälleen voitaisiin etsiä kulmaekstremaaleja ja vapaan reunan ehtoja jne. Tämä on kuitenkin hyvin samanlaista kuin ensimmäisen kertaluvun tapauksessa joten sivuutetaan lähemmät tarkastelut ja katsotaan vain pari esimerkkiä joissa luonnollisesti esiintyy toinen derivaatta Lagrangen funktiossa.

Esimerkki 4.1. *Interpolointitehtävä.* Olkoon annettu pisteet (x_i, y_i), missä $a = x_0 < x_1 < \cdots < x_n = b$. Haluttaisiin löytää funktio y siten, että $y_i = y(x_i)$ kaikilla i ja

$$J(y) = \tfrac{1}{2} \int_a^b (y'')^2 dx$$

olisi mahdollisimman pieni. Tässä siis annetaan ehtoja myös sisäpisteissä, joten etsitään kulmaekstremaalin tyyppistä ratkaisua. Huomaa, että variaatiofunktiolta vaaditaan $\varphi(x_k) = 0$. Siispä

$$\begin{aligned}\delta J_y \varphi &= \int_a^b y''\varphi'' dx = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} y''\varphi'' dx = \sum_{k=1}^n \Big|_{x_{k-1}}^{x_k} y''\varphi' - \sum_{k=1}^n \int_{x_{k-1}}^{x_k} y'''\varphi' dx \\ &= \sum_{k=1}^n \Big|_{x_{k-1}}^{x_k} y''\varphi' + \sum_{k=1}^n \int_{x_{k-1}}^{x_k} y''''\varphi dx\end{aligned}$$

Siispä vaaditaan, että $y'''' = 0$ jokaisella osavälillä, joten y on joka välillä kolmannen asteen polynomi. Jäljelle jäävä termi voidaan nyt kirjoittaa muodossa

$$\delta J_y \varphi = \sum_{k=1}^{n-1} \big(y''_-(x_k) - y''_+(x_k)\big)\varphi'(x_k) + y''(b)\varphi'(b) - y''(a)\varphi'(a)$$

Siispä y'':n täytyy olla jatkuva minkä lisäksi $y''(b) = y''(a) = 0$. Päädyttiin siis siihen, että

(i) y on kolmannen asteen polynomi osaväleillä

(ii) $y(x_j) = y_j$, $0 \le j \le n$

(iii) $y \in C^2(I)$.

(iv) $y''(b) = y''(a) = 0$

Ehdot (i) - (iii) toteuttavaa ratkaisua sanotaan *kuutiospliniksi.* Käytännön interpolointitehtävissä ehdon (iv) sijaan käytetään muitakin reunaehtoja. ✯

Yleisessä tapauksessa L voi riippua y:n derivaatoista kertalukuun q asti:

$$J(y) = \int_a^b L\big(x, y, y', \ldots, y^{(q)}\big) dx$$

Samantyyppisillä laskuilla kuin edellä saadaan Eulerin yhtälöksi tässä tapauksessa

$$\mathcal{E}_L(y) = L_y - \frac{d}{dx} L_{y'} + \ldots (-1)^q \frac{d^q}{dx^q} L_{y^{(q)}} = 0$$

Eulerin yhtälön kertaluku on siis $2q$, joten tarvitaan yhteensä $2q$ reunaehtoa.

4.2 Monen muuttujan tapaus

Monissa variaatiolaskun sovelluksissa on tarpeen tarkastella vektoriarvoisia funktioita, toisin sanoen käyriä n-ulotteisissa avaruuksissa. Yleisesti ottaen osoittautuu, että klassisen mekaniikan mallit voidaan johtaa variaatiolaskun avulla.

Muutetaan tässä hiukan merkintöjä ja toivotaan, ettei se aiheuta liikaa sekaannusta. Olkoon nyt $x : I = [t_0, t_1] \to \mathbb{R}^n$ ja tarkastellaan kuvausta

$$J(x) = \int_{t_0}^{t_1} L(t, x, x')dt$$

Tässä ajatellaan nyt, että t on aika ja x on "paikkakoordinaatti". Esimerkiksi jos $x(t)$ on pistemassan paikka hetkellä t niin $x : \mathbb{R} \to \mathbb{R}^3$. Joka tapauksessa osoittautuu, että skalaaritapaus yleistyy varsin vaivattomasti tähän tilanteeseen. Tarkastellaan variaatiota $\alpha(\varepsilon) = x + \varepsilon\varphi$ kuten ennenkin mutta nyt myös φ on vektorifunktio. Jos merkitään edelleen $g(\varepsilon) = J(x + \varepsilon\varphi)$ niin derivoimalla saadaan

$$g'(0) = \delta J_x\, \varphi = \int_{t_0}^{t_1} \sum_{j=1}^{n} \big(L_{x_j}\varphi_j + L_{x'_j}\varphi'_j\big)dx$$

Tästä edelleen osittaisintegroimalla saadaan

$$\delta J_x\, \varphi = \Big|_{t_0}^{t_1} \sum_{j=1}^{n} L_{x'_j}\varphi_j + \int_{t_0}^{t_1} \sum_{j=1}^{n} \big(L_{x_j} - \frac{d}{dx}L_{x'_j}\big)\varphi_j dx$$

Merkitään nyt

$$\mathcal{E}_L^j(x) = \frac{d}{dx}L_{x'_j} - L_{x_j}$$

Koska φ:n jokaista komponenttia voidaan muuttaa toisistaan riippumatta, niin selvästi ekstremaalin pitää toteuttaa difyhtälöt

$$\mathcal{E}_L^j(x) = 0 \quad , \quad 1 \le j \le n$$

Näitä tietysti edelleen sanotaan Eulerin yhtälöiksi.

Esimerkki 4.2. Olkoon x pistemassan paikka ja tarkastellaan seuraavaa Lagrangen funktiota

$$L = T - U = \tfrac{1}{2}\, m|x'|^2 - U(x)$$

Tässä siis T on kineettinen energia ja U on potentiaali(funktio). Eulerin yhtälöistä saadaan nyt

$$mx'' = -\nabla U$$

mikä on Newtonin toinen laki kun voimakenttä tulee potentiaalista. Tässä tapauksessa ei ole oikeastaan selvää pitäisikö ekstremaalin antaa minimi, maksimi vai satulapiste. Itse asiassa ei ole selvää onko ekstremaalin laadulla edes väliä. Edelleen reunaehdot ovat tässä oikeastaan epäoleellisia, ja tehtävää onkin mielekästä tarkastella alkuarvotehtävänä: annetaan $x(t_0)$ ja $x'(t_0)$ ja lasketaan Eulerin yhtälön avulla pistemassan liike.

Edelleen jos määritellään $H = T + U$, niin suoraan derivoimalla nähdään, että

$$\frac{d}{dt} H = 0$$

joten ratkaisun kokonaisenergia pysyy vakiona. Funktiota H sanotaan tehtävän *Hamiltonin funktioksi*. Jos sitten asetetaan $y = mx'$, niin voidaan kirjoittaa

$$H(x, y) = \tfrac{1}{2m} |y|^2 + U(x)$$

ja Eulerin yhtälö voidaan nyt kirjoittaa ensimmäisen kertaluvun systeeminä seuraavasti:

$$\begin{aligned} x' &= \nabla_y H = y/m \\ y' &= -\nabla_x H = -\nabla U \end{aligned}$$

Tätä sanotaan *Hamiltonin systeemiksi*. Huomaa, että fysikaalisesti y on tässä liikemäärä. Usein sanotaan, että x ”liikkuu” *konfiguraatioavaruudessa* (joka on tässä $\mathbb{R}^3$) ja $z = (x, y)$ *tila-avaruudessa* (joka on tässä $\mathbb{R}^6$).[1] Muuttujan z avulla Hamiltonin systeemi voidaan kätevästi kirjoittaa seuraavasti:

$$z' = \mathcal{I}\,\nabla H \quad \text{missä} \quad \mathcal{I} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

Tässä siis I on 3×3 yksikkömatriisi. $\mathcal{I}$ on *symplektinen matriisi*. Yleisesti ottaen matriisi on symplektinen, jos

$$A^T \mathcal{I} A = \mathcal{I}$$

✯

Perusyhtälöt saatiin siis hyvin samaan tapaan kuin skalaaritapauksessa. Uusi tilanne syntyy kuitenkin uudentyyppisistä rajoitusehdoista, joita esiintyy monien sovellusten yhteydessä.

Esimerkki 4.3. Tarkastellaan heiluria joka riippuu origosta. Olkoon $x : \mathbb{R} \to \mathbb{R}^2$ heilurin pään paikka ja olkoon m heilurin massa, ℓ heilurin pituus ja g maan vetovoiman kiihtyvyys. Lagrangen funktio on tällöin

$$L = T - U = \tfrac{1}{2}\, m|x'|^2 - m\mathsf{g}x_2$$

Koska heilurin pituus ℓ on vakio, niin $|x(t)|^2 - \ell^2 = 0$ kaikilla t. Ei siis voida suoraan muodostaa Eulerin yhtälöä vaan tämä rajoitusehto pitäisi ottaa jollain tavalla huomioon. Selvästi tämä rajoitusehto on hyvin erityyppinen kuin integraalimuotoinen rajoitusehto esimerkissä 2.3. Palataan tähän esimerkkiin myöhemmin. ✯

[1] Tila-avaruus on state space. Käytetään myös termiä *faasiavaruus*, phase space.

4.3 Parametriset tehtävät

Joissain tilanteissa, etenkin geometrisissa tehtävissä, on luonnollista etsiä jotain käyrää jolla on jokin tietty ominaisuus joka ei riipu käyrän parametrisoinnista. Esimerkiksi lyhimmän polun etsiminen on tällainen ongelma. Katsotaan ensin mitä ominaisuuksia Lagrangen funktiolla on näissä tapauksissa.

Selvästi L ei voi riippua muuttujasta t joten tarkastellaan tehtävää

$$J(x) = \int_{t_0}^{t_1} L(x, x')dt$$

Olkoon sitten $h : [t_0, t_1] \to [t_0, t_1]$ diffeomorfismi ja $x = z \circ h$. Tällöin siis

$$\int_{t_0}^{t_1} L(x, x')dt = \int_{t_0}^{t_1} L\big(z(s), \frac{z'(s)}{h'(s)}\big)h'(s)ds$$

Selvästi integraalin arvo ei riipu h:sta jos

$$L(x, c\,x') = c\,L(x, x') \quad \text{kaikilla} \quad c > 0 \tag{4.2}$$

Sanotaan, että L on positiivisesti homogeeninen astetta yksi muuttujan x' suhteen. Esimerkiksi muotoa $L = f(x)|x'|$ olevat Lagrangen funktiot toteuttavat tämän ehdon.

Määritelmä 4.2. *Variaatiotehtävää ja sitä vastaavaa Lagrangen funktiota L sanotaan* parametriseksi*, jos se toteuttaa ehdon* (4.2)*.*

Helposti voidaan tarkistaa

Lemma 4.1. *Jos L toteuttaa ehdon* (4.2)*, niin*

$$L = \langle \nabla_{x'} L, x' \rangle$$

Todistus. Harjoitustehtävä. ✧

Tarkastellaan sitten parametrista variaatiotehtävää

$$J(x) = \int_{t_0}^{t_1} L(x, x')dt \tag{4.3}$$

Koska ratkaisu ei riipu parametrisoinnista niin luonnollisesti ratkaisu ei voi olla yksikäsitteinen joten voisi olettaa, että Eulerin yhtälöt eivät ole riippumattomia.

Lemma 4.2. *Jos x on mielivaltainen käyrä niin*

$$\sum_{j=1}^{n} x_j' \mathcal{E}_L^j(x) = 0$$

Todistus. Suoraviivainen joskin hivenen työläs lasku osoittaa, että

$$\sum_{j=1}^{n} x_j' \mathcal{E}_L^j(x) = \frac{d}{dt}\big(\langle \nabla_{x'} L, x' \rangle - L\big)$$

Toisaalta edellisen lemman perusteella $\langle \nabla_{x'} L, x' \rangle - L = 0$. ✧

Siis yhtälöistä $\mathcal{E}_L^j(x) = 0$ saadaan vain $n-1$ riippumatonta yhtälöä. Parametriset tehtävät voidaan myös karakterisoida sisäisen variaation avulla.

Lemma 4.3. *Variaatiotehtävä on parametrinen jos ja vain jos $\partial J_x = 0$ kaikilla x.*

Todistus. Laskemalla kuten skalaaritapauksessa saadaan

$$\partial J_x \psi = \int_{t_0}^{t_1} \Big(L_t \psi + \big(L - \langle \nabla_{x'} L, x' \rangle\big)\psi' \Big) dt$$

Toisaalta parametriselle tehtävälle pätee $L_t = 0$ ja $L - \langle \nabla_{x'} L, x' \rangle = 0$. ✧

Joskus on kätevää analysoida tehtävää (4.3) seuraavan tehtävän avulla:

$$K(x) = \int_{t_0}^{t_1} N(x, x') dt = \tfrac{1}{2} \int_{t_0}^{t_1} \big(L(x, x')\big)^2 dt$$

Selvästi N ei ole parametrinen Lagrangen funktio: se on homogeeninen astetta kaksi. Kuitenkin K:n ja J:n ekstremaalit ovat "melkein samoja" seuraavassa mielessä.

Lemma 4.4.

(i) jos x on K:n ekstremaali ja $L(x, x') =$ vakio*, niin x on myös J:n ekstremaali.*

(ii) jos x on J:n ekstremaali ja $L(x, x') =$ vakio*, niin x on myös K:n ekstremaali.*

Todistus. Väitteet seuraavat seuraavasta kaavasta jonka tarkistaminen on jälleen suora mutta hiukan työläs lasku.

$$\mathcal{E}_N^j(x) = L\, \mathcal{E}_L^j(x) + L_{x_j'} \frac{d}{dt} L$$

✧

Esimerkki 4.4. Katsotaan sitten miten brakistokroniongelma voidaan ratkaista ilman arvauksia. Siinähän loppujen lopuksi päädyttiin parametriesitykseen joten tuntuisi luonnolliselta etsiä suoraan parametriratkaisua. Käytetään edellisen lemman ideaa ja kirjoitetaan

$$L = \frac{|x'|}{\sqrt{1 - x_2}} \quad , \quad N = \frac{|x'|^2}{2(1 - x_2)}$$

Tästä saadaan ensin

$$\mathcal{E}_N^1(x) = \frac{d}{dt} \frac{x_1'}{1 - x_2} = 0 \quad \Rightarrow \quad x_1' = c_0(1 - x_2)$$

Jos siis $N = c_1$ niin

$$(x_2')^2 = 2c_1(1-x_2) - c_0^2(1-x_2)^2$$

Edelleen

$$\begin{aligned}\mathcal{E}_N^2(x) &= \frac{d}{dt}\frac{x_2'}{1-x_2} - \frac{|x'|^2}{2(1-x_2)^2} = \frac{d}{dt}\frac{x_2'}{1-x_2} - \frac{c_1}{1-x_2} \\ &= \frac{(1-x_2)x_2'' + (x_2')^2 - c_1(1-x_2)}{(1-x_2)^2} = 0\end{aligned}$$

Saadaan siis lopulta

$$x_2'' + c_0^2 x_2 + c_1 - c_0^2 = 0$$

Tämähän voidaan helposti ratkaista:

$$x_2(t) = c_2\cos(c_0 t) + c_3\sin(c_0 t) + 1 - c_1/c_0^2$$

Otetaan uusi muuttuja $s = c_0 t$ ja merkitään $a = c_1/c_0^2$ jolloin

$$x_2(s) = c_2\cos(s) + c_3\sin(s) + 1 - a$$

Muistetaan sitten alkuehto $x_2(0) = 1$, joten voidaan valita $c_2 = a$ ja $c_3 = 0$. Sitten

$$x_1'(s) = x_1'(t)/c_0 = 1 - x_2(s)$$

Tästä integroimalla ja ottamalla huomioon ehto $x_1(0) = 0$ saadaan lopulta

$$\begin{aligned}x_1(s) =& a\big(s - \sin(s)\big) \\ x_2(s) =& 1 - a\big(1 - \cos(s)\big)\end{aligned}$$

"Loppuaika" s_1 ja a pitää laskea numeerisesti yhtälöistä $x_1(s_1) = 1$ ja $x_2(s_1) = 0$. ✩

Esimerkki 4.5. Parametrisissä tehtävissä tulee myös luontevasti vastaan rajoitusehdot. Olkoon $M \subset \mathbb{R}^3$ jokin pinta ja $x : [0,1] \to M$ jokin käyrä pisteitten $p = x(0)$ ja $q = x(1)$ välillä. Mikä on tässä lyhin polku? Käyrän pituus saadaan taas kaavalla

$$J(x) = \int_0^1 |x'(t)|\, dt$$

Koska käyrän pitää pysyä pinnalla M niin lyhin polku ei kuitenkaan ole p:n ja q:n välinen suora ja tämä pitäisi jollain tavalla ottaa huomioon Eulerin yhtälöissä. Tässä rajoitusehdon luonne on samanlainen kuin esimerkissä 4.3. Palataan tähänkin myöhemmin. ✩

4.4 Konveksisuus ja riittävät ehdot

Amazingly convex functionals do not play any role
in the classical calculus of variations before the turn of century,
although they seem to be predestined to assume a central place.
Only Minkowski recognized the notions of convex set and convex function
to be central concepts in mathematics although the concepts of
convex curve and convex surface were well-known and often used
in ancient times.

Giaquinta & Hildebrandt [5]

Kaikki edellä saadut välttämättömät ehdot ovat luonteeltaan lokaaleja: niiden avulla on mahdotonta sanoa mitään siitä onko löydetty ekstremaali globaali minimi tai maksimi. Tarvittaisiin siis jotain tietoa minimoitavan kuvauksen J globaaleista ominaisuuksista. Konveksisuus on eräs tällainen käsite joka on erityisen tärkeä kaikissa optimointiin liittyvissä tehtävissä ja sen avulla päästäänkin käsiksi riittäviin ehtoihin.

Tarkastellaan seuraavassa merkintöjen yksinkertaisuuden vuoksi vain ensimmäisen kertaluvun tapausta, mutta kaikki tulokset pätevät sellaisenaan yleisessä tapauksessa. Tarkastellaan siis tehtävää

$$J(y) = \int_a^b L(x, y, y')dx$$

Oletetaan, että L on konveksi (määritelmä C.2) seuraavassa mielessä:

$$L\big(x, (1-\mu)y_1 + \mu y_2, (1-\mu)z_1 + \mu z_2\big) \leq (1-\mu)L\big(x, y_1, z_1\big) + \mu L\big(x, y_2, z_2\big)$$

kaikilla x, y_j, z_j ja $0 \leq \mu \leq 1$. Tässä siis x on tämän konveksisuusehdon kannalta vain parametri. Mutta tästä saadaan heti

Lemma 4.5. *Jos L on konveksi, niin J ja $g(\varepsilon) = J(y + \varepsilon\varphi)$ ovat konvekseja.*

Todistus. J:n konveksisuuden tarkistaminen jätetään harjoitustehtäväksi.

$$\begin{aligned} g\big((1-\mu)\varepsilon_1 + \mu\varepsilon_2\big) &= J\big((1-\mu)(y + \varepsilon_1\varphi) + \mu(y + \varepsilon_2\varphi)\big) \\ &\leq (1-\mu)J(y + \varepsilon_1\varphi) + \mu J(y + \varepsilon_2\varphi) = (1-\mu)g(\varepsilon_1) + \mu g(\varepsilon_2) \end{aligned}$$

✧

Mutta tästä saadaan välittömästi

Lause 4.1. *Jos L on konveksi ja y on sitä vastaava ekstremaali, niin y antaa globaalin minimin.*

Todistus. Edellisen lemman perusteella g on konveksi joten lemman C.2 perusteella $g(\varepsilon) \geq g(0) + g'(0)\varepsilon$. Siis jos y on ekstremaali, niin

$$J(y + \varepsilon\varphi) \geq J(y) + \varepsilon\, \delta J_y\, \varphi = J(y)$$

✧

Tämän perusteella saadaan siis heti, että kahden pisteen välinen lyhin polku tosiaankin on suora koska helposti voidaan tarkistaa, että $L = \sqrt{1 + (y')^2}$ on konveksi. Samoin esimerkissä 4.1 splini tosiaankin minimoi annetun kuvauksen koska $L = \frac{1}{2}(y'')^2$ on konveksi.

Huomaa, että Weierstrassin esimerkissä 3.3 $L = \frac{1}{2}x^2(y')^2$ on konveksi mutta ratkaisua ei siitä huolimatta ole olemassa.

Valitettavasti Lagrangen funktiot eivät useinkaan ole konvekseja. Tämä on johtanut tiettyihin konveksisuuden yleistyksiin jotka kuitenkin ovat riittävän vahvoja takaamaan että ekstremaali todella antaa minimin. Näihin yleistyksiin ei kuitenkaan tällä kurssilla voida puuttua.

4.5 Harjoitustehtäviä

1. Tarkista lemman C.2 avulla, että $L = \sqrt{1 + (y')^2}$ on konveksi.

2. Täydennä lemman 4.5 todistus.

3. Osoita, että $L = y\sqrt{1 + (y')^2}$ ei ole konveksi.

Luku 5

Rajoitusehdot

5.1 Integraalirajoitteet

Esimerkissä 2.3 tarkasteltu Didon ongelma voidaan kirjoittaa

$$\begin{aligned} \text{max tai min} \quad : \quad J(y) &= \int_a^b L(x,y,y')dx \quad \text{kun} \\ K(y) &= \int_a^b N(x,y,y')dx = \ell \end{aligned} \tag{5.1}$$

Tämän lisäksi tarvitaan tietenkin jotkin sopivat reunaehdot. Variaation muodostamisessa on se ongelma, että jos määritellään $\alpha(\varepsilon) = y + \varepsilon\varphi$, niin yleisesti ottaen rajoitusehto ei toteudu kaikilla ε. Ideana on käyttää kahta varioivaa funktiota: olkoon $\varepsilon = (\varepsilon_1, \varepsilon_2)$ ja asetetaan

$$\alpha(\varepsilon) = y + \varepsilon_1\varphi_1 + \varepsilon_2\varphi_2$$

Olkoon sitten

$$\begin{aligned} f(\varepsilon) &= J\big(y + \varepsilon_1\varphi_1 + \varepsilon_2\varphi_2\big) \\ g(\varepsilon) &= K\big(y + \varepsilon_1\varphi_1 + \varepsilon_2\varphi_2\big) - \ell \end{aligned}$$

Kiinnitetään funktiot y, φ_1 ja φ_2 ja oletetaan, että y toteuttaa rajoitusehdon. Olkoon $M = g^{-1}(0) \subset \mathbb{R}^2$ on jokin käyrä $(\varepsilon_1, \varepsilon_2)$ tasossa. Koska y toteuttaa rajoitusehdon, niin $0 \in M$. Jos sitten f saa miniminsä origossa, ja $\nabla g(0) \neq 0$, niin Lemman C.5 mukaan on olemassa jokin λ siten, että

$$\nabla f(0) + \lambda \nabla g(0) = 0 \tag{5.2}$$

Lemma 5.1. *Jos y ei ole K:n ekstremaali, niin $\nabla g(0) \neq 0$.*

Todistus. Harjoitustehtävä. ✧

Oletetaan jatkossa aina, että haluttu ekstremaali ei ole rajoitusehdon ekstremaali. Tästä saadaankin sitten

Lause 5.1. *Jos y minimoi tai maksimoi tehtävän* (5.1), *niin*

$$\mathcal{E}_{\tilde{L}}(y) = 0$$

missä $\tilde{L} = L + \lambda N$. ❖

Todistus. Yhtälö (5.2) aukikirjoitettuna antaa

$$\begin{aligned} \delta J_y \varphi_1 + \lambda\, \delta K_y \varphi_1 &= 0 \\ \delta J_y \varphi_2 + \lambda\, \delta K_y \varphi_2 &= 0 \end{aligned}$$

Oletetaan, että $\delta K_y \varphi_2 \neq 0$ jolloin voidaan ratkaista $\lambda = -\delta J_y \varphi_2 / \delta K_y \varphi_2$. Huomaa, että λ ei riipu funktion φ_1 valinnasta. Toisin sanoen

$$\big(\delta J_y + \lambda\, \delta K_y\big)\varphi_1 = 0$$

Koska tämän pitää päteä olipa φ_1 valittu miten tahansa, niin tästä saadaan väite. ✧

Huomataan, että kaikissa kiinnostavissa tapauksissa $\lambda \neq 0$. Jos $\lambda = 0$, niin sehän tarkoittaisi, että y olisi rajoittamattoman tehtävän ekstremaali joka ”sattumalta” toteuttaa myös rajoitusehdon.

Määritelmä 5.1. $\tilde{L} = L + \lambda N$ *on tehtävän* (5.1) laajennettu Lagrangen funktio. ✰

Tämä on analoginen yhtälössä (C.1) määritellyn laajennetun kohdefunktion kanssa. Tehtävän ratkaisu etenee nyt lähes samoin kuten rajoittamattomassa tapauksessa:

(i) muodostetaan ensin laajennettu Lagrangen funktio ja ratkaistaan tätä vastaavat ekstremaalit.

(ii) otetaan huomioon asianmukaiset reunaehdot, ja mahdollisesti tarkistetaan voiko olla kulmaekstremaaleja.

(iii) valitaan λ siten, että rajoitusehto on voimassa

Esimerkki 5.1. Ratkaistaan edellisten tulosten avulla esimerkin 2.3 ongelma:

$$\begin{aligned} \max \quad : \quad J(y) &= \int_0^b y(x)dx \quad \text{kun} \\ K(y) &= \int_0^b \sqrt{1+(y')^2}dx = \ell > 0 \end{aligned}$$

Reunaehtona on siis $y(0) = y(b) = 0$, mutta b on vapaa. Nyt $\tilde{L} = y + \lambda\sqrt{1+(y')^2}$ joten

$$\mathcal{E}_{\tilde{L}}(y) = \frac{d}{dx}\tilde{L}_{y'} - \tilde{L}_y = \lambda \frac{d}{dx}\frac{y'}{\sqrt{1+(y')^2}} - 1 = 0$$

$$\frac{\lambda y'}{\sqrt{1+(y')^2}} = x - c_1 \quad \Rightarrow \quad (y')^2 = \frac{(x-c_1)^2}{\lambda^2 - (x-c_1)^2} \quad \Rightarrow$$

$$(x-c_1)^2 + (y-c_2)^2 = \lambda^2$$

Tehtävän ekstremaalit ovat siis ympyrän kaaria. Reunaehtojen perusteella saadaan ensin $c_1 = b/2$. Vapaan reunan ehdoksi tulee

$$\big(\tilde{L} - y'\tilde{L}_{y'}\big)(b) = y(b) + \frac{\lambda}{\sqrt{1+(y'(b))^2}} = \frac{\lambda}{\sqrt{1+(y'(b))^2}} = 0$$

Nyt näyttäisi siltä, ettei ratkaisua ole olemassa. Kuitenkin jos ympyrä leikkaa x- akselin kohtisuoraan, niin $\lim_{x\to b} y'(x) = -\infty$ joten tulkitaan vapaan reunan ehto kohtisuoruusehtona. Tällöin siis $c_2 = 0$ ja $\lambda = -b/2$. Lopuksi lasketaan vielä pinta-ala:

$$K(y) = \tfrac{1}{2}\pi b = \ell \quad \Rightarrow \quad b = \frac{2\ell}{\pi} \quad \Rightarrow \quad J(y) = \frac{\ell^2}{2\pi}$$

✰

Tarkastellaan sitten ongelman (5.1) duaalista tehtävää

$$\begin{aligned} \text{max tai min} \quad : \quad K(y) &= \int_a^b N(x,y,y')dx \quad \text{kun} \\ J(y) &= \int_a^b L(x,y,y')dx = \hat{\ell} \end{aligned}$$

Tämän laajennettu Lagrangen funktio on $\tilde{K} = K + \hat{\lambda}L$ jolla selvästi on samat ekstremaalit kuin $\tilde{L}$:llä. Didon ongelman tapauksessa on intuitiivisesti selvää, että kun siirrytään duaaliseen ongelmaan, niin minimointi muuttuu maksimoinniksi ja päinvastoin. Yleisesti ottaen tämä ei ehkä ole niin selvää, eikä se seuraa edellisistä tarkasteluista koska siinähän katsottiin vain ehtoja jotka ovat voimassa sekä minimoinnissa että maksimoinnissa. Tällainen duaalisuus on kuitenkin erittäin tärkeä periaate optimoinnissa.

Voitaisiin myös antaa useampi rajoitusehto:

$$K_j(y) = \int_a^b N_j(x,y,y')dx = \ell_j \quad , \quad j = 1,\dots,m$$

Edellä käsitelty yhden rajoitteen tapaus yleistyy helposti tähänkin tapaukseen, jonka tarkempi muotoilu jätetään harjoitustehtäväksi.

5.2 Yhtälörajoitteet

Tarkastellaan sitten esimerkkien 4.3 ja 4.5 tyyppisiä tapauksia. Näissä tehtävissä muuttuja on aina vektori. Olkoon $x \in \mathbb{R}^n$ ja $g : \mathbb{R}^n \to \mathbb{R}^k$ missä $k < n$. Oletetaan, että nolla on g:n säännöllinen arvo ja $M = \{x \in \mathbb{R}^n \,|\, g(x) = 0\}$ (määritelmä C.3). Tarkastellaan tehtävää

$$\begin{aligned} J(x) &= \int_{t_0}^{t_1} L(t,x,x')dx \\ x(t) &\in M \quad \text{kaikilla} \quad t \in I = [t_0,t_1] \end{aligned} \tag{5.3}$$

Tämäntyyppistä rajoitusehtoa sanotaan *holonomiseksi rajoitusehdoksi.* Tällä tarkoitetaan sitä, että rajoitukset liittyvät x:n arvoihin, mutta ei derivaattojen arvoihin.

Olkoon sitten $x : I \to M$ jokin käyrä ja sanotaan mitä tahansa käyrää joka toteuttaa rajoitusehdot *sallituksi käyräksi*. Ongelmana tässä on, että sallitun käyrän mielivaltainen variaatio ei yleensä ole sallittu käyrä. Jos katsotaan asiaa lokaalisti, tai "infinitesimaalisesti", niin on selvää, että sallittua käyrää voidaan varioida "sallitusti" vain tangenttiavaruuden suuntaan. Merkitään T_pM:llä moniston M tangenttiavaruutta pisteessä $p \in M$

Määritelmä 5.2. *Olkoon x sallittu käyrä. φ on käyrän x sallittu variaatio, jos*

$$\varphi(t) \in T_{x(t)}M \quad \text{kaikilla} \quad t \in I$$

x on tehtävän (5.3) *ekstremaali, jos $\delta J_x \varphi = 0$ kaikilla sallituilla variaatioilla φ.* ✰

Olkoon edelleen N_pM moniston M normaaliavaruus pisteessä p ja merkitään

$$\mathcal{E}_L(x) = \big(\mathcal{E}_L^1(x), \ldots, \mathcal{E}_L^n(x)\big)$$

Lemma 5.2. *x on tehtävän* (5.3) *ekstremaali, jos $\mathcal{E}_L(x) \in N_{x(t)}M$ kaikilla $t \in I$.*

Todistus. Jos muodostetaan δJ tavalliseen tapaan ja oletetaan, että alku- ja loppupiste on kiinnitetty niin saadaan

$$\delta J_x \varphi = \int_{t_0}^{t_1} \sum_{j=1}^{n} \Big(L_{x_j}\varphi_j + L_{x'_j}\varphi'_j\Big) dt = -\int_{t_0}^{t_1} \sum_{j=1}^{n} \mathcal{E}_L^j(x)\varphi_j dt = -\int_{t_0}^{t_1} \big\langle \mathcal{E}_L(x), \varphi \big\rangle dt$$

Jos $\delta J_x \varphi = 0$ kaikilla sallituilla φ, niin $\mathcal{E}_L(x)$ on kohtisuorassa tangenttiavaruutta vastaan. ✧

Siis jos x on ekstremaali niin Lemman C.5 mukaan on olemassa jokin käyrä $\lambda : I \to \mathbb{R}^k$ siten, että

$$\mathcal{E}_L(x) + (dg_{x(t)})^T \lambda = 0 \tag{5.4}$$

Tätä λ:a sanotaan edelleen Lagrangen kertojaksi.

Esimerkki 5.2. Jatketaan esimerkkiä 4.3. Siis $x : \mathbb{R} \to \mathbb{R}^2$ ja

$$L = \tfrac{1}{2}\, m|x'|^2 - m\mathsf{g}x_2$$

Rajoitusehto on $g(x) = |x|^2 - \ell^2 = 0$. Kun supistetaan m pois, niin systeemistä (5.4) saadaan siis yhtälöt

$$\begin{aligned} x_1'' + 2\lambda x_1 &= 0 \\ x_2'' + \mathsf{g} + 2\lambda x_2 &= 0 \\ x_1^2 + x_2^2 - \ell^2 &= 0 \end{aligned}$$

Ongelma on nyt saatu muotoiltua järkevästi, mutta sen ratkaiseminen on edelleen haasteellista, myös numeerisesti. Tähän on ainakin 2 syytä. Ensinnäkin on intuitiivisesti selvää, että muuttuja λ on jollain tavalla "eri asemassa" kuin x muuttujat. Tämä pitää ottaa huomioon numeerisessa ratkaisussa. Toisekseen tehtävä on siinä mielessä ylimääräytynyt, että ylläolevien 3 yhtälön lisäksi tarkka ratkaisu toteuttaa myös

$$\langle x, x' \rangle = 0 \quad \text{ja} \quad \langle x, x'' \rangle + |x'|^2 = 0$$

Vaikka on toisaalta selvää, että nämäkin yhtälöt ovat tärkeitä ratkaisun kannalta, niin ei ole selvää miten kaikkia 5 yhtälöä voisi käyttää numeerisessa laskennassa. ✯

Esimerkki 5.3. Palataan sitten esimerkkiin 4.5. Olkoon $M = g^{-1}(0) \subset \mathbb{R}^3$ jokin pinta ja $x : [0,1] \to M$ jokin käyrä pisteitten $p = x(0)$ ja $q = x(1)$ välillä ja $L = |x'|$. Koska tämä on parametrinen tehtävä niin olkoon $N = \frac{1}{2}|x'|^2$ ja $|x'| = 1$ jolloin ratkaisu on parametrisoitu kaarenpituuden avulla. Nyt yhtälöstä (5.4) saadaan

$$x'' + \lambda \nabla g = 0$$

Tässä siis $x''(t) \in N_{x(t)}M$ kaikilla t kun parametrisoidaan kaarenpituuden avulla. Tämän tehtävän ekstremaaleja sanotaan *geodeeseiksi*. Toisin sanoen pinnan M geodeesit ovat käyriä joille $x''(t) \in N_{x(t)}M$ ja jotka on parametrisoitu kaarenpituuden avulla.

Tätä yhtälöä voi numeerisesti ratkaista eliminoilla λ:n. Ottamalla sisätulo saadaan ensin

$$\lambda = -\frac{\langle \nabla g, x'' \rangle}{|\nabla g|^2}$$

Toisaalta derivoimalla yhtälöä $g = 0$ saadaan

$$\langle \nabla g, x' \rangle = 0 \quad \Rightarrow \quad \langle \nabla g, x'' \rangle + \langle x', d^2 g\, x' \rangle = 0$$

Siispä saadaan alkuarvotehtävä

$$\begin{aligned} & x'' + \frac{\langle x', d^2 g\, x' \rangle}{|\nabla g|^2} \nabla g = 0 \\ & x(0) = p \in M \quad , \quad x'(0) \in T_p M \end{aligned}$$

✯

Katsotaan vielä yhtälön (5.4) rakennetta. Olkoon $z = (x, \lambda) : I \to \mathbb{R}^{n+k}$ ja määritellään (vertaa (C.1))

$$\hat{L}(t, z, z') = L(t, x, x') - \langle \lambda, g(x) \rangle \tag{5.5}$$

Nyt suoraviivainen lasku osoittaa, että

$$\mathcal{E}_{\hat{L}}(z) = 0 \quad \longleftrightarrow \quad \begin{cases} \mathcal{E}_L(x) + (dg_{x(t)})^T \lambda = 0 \\ g(x) = 0 \end{cases}$$

Määritelmä 5.3. *Funktio $\hat{L}$ on tehtävän* (5.3) laajennettu Lagrangen funktio.

Siis laajennetun Lagrangen funktion tavanomaiset Eulerin yhtälöt antavat oikeat, rajoitusehdot huomioon ottavat yhtälöt.

5.3 Harjoitustehtäviä

1. Todista lemma 5.1.

2. Totea, että jos esimerkissä 5.1 valitaan $b = 2$ ja $\ell > \pi$, niin tätä vastaavaa ekstremaalia ei voi esittää x:n funktiona koko välillä $[0, 2]$. Voidaanko tätä kuitenkin pitää tehtävän mielekkäänä ratkaisuna?

Luku 6

Hamiltonin systeemi

6.1 Mekaniikka

On osoittautunut, että Eulerin yhtälöt kannattaa joskus kirjoittaa tietyllä tavalla ensimmäisen kertaluvun systeeminä. Tietenkin jokainen toisen kertaluvun systeemi voidaan palauttaa ensimmäisen kertaluvun systeemiksi vaikkapa määrittelemällä uusi muuttuja $y = x'$, mutta Eulerin yhtälöillä on sellainen rakenne, että tämä palautus on hyödyllistä tehdä toisella tavalla.

Tarkastellaan uudestaan esimerkkiä 4.2. Tällöin

$$L = T - U = \tfrac{1}{2}\, m|x'|^2 - U(x)$$

ja määriteltiin liikemäärä $y = mx'$ ja sen avulla vastaava Hamiltonin systeemi ja Hamiltonin funktio $H(x, y) = T + U$. Tässä siis uusi muuttuja oli "melkein" se standardi valinta joten ei ole helppo arvata miten yleisessä tapauksessa pitäisi toimia.

Olkoon $L : \mathbb{R}^{2n+1} \to \mathbb{R}$ jokin Lagrangen funktio ja valitaan

$$y = L_{x'}(t, x, x') = f(t, x, x') = f_{t,x}(x')$$

Tämä pitää ajatella niin, että t ja x ovat "vain" parametreja jolloin kiinteillä t ja x funktio $f_{t,x}$ on (lokaali) diffeomorfismi muuttujien x' ja y välillä. Käänteisfunktiolauseen perusteella $f_{t,x}$ on lokaali diffeomorfismi, jos

$$\det(d_{x'} f_{t,x}) = \det(L_{x'x'}) \neq 0$$

Jos tämä ehto on voimassa niin ainakin lokaalisti on olemassa käänteisfunktio $\hat{f} = f^{-1}$ ja voidaan kirjoittaa

$$x' = \hat{f}_{t,x}(y) = \hat{f}(t, x, y) \tag{6.1}$$

Äskeisessä esimerkissä siis $f_{t,x}(x') = mx'$, $\hat{f}_{t,x}(y) = y/m$ ja $d_{x'} f_{t,x} = m\, I$. Nyt pitäisi sitten kirjoittaa jotenkin ensimmäisen kertaluvun systeemi muuttujien x ja y avulla.

Määritelmä 6.1. *Lagrangen funktiota L vastaava Hamiltonin funktio on*

$$H = \langle x', L_{x'} \rangle - L$$

Nyt pitää ajatella, että H on x:n ja y:n funktio; siis tarkemmin

$$H(t,x,y) = \langle x', y\rangle - L(t,x,x') = \langle \hat{f}(t,x,y), y\rangle - L(t,x,\hat{f}(t,x,y))$$

Huomattakoon, että parametrisille tehtäville $\det(L_{x'x'}) = 0$ joten niitä vastaavan Hamiltonin funktion määrittely on hankalampaa eikä siihen puututa tällä kurssilla. Äskeisen esimerkin tapauksessa tämä siis antaa

$$H = T + U = \tfrac{1}{2m}\,|y|^2 + U(x)$$

Sitten lasketaan

$$\nabla_y H = \hat{f}(t,x,y) + (d_y\hat{f})^T\, y - (d_y\hat{f})^T\, L_{x'} = \hat{f}(t,x,y)$$

Vertaamalla yhtälöön (6.1) voidaan siis kirjoittaa $x' = \nabla_y H$. Sitten y:lle saadaan difyhtälö kun ensin huomataan, että

$$y' = \frac{d}{dt} L_{x'}(t,x,x') = L_x$$

koska x toteuttaa Eulerin yhtälön. Toisaalta

$$\nabla_x H = (d_x\hat{f})^T\, y - L_x - (d_x\hat{f})^T\, L_{x'} = -L_x$$

Edelleen voidaan taas määritellä $z = (x,y)$.

Määritelmä 6.2. *Olkoon L jokin Lagrangen funktio. Eulerin yhtälöitä $\mathcal{E}_L(x) = 0$ vastaava Hamiltonin systeemi on*

$$\begin{cases} x' = \nabla_y H \\ y' = -\nabla_x H \end{cases} \qquad \longleftrightarrow \qquad z' = \mathcal{I}\,\nabla H \quad \text{missä} \quad \mathcal{I} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

Vektorikenttä $X_H = \mathcal{I}\,\nabla H$ on funktiota H vastaava Hamiltonin vektorikenttä.

Kun on päästy näin pitkälle niin voidaan myös ”unohtaa” L ja määritellä tällainen difyhtälösysteemi mielivaltaiselle funktiolle H. Tässäkin tapauksessa puhutaan Hamiltonin funktiosta, Hamiltonin systeemistä ja Hamiltonin vektorikentästä.

Difyhtälöitten analyysissä on erittäin hyödyllistä jos voidaan löytää *invariantteja*, siis funktioita jotka ovat vakioita ratkaisua pitkin. Hamiltonin funktio itse on invariantti jos H ei suoraan riipu ajasta.

Lemma 6.1. *Olkoon $H : \mathbb{R}^{2n} \to \mathbb{R}$ mielivaltainen ja z vastaavan Hamiltonin systeemin ratkaisu. Tällöin*

$$\frac{d}{dt} H(z(t)) = 0$$

Todistus. Harjoitustehtävä. ✧

Miten sitten invariantteja voisi löytää? Luonnollisesti kaikilla difyhtälöillä ei ole invariantteja, mutta koska Hamiltonin systeemit ovat aivan erityistä muotoa, niin niille voidaan ainakin jossain määrin etsiä konstruktiivisesti uusia invariantteja. Ensin tarvitaan

Määritelmä 6.3. *Olkoon* f, $g \in C^\infty(\mathbb{R}^{2n})$. Poissonin kuvaus[1] $C^\infty(\mathbb{R}^{2n}) \times C^\infty(\mathbb{R}^{2n}) \to C^\infty(\mathbb{R}^{2n})$ *on*

$$[f,g] = \sum_{j=1}^{n} f_{x_j} g_{y_j} - f_{y_j} g_{x_j}$$

Helposti voidaan tarkistaa, että Poissonin kuvauksella on seuraavat ominaisuudet:

(i) kuvaus on bilineaarinen

(ii) $[f,g] = \langle \nabla f, \mathcal{I} \nabla g \rangle = \langle X_f, \mathcal{I} X_g \rangle$

(iii) $[f,g] = -[g,f]$

(iv) $[f_1 f_2, g] = f_1[f_2, g] + f_2[f_1, g]$

(v) $[f,[g,h]] + [h,[f,g]] + [g,[h,f]] = 0$ (Jacobin identiteetti)

Avaruus $C^\infty(\mathbb{R}^{2n})$ varustettuna Poissonin kuvauksella on *Lie-algebra* (määritelmä B.8). Tämän avulla saadaankin

Lemma 6.2. *(i) Hamiltonin systeemi voidaan kirjoittaa muodossa*

$$z_j' = [z_j, H] \quad , \quad 1 \le j \le 2n$$

(ii) f on Hamiltonin systeemin invariantti, jos $[f,H] = 0$.
(iii) Jos f ja g ovat invariantteja, niin myös $[f,g]$ on invariantti.

Todistus. (i) Harjoitustehtävä.

(ii) Suora derivointi osoittaa, että $\frac{d}{dt} f(z(t)) = [f,H]$.

(iii) Jacobin identiteetin ja kohdan (ii) perusteella

$$[H,[f,g]] = -[g,[H,f]] - [f,[g,H] = 0$$

✧

Siis jos on löydetty itse Hamiltonin funktion lisäksi 2 muuta invarianttia, niin näitten avulla voi sitten periaatteessa tuottaa uusia invariantteja Poissonin kuvauksen avulla. Valitettavasti kuitenkaan mikään ei takaa, että tällä prosessilla löytyisi uusia riippumattomia invariantteja.

[1] Englanniksi *Poisson bracket*.

Esimerkki 6.1. Olkoon x pistemassa $\mathbb{R}^3$:ssa ja oletetaan, että se liikkuu voimakentässä joka osoittaa kohti origoa. Siis jos ajatellaan, että origossa on pistemassa, niin sen aiheuttama potentiaali on $U(x) = -\alpha/|x|$ missä $\alpha > 0$. Saadaan siis Hamiltonin systeemi

$$\begin{aligned} x' =& \nabla_y H = y \\ y' =& -\nabla_x H = -\tfrac{\alpha}{|x|^3}\, x \end{aligned}$$

Siis derivoimalla

$$\frac{d}{dt}\, x \times y = x' \times y + x \times y' = y \times y - \tfrac{\alpha}{|x|^3}\, x \times x = 0$$

Siis kulmaliikemäärä $K = x \times y$ on vakio. Koska K on vektori, niin sen jokainen komponentti K_j on invariantti. Nyt voidaan tarkistaa, että $K_3 = [K_1, K_2]$. ✫

Esimerkki 6.2. Tarkastellaan edellisen esimerkin tilannetta, mutta oletetaan, että massapiste liikkuu tasossa $x_3 = 0$. Siirrytään napakoordinaatteihin: $x_1 = r\cos(\theta)$, $x_2 = r\sin(\theta)$. Tällöin kulmaliikemäärän kolmannesta komponentista saadaan

$$K_3 = x_1 x_2' - x_1' x_2 = r^2\theta' = c$$

Tästä saadaan Keplerin pinta-alalaki: planeetasta aurinkoon piirretty jana pyyhkäisee tietyssä ajassa aina saman pinta-alan. Olkoon t_0 alkuhetki ja t_1 loppuhetki ja $\theta_j = \theta(t_j)$. Tällöin pyyhkäisty ala on

$$A(\theta_0, \theta_1) = \int dxdy = \int_{\theta_0}^{\theta_1} \int_0^{r(\theta)} r dr d\theta = \tfrac{1}{2} \int_{\theta_0}^{\theta_1} r^2 d\theta = \tfrac{1}{2} \int_{t_0}^{t_1} r^2\theta' dt = \tfrac{c}{2}(t_1 - t_0)$$

Tämän avulla saadaan myös Keplerin kolmas laki. Oletetaan, että rata on jaksollinen ja olkoon τ jakso. Tällöin $A = A(0, 2\pi)$ on radan sisään jäävän alueen pinta-ala ja siispä

$$\tau = \frac{2A}{c}$$

Pinta-ala A saadaan itse asiassa eksplisiittisesti laskettua, kun huomataan tehdä seuraava muuttujan vaihto: $x_1 = \cos(\theta)/u$, $x_2 = \sin(\theta)/u$. Lagrangen funktio ja vastaava Hamiltonin funktio ovat tällöin

$$\begin{aligned} L =& \frac{1}{2u^4}\big((u')^2 + u^2(\theta')^2\big) + \alpha u \\ H =& \frac{1}{2}\big(u^4 y_1^2 + u^2 y_2^2\big) - \alpha u \end{aligned}$$

Hamiltonin systeemi on siis

$$\begin{cases} u' = u^4 y_1 \\ \theta' = u^2 y_2 \\ y_1' = -2u^3 y_1^2 - u y_2^2 + \alpha \\ y_2' = 0 \end{cases} \qquad \Rightarrow \qquad \begin{cases} u' = u^4 y_1 \\ \theta' = c\, u^2 \\ y_1' = -2u^3 y_1^2 - c^2 u + \alpha \\ y_2 = c \end{cases}$$

Ajatellaan sitten u:ta θ:n funktiona. Tällöin

$$\frac{du}{d\theta} = \frac{du/dt}{d\theta/dt} = \frac{u^2 y_1}{c}$$

Edelleen

$$\frac{d^2u}{d\theta^2} = \frac{2u^3y_1^2}{c^2} + \frac{u^2 dy_1/d\theta}{c} = -u + \frac{\alpha}{c^2}$$

Tämähän ratkeaa helposti ja ratkaisuksi saadaan

$$u(\theta) = \alpha\beta\cos(\theta + \theta_0) + \frac{\alpha}{c^2}$$

missä $\beta > 0$. Voidaan valita $\theta_0 = 0$ ja tälle ratkaisulle siis

$$r(\theta) = \frac{1}{u(\theta)} = \frac{c^2/\alpha}{1 + c^2\beta\cos(\theta)}$$

Nyt r pysyy rajoitettuna jos $c^2\beta < 1$ ja alkeisgeometriasta tiedetään, että tällöin käyräksi tulee ellipsi jonka eksentrisyys on $\varepsilon = c^2\beta$. Ratkaisun kokonaisenergia on

$$H = \frac{c^2}{2}\big((du/d\theta)^2 + u^2\big) - \alpha u = -\frac{\alpha^2(1-\varepsilon^2)}{2c^2}$$

Siis jos kokonaisenergia on negatiivinen, niin saadaan ellipsi. Pinta-alaksi saadaan

$$A = \tfrac{1}{2}\int_0^{2\pi} r^2 d\theta = \frac{\pi c^2}{\alpha^2(1-\varepsilon^2)^{3/2}}$$

Alkeisgeometriasta tiedetään, että ellipsin isompi puoliakseli on $a = c^2/\alpha(1-\varepsilon^2)$ joten lopulta saadaan

$$\tau^2 = \frac{4\pi^2 a^3}{\alpha}$$

Tämä on Keplerin kolmas laki. Huomaa, ettei tämä riipu kulmaliikemäärästä c vaan ainoastaan vakiosta α joka riippuu edelleen kappaleitten massoista ja gravitaatiovakiosta. ✮

Katsotaan vielä eräs Hamiltonin systeemien ominaisuus. Tässä tarvitaan seuraavaa.

Määritelmä 6.4. *Olkoon $x' = f(x)$ difyhtälösysteemi jossa f on sileä vektorikenttä. Tätä vastaava* virtaus *on kuvaus Φ jolle pätee*

(i) $\Phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ *ja merkitään* $\Phi^t(x) = \Phi(x,t)$.

(ii) jos x on systeemin $x' = f(x)$ ratkaisu ja $x(0) = p$ niin $x(t) = \Phi(p,t) = \Phi^t(p)$.

✩

Voidaan osoittaa, että Φ on olemassa, ainakin lokaalisti, ja se on sileä jos f on sileä. Edelleen

(i) $\Phi^0(p) = p$ ja

(ii) $\Phi^{s+t} = \Phi^s \circ \Phi^t$

joten kuvaukset Φ^t muodostavat yksiparametrisen ryhmän. Olkoon sitten $\Omega \subset \mathbb{R}^n$ ja $\Omega_t = \Phi^t(\Omega)$. Tällöin tunnetusti

$$\mathsf{vol}(\Omega_t) = \int_{\Omega_t} dx = \int_\Omega \det\big(d\Phi^t\big) dx$$

Siis virtaus säilyttää tilavuuden jos $\det\big(d\Phi^t\big) = 1$. Koska virtausta ei kuitenkaan eksplisiittisesti tunneta niin haluttaisiin jokin kriteeri tälle vektorikentän f avulla joka tunnetaan.

Lemma 6.3. *Virtaus säilyttää tilavuuden jos ja vain jos* $\mathsf{div}(f) = 0$.

Tämän perusteella saadaan heti

Seuraus 6.1. *Hamiltonin systeemin virtaus säilyttää tilavuuden.*

Todistus. Jos $z' = X_H(z) = \mathcal{I}\nabla H$ niin on helppo suoraan tarkistaa, että $\mathsf{div}(X_H) = 0$. ✧

Esimerkki 6.3. Lineaarisen systeemin tapauksessa virtaus tunnetaan. Jos $x' = f(x) = Ax$, niin tunnetusti $\Phi^t = \exp(At)$. Nyt voidaan helposti tarkistaa, että $\mathsf{div}(f) = 0$, jos $\mathsf{tr}(A) = 0$. ✯

Todistus. (Lemma 6.3) Olkoon $v(t) = \mathsf{vol}(\Omega_t)$ ja $h(x,t) = h_x(t) = \det\big(d\Phi^t\big)$. Jos v on vakio niin

$$v'(t) = \int_\Omega h_x'(t) dx = 0$$

Virtauksen määritelmän perusteella $\frac{d}{dt}\Phi^t(x) = f\big(\Phi^t(x)\big)$ joten

$$\frac{d}{dt} d\Phi^t(x) = df\big(\Phi^t(x)\big) d\Phi^t(x)$$

Merkitään $A(t) = d\Phi^t(x)$, jolloin voidaan kirjoittaa

$$A' A^{-1} = df\big(\Phi^t(x)\big)$$

Nyt tiedetään [7] että[2]

$$h_x'(t) = \mathsf{tr}\big(A' A^{-1}\big) h_x(t)$$

Toisaalta suoran laskun perusteella $\mathsf{tr}\big(df\big) = \mathsf{div}(f)$. Siispä

$$v'(t) = \int_\Omega \mathsf{div}(f) h_x(t) dx$$

Tästä väite seuraa. ✧

[2] Tätä tulosta on tapana sanoa *Jacobin kaavaksi*.

6.2 Optimisäätö

Optimisäädössä päädytään mutkan kautta Hamiltonin systeemeihin. Tässä ajatellaan, että on kahdenlaisia muuttuujia: *tilamuuttujat* x ja *säätömuuttujat* u. Nämä toteuttavat systeemin

$$x' = f(x, u)$$

ja tavoitteena on valita säätö u siten, että näin saatu x on jossain mielessä optimaalinen. Tarkastellaan vaikkapa seuraavaa kohdefunktiota:

$$J(u) = F(x(\bar{t})) + \int_0^{\bar{t}} L(x, u)dt$$

Oletetaan siis, että f, F ja L eivät riipu eksplisiittisesti ajasta. Oletetaan, että $x(0) = p$ annettu, jolloin voidaan tavalliseen tapaan siirtyä perusmuotoon

$$\tilde{J}(u) = \int_0^{\bar{t}} \tilde{L}(x, u)dt = \int_0^{\bar{t}} \big(\langle \nabla F, x' \rangle + L(x, u)\big)dt$$

Ajatellaan yhtälöä $x' = f(x, u)$ rajoitusehtona ja yritetään ottaa se huomioon Lagrangen kertoimien avulla kuten yhtälössä (5.5). Määritellään siis

$$\hat{J}(u) = \int_0^{\bar{t}} \hat{L}(x, u)dt = \int_0^{\bar{t}} \big(\tilde{L}(x, u) + \langle \lambda, x' - f(x, u)\rangle\big)dt$$

Määritellään vielä Hamiltonin funktio $H(x, u, \lambda) = \langle \lambda, f(x, u)\rangle - L(x, u)$ jolloin siis voidaan kirjoittaa

$$\hat{J}(u) = \int_0^{\bar{t}} \big(\langle \nabla F + \lambda, x' \rangle - H\big)dt$$

Oletetaan, että u minimoi $\hat{J}$:n ja olkoon x tätä vastaava optimaalinen tila, siis $x'(t) = f(x(t), u(t))$. Olkoon edelleen $g(\varepsilon) = \hat{J}(u + \varepsilon\varphi)$. Kuten ennenkin g:llä on tällöin minimi origossa ja tavoitteena on laskea $g'(0)$. Nyt täytyy ottaa myös huomioon u:n varioinnista aiheutuvat muutokset x:ssä. Difyhtälöitten teoriasta tiedetään, että u:n variointi muuttaa x:ä vain "vähän" ja voidaan kirjoittaa

$$x_\varepsilon = x + \varepsilon\eta + \mathcal{O}(\varepsilon^2)$$

Tässä siis η on jokin funktio joka riippuu φ:n valinnasta. Lisäksi $\eta(0) = 0$ koska alkutila on annettu. Näillä merkinnöillä saadaan

$$\begin{aligned} g(\varepsilon) &= \int_0^{\bar{t}} \big(\langle \nabla F, x'_\varepsilon \rangle + \langle \lambda, x'_\varepsilon \rangle - H(x_\varepsilon, u + \varepsilon\varphi, \lambda)\big)dt \\ &= \Big|_0^{\bar{t}} F(x_\varepsilon) + \langle \lambda, x_\varepsilon \rangle - \int_0^{\bar{t}} \big(\langle \lambda', x_\varepsilon \rangle + H(x_\varepsilon, u + \varepsilon\varphi, \lambda)\big)dt \end{aligned}$$

Edelleen derivoimalla saadaan

$$g'(0) = \delta\hat{J}_u\varphi = \big\langle \nabla F(x(\bar{t})) + \lambda(\bar{t}), \eta(\bar{t}) \big\rangle - \int_0^{\bar{t}} \big(\langle \lambda' + \nabla_x H, \eta \rangle + \langle \nabla_u H, \varphi \rangle\big)dt$$

Tämän pitäisi olla nolla kaikilla φ. Toisaalta η ei siis ole riippumaton muuttuja, mutta toisaalta voidaan ajatella, että se on varsin mielivaltainen kun φ:tä varioidaan. Nyt kuitenkin voidaan käyttää hyväksi sitä, että λ:t voidaan valita vapaasti. Vaaditaan siis, että

$$\begin{cases} \lambda' = -\nabla_x H \\ \lambda(\bar{t}) = -\nabla F(x(\bar{t})) \end{cases}$$

Tällä valinnalla

$$\delta \hat{J}_u \varphi = -\int_0^{\bar{t}} \langle \nabla_u H, \varphi \rangle dt$$

Saatiin siis

Lause 6.1. *Välttämätön ehto optimille on, että (x, u, λ) toteuttaa seuraavat yhtälöt:*

$$\begin{cases} x' = \nabla_\lambda H \\ \lambda' = -\nabla_x H \\ \nabla_u H = 0 \\ x(0) = p \\ \lambda(\bar{t}) = -\nabla F(x(\bar{t})) \end{cases}$$

❖

Huomaa, että vaikka t on luonteeltaan aikamuuttuja, niin saatu systeemi ei ole alkuarvotehtävä loppuehdon $\lambda(\bar{t}) = -\nabla F(x(\bar{t}))$ takia. Lisäksi yhtälöt $\nabla_u H = 0$ ovat algebrallisia yhtälöitä. Pitäisi siis tästä ensin ratkaista u minkä jälkeen saataisiin difyhtälösysteemi x:lle ja λ:lle. Kun u on eliminoitu, niin saadaan Hamiltonin systeemi muuttujille $z = (x, \lambda)$.

Lemma 6.4. *Hamiltonin funktio on vakio ratkaisua pitkin.*

Todistus. Kun u on eliminoitu, niin H on vakio lemman 6.1 perusteella. Jos ajatellaan, että u:ta ei ole eliminoitu, niin suoraan derivoimalla

$$\begin{aligned} \frac{d}{dt} H =& \langle \nabla_x H, x' \rangle + \langle \nabla_u H, u' \rangle + \langle \nabla_\lambda H, \lambda' \rangle \\ =& \langle \nabla_x H, \nabla_\lambda H \rangle + \langle \nabla_u H, u' \rangle - \langle \nabla_\lambda H, \nabla_x H \rangle = 0 \end{aligned}$$

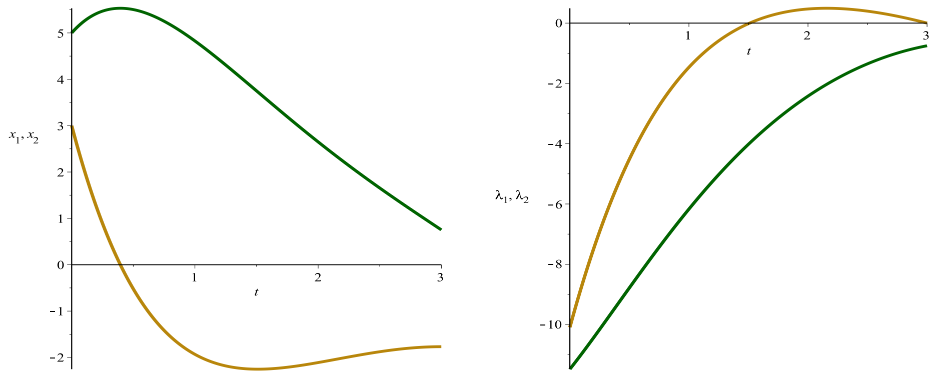
✧

Toisin sanoen ehtoon $\nabla_u H = 0$ päädytään myös sitä kautta, että vaaditaan, että H on vakio ratkaisua pitkin.

Esimerkki 6.4. Tarkastellaan tehtävää

$$J(u) = \tfrac{1}{2} x_1^2 + \tfrac{1}{2} \int_0^{\bar{t}} \left(|x|^2 + u^2 \right) dt$$

$$\begin{cases} x_1' = x_2 \\ x_2' = u \end{cases}$$



Kuva 6.1: Esimerkin 6.4 ratkaisut. Vasemmalla x_1 ja x_2, oikealla λ_1 ja λ_2.

Nyt siis $H = \lambda_1 x_2 + \lambda_2 u - \frac{1}{2}\big(|x|^2 + u^2\big)$ joten saadaan systeemi

$$\begin{cases} x_1' = x_2 \\ x_2' = u \\ \lambda_1' = x_1 \\ \lambda_2' = x_2 - \lambda_1 \\ \lambda_2 - u = 0 \\ x(0) = p \\ \lambda_1(\bar{t}) = -x_1(\bar{t}) \\ \lambda_2(\bar{t}) = 0 \end{cases}$$

Siis $u = \lambda_2$ jolloin $H = \lambda_1 x_2 + \frac{1}{2}\big(\lambda_2^2 - |x|^2\big)$. Kun u on eliminoitu, niin kyseessä on tosiaankin Hamiltonin systeemi määritelmän 6.2 mielessä. Kuvassa 6.1 on ratkaisut tapauksessa $\bar{t} = 3$ ja $p = (5, 3)$. ✯

Äskeinen esimerkki on paljon tutkittua tyyppiä jossa difyhtälöt on lineaarisia ja kohdefunktio on neliöllinen: tämä tunnetaan nimellä *LQR-tehtävä*, siis *linear quadratic regulator*. Nämä voidaan kirjoittaa muodossa

$$\begin{aligned} J(u) =& \tfrac{1}{2}\langle x, Sx\rangle + \tfrac{1}{2}\int_0^{\bar{t}} \big(\langle x, Qx\rangle + \langle u, Ru\rangle\big) dt \\ x' =& Ax + Bu \end{aligned}$$

Tässä siis S, Q ja R ovat symmetrisiä, ja lisäksi S ja Q ovat positiivisemidefiniittejä ja R on positiividefiniitti. Nyt $H = \langle \lambda, Ax + Bu\rangle - \frac{1}{2}\big(\langle x, Qx\rangle + \langle u, Ru\rangle\big)$ joten

$$\nabla_u H = B^T \lambda - Ru = 0$$

Koska R on positiividefiniitti, niin R^{-1} on olemassa joten $u = R^{-1}B^T\lambda$. Siis saadaan systeemi

$$\begin{cases} x' = Ax + BR^{-1}B^T\lambda \\ \lambda' = Qx - A^T\lambda \\ x(0) = p \\ \lambda(\bar{t}) = -Sx(\bar{t}) \end{cases}$$

Luonnollisesti muitakin reunaehtoja voi esiintyä. Esimerkiksi voitaisiin vaatia, että $x(\bar{t}) \in M \subset \mathbb{R}^n$. Tässä jouduttaisiin samoihin tarkasteluihin kuin transversaalisuusehtojen yhteydessä. Ääritapauksessa jossa M on yksi piste, niin tämähän antaa suoraan reunaehdon hetkellä $\bar{t}$ ja tällöin λ:lle ei enää tarvitakaan reunaehtoja.

Jos merkitään $z = (x, \lambda)$ niin äsken saatu difyhtälö voidaan kirjoittaa muodossa

$$z' = \mathcal{H}z \quad \text{missä} \quad \mathcal{H} = \begin{pmatrix} A & BR^{-1}B^T \\ Q & -A^T \end{pmatrix}$$

Nyt voidaan varsin helposti osoittaa:

Lemma 6.5.

(i) $\mathcal{H}^T\mathcal{I} + \mathcal{I}\mathcal{H} = 0$

(ii) jos μ on $\mathcal{H}$:n ominaisarvo, niin myös $-\mu$ on $\mathcal{H}$:n ominaisarvo.

(iii) olkoon $\mathcal{S} = e^{\mathcal{H}}$. Tällöin $\mathcal{S}^T\mathcal{I}\mathcal{S} = \mathcal{I}$.

Todistus. Harjoitustehtävä. ✧

Määritelmä 6.5. *Matriisi M on* Hamiltonin matriisi*, jos $M^T\mathcal{I} + \mathcal{I}M = 0$. M on* symplektinen*, jos $M^T\mathcal{I}M = \mathcal{I}$.* ✰

Joissain säätötehtävissä on luontevaa, että loppuaika ei ole kiinnitetty. Analysoidaan vielä tämä tilanne. Olkoon $y = (x, u, \lambda)$ jolloin voidaan kirjoittaa

$$\hat{J}(u) = \int_0^{\bar{t}} \hat{L}(y, y')dt = \int_0^{\bar{t}} \big(\langle \nabla F + \lambda, x' \rangle - H\big)dt$$

Lemma 6.6. *Vapaan loppuajan tapauksessa ratkaisun (x, u, λ) pitää toteuttaa ehto $H\big(x(\bar{t}), \lambda(\bar{t})\big) = 0$.*

Todistus. Olkoon taas $h_\varepsilon(x) = x + \varepsilon\psi$ perhe diffeomorfismeja, missä $\psi(0) = 0$ koska alkuaikaa ei varioida. Laskemalla sisäinen variaatio kuten määritelmässä 3.1 saadaan

$$\partial \hat{J}_y \psi = \int_0^{\bar{t}} \big(\hat{L}_t\psi + (\hat{L} - \langle y', \hat{L}_{y'} \rangle)\psi'\big)dt$$

Koska $\hat{L}$ ei riipu eksplisiittisesti t:stä niin $\hat{L}_t = 0$. Toisaalta selvästi $\hat{L} - \langle y', \hat{L}_{y'} \rangle = -H$ joten

$$\partial \hat{J}_y \psi = -\int_0^{\bar{t}} H\psi' dt = -\Big|_0^{\bar{t}} H\,\psi + \int_0^{\bar{t}} \psi \frac{d}{dt} H\,dt = -H\,\psi(\bar{t})$$

Koska $\psi(\bar{t})$ on mielivaltainen niin tästä saadaan väite. ✧

Esimerkki 6.5. Tarkastellaan tehtävää

$$J(u) = \frac{1}{2}\int_0^{\bar{t}} \left(1 + u^2\right) dt$$

$$\begin{cases} x_1' = x_2 \\ x_2' = u \end{cases}$$

Valitaan reunaehdoiksi $x(0) = (1, 0)$ ja $x(\bar{t}) = (0, 0)$. Tämä voidaan tulkita niin, että pistemassa on levossa pisteessä $x_1 = 1$ ja se halutaan siirtää origoon. Tämä halutaan tehdä mahdollisimman nopeasti eli $\int_0^{\bar{t}} dt = \bar{t}$ on pieni, mutta toisaalta ”taloudellisesti” eli $\int_0^{\bar{t}} u^2 dt$ on pieni.

Nyt siis $H = \lambda_1 x_2 + \lambda_2 u - \frac{1}{2}\left(1 + u^2\right)$ joten $H_u = \lambda_2 - u = 0$. Siis $H(x, \lambda) = \lambda_1 x_2 + \frac{1}{2}\left(\lambda_2^2 - 1\right)$. Vapaan loppuajan ehto on siis $H = \frac{1}{2}\left(\lambda_2(\bar{t})^2 - 1\right) = 0$ joten $\lambda_2(\bar{t}) = \pm 1$. Systeemistä saadaan

$$\begin{cases} x_1' = x_2 \\ x_2' = \lambda_2 \\ \lambda_1' = 0 \\ \lambda_2' = -\lambda_1 \\ x(0) = (1, 0) \\ x(\bar{t}) = (0, 0) \end{cases} \quad \Rightarrow \quad \begin{cases} x_1 = 2t^3/\bar{t}^3 - 3t^2/\bar{t}^2 + 1 \\ x_2 = 6t^2/\bar{t}^3 - 6t/\bar{t}^2 \\ \lambda_1 = -12/\bar{t}^3 \\ \lambda_2 = 12t/\bar{t}^3 - 6/\bar{t}^2 \end{cases}$$

Nyt $\lambda_2(\bar{t}) = 6/\bar{t}^2 = \pm 1$ joten $\bar{t} = \sqrt{6}$. Optimisäätö on siis $u = \lambda_2 = -1 + \sqrt{2/3}t$. ✪

Liite A

Differentiaalilaskentaa

Perusmerkinnät

Oletetaan, että tarkasteltavat funktiot ovat sen verran säännöllisiä että kaikki esiintyvät (osittais)derivaatat ovat jatkuvia. Seuraavassa Ω on jokin sopiva $\mathbb{R}^n$:n osajoukko.

Määritelmä A.1. *Olkoon* $u : \Omega \to \mathbb{R}$.

(i) u*:n* gradientti *tai* (ensimmäinen) differentiaali *on*

$$du = \mathrm{grad}(u) = \nabla u = \big(\partial u/\partial x_1, \ldots, \partial u/\partial x_n\big)$$

(ii) u*:n* toinen differentiaali *(tai Hessen matriisi) on*

$$d^2u = \begin{pmatrix} \partial^2 u/\partial x_1^2 & \ldots & \partial^2 u/\partial x_1 \partial x_n \\ \vdots & \ddots & \vdots \\ \partial^2 u/\partial x_1 \partial x_n & \ldots & \partial u/\partial x_n^2 \end{pmatrix}$$

(iii) Olkoon $f : \Omega \to \mathbb{R}^k$*. Tällöin* f*:n* Jacobin matriisi *eli* (ensimmäinen) differentiaali *on*

$$df = \begin{pmatrix} \partial f_1/\partial x_1 & \ldots & \partial f_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial f_k/\partial x_1 & \ldots & \partial f_n/\partial x_n \end{pmatrix}$$

Jos lasketaan df:n tai d^2u:n arvo pisteessä p, niin merkitään df_p tai d^2u_p. Huomaa, että d^2u_p on symmetrinen matriisi.

Esitetään alla vain yksinkertainen versio Taylorin lauseesta, joka riittää tässä yhteydessä. Olkoon $g : \mathbb{R}^n \to \mathbb{R}$ jokin kuvaus. Merkintä

$$g(x) = \mathcal{O}(|x|^q)$$

tarkoittaa, että on olemassa jokin $c > 0$ ja $\varepsilon > 0$ siten, että

$$|g(x)| \leq c\,|x|^q \quad \text{kun} \quad |x| < \varepsilon$$

Lause A.1. *Olkoon $u : \Omega \to \mathbb{R}$ ja $p \in \Omega$. Tällöin*

$$u(p+h) = u(p) + \langle \nabla u(p), h \rangle + \tfrac{1}{2} \langle h, d_p^2 u\, h \rangle + \mathcal{O}(|h|^3)$$

Olkoon $f : \Omega \to \mathbb{R}^n$ missä $\Omega \subset \mathbb{R}^n$.

Määritelmä A.2. *f on* Lipschitz-jatkuva *(tai toteuttaa* Lipschitz-ehdon*) alueessa Ω, jos on olemassa vakio L siten, että*

$$|f(x) - f(y)| \leq L|x-y| \quad , \quad \text{kaikilla} \;\; x, y \in \Omega$$

Lemma A.1. *Olkoon Ω konveksi ja kompakti, ja olkoon $f \in C^1(\Omega, \mathbb{R}^n)$. Tällöin f on Lipschitz-jatkuva ja vakioksi L voidaan valita*

$$L = \max_{p \in \Omega} \|df_p\|$$

Todistus. Koska Ω on kompakti, niin näin määritelty $L < \infty$. Olkoon x, $y \in \Omega$ ja määritellään $\alpha : [0,1] \to \Omega$ kaavalla $\alpha(s) = y + s(x-y)$. Koska Ω on konveksi, $\alpha(s) \in \Omega$. Olkoon edelleen $g(s) = \big(f \circ \alpha\big)(s)$. Tällöin

$$\begin{aligned} |f(x) - f(y)| =& |g(1) - g(0)| = \Big| \int_0^1 g'(s) ds \Big| = \Big| \int_0^1 df(x-y) ds \Big| \\ &\leq \int_0^1 |df(x-y)| ds \leq \int_0^1 \|df\| \, |x-y| ds \leq L|x-y| \end{aligned}$$

✧

Lipschitz-ehto on heikompi vaatimus kuin derivoituvuus: esimerkiksi $f(x) = |x|$ toteuttaa Lipschitz-ehdon välillä $[-1,1]$, mutta ei ole jatkuvasti derivoituva.

Implisiittifunktiolause

Olkoon

$$F : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^n \quad , \quad (x,y) \to F(x,y)$$

Tarkastellaan F:n osittaisderivaattoja pelkästään muuttujan y suhteen, ja merkitään

$$d_y F = \begin{pmatrix} \partial F_1/\partial y_1 & \dots & \partial F_1/\partial y_n \\ \vdots & \ddots & \vdots \\ \partial F_n/\partial y_1 & \dots & \partial F_n/\partial y_n \end{pmatrix}$$

Lause A.2. *Oletetaan, että* $F(x_*, y_*) = 0$ *ja* $\det(d_y F(x_*, y_*)) \neq 0$. *Tällöin on olemassa* x_**:n ympäristö* Ω *ja* $g : \Omega \to \mathbb{R}^n$ *siten että*

$$F(x, g(x)) = 0 \quad \text{kaikilla} \quad x \in \Omega$$

Olkoon $M = \{(x, y) \,|\, F(x, y) = 0\}$. Geometrisesti implisiittifunktiolause siis sanoo, että paikallisesti M voidaan esittää kuvauksen g graafina: $(x, g(x)) \in M$ kaikilla $x \in \Omega$.

Tämän seurauksena saadaan heti *käänteisfunktiolause*.

Lause A.3. *Olkoon* $f : \mathbb{R}^n \to \mathbb{R}^n$, $q = f(p)$ *ja oletetaan, että* $\det(df_p) \neq 0$. *Tällöin on olemassa* p*:n ympäristö* Ω_p *ja* q*:n ympäristö* Ω_q *siten, että* f*:llä on käänteisfunktio* f^{-1} *siten, että*

$$(f^{-1} \circ f)(x) = x \quad \text{kaikilla} \quad x \in \Omega_p \quad \text{ja} \quad (f \circ f^{-1})(x) = x \quad \text{kaikilla} \quad x \in \Omega_q$$

Lisäksi pätee $df_q^{-1} = (df_p)^{-1}$.

Todistus. Valitaan edellisessä lauseessa $F(x, y) = x - f(y)$. Tällöin saatu g on f:n käänteisfunktio. Derivoimalla yhtälöä $(f^{-1} \circ f)(x) = x$ saadaan jälkimmäinen väite. ✧

Määritelmä A.3. *Olkoon* $\Omega_j \subset \mathbb{R}^n$ *ja* $f : \Omega_1 \to \Omega_2$. *Tällöin* f *on* homeomorfismi, *jos*

(i) f *on bijektio*

(ii) f *ja* f^{-1} *ovat jatkuvia*

Jos lisäksi pätee

(iii) df *ja* df^{-1} *ovat jatkuvia*

niin f *on* diffeomorfismi.

Käänteisfunktiolause siis sanoo, että jos $\det(df_p) \neq 0$, niin f on diffeomorfismi jossain p:n ympäristössä.

Liite B

Funktionaalianalyysiä

Der Krieg ist eine bloße Fortsetzung der Politik mit anderen Mitteln.[1]

Carl von Clausewitz

Clausewitzia mukaillen voisi sanoa, että (lineaarinen) funktionaalianalyysi on lineaarialgebran jatkamista toisin keinoin. Funktionaalianalyysissä tarkastellaan ääretönulotteisia avaruuksia, ja tavoitteena on toisaalta tutkia mitkä tuttujen äärellisulotteisten avaruuksien (kuten $\mathbb{R}^n$) ominaisuuksista pätevät ääretönulotteisissa avaruuksissa, ja toisaalta sitten löytää uusia ilmiöitä, jotka esiintyvät vain ääretönulotteisessa tapauksessa. Erinomainen johdatus funktionaalianalyysiin nimenomaan (osittais)differentiaaliyhtälöitten näkökulmasta on [2].

Tarvittavat vektoriavaruudet ovat joko reaali- tai kompleksikertoimisia. Jos puhutaan $\mathbb{K}$-kertoimisista vektoriavaruuksista, niin tarkoitetaan sitä, että kerroinkunta voi olla joko $\mathbb{R}$ tai $\mathbb{C}$.

Joukko V on *vektoriavaruus* jos

(V_1) $x,\, y \in V \quad \Rightarrow \quad x + y \in V$ kaikilla x ja y

(V_2) $c \in \mathbb{K},\, x \in V \quad \Rightarrow \quad cx \in V$ kaikilla c ja x

Olkoon V vektoriavaruus. Äärellinen joukko $S \subset V$ on V:n *kanta* jos S on lineaarisesti riippumaton ja S virittää V:n. Tällöin $\mathsf{dim}(V) = n$, missä n on S:n alkioitten lukumäärä. Jos mikään äärellinen joukko ei viritä V:tä, niin sanotaan, että V on ääretönulotteinen ja merkitään $\mathsf{dim}(V) = \infty$.

Määritelmä B.1. *Vektoriavaruus V on* normiavaruus*, jos siellä on annettu kuvaus $n : V \to \mathbb{R}$ joka toteuttaa*

(N_1) $n(x) \geq 0$ *kaikilla* $x \in V$

(N_2) $n(x + y) \leq n(x) + n(y)$ *kaikilla* $x,\ y \in V$

(N_3) $n(cx) = |c|n(x)$ *kaikilla* $x \in V$ *ja* $c \in \mathbb{K}$

(N_4) $n(x) = 0 \iff x = 0$

[1] Sota on politiikan jatkamista toisin keinoin.

Kuvaus n on normi *ja yleensä merkitään $n(x) = \|x\|$.*

Usein normi määritellään sisätulon avulla. Muistetaan, että jos $z \in \mathbb{C}$ niin $\bar{z}$ on liittoluku.

Määritelmä B.2. *Olkoon V jokin $\mathbb{K}$-kertoiminen vektoriavaruus. Tällöin kuvaus $b : V \times V \to \mathbb{K}$ on* sisätulo, *jos*

(S_1) $b(ax + cy, z) = a\, b(x, z) + c\, b(y, z)$ *kaikilla x, y, $z \in V$ ja a, $c \in \mathbb{K}$*

(S_2) $b(x, x) \geq 0$ *kaikilla $x \in V$*

(S_3) $b(x, y) = b(y, x)$ *kaikilla $x\,, y \in V$, jos $\mathbb{K} = \mathbb{R}$ ja*
$b(x, y) = \overline{b(y, x)}$ *kaikilla $x\,, y \in V$, jos $\mathbb{K} = \mathbb{C}$*

(S_4) $b(x, x) = 0 \quad \Rightarrow \quad x = 0.$

Yleensä merkitään $b(x, y) = \langle x, y\rangle$.

Nyt on helppo tarkistaa

Lemma B.1. *Olkoon V vektoriavaruus, jossa on määritelty sisätulo. Tällöin $\|x\| = \sqrt{\langle x, x\rangle}$ on normi ja*

$$|\langle x, y\rangle| \leq \|x\|\, \|y\| \tag{B.1}$$

Epäyhtälöä (B.1) on tapana sanoa *Cauchyn epäyhtälöksi.*

Määritelmä B.3. *Vektoriavaruus jonka normi on annettu sisätulon avulla on* sisätuloavaruus.

Normin avulla voidaan yleistää analyysistä tutut käsitteet suppeneminen, Cauchy-jono jne yleisiin normiavaruuksiin. Perinteisissä määritelmissä ja todistuksissa itseisarvot vain korvataan sopivalla normilla. Erityisen tärkeitä ovat normiavaruudet, jotka toteuttavat seuraavan lisäehdon.

Määritelmä B.4. *Normiavaruus on* Banachin avaruus *(eli täydellinen normiavaruus), jos sen jokainen Cauchy-jono suppenee. Täydellinen sisätuloavaruus on* Hilbertin avaruus.

Yksinkertaisin Hilbertin avaruus on $\mathbb{R}^n$ varustettuna normilla

$$|x| = \sqrt{\langle x, x\rangle} = \left(x_1^2 + \cdots + x_n^2\right)^{1/2}$$

Määritelmä B.5. *Olkoon $\Omega \subset \mathbb{R}^n$ ja asetetaan*

$$\begin{aligned} L^1(\Omega) =& \big\{ f : \Omega \to \mathbb{C} \mid \int_\Omega |f(x)| dx < \infty \big\} \\ L^2(\Omega) =& \big\{ f : \Omega \to \mathbb{C} \mid \int_\Omega |f(x)|^2 dx < \infty \big\} \end{aligned}$$

Nämä ovat normiavaruuksia, joilla on normit

$$\|f\|_1 = \int_\Omega |f(x)|dx$$
$$\|f\|_2 = \sqrt{\int_\Omega |f(x)|^2 dx}$$

$L^2(\Omega)$ *on vieläpä sisätuloavaruus jossa on sisätulo*

$$\langle f, g\rangle = \int_\Omega f(x)\overline{g(x)}dx$$

Kun Lebesguen integraali on saatu määriteltyä, niin voidaan osoittaa

Lause B.1. $L^2(\Omega)$ *on Hilbertin avaruus ja* $L^1(\Omega)$ *on Banachin avaruus.*

Moni-indeksi on yksinkertaisesti vektori, jonka alkiot ovat einegatiivisia kokonaislukuja: siis $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$. Moni-indeksin kertaluku on sen alkioitten summa: $|\alpha| = \alpha_1 + \cdots + \alpha_n$. Olkoon sitten $\Omega \subset \mathbb{R}^n$ ja $u : \Omega \to \mathbb{R}$. Tällöin u:n yleinen osittaisderivaatta voidaan kirjoittaa seuraavasti:

$$\partial^\alpha u = \frac{\partial^{|\alpha|}u}{\partial x^\alpha} = \frac{\partial^{|\alpha|}u}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}$$

Luonnollisesti sitten sanotaan myös, että osittaisderivaatan $\partial^\alpha u$ kertaluku on $|\alpha|$.

Määritelmä B.6.

$$C^k(\Omega) = \{u : \Omega \mapsto \mathbb{R} \,|\, \partial^\alpha u \text{ on jatkuva} \quad \forall\ |\alpha| \leq k\}$$
$$C^\infty(\Omega) = \{u : \Omega \mapsto \mathbb{R} \,|\, u \in C^k(\Omega) \quad \forall\ k\}$$

Avaruuden $C^\infty(\Omega)$ funktioita on tapana sanoa *sileiksi* ja yleensä merkitään $C^0(\Omega) = C(\Omega)$. Olkoon sitten Ω kompakti (suljettu ja rajoitettu). Tällöin avaruuteen $C(\Omega)$ voidaan määritellä normi seuraavasti:

$$\|u\|_\infty = \max_{x\in\Omega} |u(x)|$$

Ei liene yllätys, että tätä normia on tapana sanoa *maksiminormiksi.* Tämän avulla saadaan normi myös avaruuksiin $C^k(\Omega)$:

$$\|u\|_{\infty,k} = \sum_{|\alpha|\leq k} \|\partial^\alpha u\|_\infty$$

On varsin suoraviivaista tarkistaa, että nämä todella määrittelevät normit. Nyt voidaan osoittaa:

Lause B.2. *Avaruudet $C^k(\Omega)$ ovat Banachin avaruuksia.*

Olkoon sitten V ja W vektoriavaruuksia. Kuvaus $T : V \to W$ on lineaarikuvaus, jos

(L_1) $T(x+y) = Tx + Ty$ kaikilla x ja y

(L_2) $T(c\,x) = c\,Tx$ kaikilla c ja x

Määritelmä B.7. *Olkoon $T : V \to W$ lineaarikuvaus missä V ja W ovat normiavaruuksia ja asetetaan*

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}$$

T on rajoitettu, *jos $\|T\| < \infty$.*

Nyt voidaan osoittaa

Lemma B.2. *T on rajoitettu jos ja vain jos T on jatkuva.*

Määritelmä B.8. *Reaalikertoiminen vektoriavaruus V on* Lie-algebra, *jos siellä on määritelty kuvaus $b : V \times V \to V$ jolla on seuraavat ominaisuudet:*

(i) *b on bilineaarinen*

(ii) $b(x,y) = -b(y,x)$

(iii) $b\big(x, b(y,z)\big) + b\big(z, b(x,y)\big) + b\big(y, b(z,x)\big) = 0$ *(Jacobin identiteetti)*

Yleensä merkitään $b(x,y) = [x,y]$.[2]

Lie-algebra on siis vektoriavaruus, jossa on lisäksi määritelty eräänlainen kertolasku.

(i) $\mathbb{R}^3$ varustettuna vektorien ristitulolla on Lie-algebra.

(ii) $n \times n$ matriisit operaatiolla $[A,B] = AB - BA$ on Lie-algebra. Tätä operaatiota sanotaan joskus *kommutaattoriksi*. Huomaa, että tässä tapauksessa on siis käytössä 2 erityyppistä kertolaskua.

(iii) Vektorikenttä voidaan samaistaa differentiaalioperaattoriin. Jos siis $X = (a_1, \ldots, a_n)$ on vektorikenttä ja f jokin funktio, niin voidaan kirjoittaa

$$X(f) = \sum_{j=1}^{n} a_j(x) \frac{\partial f}{\partial x_j}$$

Siis $X(f)$ on f:n suunnattu derivaatta tai Lie-derivaatta vektorikentän X suuntaan. Olkoon sitten Y jokin toinen vektorikenttä ja asetetaan

$$[X,Y](f) = X(Y(f)) - Y(X(f))$$

Nyt voidaan tarkistaa, että $[X,Y]$ on edelleen vektorikenttä ja että tämä tekee vektorikenttien avaruudesta Lie-algebran. Tätä operaatiota sanotaan usein Lien kuvaukseksi (englanniksi Lie bracket).

[2] Englanniksi kuvauksesta käytetään usein nimeä *bracket*.

(iv) Määritelmässä 6.3 esiteltiin Poissonin kuvaus joka teki avaruudesta $C^\infty(\mathbb{R}^{2n})$ Lie-algebran. Poissonin ja Lien kuvausten välillä on seuraava yhteys

$$X_{[f,g]} = [X_f, X_g]$$

Tässä siis X_f on funktiota f vastaava Hamiltonin vektorikenttä. Kuvaus $f \to X_f$ on siis kahden Lie-algebran välinen homomorfismi.

B.1 Harjoitustehtäviä

1. Olkoon $I = [a, b]$, $V_0 = \{f \in C^1(I) \,|\, f(a) = f(b) = 0\}$ ja $V = \{f \in C^1(I) \,|\, f(a) = c\ ,\ f(b) = d\}$. Tarkista, että V_0 on vektoriavaruus, mutta V ei ole vektoriavaruus.

2. Tarkista, että $\|y\|_\infty$ ja $\|y\|_{1,\infty}$ määrittelevät normin.

3. Olkoon $f_n(x) = \sin(nx)/n$. Näytä, että $\|f_n\|_\infty \to 0$, kun $n \to \infty$, mutta $\|f_n\|_{1,\infty}$ ei lähesty nollaa. Siis funktiojono f_n suppenee kohti nollafunktiota avaruudessa $C(I)$, mutta ei suppene avaruudessa $C^1(I)$.

4. Olkoon $I = [a, b]$ ja määritellään kuvaus

$$\mathcal{I} : C(I) \to \mathbb{R} \quad , \quad \mathcal{I}(f) = \int_a^b f(x)dx$$

Tarkista, että $\mathcal{I}$ on jatkuva lineaarikuvaus. Mikä on $\|\mathcal{I}\|$?

5. Olkoon $I = [a, b]$ ja määritellään kuvaus

$$\mathcal{D} : C^1(I) \to C(I) \quad , \quad \mathcal{D}(f) = f'$$

Tarkista, että $\mathcal{D}$ on jatkuva lineaarikuvaus. Mikä on $\|\mathcal{D}\|$?

Liite C

Optimointi

Optimoinnista on kirjoitettu monia kirjoja. Eräs uudempi kirja on [1], jossa korostetaan erityisesti konveksisuuden merkitystä.

C.1 Ilman rajoitusehtoja

Määritelmä C.1. *Olkoon* $\Omega \subset \mathbb{R}^n$, $u : \Omega \to \mathbb{R}$ *ja* $B_r(p) = \{x \in \mathbb{R}^n \mid |x - p| < r\}$.

1. p *on* u*:n* kriittinen piste, *jos* $\nabla u(p) = 0$. *Kriittinen piste* p *on* säännöllinen, *jos* $\det(d^2 u_p) \neq 0$.
2. p *on* u*:n* (aito) paikallinen minimi, *jos on olemassa* r *siten, että* $u(x) \geq u(p)$ $(u(x) > u(p))$ *kaikilla* $x \in B_r(p)$.
3. p *on* u*:n* (aito) paikallinen maksimi, *jos on olemassa* r *siten, että* $u(x) \leq u(p)$ $(u(x) < u(p))$ *kaikilla* $x \in B_r(p)$.
4. p *on* u*:n* satulapiste, *jos kaikilla* r *on olemassa* x, $y \in B_r(p) \cap \Omega$ *siten, että* $u(x) < u(p) < u(y)$.
5. p *on* u*:n* globaali minimi, *jos* $u(x) \geq u(p)$ *kaikilla* $x \in \Omega$.
6. p *on* u*:n* globaali maksimi, *jos* $u(x) \leq u(p)$ *kaikilla* $x \in \Omega$.

Funktion u toisen differentiaalin avulla voidaan tutkia kriittisen pisteen laatua. Palautetaan matriisilaskusta mieleen, että

- symmetrisen matriisin ominaisarvot ovat reaalisia
- symmetrinen matriisi A on *positiividefiniitti*, jos

$$\langle x, Ax \rangle > 0 \quad \text{kaikilla} \quad x \neq 0$$

- A positiividefiniitti $\quad \Leftrightarrow \quad$ A:n ominaisarvot ovat positiivisia

Lemma C.1. *Olkoon p funktion u säännöllinen kriittinen piste. Tällöin p on aito paikallinen minimi, jos d^2u_p:n ominaisarvot ovat positiivisia, ja aito paikallinen maksimi, jos ominaisarvot ovat negatiivisia. Jos d^2u_p:llä on sekä positiivisia että negatiivisia ominaisarvoja, niin p on satulapiste.*

Todistus. Lauseessa A.1 esitetyn sarjan toinen termi on kriittisessä pisteessä nolla. Tarkastelemalla termiä $\langle h, d^2u_p h\rangle$ saadaan väitteet. ✧

Paikallisesta minimistä päästään globaaliin minimiin, jos minimoitava funktio sattuu olemaan konveksi.

Määritelmä C.2.

(i) $\Omega \subset \mathbb{R}^n$ *on* konveksi*, jos*

$$x\,, y \in \Omega \quad \Rightarrow \quad (1-\mu)x + \mu y \in \Omega \quad \text{kaikilla} \quad 0 \le \mu \le 1$$

(ii) Jos Ω *konveksi, niin funktio* $f : \Omega \to \mathbb{R}$ *on* konveksi*, jos*

$$f\big((1-\mu)x + \mu y\big) \le (1-\mu)f(x) + \mu f(y) \quad \text{kaikilla} \quad 0 \le \mu \le 1$$

(iii) f *on* aidosti konveksi*, jos*

$$f\big((1-\mu)x + \mu y\big) < (1-\mu)f(x) + \mu f(y) \quad \text{kaikilla} \quad 0 < \mu < 1$$

(iv) f *on* (aidosti) konkaavi*, jos* $-f$ *on (aidosti) konveksi.*

Konveksin funktion ei siis yleisesti ottaen tarvitse olla derivoituva. Jos derivaattoja on olemassa, niin konveksisuus voidaan karakterisoida seuraavasti.

Lemma C.2.

(i) Olkoon $f \in C^1(\Omega)$. *Tällöin* f *on konveksi jos ja vain jos*

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle \quad \text{kaikilla } x\,,\ y \in \Omega$$

(ii) Olkoon $f \in C^2(\Omega)$. *Tällöin* f *on konveksi jos ja vain jos* $d^2 f_p$ *on positiivisemidefiniitti kaikilla* $p \in \Omega$.

Todistus. Harjoitustehtävä. ✧

Tästä saadaan välittömästi

Lemma C.3. *Olkoon p konveksin funktion f kriittinen piste. Tällöin p on funktion f globaali minimi.*

Todistus. Edellisen lemman mukaan $f(y) \ge f(p) + \langle \nabla f(p), y - p\rangle$ kaikilla y. Toisaalta kriittisessä pisteessä $\nabla f(p) = 0$. ✧

C.2 Rajoitusehdot

Usein optimointithetäviä ei haluta ratkaista jossain $\mathbb{R}^n$ avoimessa osajoukossa vaan lisäksi on annettu joitain rajoitusehtoja. Tarkastellaan seuraavassa vain erästä tyypillistä tapausta.

Olkoon $g : \mathbb{R}^n \to \mathbb{R}^k$ missä $k < n$.

Määritelmä C.3.

1. *g:n* rangi *pisteessä p,* $\mathsf{rank}(g_p) = \mathsf{rank}(dg_p)$.

2. *p on* säännöllinen piste, *jos* $\mathsf{rank}(g_p) = k$.

3. *Olkoon $M = \{x \in \mathbb{R}^n \mid g(x) = 0\}$. Nolla on* säännöllinen arvo, *jos kaikki M:n pisteet ovat säännöllisiä.*

Nyt voidaan osoittaa

Lemma C.4. *Jos nolla on g:n säännöllinen arvo, niin M on sileä monisto.*

Sileä monisto voidaan tässä yhteydessä ajatella sellaisena osajoukkona, joka voidaan lokaalisti esittää jonkin sileän kuvauksen graafina, vertaa lause A.2. Tällöin M:n *tangenttiavaruus* on hyvin määritelty, ja olkoon T_pM tangenttiavaruus pisteessä p. Tangenttiavaruuden ortogonaalikomplementti on *normaaliavaruus*; merkitään N_pM.

Olkoon sitten $\alpha : \mathbb{R} \to M$ jokin käyrä, $\alpha(0) = p$ ja $h = g \circ \alpha$. Siis h on nollakuvaus, joten

$$g'(s) = dg\, \alpha' = 0$$

Nyt $\alpha'(0) \in T_pM$ joten dg_p rivit ovat kohtisuorassa T_pM:n kanssa. Siis $(dg_p)^T$:n sarakkeet virittävät N_pM:n.

Olkoon sitten nolla g:n säännöllinen arvo, $M = \{x \in \mathbb{R}^n \mid g(x) = 0\}$ ja $f : M \to \mathbb{R}$.

Lemma C.5. *Jos p minimoi f:n, niin $\nabla f(p) \in N_pM$. On siis olemassa vektori $\lambda = (\lambda_1, \ldots, \lambda_k)$ siten, että*

$$\nabla f(p) + (dg_p)^T \lambda = 0$$

Vektorin λ komponentteja sanotaan *Lagrangen kertoimiksi.*

Todistus. Olkoon jälleen $\alpha : \mathbb{R} \to M$ jokin käyrä ja $\alpha(0) = p$. Jos p minimoi f:n niin tällöin p on myös funktion $h = f \circ \alpha$ minimi. Siis

$$h'(0) = \langle \nabla f(p), \alpha' \rangle = 0$$

olipa α mikä tahansa. Koska $\alpha'(0) \in T_pM$, niin tämä on mahdollista vain jos $\nabla f(p) \in N_pM$. ✧

Määritellään kuvaus

$$F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n \times \mathbb{R}^k \quad , \quad F(x,\lambda) = \begin{pmatrix} \nabla f + (dg)^T \lambda \\ g \end{pmatrix}$$

Äskeisten laskujen perusteella siis jos (p, λ_*) on minimi niin $F(p, \lambda_*) = 0$. Tämähän on algebrallinen yhtälö joka voidaan ratkaista, ainakin numeerisesti. Mutta Lagrange huomasi, että F on aivan erityistä muotoa:

$$\tilde{f} = f + \langle \lambda, g \rangle \quad \Rightarrow \quad F = \nabla \tilde{f} \tag{C.1}$$

Toisin sanoen jos määritellään *laajennettu kohdefunktio* $\tilde{f}$, niin rajoitetun optimointitehtävän välttämätön ehto on muodollisesti samanlainen kuin rajoittamattoman optimointitehtävän.

C.3 Harjoitustehtäviä

1. Todista lemma C.2.

2. Näytä, että laajennetun kohdefunktion (C.1) kriittiset pisteet ovat aina satulapisteitä.

Kirjallisuutta

[1] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

[2] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[3] G. Buttazzo, M. Giaquinta, and S. Hildebrandt, *One-dimensional variational problems*, Oxford Lecture Series in Mathematics and its Applications, vol. 15, The Clarendon Press Oxford University Press, New York, 1998.

[4] I. M. Gelfand and S. V. Fomin, *Calculus of variations*, Dover, 2000.

[5] M. Giaquinta and S. Hildebrandt, *Calculus of variations. I*, Grundlehren der Mathematischen Wissenschaften, vol. 310, Springer-Verlag, Berlin, 1996, The Lagrangian formalism.

[6] ———, *Calculus of variations. II*, Grundlehren der Mathematischen Wissenschaften, vol. 311, Springer-Verlag, Berlin, 1996, The Hamiltonian formalism.

[7] M. A. Golberg, *The derivative of a determinant*, Amer. Math. Monthly **79** (1972), 1124–1126.

[8] D. Liberzon, *Calculus of variations and optimal control theory*, Princeton University Press, Princeton, NJ, 2012.

[9] M. Mesterton-Gibbons, *A primer on the calculus of variations and optimal control theory*, Student Mathematical Library, vol. 50, American Mathematical Society, Providence, RI, 2009.

[10] B. van Brunt, *The calculus of variations*, Universitext, Springer-Verlag, New York, 2004.

[11] L. C. Young, *Lectures on the calculus of variations and optimal control theory*, AMS Chelsea Publishing, 1980.

[12] E. Zeidler, *Nonlinear functional analysis and its applications. III*, Springer-Verlag, New York, 1985, Variational methods and optimization.