

## Aikariippuvuus pienimmän neliön elementtimenetelmässä

Jorma Kinnunen ja Eero-Matti Salonen

**Tiivistelmä.** Artikkelissa selostetaan tiettyjä aikariippuvaan pienimmän neliön elementtimenetelmään liittyviä piirteitä. Lisäksi esitetään eräs — kirjoittajien tietämyksen mukaan — uusi sovellusversio. Versiota kuvataan yksidimensioisen lineaarisen konvektiotehtävän yhteydessä. Von Neumannin analyysiä käytetään vertailussa kahteen vaihtoehtoiseen formulaatioon. Joitakin numeerisia esimerkkituloksia esitetään.

*Avainsanat:* pienimmän neliön keino, aikariippuvuus, elementtimenetelmä, konvektioyhtälö.

### Johdanto

Artikkelissa selostetaan tiettyjä aikariippuvaan pienimmän neliön elementtimenetelmään liittyviä piirteitä. Lisäksi esitetään eräs sovellusversio, jota ei kirjoittajien tietämyksen mukaan ole esitetty aikaisemmin alan kirjallisuudessa.

Pienimmän neliön keinoa ei välttämättä pidetä sopivana aikariippuvien tehtävien käsittelyssä. Tähän liittyy seuraava lainaus [5, s. 67]: ”Whereas the Galerkin method has almost no restriction on the type of problem that can be considered, the least-squares method is not well-suited to evolutionary or eigenvalue problems.” Tämä viite on kuitenkin elementtimenetelmän historian kannalta melko vanha ja täysin käyttökelpoisia pienimmän neliön keinojen sovelluksia aikariippuviin tehtäviin on esitetty myöhemmin alan kirjallisuudessa.

Käsittely pelkistetään tässä koskemaan seuraavassa kappaleessa esitettyä yksinkertaista demonstraatiotapausta.

Tarkastellaan yksidimensioista *aikariippuvaa lineaarista konvektioyhtälöä*

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} - s = 0, \quad (1)$$

joka on voimassa paikka-alueessa  $\Omega = (0, L)$  ja aika-alueessa  $t > 0$  tiettyjen reuna- ja alkuehtojen alaisena. Tuntemattomana on suure  $u = u(x, t)$ ; esimerkiksi jonkin aineosuuden konsentraatio. Annetut suureet yhtälössä (1) ovat konvektionopeus  $a = a(x, t)$  ja lähdetermi  $s = s(x, t)$ . Alkuehto on muotoa

$$u(x, 0) = u_0(x), \quad (2)$$

jossa  $u_0$  on annettu. Reunaehto annetaan vain sisäänvirtausreunalla. Jos  $a$  on esimerkiksi positiivinen, sisäänvirtausreuna on viiva  $x = 0$ . Dirichlet-tyyppinen reunaehto on tällöin muotoa

$$u(0, t) = u_D(t), \quad (3)$$

jossa  $u_D$  on annettu. Nimityksen konvektio (engl. convection) sijasta kirjallisuudessa käytetään paljon myös termiä advektio (engl. advection), josta seuraa konvektionopeuden ehkä tavanomaisin tunnus  $a$ .

Kaikkein yleisin lähestymistapa aikariippuvissa tehtävissä on asettaa paikkakoordinaatit ja aika ainakin näennäisesti samanarvoisiin rooleihin. Elementtimenetelmässä se merkitsee paikka-aikamuotofunktioiden käyttöä. Erityisesti tässä kirjoitettaisiin siis tyyppiä

$$\tilde{u}(x, t) = \sum_j N_j(x, t) u_j \quad (4)$$

oleva approksimaatio. Termit  $N_j$  ovat muotofunktioita ja termit  $u_j$  solmuarvoja. Verrattain tuore teos [2] käsittelee syvällisesti pienimmän neliön elementtimenetelmää. Paikka-aikaformulaatiota kosketellaan siinä kuitenkin vain lyhyesti, vaikka sen todetaan omaavan tiettyjä lupaavia piirteitä. Tässä artikkelissa jätetään myös paikka-aikaformulaatio käsittelemättä.

## Aika-paikka diskretointijärjestys

Tavanomaisin tapa edetä elementtimenetelmässä aikariippuvissa tehtävissä on ns. semidiskretointi (engl. semi-discretization): diskretointi suoritetaan kahdessa vaiheessa erikseen paikan ja erikseen ajan suhteen. Ehkä vielä kaikkein luontevin tapa on diskretoida ensin paikan suhteen. Saadaan tavallisten differentiaaliyhtälöiden systeemi. Se ratkaistaan sitten edelleen jollain diskretointimenetelmällä aikakoordinaatin suhteen askeltamalla. Menettelystä käytetään enlanninkielisessä kirjallisuudessa usein myös nimitystä method of lines (MOL). Seuraavassa luvussa selostetaan tällaiseen käsittelytapaan perustuvaa pienimmän neliön keinon versiota, jota emme ole nähneet sovelletun aikaisemmin.

Teoksessa [2] on kuvattu kuitenkin melko paljon juuri päinvastaisessa järjestyksessä etenevää diskretointia. Vertailutulosten saavuttamiseksi tullaan tässä esittämään siten ensin lyhyesti viitteen [2] käsittelytapaa sovellettuna yhtälöön (1). Itse asiassa juuri tätä sovellusta on käsitelty vielä tarkemmin viitteessä [3, s. 121].

Tehtävä diskretoidaan ensin aikasuunnassa käyttäen yhtälön (1) aikaderivaatalle differenssiapproksimaatiota

$$\frac{\partial u}{\partial t} \approx \frac{1}{\Delta t} (u^{n+1} - u^n) \quad (5)$$

jossa yläindeksit  $n$  ja  $n+1$  viittaavat tavanomaiseen tapaan aikatasoihin  $t = t_n$  ja  $t = t_{n+1} = t_n + \Delta t$ . Lisäksi esimerkiksi Crank-Nicolsonin menetelmässä (ns.  $\theta$ -menetelmän erikoistapaus) yhtälön (1) muut termit approksimoidaan keskiarvoina aikatasoilta  $t_n$  ja  $t_{n+1}$ :

$$a \frac{\partial u}{\partial x} - s \approx \frac{1}{2} \left( a^{n+1} \frac{\partial u^{n+1}}{\partial x} + a^n \frac{\partial u^n}{\partial x} - s^{n+1} - s^n \right). \quad (6)$$

Syntyvä yhtälö on

$$\frac{1}{2} a^{n+1} \frac{\partial u^{n+1}}{\partial x} + \frac{1}{\Delta t} u^{n+1} - \left( \frac{1}{2} s^{n+1} + \frac{1}{2} s^n - \frac{1}{2} a^n \frac{\partial u^n}{\partial x} + \frac{1}{\Delta t} u^n \right) = 0. \quad (7)$$

Suluissa oleva termi tunnetaan askelletun ratkaisun perusteella aikaisemmalta aikatasolta tai aivan alussa alkuehdon perusteella (lähdetermi  $s$  on annettu paikan ja ajan funktio, joten  $s^{n+1}$  on myös tunnettu).

Jotta jatkoon kaavat saataisiin pelkistettyyn ulkoasuun, jätetään yläindeksi  $n+1$  tilapäisesti pois ja sulussa olevaa termiä merkitään lyhyesti tunnuksella  $\bar{s}$ . On päädytty differentiaaliyhtälöön

$$\mathcal{R}(u) \equiv \frac{1}{2} a \frac{\partial u}{\partial x} + \frac{1}{\Delta t} u - \bar{s} = 0. \quad (8)$$

Tämä on muodollisesti *stationaarinen lineaarinen konvektio-reaktioyhtälö* (aikatasolla  $t = t_{n+1}$ ). Lyhennysmerkintä  $\mathcal{R}$  viittaa ns. yhtälöjäännökseen (engl. equation residual). On huomattava, että yhtälö (8) sisältää aikadiskretoinnin johdosta jo approksimaatioita, vaikka niitä ei ole tässä merkinnällisesti osoitettu.

Pienimmän neliön menetelyn lähtökohtana on pienimmän neliön funktionaali

$$\Pi(u) = \frac{1}{2} \int_{\Omega} \mathcal{R}^2(u) d\Omega. \quad (9)$$

Vaaditaan tämän variaation häviämistä:

$$\delta \Pi \equiv \int_{\Omega} \mathcal{R} \delta \mathcal{R} d\Omega \equiv \int_{\Omega} \delta \mathcal{R} \mathcal{R} d\Omega = 0. \quad (10)$$

Lausekkeen (8) perusteella on yksityiskohtaisemmin

$$\delta \mathcal{R} = \frac{1}{2} a \frac{\partial \delta u}{\partial x} + \frac{1}{\Delta t} \delta u. \quad (11)$$

On siis päädytty heikkoon muotoon

$$\int_{\Omega} \left( \frac{1}{2} a \frac{\partial \delta u}{\partial x} + \frac{1}{\Delta t} \delta u \right) \left( \frac{1}{2} a \frac{\partial u}{\partial x} + \frac{1}{\Delta t} u - \bar{s} \right) d\Omega = 0. \quad (12)$$

Elementtimenetelmässä otetaan (aikatasolla  $t = t_{n+1}$ ) approksimaatio

$$\tilde{u}(x) = \sum_j N_j(x) u_j \quad (13)$$

ja tämän variaatio saadaan varioimalla solmuarvoja:

$$\delta \tilde{u}(x) = \sum_j N_j(x) \delta u_j. \quad (14)$$

Yhtälön (12) diskreeteiksi vastineiksi saadaan (korvataan  $u$   $\tilde{u}$  :lla, otetaan peräkkäisesti  $\delta\tilde{u} = N_i\delta u_i$ ,  $i = 1, 2, \dots$  ja jaetaan aina vielä termillä  $\delta u_i$ )

$$\int_{\Omega} \left( \frac{1}{2} a \frac{dN_i}{dx} + \frac{1}{\Delta t} N_i \right) \left( \frac{1}{2} a \sum \frac{dN_j}{dx} u_j + \frac{1}{\Delta t} \sum N_j u_j - \bar{s} \right) d\Omega = 0 \quad (15)$$

tai yksityiskohtaisemmin

$$\begin{aligned} \sum_j \int_{\Omega} \left( \frac{a^2}{4} \frac{dN_i}{dx} \frac{dN_j}{dx} + \frac{a}{2\Delta t} \frac{dN_i}{dx} N_j + \frac{a}{2\Delta t} N_i \frac{dN_j}{dx} + \left( \frac{1}{\Delta t} \right)^2 N_i N_j \right) d\Omega u_j \\ = \int_{\Omega} \left( \frac{1}{2} a \frac{dN_i}{dx} + \frac{1}{\Delta t} N_i \right) \bar{s} d\Omega. \end{aligned} \quad (16)$$

Tämä kehitetään nyt vielä pitemmälle palauttamalla takaisin yläindeksit  $n$  ja  $n+1$  ja termin  $\bar{s}$  sisältö:

$$\begin{aligned} \sum_j \int_{\Omega} \left( \frac{(a^{n+1})^2}{4} \frac{dN_i}{dx} \frac{dN_j}{dx} + \frac{a^{n+1}}{2\Delta t} \frac{dN_i}{dx} N_j + \frac{a^{n+1}}{2\Delta t} N_i \frac{dN_j}{dx} + \left( \frac{1}{\Delta t} \right)^2 N_i N_j \right) d\Omega u_j^{n+1} \\ = \sum_j \int_{\Omega} \left( -\frac{a^{n+1} a^n}{4} \frac{dN_i}{dx} \frac{dN_j}{dx} + \frac{a^{n+1}}{2\Delta t} \frac{dN_i}{dx} N_j - \frac{a^n}{2\Delta t} N_i \frac{dN_j}{dx} + \left( \frac{1}{\Delta t} \right)^2 N_i N_j \right) d\Omega u_j^n \\ + \int_{\Omega} \left( \frac{1}{2} a^{n+1} \frac{dN_i}{dx} + \frac{1}{\Delta t} N_i \right) \left( \frac{1}{2} s^{n+1} + \frac{1}{2} s^n \right) d\Omega. \end{aligned} \quad (17)$$

Näistä *linearisista algebrallisista systeemyhtälöistä* ( $i = 1, 2, \dots$ ) ratkaistaan solmuarvot  $u_j^{n+1}$ . Vastaava askel toistetaan seuraavalla aikatasolla jne.

Tämän luvun formulaatioon ja siihen liittyviin suureisiin tullaan viittaamaan jatkossa tarvittaessa lyhyesti tunnuksella I.

## Paikka-aika diskreetointijärjestys

Aloitetaan suoraan yhtälöstä (1) ja kirjoitetaan se muotoon (lyhennysmerkinnällä  $\mathcal{R}$  on nyt siis eri merkitys kuin edellä)

$$\mathcal{R}(u) \equiv \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} - s = 0. \quad (18)$$

Pienimmän neliön funktionaalin

$$\Pi(u) = \frac{1}{2} \int_{\Omega} \mathcal{R}^2(u) d\Omega \quad (19)$$

variaatio antaa vastaavasti kuin edellisessä luvussa heikon muodon

$$\int_{\Omega} \delta \mathcal{R} \mathcal{R} d\Omega = 0, \quad (20)$$

jossa

$$\delta \mathcal{R} = \frac{\partial \delta u}{\partial t} + a \frac{\partial \delta u}{\partial x}. \quad (21)$$

Tämä olisi  $\mathcal{R}$  :n variaatio täydessä paikka-aikaformulaatiossa. Nyt kuitenkin muutetaan asetelmaa siten, että rajoitetaan variaatiota  $\delta u$  siten, että se *ei riipu ajasta*:

$$\delta u = \delta u(x) \quad (22)$$

ja saadaan lausekkeen (21) sijasta

$$\delta \mathcal{R} = a \frac{d\delta u}{dx}. \quad (23)$$

Ilmeisesti mikään ei estä tekemästä esitettyä rajoitusta. Täten päädytään heikkoon muotoon

$$\int_{\Omega} a \frac{d\delta u}{dx} \left( \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} - s \right) d\Omega = 0. \quad (24)$$

Esityksen (13) sijasta elementtiapproksimaatio kirjoitetaan nyt muotoon

$$\tilde{u}(x, t) = \sum_j N_j(x) u_j(t). \quad (25)$$

Vastaava variaatio voidaan kirjoittaa  $\delta \tilde{u}(x) = \sum_j N_j(x) \delta u_j$  kuten kaavoissa (14) rajoittamalla vain solmuarvojen variaatiot  $\delta u_j$  ei-aikariippuviksi. Yhtälön (24) diskreetit vastineet ovat (korvataan jälleen  $u$   $\tilde{u}$  :lla ja otetaan peräkkäisesti  $\delta \tilde{u} = N_i \delta u_i$ ,  $i = 1, 2, \dots$ ):

$$\int_{\Omega} a \frac{dN_i}{dx} \left( \sum_j N_j \frac{du_j}{dt} + a \sum_j \frac{dN_j}{dx} u_j - s \right) d\Omega = 0 \quad (26)$$

tai yksityiskohtaisemmin

$$\sum_j \int_{\Omega} a \frac{dN_i}{dx} N_j d\Omega \frac{du_j}{dt} + \sum_j \int_{\Omega} a^2 \frac{dN_i}{dx} \frac{dN_j}{dx} d\Omega u_j - \int_{\Omega} a \frac{dN_i}{dx} s d\Omega = 0. \quad (27)$$

Kyseessä on *lineaarinen tavallinen* (ajan suhteen ensimmäisen kertaluvun) *differentiaaliyhtälösystemi* ( $i = 1, 2, \dots$ ). Näiden systeemien numeerista käsittelyä (eli siis jälleen diskretointia) koskevaa kirjallisuutta on pitkältä ajalta valtaisa määrä. Tässä sovelletaan taas Crank-Nicolsonin menetelmää. Yhtälöitä (27) tarkastellaan aikatasoilla  $t = t_n$  and  $t = t_{n+1} = t_n + \Delta t$ . Aikaderivaatalle käytetään esitystä

$$\frac{du_j}{dt} \approx \frac{1}{\Delta t} (u_j^{n+1} - u_j^n) \quad (28)$$

ja loput termeistä korvataan keskiarvoilla. Saadaan

$$\begin{aligned}
& \sum_j \left( \frac{1}{2} \int_{\Omega} (a^{n+1})^2 \frac{dN_i}{dx} \frac{dN_j}{dx} d\Omega + \frac{1}{\Delta t} \int_{\Omega} \left( \frac{a^{n+1} + a^n}{2} \right) \frac{dN_i}{dx} N_j d\Omega \right) u_j^{n+1} \\
&= \sum_j \left( -\frac{1}{2} \int_{\Omega} (a^n)^2 \frac{dN_i}{dx} \frac{dN_j}{dx} d\Omega + \frac{1}{\Delta t} \int_{\Omega} \left( \frac{a^{n+1} + a^n}{2} \right) \frac{dN_i}{dx} N_j d\Omega \right) u_j^n \\
&\quad + \frac{1}{2} \int_{\Omega} a^{n+1} \frac{dN_i}{dx} s^{n+1} d\Omega + \frac{1}{2} \int_{\Omega} a^n \frac{dN_i}{dx} s^n d\Omega. \tag{29}
\end{aligned}$$

Nämä ovat samaa tyyppiä kuin yhtälöt (17).

Lisävalaistusta tässä tarkastellun pienimmän neliön version sisältöön saadaan seuraavasti. Yhtälö (24) on hieman kehitettynä

$$\int_{\Omega} \frac{d\delta u}{dx} \left( a \frac{\partial u}{\partial t} + a^2 \frac{\partial u}{\partial x} - as \right) d\Omega = 0. \tag{30}$$

Osittaisintegrointi antaa

$$\begin{aligned}
\int_{\Omega} \frac{d\delta u}{dx} \left( a \frac{\partial u}{\partial t} + a^2 \frac{\partial u}{\partial x} - as \right) d\Omega &= - \int_{\Omega} \delta u \frac{d}{dx} \left( a \frac{\partial u}{\partial t} + a^2 \frac{\partial u}{\partial x} - as \right) d\Omega \\
&\quad + \delta u \left( a \frac{\partial u}{\partial t} + a^2 \frac{\partial u}{\partial x} - as \right) \Big|_0^L = 0. \tag{31}
\end{aligned}$$

Koska  $\delta u$  mielivaltainen, saadaan kenttäyhtälö (ns. Eulerin yhtälö)

$$\frac{\partial}{\partial x} \left( a \frac{\partial u}{\partial t} \right) + \frac{\partial}{\partial x} \left( a^2 \frac{\partial u}{\partial x} \right) - \frac{\partial}{\partial x} (as) = 0. \tag{32}$$

Otaksutaan lisäksi, että voimassa on (oleellinen) reunaehto (3). Tämän perusteella asetetaan  $\delta u(0, t) = 0$ , mutta  $\delta u$  on mielivaltainen reunalla  $x = L$ . Yhtälö (31) antaa nyt lisäksi (luonnollisen) reunaehdon

$$a \frac{\partial u}{\partial t} + a^2 \frac{\partial u}{\partial x} - as = 0, \quad x = L. \tag{33}$$

Yhtälöt (33) ja (32) ovat selvästi konsistentteja yhtälön (18) kanssa, koska ne saadaan manipuloimalla sitä: reunaehto syntyy kertomalla konvektionopeudella  $a$  ja kenttäyhtälö edelleen derivoimalla  $x$ :n suhteen. Täten esitetty pienimmän neliön versio voidaan tulkita menettelyksi, jossa ratkaistaan kenttäyhtälön (18) sijasta kenttäyhtälöä (32) annetun sisäänvirtausreunaehdon ja alkuehdon alaisena täydennettynä vielä reunaehdolla (33) ulosvirtausreunalla.

Ylläesitetyn tapainen lisätarkastelu voidaan suorittaa myös lähtien liikkeelle formulaation I heikosta muodosta (12). Myös siinä tulee Eulerin yhtälöön mukaan — poiketen siis alkuperäisestä konvektiohtälöstä — toinen paikkaderivaatta. Koska toisella derivaatalla on epätieteellisesti sanottuna ratkaisuja siloittava ominaisuus, tämä piirre voi tulla esille tietyllä tavalla numeerisissa ratkaisuisissa.

Tämän luvun formulaatioon ja siihen liittyviin suureisiin tullaan viittaamaan jatkossa tarvittaessa lyhyesti tunnuksella II.

## Galerkinin menettely

Käsitellään vielä lyhyesti Galerkinin menetelmän mukaista elementtimenetelmäversiota sovellettuna yhtälöön (1). Tämän tarkoituksena on lähinnä saada vertailutuloksia edellä esitettyihin kahteen pienimmän neliön elementtimenetelmäversioon.

Tarkastellaan tavanomaista semidiskreettiä Galerkinin menetelmää. Kirjoitetaan heikko muoto

$$\int_{\Omega} w(x) \left( \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} - s \right) d\Omega = 0. \quad (34)$$

On oleellista, että vaikka  $u$  riippuu sekä paikasta että ajasta, paino- eli testifunktio  $w$  on valittu *ei-aikariippuvaksi*. Itse asiassa juuri tämä tavanomaisen semidiskreetin Galerkinin lähestymistapa on ollut jonkinlaisena tukena valintaan (23). Jos vielä tehdään tulkinta  $w = \delta u$  askeleessa kohti Galerkinin formulaatiota, saadaan

$$\int_{\Omega} \delta u(x) \left( \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} - s \right) d\Omega = 0. \quad (35)$$

Heikkojen muotojen (24) and (35) nähdään eroavan vain testifunktioidensa suhteen.

Kun sovelletaan approksimaatiota (25), tavalliseksi differentiaaliyhtälösystemiksi saadaan ( $i = 1, 2, \dots$ )

$$\sum_j \int_{\Omega} N_i N_j d\Omega \frac{du_j}{dt} + \sum_j \int_{\Omega} a N_i \frac{dN_j}{dx} d\Omega u_j - \int_{\Omega} N_i s d\Omega = 0. \quad (36)$$

Etenemällä sitten kuten yhtälösystemin (27) suhteen päädytään algebralliseen yhtälösystemiin

$$\begin{aligned} & \sum_j \left( \frac{1}{2} \int_{\Omega} a^{n+1} N_i \frac{dN_j}{dx} d\Omega + \frac{1}{\Delta t} \int_{\Omega} N_i N_j d\Omega \right) u_j^{n+1} \\ & = \sum_j \left( -\frac{1}{2} \int_{\Omega} a^n N_i \frac{dN_j}{dx} d\Omega + \frac{1}{\Delta t} \int_{\Omega} N_i N_j d\Omega \right) u_j^n \\ & \quad + \int_{\Omega} \frac{dN_i}{dx} \frac{s^{n+1} + s^n}{2} d\Omega. \end{aligned} \quad (37)$$

Tämän luvun formulaatioon ja siihen liittyviin suureisiin tullaan viittaamaan jatkossa tarvittaessa lyhyesti tunnuksella G.

## Von Neumann analyysi

Ns. von Neumannin tai Fourier'n analyysi antaa usein melko vähäisellä työmäärällä arvokasta tietoa eri aikariippuvien probleemien numeeristen formulaatioiden luonteista. Tässä tätä analyysia sovelletaan edellä esitettyihin formulaatioihin I, II ja G. Taustalla olevaa teoriaa on selostettu mm. viitteissä [1] and [3] ja tässä kuvataan vain analyysin soveltamisen pääkohdat.

Tarkastellaan yhtälöä (1) tapauksessa, jossa konvektionopeus  $a$  on vakio ja lähde-termi  $s$  häviää:

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0. \quad (38)$$

Otaksutaan muuttujien tulomuotoista tyyppiä oleva analyttinen ratkaisu

$$u(x, t) = A(t) e^{ikx}. \quad (39)$$

Tässä  $i$  on imaginaariyksikkö ja  $k$  reaalimuuttuja (ns. aaltoluku, joka yleisesti käytetty nimitys on sinänsä huono, koska kyseessä ei ole luku vaan dimension omaava suure). Kun lauseke (39) sijoitetaan yhtälöön (38), saadaan johdettua yksityiskohtaisempi muoto

$$u(x, t) = A_0 e^{-iut} e^{ikx}, \quad (40)$$

jossa  $A_0$  on mielivaltainen vakio. Ns. eksakti vahvennuskerroin (engl. exact amplification factor) määritellään seuraavasti:

$$G_E \equiv \frac{u(x, t + \Delta t)}{u(x, t)} = \frac{A_0 e^{-iu(t+\Delta t)} e^{ikx}}{A_0 e^{-iut} e^{ikx}} = e^{-iu\Delta t}. \quad (41)$$

Jatkossa seuraavien kaavojen pelkistämiseksi määritellään lisäksi dimensioton aaltoluku

$$\beta = kh \quad (42)$$

ja Courantin luku

$$co = \frac{a\Delta t}{h}, \quad (43)$$

jossa  $h$  on elementin pituus. Suure (41) voidaan kirjoittaa muotoon

$$G_E = e^{-ico\cdot\beta}. \quad (44)$$

Jatkoa silmälläpitäen vahvennuskerroin esitetään vielä vaihtoehtoisessa muodossa

$$G_E = |G_E| e^{i\phi_E} \quad (45)$$

jossa itseisarvo

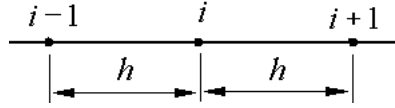
$$|G_E| = 1$$

ja vaihekulma



$$\phi_E = -\text{co} \cdot \beta \quad (46)$$

Seuraavaksi tarkastellaan diskreettejä ratkaisuja. Rajoitutaan soveltamaan vain kaksisolmuisia (linearisia) elementtejä. Kuva 1 esittää vastaavan säännöllisen elementtiverkon solmun  $i$  lähiympäristöä.



Kuva 1. Osa säännöllistä elementtiverkkoa.

Kuvan 1 solmuun  $i$  liittyvä diskreetti yhtälö saa formulaatiossa I muodon (yhtälön (17) erikoistapaus)

$$\begin{aligned} & \left( \frac{h}{6(\Delta t)^2} - \frac{a^2}{4h} \right) u_{i-1}^{n+1} + \left( \frac{2h}{3(\Delta t)^2} + \frac{a^2}{2h} \right) u_i^{n+1} + \left( \frac{h}{6(\Delta t)^2} - \frac{a^2}{4h} \right) u_{i+1}^{n+1} \\ &= \left( \frac{h}{6(\Delta t)^2} + \frac{u}{2\Delta t} + \frac{u^2}{4h} \right) \phi_{i-1}^n + \left( \frac{2h}{3(\Delta t)^2} - \frac{u^2}{2h} \right) \phi_i^n \\ & \quad + \left( \frac{h}{6(\Delta t)^2} - \frac{u}{2\Delta t} + \frac{u^2}{4h} \right) \phi_{i+1}^n. \end{aligned} \quad (47)$$

Formulaatiossa II saadaan yhtälö (yhtälön (29) erikoistapaus)

$$\begin{aligned} & \left( \frac{a}{2\Delta t} - \frac{a^2}{2h} \right) u_{i-1}^{n+1} + \frac{a^2}{h} u_i^{n+1} + \left( -\frac{a}{2\Delta t} - \frac{a^2}{2h} \right) u_{i+1}^{n+1} \\ &= \left( \frac{a}{2\Delta t} + \frac{a^2}{2h} \right) u_{i-1}^n - \frac{a^2}{h} u_i^n + \left( -\frac{a}{2\Delta t} + \frac{a^2}{2h} \right) u_{i+1}^n. \end{aligned} \quad (48)$$

Formulaatiossa G saadaan yhtälö (yhtälön (37) erikoistapaus)

$$\begin{aligned} & \left( \frac{h}{6\Delta t} - \frac{a}{4} \right) u_{i-1}^{n+1} + \frac{2h}{3\Delta t} u_i^{n+1} + \left( \frac{h}{6\Delta t} + \frac{a}{4} \right) u_{i+1}^{n+1} \\ &= \left( \frac{h}{6\Delta t} + \frac{a}{4} \right) u_{i-1}^n + \frac{2h}{3\Delta t} u_i^n + \left( \frac{h}{6\Delta t} - \frac{a}{4} \right) u_{i+1}^n. \end{aligned} \quad (49)$$

Von Neumannin analyysissä otetaan lauseke

$$\hat{u}(x, t) = \hat{A}_0 e^{\alpha t} e^{ikx}, \quad (50)$$

joka on samantyyppinen kuin (40). Paikkatermi  $\exp ikx$  on sama kuin lausekkeessa (40) mutta aikatermiä  $\exp \alpha t$  ei ole vielä kiinnitetty. Lausekkeen (50) arvo lasketaan yhtälöiden (47), (48) and (49) solmuissa. Tekemällä nämä sijoitukset havaitaan, että kyseisistä yhtälöistä voidaan määrittää suureen  $\exp \alpha \Delta t$  arvo. Kirjallisuudessa tätä suuretta, joka esiintyy myös määritelmässä

$$G \equiv \frac{\hat{u}(x, t + \Delta t)}{\hat{u}(x, t)} = \frac{\hat{A}_0 e^{\alpha(t+\Delta t)} e^{ikx}}{\hat{A}_0 e^{\alpha t} e^{ikx}} = e^{\alpha \Delta t} \quad (51)$$

nimitetään algoritmiseksi vahvennuskertoimeksi (engl. algorithmic amplification factor)  $G$  voidaan esittää vastaavasti kuin  $G_E$  muodossa

$$G = |G| e^{i\theta}. \quad (52)$$

Tietyssä mielessä diskretoinnin “hyvyyttä” voidaan mitata tarkastelemalla kuinka lähellä suure  $G$  on suuretta  $G_E$ .

Suorittamalla laskelmat (nämä olisivat käsin laskien melko työläitä, mutta hoituvat tässä helposti Mathematica-ohjelmiston [9] avulla) saadaan formulaatiossa I

$$G \equiv G_I = \frac{4 - 3\text{co}^2 + (2 + 3\text{co}^2)\cos\beta - 6i\text{co}\cdot\sin\beta}{4 + 3\text{co}^2 + (2 - 3\text{co}^2)\cos\beta}, \quad (53)$$

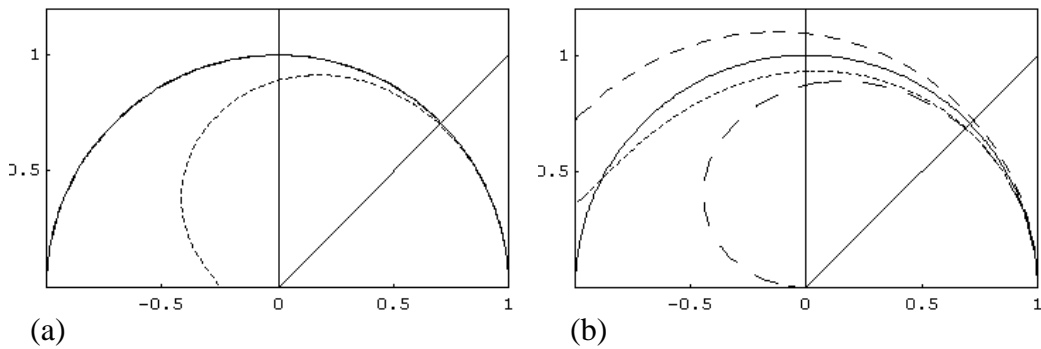
formulaatiossa II

$$G \equiv G_{II} = \frac{1 + \text{co} + (1 - \text{co})e^{i\beta}}{1 - \text{co} + (1 + \text{co})e^{i\beta}} \quad (54)$$

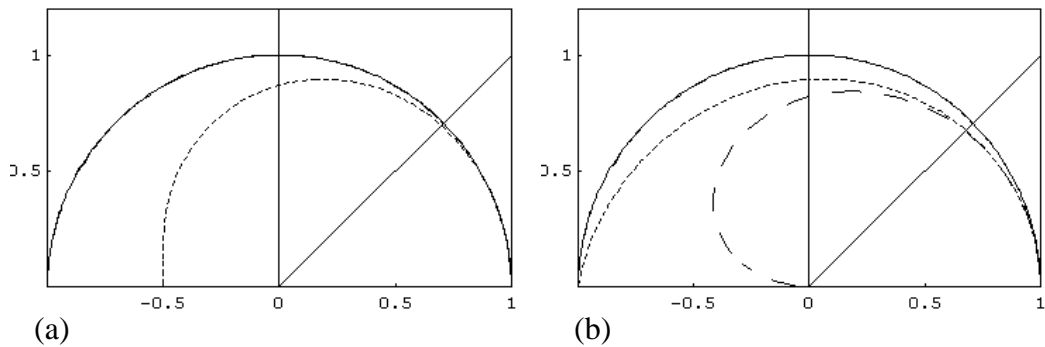
ja formulaatiossa G

$$G \equiv G_G = \frac{4 + 2\cos\beta - 3i\text{co}\cdot\sin\beta}{4 + 2\cos\beta + 3i\text{co}\cdot\sin\beta}. \quad (55)$$

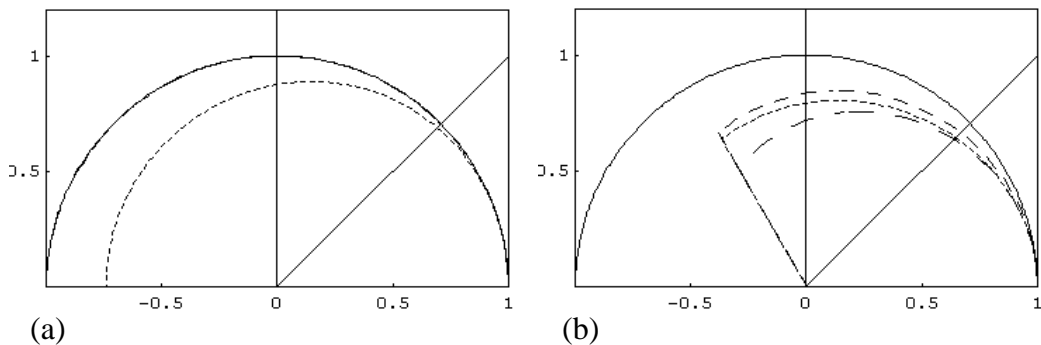
Joitakin saatuja tuloksia on esitetty graafisessa muodossa kuvissa 2, 3 and 4. Esitystapa on tyyppiä, jota on käytetty viitteessä [6]. Viitteessä [4] on todettu tämänäpapisistä diagrammeista osuvasti seuraavaa: ”These diagrams can be viewed as representing the fingerprints of the various schemes, in that they allow identifying the different approximations and recognizing their mutual relationships.” Käyrät on esitetty napakoordinaatistossa napakulman  $\beta$  funktiona (tässä siis dimensioton aaltoluku  $\beta$  on napakoordinaatiston napakulma eikä kompleksisen vahvennuskertoimen vaihekulma  $\theta$ ). Säteen pituutena on suhde  $|G_{DIS}|/|G_E|$  tai  $\theta_{DIS}/\theta_E$ . DIS viittaa indekseihin I, II tai G. Tässä suhteita  $|G_{DIS}|/|G_E|$  ja  $\theta_{DIS}/\theta_E$  tullaan nimittämään vastaavasti amplituudisuhteeksi ja vaihesuhteeksi. Käyrät on katkoviivoitettu siten, että formulaation I viivoitustiheys on suurin ja formulaation G pienin. Yhtenäinen puoliympyrä, jolla on säteenä arvo 1, kuvaa virheetöntä tulosta. Viiva  $\beta = \pi/4$  esittää tiettyä rajaa. Tarkkuuden kannalta nimittäin karkeasti vain alue  $0 \leq \beta \leq \pi/4$  on kiinnostava, koska arvo  $\beta = \pi/4$  vastaa tapausta, jossa aallonpituutta kohti on kahdeksan elementtiä, [3]. Suuremmilla  $\beta$ :n arvoilla verkko ei enää kykene kunnolla kuvaamaan paikan suhteen voimakkaammin vaihtelevaa käyttäytymistä. Kuitenkin etenkin amplituudisuhteen tarkastelu arvoon  $\beta = \pi$  asti on tärkeää formulaation stabiiliuden selvittämisen kannalta. Kuvat 2, 3 ja 4 on esitetty vastaavasti Courantin luvun arvoilla  $\text{co} = 0.75$ ,  $\text{co} = 1.0$  ja  $\text{co} = 1.5$ .



Kuva 2.  $co = 0.75$ . (a) Amplituudisuhde. (b) Vaihesuhde.



Kuva 3.  $co = 1.0$ . (a) Amplituudisuhde. (b) Vaihesuhde.



Kuva 4.  $co = 1.5$ . (a) Amplituudisuhde. (b) Vaihesuhde.

Diagrammeista voidaan todeta yleisesti seuraavaa. Amplituudisuhteen poikkeaminen arvosta yksi (alle yhden) merkitsee diskreetin ratkaisun amplituudien epärealistista vaimenemista ajan mukana. Päinvastainen poikkeaminen (yli yhden) merkitsee diskreetin formulaation olevan luultavasti epästabiili; amplituudit kasvavat holtittomasti. Vaihesuhteen poikkeaminen arvosta yksi merkitsee diskreetin ratkaisun siirtyvän paikan suhteen ajan mukana eri nopeudella kuin tarkka ratkaisu. Jos arvo on yli (alle) yhden, diskreetti ratkaisu liikkuu nopeammin (hitaammin) kuin tarkka ratkaisu.

Lisäksi tässä; ensinnäkin formulaatiot II ja G antavat aina oikean amplituudisuhteen. Formulaatio I sisältää amplituudin vaimennusta. Amplituudisuhteen arvo ei ylitä arvoa yksi. Toiseksi, formulaatiot sisältävät vaihekulmavirheellisyttä paitsi formulaatio II, joka sattuu antamaan oikean arvon tapauksessa  $co = 1.0$ . Kun  $co = 1.5$ ,

vaihekulmavirheet ovat kaikilla formulaatioilla jo verrattain suuria. Kolmanneksi, käyrät kuvassa 4 (b) loppuvat arvoon  $\beta = 2\pi/3$ , koska Mathematica-ohjelmisto rajoittaa kompleksiluvun vaihekulman välille  $[-\pi, \pi]$  ja tämä heijastuu kuvan piirtämisessä.

## Käytännön näkökohtia

Insinööriyössä on normaalisti helpointa käyttää elementtimenetelmän sovelluksissa valmisohjelmistoja. Niiden laatimiseen ja ylläpitoon on panostettu resursseja, joihin ei ole yksilötasolla tietenkään mitään käytännön edellytyksiä. Seuraavassa luvussa esitetyt kaksi esimerkkitapausta ovat kuitenkin niin pelkistettyjä, että niiden laskenta on onnistunut vielä siedettävällä työmäärällä soveltaen ”kotitekoisia” Mathematica ohjelmistoon [9] perustuvia ohjelmia. Osa esimerkkilaskelmista on suoritettu vertailutulosten saamiseksi myös laajalti käytössä olevan Comsol Multiphysics ohjelmiston (seuraavassa lyhyesti CM-ohjelmisto) [8] avulla. CM-ohjelmistossa on arvokkaana ominaisuutena mahdollisuus syöttää melko helposti ratkaistavaksi elementtimenetelmän avulla miltei mielivaltaisia differentiaaliyhtälösystemejä.

Soveltaja voi antaa CM-ohjelmistossa differentiaaliyhtälöitä koskevan ongelmansa kolmella perustavalla, joista yksi liittyy juuri heikkoihin muotoihin. Tätä versiota on käytetty hyväksi. Osoittautuu, että formulaation II ja G käyttö onnistuu (tiettyyn pisteeseen asti) juohevasti; lähdetään liikkeelle vastaavasti heikoista muodoista (24) ja (35). Sen sijaan formulaation I käsittely menee kömpelöksi. Selityksenä tähän on, että CM-ohjelmisto työstää aikariippuvia probleemeja suoraan vain MOL-menettelyllä. Täten muodot (24) ja (35) sopivat alustavasti hyvin lähtökohdiksi. Esitetty pienimmän neliön versio II (mahdollisine yleistyksineen) tuntuu siis käytännössä versiota I käyttökelpoisemmalta ainakin CM-ohjelmistoa käytettäessä.

## Esimerkkituloksia

### *Yleistä*

Numeerisessa käsittelyssä on syytä siirtyä dimensiottomiin muuttujiin. Otetaan käyttöön jokin karakteristinen suureen  $u$  arvo  $u_c$  (esimerkiksi alku- tai reunaehtoihin liittyen), jokin karakteristinen pituus  $l_c$  kuten esimerkiksi tarkastelualueen pituus  $L$  ja jokin karakteristinen aika  $t_c$  kuten esimerkiksi  $L/a_c$ , jossa  $a_c$  on jokin karakteristinen konvektionopeuden arvo. Operoidaan sitten dimensiottomilla muuttujilla  $\hat{u} = u/u_c$ ,  $\hat{x} = x/l_c = x/L$ ,  $\hat{t} = t/t_c = ta_c/L$ . Yhtälö (1) tulee ketjuderivointia käyttäen ensin muotoon

$$\frac{\partial(u_c \hat{u})}{\partial \hat{t}} \frac{a_c}{L} + a \frac{\partial(u_c \hat{u})}{\partial \hat{x}} \frac{1}{L} - s = 0. \quad (56)$$

Sieventämällä päädytään yhtälöön

$$\frac{\partial \hat{u}}{\partial \hat{t}} + \frac{a}{a_c} \frac{\partial \hat{u}}{\partial \hat{x}} - \hat{s} = 0, \quad (57)$$

jossa dimensioton lähde-termi  $\hat{s} = sL/(u_c a_c)$ . Reuna- ja alkuehdot voidaan manipuloida vastaavaan tapaan dimensiottomiksi.

Kirjallisuudessa ehkä tavanomaisin käytäntö on siirtyä mutkattomasti ilman suurempia selittelyjä suoraan yhtälöihin, joissa näkyy vain puhtaita dimensiottomia muuttujia, mutta symbolit ovat samoja kuin dimensiollisissa muodoissa. Korrektimmin tähän päädytään suorittamalla ensin yllä esitetyn tapaiset askeleet ja vaihtamalla sitten dimensiottomien suureiden merkinnät sopivasti (poistetaan ylä-tunnus  $\wedge$ ). Tässä tullaan lisäksi käsittelemään vain tapausta, jossa konvektionopeus  $a$  on (positiivinen) vakio. Kun valitaan vielä  $a_c = a$ , päädytään siten esimerkkiyhtälöön

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} - s = 0, \quad (58)$$

jossa konvektionopeus on 1 ja jossa laskenta-alue paikan suhteen on  $(0,1)$ .

Tarkkojen ratkaisujen suhteen todettakoon, että jos lähde-termi  $s$  häviää, suureen  $u$  arvo on vakio suorilla (karakteristikoilla), jotka muodostavat kulman 45 astetta  $x$ - ja  $t$ - akselien suhteen ja arvo määräytyy siis vastaavasta alku- tai reunaehdosta. Jos lähde-termi on nolasta eroava, ratkaisu saadaan integroimalla pitkin karakteristikoita.

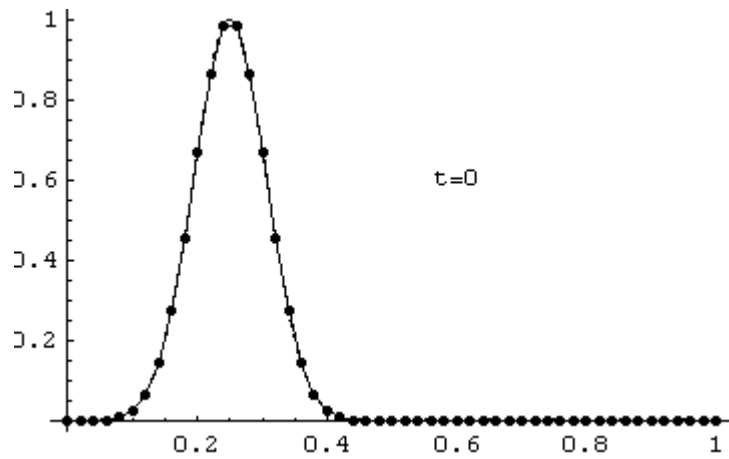
### ***Ensimmäinen esimerkki***

Esimerkki on lainattu viitteestä [7]. Lähde-termi on nolla. Alkuehdossa

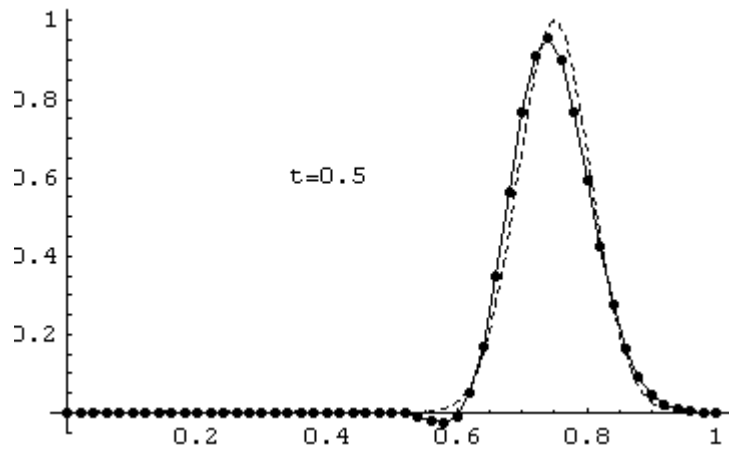
$$u_0 = e^{-10(4x-1)^2}. \quad (59)$$

Kyseessä on kellokäyrätyyppinen funktio, jonka symmetriakohta on pisteessä  $x = 0.25$  ja jonka amplituudi on 1. Reunaehdossa  $u_D = 0$ . Teoriassa — mutta ei käytännössä — tässä on pientä epäkonsistenssia, koska alkuehto antaa pisteissä  $x = 0$  hieman nollasta eroavan arvon. Tarkka ratkaisu on nopeudella 1 ajan mukana oikealle siirtyvä alkuehdon mukainen funktio.

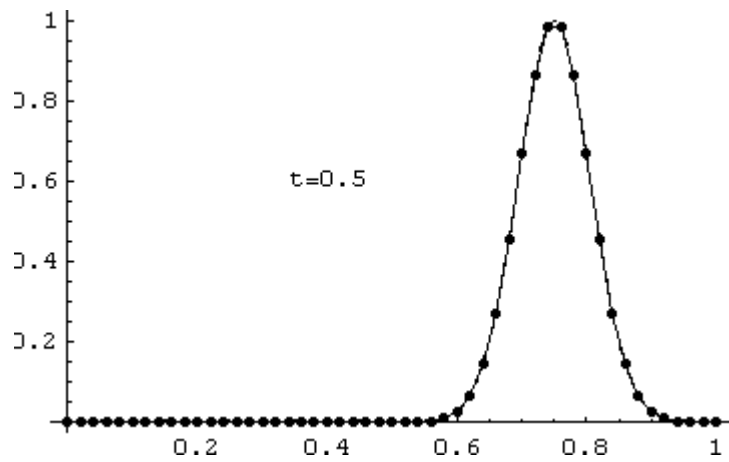
Käytetään tasaista elementtiverkkoa (50 elementtiä), joten  $h = 0.02$ . Samoin kuten lähteessä [7] Courantin luvun arvoksi valitaan  $co = 1$  eli  $\Delta t = 0.02$ . Diskreetti alkuehto otetaan valitsemalla solmuarvot lausekkeesta (59). Kuvassa 5 on esitetty alkutilanne hetkellä  $t = 0$  ja kuvissa 6, 7 ja 8 formulaatioilla I, II ja G saadut ratkaisut hetkellä  $t = 0.5$ . Tarkan ratkaisun kuvaaja on esitetty katkoviivoituksella.



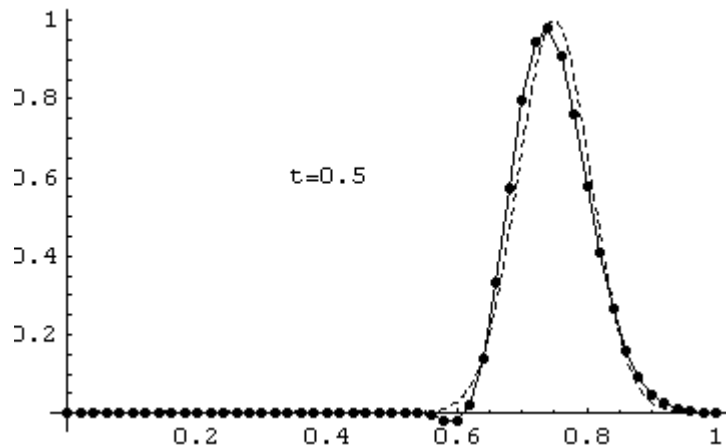
Kuva 5. Ratkaisu hetkellä  $t = 0$ .



Kuva 6. Formulaatio I. Ratkaisu hetkellä  $t = 0.5$ .



Kuva 7. Formulaatio II. Ratkaisu hetkellä  $t = 0.5$ .



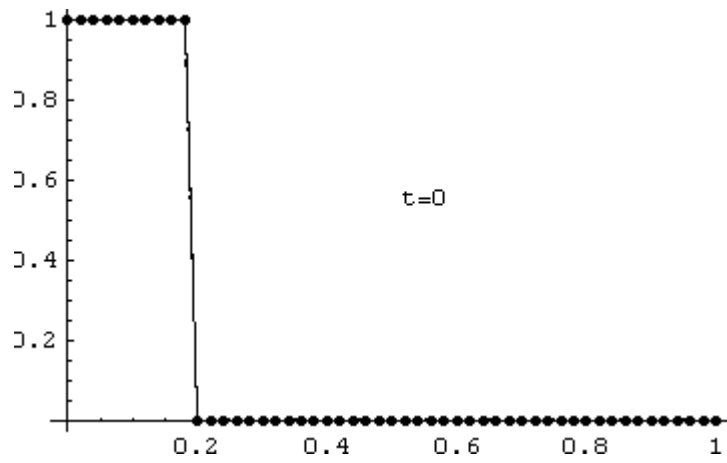
Kuva 8. Formulaatio G. Ratkaisu hetkellä  $t = 0.5$ .

Saadut tulokset ovat kohtuullisessa sopusoinnussa kuvan 3 sormenjälkien kanssa. Kuvan 6 käyrässä on nähtävissä lievää amplituudin alenemista ja samoin ehkä siirtymänopeuden alenema. Kuvan 7 ratkaisu yhtyy solmuissa tarkkaan ratkaisuun. Kuvan 8 tapauksessa on jälleen havaittavissa siirtymänopeuden alenema. CM-ohjelmisto antaa Generalized- $\alpha$  aika-askellusta käyttäen käytännössä kuvien 7 ja 8 mukaiset tulokset.

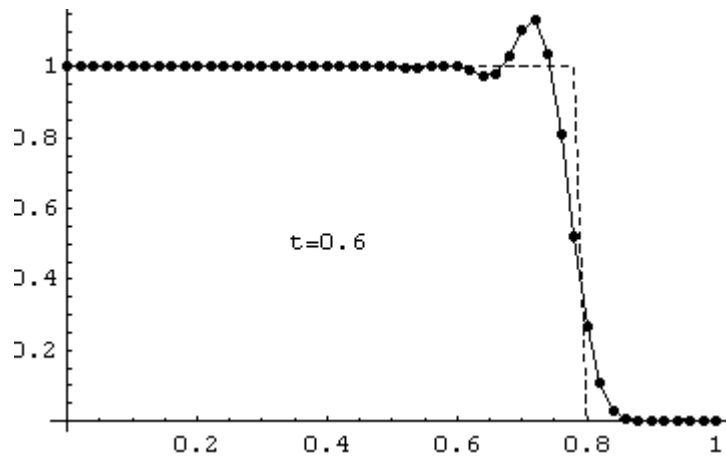
### *Toinen esimerkki*

Tämä esimerkki on lainattu viitteestä [3]. Tehtävä on muuten samantyyppinen kuin edellinen tapaus ( $s = 0$ ), mutta alkuehdossa esiintyy jyrkkä porraskorkeus kohdassa  $x = 0.2$ . Käytetään jälleen tasaista verkkoa  $h = 0.02$ . Termi  $u_0 = 1$ , kun  $0 \leq x \leq 0.18$ , laskee lineaarisesti arvoon 0, kun  $x = 0.2$  ja on 0, kun  $0.2 \leq x \leq 1$ . Reunaehdossa  $u_D = 1$ . Tarkka ratkaisu on nopeudella 1 oikealle etenevä korkeuden 1 omaava jyrkkä porraskorkeus. Tehtävä on vaativa jyrkän muutoksen kuvaamisen suhteen.

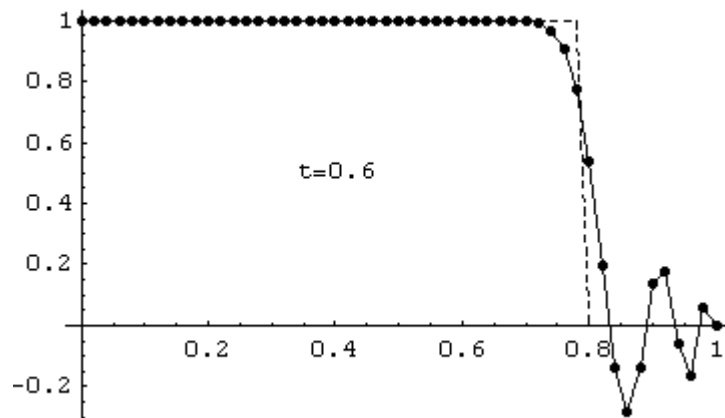
Courantin luvun arvoksi otetaan lähteen [3] mukaisesti  $co = 0.75$ , joka merkitsee, että  $\Delta t = 0.015$ . Kuvassa 9 on esitetty alkutilanne hetkellä  $t = 0$  ja kuvissa 10, 11 ja 12 formulaatioilla I, II ja G saadut ratkaisut hetkellä  $t = 0.6$ . Tarkan ratkaisun kuvaaja on esitetty katkoviivoituksella.



Kuva 9. Ratkaisu hetkellä  $t = 0$ .

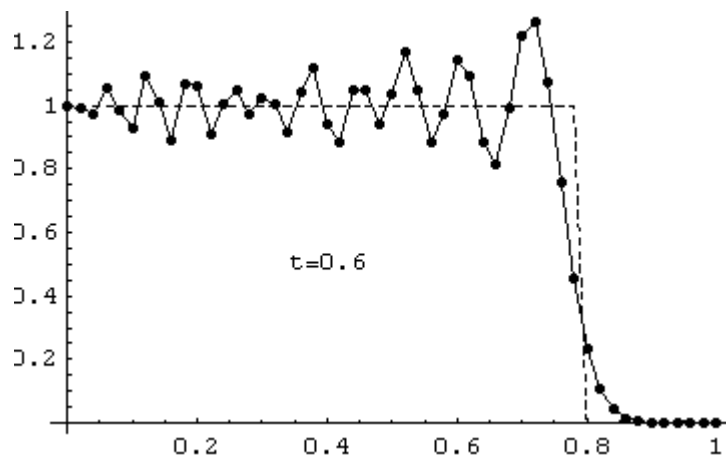


Kuva 10. Formulaatio I. Ratkaisu hetkellä  $t = 0.6$ .



Kuva 11. Formulaatio II. Ratkaisu hetkellä  $t = 0.6$ .





Kuva 12. Formulaatio G. Ratkaisu hetkellä  $t = 0.6$ .

Kuvan 2 sormenjälkien perusteella on nyt melko vaikea tehdä mitään selviä johtopäätöksiä. Galerkinin menetelmässä porras synnyttää laajalle alueella ulottuvaa heilahtelua. Pienimmän neliön menetelmissä heilahtelua esiintyy myös, mutta se vaimenee paikan suhteen. Tämä on ehkä ymmärrettävissä Eulerin yhtälöiden yhteydessä esitetyn kommentin perusteella. Portaen eksaktimpi approksimaatio vaatisi jo kirjallisuudessa esitettyjä epälineaarisia lähestymistapoja kuten esimerkiksi ns. TVD-menetelmää; esimerkiksi [3]. CM-ohjelmisto antaa käytännössä kuvien 11 ja 12 mukaiset tulokset.

## Loppuhuomautuksia

Pienimmän neliön formulaatiolla on tärkeänä periaatteellisena etuna, että siinä on mielivaltaisen differentiaaliyhtälösystemin yhteydessä aina käytettävissä automaattisesti variaatioperiaate: pienimmän neliön funktionaalinen stationaarisuusehto. Toisaalta voimakkaasti hankaloittavana piirteenä on, että jos vallitsevilla differentiaaliyhtälöissä esiintyy toisen (tai korkeamman) kertaluvun derivaattoja, tavanomaisia  $C^0$ -jatkuvia muotofunktioina ei voida käyttää. Ongelma voidaan kiertää määrittelemällä uusia muuttujia ja lisädifferentiaaliyhtälöitä. Vaadittava teoria ja käytäntö tulevat kuitenkin verrattain mutkikkaiksi. Näitä aiheita käsitellään teoksessa [2].

Tässä artikkelissa tarkasteltu differentiaaliyhtälö on ollut erityisen yksinkertainen ja ei ole sisältänyt toisen kertaluvun derivaattaa, joten pienimmän neliön elementtimenetelmän soveltaminen on ollut suoraviivaista. Lainataan vielä viitettä [3, s. 92]: “Which discretization is performed first is not an issue for linear spatial operators with constant coefficients and a Galerkin formulation. However, for future cases (nonlinear problems, stabilization techniques for transient analysis, etc.) it is preferable that the time discretization is performed before the spatial discretization.” Olemme havainneet edellä, että kun pienimmän neliön elementtimenetelmää sovelletaan lineaariseen tehtävään, saadaan kuitenkin diskretointijärjestyksestä riippuen melko erilaiset diskreetit formulaatiot I ja II. Ensin paikka- sitten aikadiskretointijärjestys tuotti formulaation II, jota voitaneen pitää von Neumannin analyysin ja esitettyjen (tosin

rajoittuneiden) esimerkkilaskelmien perusteella verrattavissa tarkkuudeltaan formulaatioon I. Formulaatio II antaa siten käytännössä mahdollisesti helpomman vaihtoehdoisen lähestymistavan valmisohjelmien yhteydessä. Ilmeisesti mikään ei ole teoriassa esteenä yrittää soveltaa ehdotettua formulaatiota II myös mutkikkaampiin aikariippuviin tehtäviin.

## Viitteet

- [1] D A. Anderson, J. C. Tannehill, R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer*, McGraw-Hill, 1984.
- [2] P. B. Bochev, M. D. Gunzburger, *Least-Squares Finite Element Methods*, Springer, 2009.
- [3] J. Donea, A. Huerta, *Finite Element Methods for Flow Problems*, Wiley, 2003.
- [4] J. Donea, L. Quartapelle, An introduction to finite element methods for transient advection problems, *Computer Methods in Applied Mechanics and Engineering*, 1992: **95**:169-203.
- [5] C.A. J. Fletcher, *Computational Galerkin Methods*, Springer, 1984.
- [6] A. Huerta, J. Donea, Time-accurate solution of stabilized convection-diffusion-reaction equations: I—time and space discretization, II—accuracy analysis and examples, *Communications in Numerical Methods in Engineering*, 2002, **18**:565-573, 575-584.
- [7] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations*, Cambridge University Press, 1994.
- [8] [www.comsol.com](http://www.comsol.com)
- [9] [www.wolfram.com/products/mathematica](http://www.wolfram.com/products/mathematica)

Jorma Kinnunen  
Metso Paper Inc.  
Wärtsilänkatu 100  
04400 Järvenpää  
[jorma.kinnunen@metso.com](mailto:jorma.kinnunen@metso.com)

Eero-Matti Salonen  
Aalto-yliopiston insinöörityötekniikan korkeakoulu  
Rakenne- ja rakennustuotantotekniikan laitos  
PL 2100, 02015 TKK  
[eero-matti.salonen@tkk.fi](mailto:eero-matti.salonen@tkk.fi)