MA2501 – Forelesningsnotater Numerisk ODE

Sist endret: March 31, 2024

Innledning

Vi begynner med skalare ordinære differensialligninger av første orden. Disse har formen

$$y'(t) = f(t, y(t)),\ t \in [t_0, T] \qquad [\text{ skrives ofte i kortform } y' = f(t, y) \text{ }]$$

Sammen med ligningen spesifiseres oftest en startverdi: $y(t_0) = y_0$.

Eksempel

$$y' = 2\,t\,y^2, \quad y(0) = 1$$

Den eksakte løsningen er enkel å finne: $y(t) = \frac{1}{1-t^2}$, $|t| < 1$. Men det er lett å lage eksempler der løsningen ikke kan uttrykkes ved hjelp av elementære funksjoner, som for eksempel

$$y' = \sin y^2, \quad y(0) = 1.$$

Svaret man får i Maple eller Wolfram Alpha er ikke veldig nyttig.

1 Prinsipp for numerisk løsning av ordinære differensialligninger

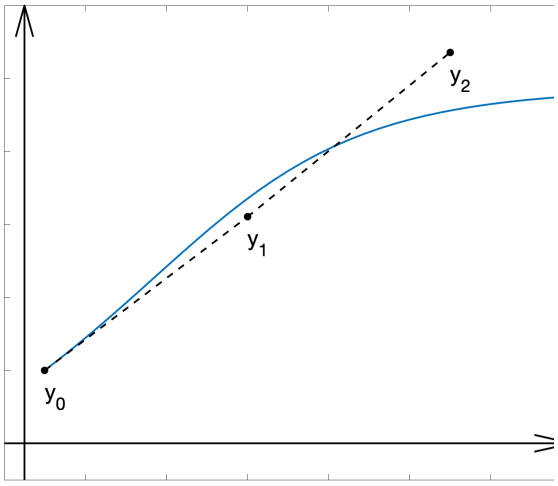
Hovedstrukturen er som følger

- Del opp t-aksen i små intervaller $t_0 < t_1 < \cdots < t_N = T$. Definier $h_i = t_{i+1} - t_i$. Noen ganger setter man $h_i = h =$ konstant for all i. Da blir $t_i = t_0 + ih$ og $h = (T - t_0)/N$
- Start med $y(t_0) = y_0$ som er oppgitt. Beregn $y_1 \approx y(t_1)$ ved en numerisk approksimasjon. Deretter beregn $y_2 \approx y(t_2)$ basert på y_1 og så videre.

2 Eulers metode

Dette er den enkleste og mest intuitive av alle metoder, og er illustrert i figuren. Vi legger en linje $L(t)$ gjennom punktet $(t_0, y0)$ med stigningstall $L'(t) = y'(t_0) = f(t_0, y_0)$, dette gir $L(t) = y_0 + tf(t_0, y_0)$. Vi setter $y_1 = L(h_0) = y_0 + h_0 f(t_0, y_0)$. Deretter lager vi en ny rett linje $L(t)$ gjennom (t_1, y_1) med stigningstall $f(t_1, y_1)$ og finner

$$y_2 = y_1 + h_1 f(t_1, y_1).$$

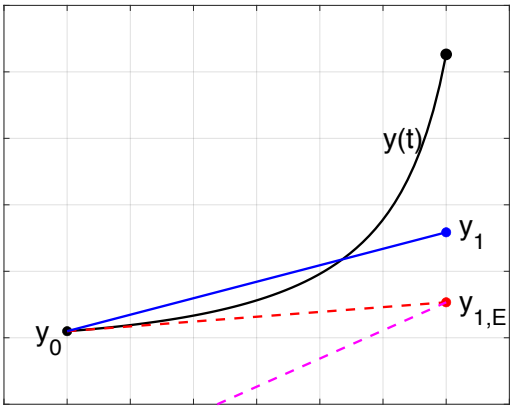


Eulers metode er generelt gitt som $y_{n+1} = y_n + h_n f(t_n, y_n)$.

3 Forbedret Eulers metode (Heuns metode, Eksplisitt trapesmetode)

Eulers metode kan tolkes som å følge tangenten til en løsning over et intervall h. Forbedret utgave kan konstrueres som følger

- Ta et skritt med Eulers metode og finn $y_{1,E} = y_0 + h_0 f(t_0, y_0) = y_0 + h_0 k_1$ der $k_1 = f(t_0, y_0)$
- Finn stigningstallet $k_2 = f(t_1, y_{1,E})$
- Finn middelverdien mellom disse to stigningstallene $\tilde{k} = \frac{1}{2}(k_1 + k_2)$
- Bruk $\tilde{k}$ som stigningstall for hele skrittet, dvs sett $y_1 = y_0 + h_0 \tilde{k}$.



Vi skriver kun h for skrittlengden og får *Forbedret Eulers metode*

$$\begin{aligned} k_1 &= f(t_n, y_n) \\ k_2 &= f(t_{n+1}, y_n + hk_1) \\ y_{n+1} &= y_n + \tfrac{1}{2}h(k_1 + k_2) \end{aligned}$$

4 Systemer av differensialligninger

I praksis ser vi sjelden på skalare ligninger, men istedet søker vi en vektor

$$\mathbf{y}(t) = [y^1(t), y^2(t), \ldots, y^m(t)]^T$$

som tilfredsstiller et system av differensialligninger

$$\begin{aligned} (y^1)' &= f^1(t, y^1, \ldots, y^m) \\ (y^2)' &= f^2(t, y^1, \ldots, y^m) \\ &\vdots \\ (y^m)' &= f^m(t, y^1, \ldots, y^m) \end{aligned} \quad \rightarrow \quad \begin{aligned} &\text{Kompakt notasjon:} \\ &\mathbf{y}' = \mathbf{f}(t, \mathbf{y}) \end{aligned}$$

Merk at vi her bruker superskript når vi refererer til komponent-indeksen i en vektor.

Lipschitz-konstant Hvis det fins en $L \in \mathbb{R}$ slik at

$$\|\mathbf{f}(t, \mathbf{y}) - \mathbf{f}(t, \mathbf{z})\| \leq L \|\mathbf{y} - \mathbf{z}\|, \quad t \in [t_0, T]$$

og alle $\mathbf{y}, \mathbf{z} \in \mathbb{R}^m$ så er L en *Lipschitz-konstant* og $\mathbf{f}$ er *Lipschitz-kontinuerlig.* Disse konseptene er viktige for eksistens og entydighet av løsningen av differensialligningen og for estimater av feil og konvergens av numerisk metode. For $\mathbf{y}' = \mathbf{f}(t, \mathbf{y})$ antar vi alltid at $\mathbf{f}$ er (minst lokal) Lipschitz.

Theorem 4.1. *Systemer av diffligninger. La* $\mathbf{f}(t, \mathbf{y})$ *være definert og kontinuerlig på stripen*

$$S = \{(t, \mathbf{y}) | a \leq t \leq b,\ \mathbf{y} \in \mathbb{R}^m\}, \quad a, b \text{ endelige}$$

Anta at $L > 0$ *fins slik at*

$$\|\mathbf{f}(t, \mathbf{y}) - \mathbf{f}(t, \mathbf{z})\| \leq L \|\mathbf{y} - \mathbf{z}\|, \quad t \in [t_0, T],\ \forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^m$$

For alle $t_0 \in [a, b]$, $\mathbf{y}_0 \in \mathbb{R}^m$ *fins entydig* $\mathbf{y}(t)$ *slik at*

(a) $\mathbf{y}(t)$ *er kontinuerlig og kontinuerlig deriverbar for* $t \in [a, b]$

(b) $\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t))$ *for* $t \in [a, b]$

(c) $\mathbf{y}(t_0) = \mathbf{y}_0$

Bevis er ikke pensum

Teoremet kan generaliseres ved å anta kun lokal Lipschitz-betingelse, men da får man kun lokal løsning nær t_0.

5 Autonome systemer av differensialligninger

Hvis $\mathbf{y}' = \mathbf{f}(\mathbf{y})$, det vil si at $\mathbf{f}$ ikke avhenger eksplisitt av t, kalles differensialligningen autonom. Ikke-autonome differensialligninger er da selvsagt de som er slik at $\mathbf{f}$ avhenger eksplisitt av t. En ikke-autonom differensialligning kan gjøres autonom ved at en innfører en ekstra komponent og dermed øker systemets dimensjon fra m til $m+1$. Vi starter med det ikke-autonome systemet

$$\mathbf{u}' = \mathbf{g}(t, \mathbf{u})$$

og definerer så

$$\mathbf{y} = [\mathbf{u}, t]^T = [u^1, \ldots, u^m, t]^T, \quad y^i = u^i,\ i = 1, \ldots, m,\ y^{m+1} = t.$$

Vi får den ekstra diffligningen $(y^{m+1})' = 1$. Vi kan dermed skrive

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}) = \begin{bmatrix} \mathbf{g}(y^{m+1}, \mathbf{u}) \\ 1 \end{bmatrix}$$

I mange sammenhenger når man ser på systemer av differensialligninger antar man derfor uten tap av generalitet at systemet er autonomt.

6 Lokal og global avbruddsfeil, konvergens

Lokal avbruddsfeil. En numerisk metode gir fra $\mathbf{y}_0$ approksimasjoner $\mathbf{y}_1, \mathbf{y}_2, \ldots$ som tilnærming til eksakt løsning, $\mathbf{y}_i = \mathbf{y}(t_i)$. Siden $\mathbf{y}_{n+1}$ regnes ut som funksjon av $\mathbf{y}_n$ kan vi tenke på den numeriske metoden som en avbildning (funksjon) $\Psi_{h,\mathbf{f}}$ som for en gitt diffligning ($\mathbf{f}$) og skrittlengde h beregner

$$\mathbf{y}_{n+1} = \Psi_{h,\mathbf{f}}(\mathbf{y}_n)$$

For eksempel i Eulers metode er $\Psi_{h,\mathbf{f}}(\mathbf{y}) = \mathbf{y} + h\mathbf{f}(\mathbf{y})$. Den eksakte løsningen kan også fremstilles som en lignende avbildning $\varphi_{t,\mathbf{f}}$ som for en gitt startverdi $\mathbf{y}_0$ returnerer den eksakte løsningen $\mathbf{y}(t)$ til $\mathbf{y}' = \mathbf{f}(\mathbf{y})$, dvs

$$\mathbf{y}(t) = \varphi_{t,\mathbf{f}}(\mathbf{y}_0)$$

Lokal avbruddsfeil for en metode $\Psi_{h,\mathbf{f}}$ er definert som

$$\Psi_{h,\mathbf{f}}(\mathbf{y}) - \varphi_{h,\mathbf{f}}(\mathbf{y})$$

Den er lokal fordi både eksakt løsning og numerisk løsning starter med samme utgangspunkt $\mathbf{y}$. Dette står i kontrast til *global avbruddsfeil* som tar hensyn til at feilen har akkumulert seg opp over mange skritt.

Global avbruddsfeil. Dette er den virkelige feilen

$$\mathbf{y}_n - \mathbf{y}(t_n)$$

der $\mathbf{y}_n$ er beregnet ved $\mathbf{y}_i = \Psi_{h,\mathbf{f}}(\mathbf{y}_{i-1}),\ i = 1, \ldots, n$, mens $\mathbf{y}(t)$ er den eksakte løsningen av $\mathbf{y}' = \mathbf{f}(\mathbf{y}),\ \mathbf{y}(t_0) = \mathbf{y}_0$. Man kan lett la seg forlede til å tenke at globalfeil kun er summen av alle lokalfeilene. Dette er ikke riktig. Hvorfor?

Konvergens konstant skrittlengde. La $t \in [t_0, T]$ være vilkårlig. La nå $n \to \infty$ og $h \to 0$ være slik at $t = t_n = t_0 + nh$ holdes konstant. Hvis

$$\lim_{\substack{n\to\infty \\ h\to 0}} \|\mathbf{y}_n - \mathbf{y}(t_n)\| = 0$$

så sier vi at metoden er konvergent. Definisjonen generaliseres enkelt til variabel skrittlengde ved å kreve at $\max_i h_i \to 0$ (fremdeles holdes punktet t fast).

Lokal avbruddsfeil i Eulers metode. La $t = t_n \in [a, b]$. Anta at $\mathbf{y}(t_n) = \mathbf{y}_n$, slik at vi ser på integralkurven gjennom $(t_n, \mathbf{y}_n)$. Lokal avbruddsfeil er nå

$$\boldsymbol{\tau}_{n+1} = \mathbf{y}_{n+1} - \mathbf{y}(t_{n+1}) = \Psi_{h,\mathbf{f}}(\mathbf{y}_n) - \varphi_{h,\mathbf{f}}(\mathbf{y}_n)$$

Taylorutvikling av eksakt løsning gir

$$\varphi_{h,\mathbf{f}}(\mathbf{y}_n) = \mathbf{y}(t_n + h) = \mathbf{y}(t_n) + h\mathbf{y}'(t_n) + \tfrac{1}{2}h^2\mathbf{y}''(\xi_n) = \mathbf{y}_n + h\mathbf{f}(\mathbf{y}_n) + \tfrac{1}{2}h^2\mathbf{y}''(\xi_n)$$

for $\xi_n \in (t_n, t_{n+1})$. Dermed blir

$$\boldsymbol{\tau}_{n+1} = \mathbf{y}_{n+1} - \mathbf{y}(t_{n+1}) = \mathbf{y}_n + h\mathbf{f}(\mathbf{y}_n) - (\mathbf{y}_n + h\mathbf{f}(\mathbf{y}_n) + \tfrac{1}{2}h^2\mathbf{y}''(\xi_n)) = -\tfrac{1}{2}h^2\mathbf{y}''(\xi_n))$$

Hvis vi kan anta at $\|\mathbf{y}''(t)\| \leq M$ globalt kan vi konkludere med at

$$\|\boldsymbol{\tau}_{n+1}\| \leq \tfrac{1}{2}h^2 M \quad \text{for alle } n.$$

Global avbruddsfeil i Eulers metode. Vi antar nå at $\mathbf{f}$ i differensialligningen har en Lipschitz-konstant L. Samtidig antar vi at lokalfeilen kan begrenses som ovenfor. Definer den globale feilen ved

$$\mathbf{E}_n = \mathbf{y}_n - \mathbf{y}(t_n) = \mathbf{y}_n - \varphi_{t_n,\mathbf{f}}(\mathbf{y}_0)$$

Vi forsøker å få til en rekursjonsformel for globalfeilen

$$\mathbf{E}_{n+1} = \mathbf{y}_{n+1} - \mathbf{y}(t_{n+1}) = \mathbf{y}_n + h\mathbf{f}(\mathbf{y}_n) - \mathbf{y}(t_n) - h\mathbf{f}(\mathbf{y}(t_n)) - \boldsymbol{\tau}_{n+1}$$

Vi tar norm og bruker trekantulikhet og Lipschitz-betingelse

$$\begin{aligned}\|\mathbf{E}_{n+1}\| &\leq \|\mathbf{y}_n - \mathbf{y}(t_n)\| + h\|\mathbf{f}(\mathbf{y}_n) - \mathbf{f}(\mathbf{y}(t_n))\| + \|\boldsymbol{\tau}_{n+1}\| \\ &\leq \|\mathbf{y}_n - \mathbf{y}(t_n)\| + hL\|\mathbf{y}_n - \mathbf{y}(t_n)\| + \tfrac{1}{2}h^2 M = (1 + hL)\|\mathbf{E}_n\| + \tfrac{1}{2}h^2 M\end{aligned}$$

Denne rekursjonsulikheten lar seg løse, og man finner at

$$\|\mathbf{E}_n\| \leq (1 + hL)^n\|\mathbf{E}_0\| + \tfrac{1}{2}h^2 M \sum_{k=0}^{n-1}(1 + hL)^k$$

Den endelige geometriske rekken kjenner vi summen av

$$\sum_{k=0}^{n-1}(1+hL)^k = \frac{(1+hL)^n - 1}{1+hL-1} = \frac{1}{hL}\left((1+hL)^n - 1\right)$$

Nå husker vi at $t_n = t_0 + nh$ er fiksert slik at $h = \frac{t_n - t_0}{n}$. Dermed har vi

$$(1+hL)^n = (1+\frac{\alpha}{n})^n \quad \text{med} \quad \alpha = L(t_n - t_0)$$

Nå er $(1+\frac{\alpha}{n})^n \leq e^\alpha$ for alle $n \in \mathbb{N}$ og følgen er monotont voksende og konvergerer mot e^α. Vi antar at $\mathbf{E}_0 = 0$ siden vi bruker eksakt startverdi. Da kan vi konkludere med at

$$\|\mathbf{E}_n\| \leq \tfrac{1}{2}h^2 M \frac{1}{hL}(e^{L(t_n - t_0)} - 1) = \frac{M}{2L}(e^{L(t_n-t_0)} - 1)h$$

så den globale feilen i Eulers metode er $\mathcal{O}(h)$ mens den lokale er $\mathcal{O}(h^2)$.

Konvergensorden. Hvis en metode er slik at globalfeilen for en $h^* > 0$ oppfyller

$$\|\mathbf{E}_n\| \leq Ch^p, \qquad |h| \leq h^*$$

så har metoden *orden* p. Eulers metode har derfor orden 1. Typisk gjelder:

$$\text{Lokalfeil } \|\boldsymbol{\tau}_n\| \leq Dh^{p+1} \quad \Rightarrow \quad \text{Globalfeil } \|\mathbf{E}_n\| \leq C(D,L)h^p$$

Forbedret Euler har orden $p = 2$.

7 Runge–Kutta metoder

Eulers metode kan skrives

$$\begin{aligned} \mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\ \mathbf{y}_{n+1} &= \mathbf{y}_n + h\mathbf{k}_1 \end{aligned}$$

og vi minner om Heuns metode

$$\begin{aligned} \mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\ \mathbf{k}_2 &= \mathbf{f}(t_n + h, \mathbf{y}_n + h\mathbf{k}_1) \\ \mathbf{y}_{n+1} &= \mathbf{y}_n + \frac{h}{2}(\mathbf{k}_1 + \mathbf{k}_2) \end{aligned}$$

Eksplisitte Runge–Kutta metoder. Vi kan ane oss en generalisering av dette. Vi innfører skalare koeffisienter c_i, b_i og a_{ij} og forsøker oss med

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\
\mathbf{k}_2 &= \mathbf{f}(t_n + c_2 h, \mathbf{y}_n + h a_{21}\mathbf{k}_1) \\
&\vdots \\
\mathbf{k}_s &= \mathbf{f}(t_n + c_s h, \mathbf{y}_n + h \sum_{j=1}^{s-1} a_{sj}\mathbf{k}_j) \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + h \sum_{i=1}^{s} b_i \mathbf{k}_i
\end{aligned}$$

Tallet s er *antall nivåer* i metoden (eng: stages). Åpenbart har Heuns metode $s = 2$, mens Eulers metode har $s = 1$. En berømt eksplisitt Runge–Kutta metode med $s = 4$ ble laget av Kutta og publisert i 1901.

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\
\mathbf{k}_2 &= \mathbf{f}(t_n + \frac{h}{2}, \mathbf{y}_n + \frac{h}{2}\mathbf{k}_1) \\
\mathbf{k}_3 &= \mathbf{f}(t_n + \frac{h}{2}, \mathbf{y}_n + \frac{h}{2}\mathbf{k}_2) \\
\mathbf{k}_4 &= \mathbf{f}(t_n + h, \mathbf{y}_n + h\mathbf{k}_3) \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + \frac{h}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4)
\end{aligned}$$

8 Skrittlengdekontroll

Man kan estimere lokal avbruddsfeil i en Runge–Kutta metode ved å benytte et såkalt innbygd par (embedded pair). Anta at vi har en Runge–Kutta metode med nivåer $\mathbf{k}_1, \ldots, \mathbf{k}_s$ og at disse kan kombineres med to ulike sett av vekter $b_1, \ldots, b_s$ og $\hat{b}_1, \ldots, \hat{b}_s$. Vi kan da beregne

$$\begin{aligned}
\mathbf{y}_{n+1} &= \mathbf{y}_n + h \sum_{i=1}^{s} b_i \mathbf{k}_i && \longleftarrow \text{ orden } p \\
\hat{\mathbf{y}}_{n+1} &= \mathbf{y}_n + h \sum_{i=1}^{s} \hat{b}_i \mathbf{k}_i && \longleftarrow \text{ orden } \hat{p}
\end{aligned}$$

Approksimasjonen $\mathbf{y}_{n+1}$ er den som danner grunnlag for neste skritt, mens $\hat{\mathbf{y}}_{n+1}$ er en hjelpeapproksimasjon som kun brukes til feilestimering. Vi har for lokal avbruddsfeil

$$\begin{aligned}
\|\mathbf{y}(t_{n+1}) - \mathbf{y}_{n+1}\| &= K_R\, h^{p+1} + \mathcal{O}(h^{p+2}) \\
\|\mathbf{y}(t_{n+1}) - \hat{\mathbf{y}}_{n+1}\| &= \hat{K}_R\, h^{\hat{p}+1} + \mathcal{O}(h^{\hat{p}+2})
\end{aligned}$$

Anta nå i første omgang at $\hat{p} > p$ og sett $\mathbf{est}_{n+1} = \hat{\mathbf{y}}_{n+1} - \mathbf{y}_{n+1}$. Da finner vi at

$$\begin{aligned}\|\mathbf{est}_{n+1}\| &= \|\mathbf{y}(t_{n+1}) - \mathbf{y}_{n+1} - (\mathbf{y}(t_{n+1}) - \hat{\mathbf{y}}_{n+1})\| = \|\mathbf{y}(t_{n+1}) - \mathbf{y}_{n+1}\| + \mathcal{O}(h^{p+2}) \\ &= \text{Lokalfeil i } \mathbf{y}_{n+1} + \text{høyere ordens ledd}\end{aligned}$$

Anta nå at vi har en brukerspesifisert toleranse `tol`. Strategien er å velge skrittlengden $h = h_n$ i skrittet fra t_n til t_{n+1} slik at vi får

$$\|\mathbf{est}_{n+1}\| \approx \texttt{tol}$$

dvs vi forsøker å kontrollere lokal avbruddsfeil. Algoritmisk tenker vi slik

if $\mathbf{est}_{n+1} \leq \texttt{tol}$
    Aksepter skrittet, og øk h til neste skritt
else
    Forkast skrittet, reduser h, og forsøk å ta skrittet på nytt

Justeringsformel for skrittlengden. Vi har

$$t_n = t_{n-1} + h_{n-1}$$

Det følger fra diskusjonen ovenfor at vi fra dette skrittet hadde

$$\|\mathbf{est}_n\| = K_R h_{n-1}^{p+1} + \mathcal{O}(h^{p+2}) \approx K_R h_{n-1}^{p+1}$$

Hvis vi forsøker å justere neste skritt slik at

$$\|\mathbf{est}_{n+1}\| \approx K_R h_n^{p+1} = \texttt{tol}$$

så impliserer dette ved å kombinere de to siste ligningene at vi vil ha

$$h_n^{p+1} = \frac{\texttt{tol}}{K_R} = \frac{\texttt{tol}}{\|\mathbf{est}_n\|} h_{n-1}^{p+1}$$

Så vi kan da estimere den nye skrittlengden h_n til å være

$$h_n = \left(\frac{\texttt{tol}}{\|\mathbf{est}_n\|}\right)^{\frac{1}{p+1}} h_{n-1}$$

For tilfellet med at et skritt forkastes kan bruke akkurat samme ide, og får en helt lik formel. Hvis man forsøker med et skritt h fra t_{n-1} til t_n som forkastes så setter man

$$h_{ny} = \left(\frac{\texttt{tol}}{\|\mathbf{est}_n\|}\right)^{\frac{1}{p+1}} h \tag{1}$$

og forsøker å gjenta skrittet hvor skrittlengden er redusert fra h til h_{ny}.

Da denne metoden ble foreslått og eksempler ble utviklet av Fehlberg på 1960-tallet så var det flere som reagerte på at man stepper seg framover med metoden med lavest orden (husk at vi antok $p < \hat{p}$). Hvorfor kaste bort den mer nøyaktige approksimasjonen $\hat{\mathbf{y}}_{n+1}$ når man likevel har den tilgjengelig? Og egentlig så er det viktigere å kontrollere den globale enn den lokale avbruddsfeilen uansett. Det viser seg i praksis at det går like bra å basere seg på metoden av høyest orden, vår tolkning av dette er at vi velger $p > \hat{p}$, og vanligvis $p = \hat{p} + 1$. Dette får den konsekvensen at vi må modifisere litt på skrittlengdeformelen, vi setter nå

$$h_n = \left(\frac{\texttt{tol}}{\|\mathbf{est}_n\|} \right)^{\frac{1}{p}} h_{n-1}$$

Til slutt en litt viktig sak. Fordi det er kostbart å forkaste skritt så kan man legge inn en sikkerhetsmargin for å unngå at dette skjer, en såkalt *pessimistfaktor* α. Dette vil si at vi i virkeligheten, med $p > \hat{p}$ bruker formelen

$$h_n = \alpha \left(\frac{\texttt{tol}}{\|\mathbf{est}_n\|} \right)^{\frac{1}{p}} h_{n-1}, \qquad \alpha \approx 0.8 - 0.9$$

Eksempler på innbygde par.

Eksempel (Forbedret Euler, Euler) $p(\hat{p}) = 2(1)$

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\
\mathbf{k}_2 &= \mathbf{f}(t_n + h, \mathbf{y}_n + h\mathbf{k}_1) \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + \tfrac{1}{2}h(\mathbf{k}_1 + \mathbf{k}_2) \\
\hat{\mathbf{y}}_{n+1} &= \mathbf{y}_n + h\mathbf{k}_1 \\
\mathbf{est}_{n+1} &= \hat{\mathbf{y}}_{n+1} - \mathbf{y}_{n+1} = \tfrac{1}{2}h(\mathbf{k}_1 - \mathbf{k}_2)
\end{aligned}$$

Eksempel (Bogacki-Shampine par). $p(\hat{p}) = 3(2)$

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{y}_n) \\
\mathbf{k}_2 &= \mathbf{f}(t_n + \frac{h}{2}, \mathbf{y}_n) + \frac{h}{2}\mathbf{k}_1 \\
\mathbf{k}_3 &= \mathbf{f}(t_n + \tfrac{3}{4}h, \mathbf{y}_n) + \tfrac{3}{4}h\mathbf{k}_2) \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + \tfrac{1}{9}h(2\mathbf{k}_1 + 3\mathbf{k}_2 + 4\mathbf{k}_3) \\
\mathbf{k}_4 &= \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \\
\hat{\mathbf{y}}_{n+1} &= \mathbf{y}_n + \tfrac{1}{24}\, h\, (7\mathbf{k}_1 + 6\mathbf{k}_2 + 8\mathbf{k}_3 + 3\mathbf{k}_4)
\end{aligned}$$

Det er interessant, og ingen tilfeldighet, at $\mathbf{k}_4$ i skritt n er det samme som $\mathbf{k}_1$ i skritt $n + 1$. Dette innebærer at det første nivået $\mathbf{k}_1$ i hvert skritt utenom det aller første er gratis tilgjengelig. Så selv om metodeparet formelt sett har 4 nivåer og blir det i praksis bare 3.

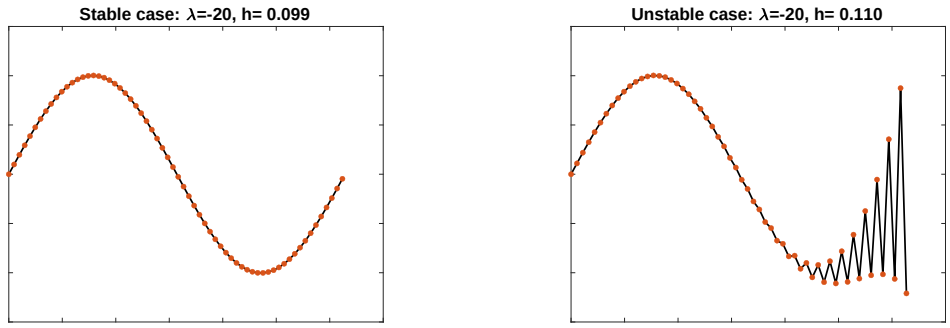
Populære innbygde par av høyere orden er Runge–Kutta–Fehlberg 4(5) og Dormand–Prince 5(4) (se side 328–329 i Sauer). Det fins også innbygde par av høyere orden enn dette, for eksempel Dormand-Prince 8(7), men jo høyere orden, jo vanskeligere blir de å konstruere, særlig hvis man vil ha så lavt antall nivåer som mulig. For eksempel metoder av orden 5 må ha minst 6 nivåer, mens for orden 8 trengs minst 11 nivåer. I 1978 konstruerte Hairer en eksplisitt Runge–Kutta metode av orden 10 med 17 nivåer.[1]

9 Stabilitet av Runge–Kutta metoder

For å forstå begrepet stabilitet kan det være nyttig å starte med et eksempel. Et spesialtilfelle av Prothero-Robinsons problem er

$$y' = \lambda(y - \sin t) + \cos t, \quad y(0) = 0.$$

Ved innsetting ser vi at funksjonen $y(t) = \sin t$ tilfredsstiller både diffligningen og initialverdien. Vi forsøker nå å kjøre to numeriske eksperimenter med Eulers metode der vi lar skrittlengden være henholdsvis $h = 0.099$ og $h = 0.110$. Resultatet ser slik ut:



Noe skjer tydeligvis når skrittlengden passerer $h = 0.1$ i dette tilfellet, til tross for at løsningen er så glatt som den kan få blitt.

Modelligning for stabilitet Vi forenkler problemet enda mer og ser kun på

$$y' = \lambda y, \qquad \lambda \in \mathbb{C} \ (\lambda \text{ er en kompleks konstant})$$

[1] Hairers konstruksjon var noe som ble sett på som litt av en bragd. Han måtte da forholde seg til intet mindre enn 1205 ikke-lineære betingelser for de 153 koeffisientene i metoden. Prestasjonen ble belønnet med plass i Guiness rekordbok.

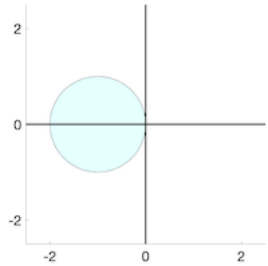
Den eksakte løsningen til dette problemet er $y(t) = e^{\lambda t} y_0$. Vi ser på oppførselen til løsningen når $t \to \infty$. Anta derfor at $\lambda = \alpha + \mathrm{i}\beta$ der $\mathrm{i} = \sqrt{-1}$, $\alpha = \operatorname{Re}\lambda$ og $\beta = \operatorname{Im}\lambda$.

$$y(t) = e^{(\alpha+\mathrm{i}\beta)t} y_0 = e^{\alpha t} e^{\mathrm{i}\beta t} y_0 \qquad y(t) \xrightarrow{t\to\infty} 0 \Leftrightarrow \alpha < 0$$

$$|y(t)| = e^{\alpha t}|y_0| \qquad y(t) \xrightarrow{t\to\infty} \infty \Leftrightarrow \alpha > 0$$

Ligningen er stabil når $\alpha = \operatorname{Re}\lambda < 0$, det vil si at $\lambda \in \mathbb{C}^-$, også kalt venstre halvplan. Men hva skjer med Eulers metode?

Stabilitet av Eulers metode.



$$y_{n+1} = y_n + hf(y_n) = y_n + h\lambda y_n = (1 + h\lambda)y_n$$

og dermed blir $y_n = (1 + h\lambda)^n y_0$.

$$|y_n| \xrightarrow{n\to\infty} 0 \quad \text{krever} \quad |1 + \lambda h| < 1$$

Vi innfører nå $z = h\lambda = h\alpha + \mathrm{i}h\beta = z_r + \mathrm{i}z_i$.

$$|1+z| < 1 \;\Leftrightarrow\; |1+z|^2 < 1, \quad \text{dvs } (1+z)(1+\bar{z}) = (1+z_r+\mathrm{i}z_i)(1+z_r-\mathrm{i}z_i) = (1+z_r)^2+z_i^2 < 1$$

Men dette er ligningen for en sirkelskive i planet, sentrert i $c = -1$ og med radius $r = 1$. Hvis for eksempel $\lambda \in \mathbb{R}$ med $\lambda \ll 0$ må vi ha

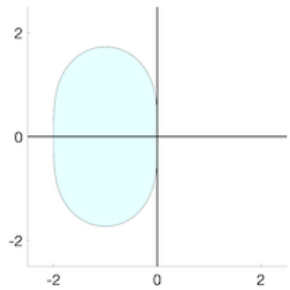
$$-2 < h\lambda < 0 \;\Rightarrow\; h < \frac{2}{(-\lambda)}$$

Vi ser at dette gir samme begrensning på h som i Prothero-Robinson eksemplet der $\lambda = -20$ i testene.

Stabilitet av Heuns metode.

Vi beregner

$$\begin{aligned} k_1 &= f(t_n, y_n) = \lambda y_n \\ k_2 &= f(t_n + h, y_n + hk_1) = \lambda(y_n + h\lambda y_n) \\ y_{n+1} &= y_n + \tfrac{1}{2}h(k_1 + k_2) = y_n + \tfrac{1}{2}h(\lambda y_n + \lambda(y_n + h\lambda y_n)) \\ &= (1 + h\lambda + \tfrac{1}{2}(h\lambda)^2)y_n = (1 + z + \tfrac{1}{2}z^2)y_n \end{aligned}$$



Nå kan vi studere hvilke verdier av z det er som gjør at $p(z) = 1 + z + \frac{1}{2}z^2$ oppfyller $|p(z)| < 1$. Det som i Eulers metode var en sirkelskive får nå en mer oval form som i figuren.

Stabilitet av Kuttas metode. Vi skriver som før $z = h\lambda$ og finner

$$\begin{aligned}
hk_1 &= zy_n \\
hk_2 &= hf(y_n + \tfrac{1}{2}hk_1) = z(1 + \tfrac{1}{2}z)y_n \\
hk_3 &= hf(y_n + \tfrac{1}{2}hk_2) = z(1 + \tfrac{1}{2}z(1 + \tfrac{1}{2}z))y_n = (\tfrac{1}{4}z^3 + \tfrac{1}{2}z^2 + z)y_n \\
hk_4 &= hf(y_n + hk_3) = z(1 + \tfrac{1}{4}z^3 + \tfrac{1}{2}z^2 + z)y_n \\
y_{n+1} &= yn + \tfrac{1}{6}(hk_1 + 2hk_2 + 2hk_3 + hk_4) = (1 + z + \tfrac{1}{2}z^2 + \tfrac{1}{6}z^3 + \tfrac{1}{24}z^4)y_n
\end{aligned}$$

Så her er $y_{n+1} = p(z)y_n$ med $p(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4$.

Generelt om stabilitetsfunksjoner For enhver eksplisitt Runge–Kutta metode finner man at når den anvendes på den skalare komplekse modelligningen så får man en formel av typen

$$y_{n+1} = p(z)y_n$$

der $p(z)$ er et polynom av grad s (antall nivåer). Området i det komplekse plan gitt ved

$$S_p = \{z \in \mathbb{C} : |p(z)| \leq 1\}$$

kalles metodens stabilitetsområde. $z = 0$ ligger alltid på randen av området, og S_p vil alltid være begrenset, men ikke nødvendigvis sammenhengende. Vi skal etterhvert introdusere implisitte Runge–Kutta metoder. De har også en stabilitetsfunksjon med et tilhørende stabilitetsområde, men nå kan funksjonen være rasjonal, dvs en kvotient av to polynomer p og q.

10 Implisitte metoder og stive ligninger

Baklengs Euler. Vi ser nå igjen på systemer av differensialligninger

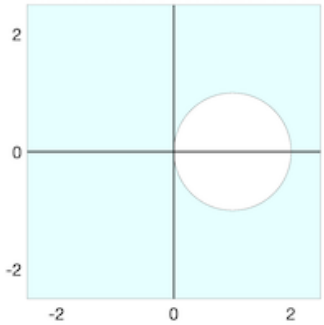
$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}_0, \ \mathbf{f} : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$$

Vi gjør nå en vri på Eulers metode og tar tangenten i høyre endepunkt istedetfor venstre, det vil si

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{f}(t_{n+1}, \mathbf{y}_{n+1})$$

Denne ligningen må vi løse med hensyn på $\mathbf{y}_{n+1}$ i hvert eneste skritt. Generelt vil dette være et ikke-lineært system av ligninger, slik vi så på helt på slutten i modul 1. Man må da bruke en form for iterasjon (f.eks. Newton) for å finne $\mathbf{y}_{n+1}$. Men vi starter igjen med den enkle skalare modelligningen for stabilitet

$$y' = \lambda y, \quad \lambda \in \mathbb{C}.$$



Baklengs Euler gir $y_{n+1} = y_n + h\lambda y_{n+1}$ som impliserer

$$y_{n+1} = \frac{1}{1-h\lambda}\, y_n = \frac{1}{1-z}\, y_n = R(z) y_n$$

der $R(z) = \frac{1}{1-z}$. Om vi nå krever $|R(z)| \leq 1$ så innebærer dette at $|1-z| \geq 1$, det vil si at $z = h\lambda$ må ligge *utenfor* sirkelen sentrert i $z = c = 1$ med radius $r = 1$.

En viktig observasjon her er at område vi akkurat beskrev, nemlig $S_R = \{z \in \mathbb{C} : \frac{1}{|1-z|} \leq 1\}$ inneholder venstre halvplan $\mathbb{C}^-$. Dette vil si at for alle λ slik at den eksakte løsningen av modelligningen går mot null når $t \to \infty$ vil også de numeriske approksimasjonene y_n med Baklengs Euler gå mot null når $n \to \infty$. Det betyr at de skrittlengdebegrensningene vi så med Eulers metode ikke fins for Baklengs Euler.

Løsning av ikke-lineære ligninger i hvert skritt med Baklengs Euler. Vi begynner for enkelhets skyld med det skalare autonome tilfellet

$$y' = f(y), \quad y(0) = y_0$$

Baklengs Euler

$$y_{n+1} = y_n + hf(y_{n+1}$$

Skriv $x = y_{n+1}$ slik at ligningen vi skal løse med hensyn på x er

$$x = y_n + hf(x)$$

Dette er en ikke-lineær ligning i fikspunktform så vi kan uten videre forsøke fikspunktiterasjonen fra Modul 1

$$x^{(k+1)} = y_n + hf(x^{(k)}).$$

Vi kan nå studere iterasjonsfeilen

$$e^{(k+1)} = x^{(k+1)} - x = h(f(x^{(k)}) - f(x)) = hf'(\xi^{(k)})(x^{(k)} - x) = hf'(\xi^{(k)})e^{(k)}$$

La oss anta at $|f'(\xi)| \leq M$ for alle ξ. Vi må kreve $hM < 1$, det vil si $h < \frac{1}{M}$. Dette er lite oppløftende fordi hele poenget med å bruke Baklengs Euler var jo nettopp å unngå skrittlengdebegrensninger. Går vi igjen tilbake til modelligningen så vil der $M = |\lambda|$ så dersom vi skulle brukt fikspunktiterasjon på denne ville situasjonen faktisk vært enda verre enn med Euler. Konklusjonen er at vi må bruk en mer avansert løser for det ikke-lineære problemet som for eksempel Newtons metode eller en variant av denne.

Systemer. Vi kan bruke Newtons metode for systemer av ikke-lineære ligninger som ble introdusert i Modul 1. Skriv da $\mathbf{x} = \mathbf{y}_{n+1}$ og se på ligningen

$$\mathbf{F}(\mathbf{x}) = \mathbf{x} - \mathbf{y}_n - h\mathbf{f}(\mathbf{x})$$

Følger vi notatet som ble laget i Modul 1 så finner vi at Newtoniterasjonen er

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - D\mathbf{F}(\mathbf{x}^{(k)})^{-1}\mathbf{F}(\mathbf{x}^{(k)})$$

der $D\mathbf{F}(\mathbf{x}) = I - hD\mathbf{f}(\mathbf{x})$. Det er ofte for kostbart å regenerere Jacobi-matrisen $D\mathbf{F}(\mathbf{x})$ for hver iterasjon. Løser man også systemet ved å bruke LU-faktorisering så må dessuten denne $\mathcal{O}(m^3)$-operasjonen utføres for hver eneste iterasjon. Men i praksis beregner man kun $D\mathbf{F}(\mathbf{x})$ en gang i blant, ofte kan man beholde matrisen, ikke bare over alle iterasjonene i ett skritt, men faktisk også over flere etterfølgende skritt. Da kan man også gjøre LU-faktorisering bare når $D\mathbf{F}(\mathbf{x})$ regenereres og så deretter bare bruke substitusjoner for hver iterasjon, dette er som kjent $\mathcal{O}(m^2)$-operasjoner.

Stive differensialligninger. En presis definisjon av stive differensialligninger er utenfor rammene av dette kurset, men vi kan karakterisere begrepet for lineære differensialligninger med konstante koeffisienter, det vil si

$$\mathbf{y}' = A\mathbf{y}, \quad A \in \mathbb{R}^{m\times m}$$

Vi benevner A sine egenverdier $\lambda_1, \ldots, \lambda_m$, der hver $\lambda_i \in \mathbb{C}$. Hvis det fins en egenverdi λ_s slik at $\operatorname{Re}\lambda_s \ll 0$ og det samtidig fins en eller flere andre egenverdier av moderat størrelse, så er dette lineære systemet stivt. Typisk kjennetegn er prosesser i systemet som utvikler seg på veldig ulike tidsskalaer. Et veldig enkelt eksempel med $m = 2$ er

$$A = \begin{bmatrix} -10^{-6} & 0 \\ 0 & -1 \end{bmatrix}$$

For ikke-lineære systemer er det litt mer uklart hvordan stivhet skal defineres. Men vi kan koste på oss å vise et klassisk eksempel, nemlig Van der Pol's ligning (oscillator)

$$y'' - \mu(1 - y^2)y' + y = 0$$

Her er det lurt å skrive om denne andre ordens ligningen til et system med $m = 2$, vi setter $u = y$ og $v = y'$ og får

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} v \\ \mu(1 - u^2)v - u \end{bmatrix}$$

Dette systemet blir veldig stivt når man øker μ. Man kan få en indikasjon på stivhet av et ikke-lineært problem om man lineariserer rundt et punkt (si f.eks. (u^*, v^*) i Van der Pol og beregner egenverdiene der, dvs $D\mathbf{f}(u^*, v^*)$. Mer at et system kan være stivt i ett område av faserommet (løsningsrommet) og ikke-stivt i et annet.

Andre implisitte metoder. To populære metoder av orden to er

$$\text{Trapesmetoden: } \mathbf{y}_{n+1} = \mathbf{y}_n + \frac{h}{2}(\mathbf{f}(\mathbf{y}_n) + \mathbf{f}(\mathbf{y}_{n+1}))$$
$$\text{Midtpunktmetoden: } \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{f}(\tfrac{1}{2}(\mathbf{y}_n + \mathbf{y}_{n+1}))$$

Man kan også generalisere formatet for eksplisitte Runge–Kutta metoder til å inkludere implisitte metoder

$$
\begin{aligned}
\mathbf{k}_1 &= f(\mathbf{y}_n + h\sum_{j=1}^{s} a_{1j}\mathbf{k}_j) \\
\mathbf{k}_2 &= f(\mathbf{y}_n + h\sum_{j=1}^{s} a_{2j}\mathbf{k}_j) \\
&\vdots \\
\mathbf{k}_s &= f(\mathbf{y}_n + h\sum_{j=1}^{s} a_{sj}\mathbf{k}_j) \\
\mathbf{y}_{n+1} &= \mathbf{y}_n + h\sum_{i=1}^{s} b_i\mathbf{k}_i
\end{aligned}
$$

Stabilitet av trapesmetoden. Anvender metoden på $y' = \lambda y$, $y(0) = y_0$.

$$y_{n+1} = y_n + \frac{h}{2}(f(y_n) + f(y_{n+1})) = y_n + \frac{h\lambda}{2}(y_n + y_{n+1})$$

Løser vi med hensyn på y_{n+1} og setter som vanlig $z = h\lambda$ får vi

$$y_{n+1} = \frac{1+\frac{z}{2}}{1-\frac{z}{2}}\, y_n. \quad \text{Må sjekke når } \frac{|1+\frac{z}{2}|}{|1-\frac{z}{2}|} \leq 1.$$

det vil si $|1+\frac{z}{2}|^2 - |1-\frac{z}{2}|^2 \leq 0$. Skriver vi nå $z = \alpha + \mathrm{i}\beta$ og setter inn så får vi

$$\left(1+\frac{\alpha}{2}\right)^2 + \left(\frac{\beta}{2}\right)^2 - \left(1-\frac{\alpha}{2}\right)^2 - \left(\frac{\beta}{2}\right)^2 \leq 0$$

som gir $\alpha \leq 0$. Dermed er stabilitetsområdet til trapesmetoden $S_R = \mathbb{C}^-$, det venstre halvplan. Det står godt til hvordan den eksakte løsningen oppfører seg.

11 Eksamensoppgaver fra TMA5320

August 2016, Oppgave 4

$$y''(t) = -y(t)^2, \quad y(0) = 1,\ y'(0) = 0.$$

Runge–Kutta metode

$$
\begin{aligned}
\mathbf{k}_1 &= f(t_n, \mathbf{w}_n) \\
\mathbf{k}_2 &= f(t_n + \tfrac{2}{3}h, \mathbf{w}_n + \tfrac{2}{3}h\mathbf{k}_1) \\
\mathbf{w}_{n+1} &= \mathbf{w}_n + h(\tfrac{1}{4}\mathbf{k}_1 + \tfrac{3}{4}\mathbf{k}_2)
\end{aligned}
$$

Sett $h = 0.5$ og finn tilnærmelse til løsningen i $t = 0.5$.
Merknad. I denne eksamensoppgaven brukes notasjonen $\mathbf{w}_n \approx \mathbf{y}(t_n)$ for den numeriske løsningen.
Svar.
(a) Her må man først skrive om den skalare andre ordens differensialligningen til et system av to første ordens ligninger. Vi setter $u = y$ og $v = y'$, $\mathbf{y} = [u, v]^T$ og får da

$$\mathbf{y}' = \begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} v \\ -u^2 \end{bmatrix} = \mathbf{f}(\mathbf{y}), \qquad \begin{bmatrix} u(0) \\ v(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mathbf{y}_0,$$

For numerisk approksimasjon skriver vi $\mathbf{w}_n = [u_n, v_n]^T$. Vi beregner

$$\mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{y}_0) = [v_0, -u_0^2]^T = [0, -1]^T$$

Vi forbereder beregning av $\mathbf{k}_2$ ved først å finne

$$\mathbf{Y}_2 = \mathbf{w}_0 + \tfrac{2}{3}h\mathbf{k}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \tfrac{2}{3} \cdot 0.5 \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.333 \end{bmatrix}$$

Da finner vi

$$\mathbf{k}_2 = \mathbf{f}(t_n + \tfrac{2}{3}h, \mathbf{Y}_2) = \begin{bmatrix} -0.333 \\ -1 \end{bmatrix}$$

Til slutt

$$\mathbf{w}_1 = \mathbf{w}_0 + \frac{0.5}{4}(\mathbf{k}_1 + 3\mathbf{k}_2) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.125 \cdot \left(\begin{bmatrix} 0 \\ -1 \end{bmatrix} + 3 \begin{bmatrix} -0.333 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 0.875 \\ -0.5 \end{bmatrix}$$

(b) Lokal avbruddsfeil i metoden i **(a)** er $\mathcal{O}(h^3)$. I Eulers metode er den $\mathcal{O}(h^2)$. Først bes vi om å beregne en tilnærmelse med Eulers metode fremdeles med $h = 0.5$ og de samme startverdier. Siden vi allerede har beregnet den samme $\mathbf{k}_1$ som brukes i Eulers metode, så er dette fort gjort. La oss kalle denne approksimasjonen for $\mathbf{w}_E$

$$\mathbf{w}_E = \mathbf{w}_0 + h\mathbf{k}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.5 \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$$

Nå bes vi om å vurdere lokal avbruddsfeil i Eulers metode. Da har vi situasjonen $p < \hat{p}$ med referanse til kapittel 8. Vi bruker differansen $\mathbf{w}_E - \mathbf{w}_1$ som feilestimat, og finner da

$$\mathbf{est}_1 = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix} - \begin{bmatrix} 0.875 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0 \end{bmatrix}$$

Til slutt bes vi om å estimere en skrittlengde h^* som gjør at lokal avbruddsfeil blir omtrent 10^{-4}. Da kan vi simpelthen benytte oss av (1) idet vi setter `tol` $= 10^{-4}$, $\|\mathbf{est}_1\| = 0.125$ og $h = 0.5$. Finner da $h_{ny} = h^*$.

$$h^* = \left(\frac{10^{-4}}{0.125} \right)^{1/2} \cdot 0.5 = 0.0141$$

Mai 2017, Oppgave 4. Initialverdiproblemet kan skrives (kort) som

$$y'' = k(v(t,y) - y'), \quad y(0) = \hat{y}_0,\ y'(0) = \hat{y}_1,\ k > 0,\ v \text{ gitt funksjon.}$$

(a) Skriv om ligningen til et system av førsteordens differensialligninger. Vi løser dette ved å skrive $z = y'$ og får da følgende to differensialligninger

$$\begin{aligned} y' &= z, & y(0) &= \hat{y}_0 \\ z' &= k(v(t,y) - z), & z(0) &= \hat{y}_1 \end{aligned}$$

(b) Vi anmodes om å benytte baklengs Euler (implisitt Euler) idet vi bruker $v(t,y) = \sin y$, $\hat{y}_0 = \frac{\pi}{2}$, $\hat{y}_1 = 0$, $k = 2$, $h = 0.1$. Første oppdrag er å skrive ned systemet av ikke-lineære ligninger som må løses for å ta ett skritt.

$$\mathbf{Y} := \begin{bmatrix} y \\ z \end{bmatrix}, \quad \rightarrow \quad \mathbf{Y}_1 = \mathbf{Y}_0 + h\mathbf{F}(\mathbf{Y}_1)$$

der

$$\mathbf{F}(\mathbf{Y}_1) = \begin{bmatrix} z_1 \\ k(\sin y_1 - z_1) \end{bmatrix}, \quad \mathbf{Y}_0 = \begin{bmatrix} \frac{\pi}{2} \\ 0 \end{bmatrix}$$

De to ligningene blir da

$$\begin{aligned} y_1 &= \frac{\pi}{2} + 0.1\, z_1 \\ z_1 &= 0 + 0.1 \cdot 2(\sin y_2 - z_1) = 0.2(\sin y_1 - z_1) \end{aligned}$$

eller om man vil

$$\begin{aligned} y_1 - 0.1\, z_1 &= \frac{\pi}{2} \\ 1.2\, z_1 - 0.2\, \sin y_1 &= 0 \end{aligned}$$

(c) Så skal man gjøre en iterasjon med Newtons metode for dette systemet. Som startverdi til iterasjonen skal man bruke $y_1^{(0)} = \frac{\pi}{2}$ og $z_1^{(0)} = 0$. En mulig framgangsmåte her er å redusere problemet til en skalar ligning, for eksempel ved å sette inn $y_1 = \frac{\pi}{2} + 0.1\, z_1$ fra den første ligningen inn i den andre, og så løse bare for z_1. Men siden vi har lært om Newtons metode for systemer, så kan vi forsøke å løse dette som et ikke-lineært ligningssystem. Systemet kan skrives $\mathbf{f}(y,z) = 0$ der vi nå setter opp funksjonen og dens Jacobimatrise

$$\mathbf{f}\begin{pmatrix} y \\ z \end{pmatrix} = \begin{bmatrix} y - 0.1\, z - \frac{\pi}{2} \\ 1.2\, z - 0.2\, \sin y \end{bmatrix}, \qquad D\mathbf{f}\begin{pmatrix} y \\ z \end{pmatrix} = \begin{bmatrix} 1 & -0.1 \\ -0.2\, \cos y & 1.2 \end{bmatrix}$$

Vi følger strategien fra Modul 1 og skriver

$$\boldsymbol{\delta}^{(k)} = \begin{bmatrix} y^{(k+1)} \\ z^{(k+1)} \end{bmatrix} - \begin{bmatrix} y^{(k)} \\ z^{(k)} \end{bmatrix}$$

Vi må da løse

$$\boldsymbol{\delta}^{(k)} = -\left[D\mathbf{f}(y^{(k)}, z^{(k)})\right]^{-1} \mathbf{f}(y^{(k)}, z^{(k)})$$

det vil si, løs systemet

$$D\mathbf{f}(y^{(k)}, z^{(k)}) \cdot \boldsymbol{\delta}^{(k)} = -\mathbf{f}(y^{(k)}, z^{(k)})$$

For den første iterasjonen evaluerer vi

$$\mathbf{f}(\frac{\pi}{2}, 0) = \begin{bmatrix} \frac{\pi}{2} - 0.1 \cdot 0 - \frac{\pi}{2} \\ 1.2 \cdot 0 - 0.2 \cdot \sin \frac{\pi}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix}$$

og

$$D\mathbf{f}(\frac{\pi}{2}, 0) = \begin{bmatrix} 1 & -0.1 \\ 0 & 1.2 \end{bmatrix}$$

Løser ligningssystem

$$\begin{bmatrix} 1 & -0.1 \\ 0 & 1.2 \end{bmatrix} \cdot \begin{bmatrix} \delta_1^{(0)} \\ \delta_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix} \Rightarrow \begin{bmatrix} \delta_1^{(0)} \\ \delta_2^{(0)} \end{bmatrix} = \begin{bmatrix} \frac{1}{60} \\ -\frac{1}{6} \end{bmatrix} = \begin{bmatrix} 0.01667 \\ -0.1667 \end{bmatrix}$$

Så til slutt får vi

$$y^{(1)} = y^{(0)} + \delta_1^{(0)} = \frac{\pi}{2} + \frac{1}{60} \approx 1.5875$$
$$z^{(1)} = z^{(0)} + \delta_2^{(0)} = -\frac{1}{6} \approx -0.16667$$

Det later til å være en fortegnsfeil i den offisielle fasiten på den andre komponenten.