Tidsstegning 1

Implicit (backward) Euler: Har betraktat ekvationen

$$\dot u - \nabla \cdot a \nabla u = f,$$

med $a = a(x)$ och $f = f(x,t)$, som efter multiplikation med $v = v(x)$ och integration i tid och rum och partiell integration i rumsled kunde skrivas

$$\int_\Omega v \underbrace{\int_{I_n} \dot u}_{u_n - u_{n-1}} + \int_\Omega \nabla v \cdot a \nabla \underbrace{\int_{I_n} u}_{\approx k_n u_n} = \int_\Omega v \underbrace{\int_{I_n} f}_{\approx k_n f_n},$$

där $u_n = u(x,\, t_n)$ och $f_n = f(x,\, t_n)$.

Tidsstegning 2

Med $u_n \to U_n = \sum_{j=1}^m U_{n,j}\, \phi_j$ och $v = \phi_i$ gav detta diskretiseringsmetoden

$$M\,(U_n - U_{n-1}) + k_n\, A\, U_n = k_n\, F_n,$$

där alltså $U_n \approx u_n$, $k_n = t_n - t_{n-1}$ är tidssteget, M är massmatrisen, A styvhets/diffusionsmatrisen, och F_n är lastvektorn med element $\int_\Omega \phi_i\, f_n$.

Efter multiplikation med M^{-1} kan detta skrivas

$$U_n + k\, \tilde a\, U_n = U_{n-1} + k\, \tilde f,$$

där $k = k_n$, $\tilde a = M^{-1} A$, och $\tilde f = M^{-1} F_n$.

Tidsstegning 3

För enkelhets skull betraktar vi fortsättningsvis motsvarande *skalära* ekvationen

$$\dot u + a\, u = f,$$

där a är en given konstant och $f = f(t)$. Motsvarande tidsdiskretiseringsmetod för beräkning av $U_n \approx u_n = u(t_n)$ blir

$$U_n + k\, a\, U_n = U_{n-1} + k\, f_n,$$

dvs

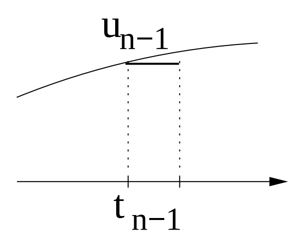
$$U_n = (1 + k\, a)^{-1}\, (U_{n-1} + k\, f_n),$$

där $f_n = f(t_n)$.

Tidsstegning 4

Explicit (forward) Euler: Vi erinrar oss också att motsvarande approximationer

$$\underbrace{\int_{I_n} \dot u}_{u_n - u_{n-1}} + a \underbrace{\int_{I_n} u}_{\approx k\, u_{n-1}} = \underbrace{\int_{I_n} f}_{\approx k\, f_{n-1}},$$



Tidsstegning 5

leder till diskretiseringsmetoden

$$U_n - U_{n-1} + k\,a\,U_{n-1} = k\,f_{n-1},$$

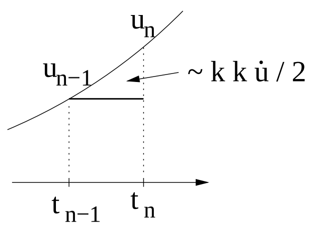
dvs

$$U_n = (1 - k\,a)\,U_{n-1} + k\,f_{n-1},$$

allmänt känd som Explicit (forward) Euler.

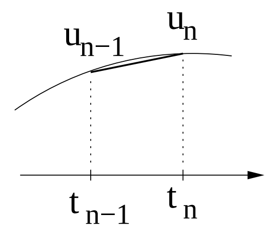
Tidsstegning 6

De båda Euler metoderna kan förväntas vara jämförbara vad gäller *noggrannhet*. De fel vi introducerar genom att ersätta $\int_{I_n} u$ och $\int_{I_n} f$ med $k\,u_n$ och $k\,f_n$ resp $k\,u_{n-1}$ och $k\,f_{n-1}$ är ju i båda fallen av storleksordning k^2, dvs felet efter *en tidsenhet*, dvs efter k^{-1} tidssteg, kan väntas ha accumulerats till storleksordning k.



Tidsstegning 7

Crank-Nicolson: Uppenbarligen finns möjligheten att approximera integralerna $\int_{I_n} u$ och $\int_{I_n} f$ bättre än ovan med t.ex. *trapetsregeln*:



Tidsstegning 8

dvs

$$\underbrace{\int_{I_n} \dot{u}}_{u_n - u_{n-1}} \; + \; a \underbrace{\int_{I_n} u}_{\approx k\,(u_{n-1}+u_n)/2} \; = \; \underbrace{\int_{I_n} f}_{\approx k\,(f_{n-1}+f_n)/2} \; ,$$

vilket leder till diskretiseringsmetoden

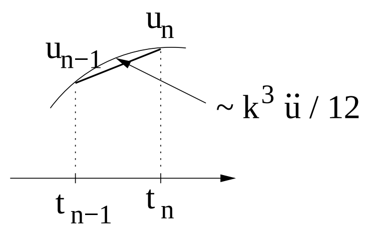
$$U_n - U_{n-1} + k\,a\,(U_{n-1} + U_n)/2 = k\,(f_{n-1} + f_n)/2,$$

dvs

$$U_n = (1 + \frac{k}{2}\,a)^{-1}(U_{n-1} - \frac{k}{2}\,a\,U_{n-1} + k\,(f_{n-1} + f_n)/2).$$

Tidsstegning 9

Felet för denna metod bör vara av storleksordning k^3 per tidssteg, dvs av storleksordning k^2 efter en tidsenhet.



Tidsstegning 10

För $f = 0$ reduceras de tre metoderna till:

$$U_n = (1 + k\,a)^{-1}\,U_{n-1}$$
$$U_n = (1 - k\,a)\,U_{n-1}$$
$$U_n = (1 + \tfrac{k}{2}\,a)^{-1}\,(1 - \tfrac{k}{2}\,a)\,U_{n-1}$$

Tidsstegning 11

Motsvarande metoder för värmeledningsproblemet ovan är

$$(M + k\,A)\,U_n = U_{n-1}$$
$$M\,U_n = (M - k\,A)\,U_{n-1}$$
$$(M + \tfrac{k}{2}\,A)\,U_n = (M - \tfrac{k}{2}\,A)\,U_{n-1}$$

Vi noterar att alla dessa metoder kräver *ekvationslösning*, men om massmatrisen M i vänsterledet ersätts med motsvarande *lumpade massmatris* $\bar{M}$ så kan metod II skrivas

$$U_n = \bar{M}^{-1}(M - k\,A)\,U_{n-1},$$

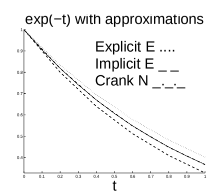
dvs metoden blir *explicit*, dvs man erhåller utan ekvations-lösning en *formel* för U_n, därav namnet *explicit Euler*.

Tidsstegning 12

Egenskaper: Inleder med att studera modellproblemet

$$\dot{u} + u = 0 \quad t > 0, \quad u(0) = 1,$$

med exakt lösning $u(t) = \exp(-t) = 1 - t + \frac{1}{2}t^2 - \frac{1}{6}t^3 + \ldots$



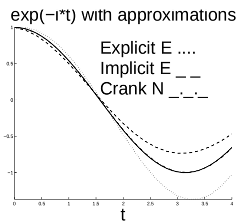
Man noterar inga väsentliga skillnader mellan Euler metoderna, medan Crank-Nicolsson som väntat är mera exakt.

Tidsstegning 13

Skillnaden mellan metoderna blir mera uppenbar om vi betraktar ett par andra *typ-problem*. Vi börjar med

$$\dot{u} + i\,u = 0 \quad t > 0, \quad u(0) = 1,$$

med (komplexvärd) lösning $u(t) = \cos(t) + i\sin(t)$, med realdel och imaginärdel motsvarande lägeskoordinat resp hastighet hos en odämpad svängande massa med massan 1 upphängd i en fjäder med fjäderkonstant 1.

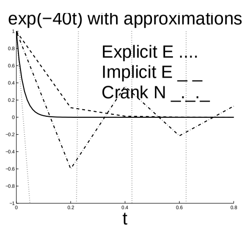


Tidsstegning 14

Slutligen betraktar vi ett *styvt* problem

$$\dot{u} + 40\,u = 0 \quad t > 0, \quad u(0) = 1,$$

med lösning $u(t) = \exp(-40\,t)$.



Vi noterar att Crank-Nicolsson nu uppför sig konstigt. Implicit Euler är bättre! Explicit Euler är helt ute!

Tidsstegning 15

Vi erinrar oss att (den preliminära) felanalysen för CN indikerade ett fel av storleksordning $k^3\,\dddot{u}/12$ per tidssteg, dvs $0.2^3\,40^2/12 \approx 1.1$ i det aktuella fallet, så resultatet borde kanske inte vara överraskande, men *oscillationen* är slående!

För implicit Euler är resultatet måhända något bättre än väntat, eftersom (den preliminära) analysen pekade på ett möjligt fel av storleksordning $k^2\,\ddot{u}/2 = 0.8$.

Tidsstegning 16

Vi kan förstå uppförandet hos de tre metoderna genom att studera ett enskilt tidssteg av längd 0.2.

$$U_n = (1 + 0.2\,40)^{-1}\,U_{n-1} = 1/9\,U_{n-1}$$
$$U_n = (1 - 0.2\,40)\,U_{n-1} = -7\,U_{n-1}$$
$$U_n = (1 + \tfrac{0.2}{2}\,40)^{-1}\,(1 - \tfrac{0.2}{2}\,40)\,U_{n-1} = -3/5\,U_{n-1}$$

med faktorer framför U_{n-1} som ju bör approximera $\exp(-k\,a) = \exp(-8)$ (eftersom för den exakta lösningen $u_n = \exp(-a\,t_n) = \exp(-a\,(t_{n-1} + k) = \exp(-k\,a)\exp(-k\,t_{n-1}) = \exp(-k\,a)\,u_{n-1}$) men som för explicit Euler och CN blir negativa, och för Euler dessutom blir stor vilket leder till instabilitet! Problemet ligger i att kvantiteten $k\,a$ här inte är tillräckligt "liten".

Tidsstegning 17

Notera att faktorn $(1 - k\,a)$ utgör de två första termerna i Taylor approximationen av exp$(-k\,a)$, som ger en okey approximation om (och endast om i detta fall) $k\,a$ är “liten”. Analogt utgör förstås faktorn $(1 + k\,a)^{-1} = 1 - k\,a + k^2\,a^2 - k^3\,a^3 + ..$ motsvarande implcit Euler en okey approximation till exp$(-k\,a)$ om $k\,a$ är liten. Notera att exp$(-k\,a) = 1 - k\,a + \frac{1}{2}\,k^2\,a^2 -$ ... En analys visar att motsvarande faktor för CN sammanfaller med denna serie upp t.o.m. k^2-termen, vilket än en gång visar att metoden är av ordning k^3 fel i varje tidssteg.

Tidsstegning 18

Vi studerar nu motsvarigheten till CN’s uppförande på ett värmeledningsproblem.

Tidsstegning 19

Samma typ av CN-reaktion kan man se som resultat av alla “plötsliga förändringar” i data. Notera att begynnelsevärdet $u(0) = 1$ i fallet $\dot{u} + 40\,u = f$ kan ses som en stationär lösning motsvarande $f = 40$, och att sedan, vid tiden $t = 0$ lasten f plötsligt ändras till $f = 0$.

Ett sätt att undvika detta uppförande för CN är förstås att helt enkelt undvika problem med denna typ av snabba “ryck”, t.ex. genom att påföra kraft/lastförändringar *succesivt*.

Tidsstegning 20

En intressant kombination av Implicit Euler’s goda dämpande egenskaper och CN högre noggrannhet är att inleda med *två* Implicita Eulersteg, för att därefter fortsätta med CN.

Tidsstegning 21

Andra ordningens tidsderivator: Vi betraktar modelproblemet

$$\ddot{u} + a\,u = 0 \quad t > 0, \quad u(0) = u_0, \quad \dot{u}(0) = v_0,$$

vilket vi skriver om som ett ekvivalent *system* med $v = \dot{u}$:

$$\begin{cases} \dot{u} - v = 0 \\ \dot{v} + a\,u = 0 \end{cases}$$

dvs med $w = [uv]^\perp$ och $A = \begin{bmatrix} 0 & -1 \\ a & 0 \end{bmatrix}$

$$\dot{w} + A\,w = 0 \quad t > 0, \quad w(0) = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}.$$

Tidsstegning 22

Heuns metod: Explicit Euler blir här

$$W_n = W_{n-1} - A\,W_{n-1}.$$

Ett försök till komponentvis implementering skulle kunna se ut:

```
while time < finaltime
 u=u+k*v;
 v=v-k*a*u;
 time=time+k;
end
```

Vad är *fel* med denna implementering?
Fungerar implementeringen? Jämför med en korrekt implementering.

Tidsstegning 23

Galerkin metoder:

Galerkinmetoder bygger på att utifrån en given *lösningsansats*, dvs "försökslösningar" U av viss typ, "testa" sig fram till en lösning U vars *residual* uppfyller lämpliga ortogonalitetsvillkor.

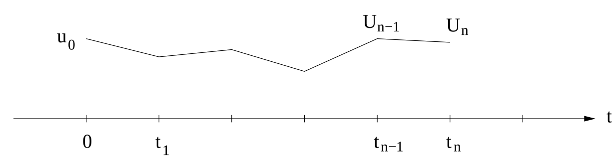
För vår modellekvation $\dot{u} + a\,u = f$ ges residualen av $\dot{U} + a\,U - f$ och ortogonaliteten kommer till uttryck som

$$\int_{I_n} v\,(\dot{U} + a\,U - f) = 0, \quad \text{dvs} \quad \int_{I_n} v\,(\dot{U} + a\,U) = \int_{I_n} v\,f,$$

för alla v av viss typ.

Tidsstegning 24

cG1: Här står c för *continuous*, G för *Galerkin*, och 1 för grad 1, eller *linjär*, dvs utgångspunkten är en *ansats* till lösning på I_n som i figur, med just dessa egenskaper:



och U_n bestäms så att

$$\int_{I_n} \dot{U} + a\,U = \int_{I_n} f,$$

motsvarande residualortogonalitet mot alla *konstanta* $v = v(t)$ på I_n.

Tidsstegning 25

Speciellt erhåll för a konstant

$$\underbrace{\int_{I_n} \dot{U}}_{U_n - U_{n-1}} + \underbrace{\int_{I_n} a\,U}_{a\,k_n\,(U_{n-1}+U_n)/2} = \int_{I_n} f,$$

dvs metoden sammanfaller väsentligen med CN (medelvärde av f i högerledet motsvaras i CN av trapetsapproximationen $k_n\,(f_{n-1}+f_n)/2$). För f linjär i t är metoderna identiska.

Tidsstegning 26

Mera allmänt kan ansatsen U på I_n representeras med hjälp av tidsbasfunktionerna $\psi_{n-1}(t) = (t_n - t)/k_n$ och $\psi_n(t) = (t - t_{n-1})/k_n$ som $U(t) = U_{n-1}\,\psi_{n-1}(t) + U_n\,\psi_n(t)$, vilket ger

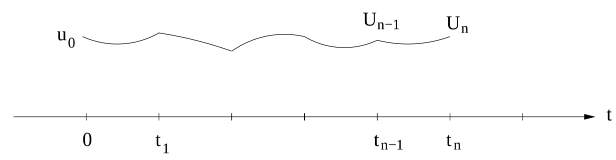
$$U_n - U_{n-1} + \underbrace{\int_{I_n} a\,\psi_{n-1}}_{\frac{k_n}{2}\,\tilde{a}_{n-1}} U_{n-1} + \underbrace{\int_{I_n} a\,\psi_n}_{\frac{k_n}{2}\,\tilde{a}_n} U_n = \int_{I_n} f,$$

vilket för a konstant ju reduceras till

$$U_n - U_{n-1} + \frac{k_n}{2}\,U_{n-1} + \frac{k_n}{2}\,U_n = \int_{I_n} f.$$

Tidsstegning 27

cG2: Analogt utgår cG2 från en styckvis *kvadratisk* ansats:



och U på I_n bestäms av att

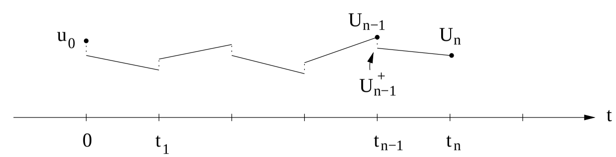
$$\int_{I_n} v\,(\dot{U} + a\,U) = \int_{I_n} v\,f,$$

för $v = \psi_{n-1}$ och $v = \psi_n$, vilket tillsammans med kontinuitetskravet ger tre ekvationer för det kvadratiska polynomet U på I_n.

Tidsstegning 28

dG1: Om vi släpper kontinuitetskravet erhålls s.k. dG metoder, där d för *discontinuous*, och G för Galerkin, som förut.

För dG1 använder vi en ansats som i figur, där vi på det nya intervallet I_n har att bestämma både U_{n-1}^+ och U_n.



Tidsstegning 29

För detta behöver vi två ekvationer och vi väljer

$$\int_{I_n} v\,(\dot U + a\,U) = \int_{I_n} v\,f,$$

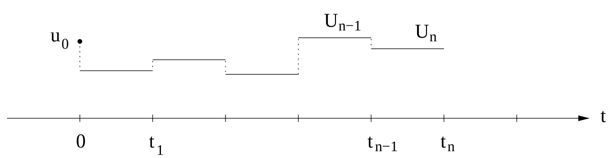
med $v = \psi_{n-1}$ och $v = \psi_n$ som förut. Notera emellertid att nu ingår i $\dot U$ på I_n även derivatan av det initiala hoppet $U_{n-1}^+ - U_{n-1}$, vilket ger bidraget

$$(U_{n-1}^+ - U_{n-1})\,v(t_{n-1})$$

till integralen. Innan vi räknar på ett konkret dG1-exempel noterar vi att även styckvis konstant ansats är möjligt utan kravet på kontinuitet:

Tidsstegning 30

dG0: Här utgår vi alltså från en ansats som i figur med U *styckvis konstant*, med $U_{n-1}^+ = U_n = U(t)$ för alla $t \in I_n$:



Tidsstegning 31

För att bestämma $U_n = U(t)$ på I_n kräver vi nu att

$$\int_{I_n} v\,(\dot U + a\,U) = \int_{I_n} v\,f,$$

för $v = v(t) = 1$, dvs

$$(U_{n-1}^+ - U_{n-1}) + \int_{I_n} (\underbrace{\dot U}_{=0} + a\,U) = \int_{I_n} f,$$

dvs

$$U_n - U_{n-1} + \int_{I_n} a\,U_n = \int_{I_n} f.$$

Vi noterar att för a och f konstanta sammanfaller denna metod med Implicit Euler.

Tidsstegning 32

Vi återgår nu till dG1 med a konstant och $f = 0$. För $v = \psi_{n-1}$ erhålls

$$U_{n-1}^+ - U_{n-1} + \int_{I_n} \psi_{n-1}\,(U_n - U_{n-1}^+) + a \int_{I_n} \psi_{n-1}\,(U_{n-1}^+\,\psi_{n-1} + U_n\,\psi_n$$

dvs

$$U_{n-1}^+ - U_{n-1} + \frac{k}{2}\,(U_n - U_{n-1}^+) + a\,(\frac{k}{3}\,U_{n-1}^+ + \frac{k}{6}\,U_n) = 0,$$

och för $v = \psi_n$

$$\frac{k}{2}\,(U_n - U_{n-1}^+) + a\,(\frac{k}{6}\,U_{n-1}^+ + \frac{k}{3}\,U_n) = 0.$$

Tidsstegning 33

Efter elimination av U_{n-1}^{+} erhålls

$$U_n = (1 + \frac{2\,a\,k}{3} + \frac{a^2\,k^2}{6})^{-1}\,(1 - \frac{a\,k}{3})\,U_{n-1},$$

vilket ger oss en ny (rationell) approximation av av exponentialfaktorn exp$(-a\,k)$.

Tidsstegning 34

A posteriori felanalys för cG1:

Vi betraktar ekvationen $\dot{u} + u = f$, $t > 0$, $u(0) = u_0$ och visar att för cG1 approximationen U till u gäller

$$|(u-U)(T)| \le \max_{[0,T]} |k(f - \dot{U} - U)|.$$

Tidsstegning 35

Vi utgår från *felekvationen*

$$\int_0^T v\,(\dot{e} + e) = 0, \quad (e = u - U)$$

här giltig för alla $v = v(t)$ som är styckvis *konstanta*. Med ϕ sådan att

$$-\dot{\phi} + \phi = 0 \quad \text{for } t < T, \quad \phi(T) = e(T),$$

fås

Tidsstegning 36

$$|e(T)|^2 = \phi(T)\,e(T) + \textstyle\int_0^T \underbrace{(-\dot{\phi} + \phi)}_{=0}\,e$$
$$= \textstyle\int_0^T \phi\,(\dot{e} + e) = \int_0^T (\phi - v)\,(\dot{e} + e) = \int_0^T (\phi - v)\,\underbrace{(f - \dot{U} - U)}_{=:r}$$
$$\le \textstyle\int_0^T k^{-1}|\phi - v|\,\max_{[0,T]}|k\,r(U)|.$$

Tidsstegning 37

Med lämpligt val av interpolant v av ϕ, känd uppskattning av interpolationsfelet $\phi - v$, och genom att utnyttja att $\dot{\phi} = \phi$ & $\phi(t) = \exp(t - T)\, e(T)$, erhålls

$$|e(T)|^2 \leq \underbrace{\int_0^T \exp(t - T)\, dt}_{=1-\exp(-T)\leq 1} \max_{[0,T]} |k\, r(U)|,$$

dvs

$$|e(T)| \leq \max_{[0,T]} |kr|.$$

Tidsstegning 38

Övning: Generalisera till fallet $\dot{u} + au = f$, med a konstant > 0.

Övning: Är samma a posteriori uppskattning giltig även för ekvationen $\dot{u} - u = f$?