NTNU
Matematisk institutt

Matematiske metoder
Numerikk
Høst 2021

Løsningsforslag - Øving 1

Fra Sauer 2E, avsnitt 0.1

1a I oppgaven skal du skrive om polynomet $P(x) = 6x^4 + x^3 + 5x^2 + x + 1$ ved hjelp av Horners metode. Regn ut vardien til polynomet for $x = 1/3$.

Løsning: $P(x) = (((6x + 1)x + 5)x + 1)x + 1$.

$P(1/3) = (((6\cdot\frac{1}{3}+1)\cdot\frac{1}{3}+5)\cdot\frac{1}{3}+1)\cdot\frac{1}{3}+1 = ((3\cdot\frac{1}{3}+5)\cdot\frac{1}{3}+1)\cdot\frac{1}{3}+1 = (6\cdot\frac{1}{3}+1)\cdot\frac{1}{3}+1 = 3\cdot\frac{1}{3}+1 = 2.$

Uten Horners metode får vi $P(1/3) = 6 \cdot \left(\frac{1}{3}\right)^4 + \left(\frac{1}{3}\right)^3 + 5 \cdot \left(\frac{1}{3}\right)^2 + \frac{1}{3} + 1$
$= 6 \cdot \frac{1}{81} + \frac{1}{27} + 5 \cdot \frac{1}{9} + \left(\frac{1}{3}\right) + 1 = \frac{2}{27} + \frac{1}{27} + \frac{5}{9} + \frac{1}{3} + 1 = \frac{2}{27} + \frac{1}{27} + \frac{15}{27} + \frac{9}{27} + 1 = 2.$

CP1 I oppgaven skal du regne ut verdien av $f(x) = \frac{x^51-1}{x-1}$ for $x = 1.00001$ ved å sette inn i formelen og å bruke at $f(x) = p(x) = x^{50} + x^{49} + x^{48} + x^{47} + x^{46} + \cdots + x^3 + x^2 + x + 1$. Følgende *python-kode* gir svaret. (Pakken numpy må være importert)

```python
from minefunksjoner import nest
import numpy
c = [1] * 51
x = 1.00001
y = nest(50,c,x)
print("Svar= %f" % y)
feil=y-(x**51-1)/(x-1)
print("Feil= %e" % feil)
```

Koden for nest ligger i filen `minefunksjoner.py`. Hvis din kode ligger i en annen fil bytter du `minefunksjoner` med roten til navnet på din fil. Koden produserer:

```
Svar= 51.012752
Feil= 4.760636e-12
```

Legg merke til at forskjellen er lik $21440\,\varepsilon_{mach}$. Den modne student vil ønske å vite hva som gir riktigst svar.

Svaret på det er at polynomet $p(x)$ gir et mye mer nøyaktig svar enn brøkuttrykket da feilen blir stor når vi trekker to svært like tall fra hverandre. Maksimal feil i $x - 1$ er $\varepsilon_{mach}/2$. Relativ feil i $x - 1$ er derfor $\varepsilon_{mach}/2/(x - 1)$.

Når $x = 1.00001$ blir teoretisk relativ feil $50000\varepsilon_{mach}$. I praksis ser du at den blir litt mindre.

Fra Sauer 2E, avsnitt 0.2

3b Gjør om $1/3$ til binært.

- Heltallsdelen er 0.

- Vi finner binærbrøkdelen:

$$\begin{aligned} 1/3 \times 2 &= 2/3 + 0 \\ 2/3 \times 2 &= 1/3 + 1 \\ 1/3 \times 2 &= 2/3 + 0 \\ &\vdots \end{aligned}$$

Vi får derfor $1/3 = (0{,}010101\ldots)_2 = (0{,}\overline{01})_2$.

5 Vi starter med å skrive π som en sum av et heltall og en ekte brøk.

$$\pi = 3 + 0.141592653589793$$

Heltallet skrevet om på binært er $3 = 2 + 1 = (11)_2$. Vi utfører gjentatt multiplikasjon med to.

$$\begin{aligned} 0.141592653589793 \cdot 2 &= 0.283185307179586 \\ 0.283185307179586 \cdot 2 &= 0.566370614359173 \\ 0,566370614359173 \cdot 2 &= 1.132741228718346 \\ 0,132741228718346 \cdot 2 &= 0.265482457436692 \\ 0,265482457436692 \cdot 2 &= 0.530964914873384 \\ 0,530964914873384 \cdot 2 &= 1.061929829746767 \\ 0,061929829746767 \cdot 2 &= 0.123859659493535 \\ 0,123859659493535 \cdot 2 &= 0.247719318987069 \\ 0,247719318987069 \cdot 2 &= 0.495438637974138 \\ 0,495438637974138 \cdot 2 &= 0.990877275948276 \\ 0,990877275948276 \cdot 2 &= 1.981754551896552 \\ 0,981754551896552 \cdot 2 &= 1.963509103793105 \\ 0,963509103793105 \cdot 2 &= 1.927018207586209 \\ 0,927018207586209 \cdot 2 &= 1.854036415172419 \end{aligned}$$

Vi leser av de binære sifrene som heltallsdelen av tallene til høyre i tabellen.

$$\pi = (11.0010010000111)_2$$

Neste bit er 1 så vi har trunkert. Avrunding oppover ville gitt

$$\pi = (11.0010010001000)_2.$$

Begge svar godkjennes.

7f Gjør om $x = (110{,}1\overline{101})_2$ til desimalt. Vi har $8x = (110110{,}1\overline{101})_2$ og $7x = 8x - x = (110000)_2 = 48$. Dvs at $x = 48/7$.

Fra Sauer 2E, avsnitt 0.3

1d Du skal skrive om $x = 0.9$ til flyttall $fl(x)$.

Vi skriver det først på normalisert form: $x = 1.8 \cdot 2^{-1}$.

Vi skal skrive 0.8 med 52 bit.

$$\begin{aligned}
0.8 \cdot 2 &= 0.6 + 1 \\
0.6 \cdot 2 &= 0.2 + 1 \\
0.2 \cdot 2 &= 0.4 + 0 \\
0.4 \cdot 2 &= 0.8 + 0 \\
\mathbf{0.8} \cdot 2 &= 0.6 + 1 \\
&\vdots
\end{aligned}$$

Vi får repetisjon med periode 4. Mantissa blir derfor kodet med

```
1100 ..44(bits).. 1100 | 1100
```

Vi må runde oppover da første bit som sløyfes er 1 og de resterende bit er ikke 1.

```
1100 ..44(bits).. 1101
```

$$fl(0.9) = 1.1100110011001100110011001100110011001100110011001101 \cdot 2^{-1}$$

3 For hvilket positive heltall kan $5 + 2^{-k}$ presenteres eksakt i flyttall?

Vi har at $5 = 1.25 \cdot 2^2 = 1.01_2 \cdot 2^2$. Det minste tallet vi kan legge til 1.01_2 i flyttall uten å få avrunding er 2^{-52}. Det gir $5 + 2^{-k} = (1.01_2 + 2^{-52}) \cdot 2^2 = 1.01_2 \cdot 2^2 + 2^{-52} \cdot 2^2 = 5 + 2^{-50}$. Det gir $k = 50$.

5 **a)** I double-presisjons aritmetikk har vi 52 binære siffer etter ledende siffer. Da er

$$(1 + (2^{-51} + 2^{-53})) = 1,000 \ldots \text{i alt 50 nuller} \ldots 0010|100 \ldots$$

Bit 53 er en og alle bits til høyre for dette er 0. Bit 52 er null og derfor skal IKKE en legges til i bit 52. Svaret er derfor

$$(1 + (2^{-51} + 2^{-53})) - 1 \approx 2^{-51}$$

b) I $(1 + (2^{-51} + 2^{-52} + 2^{-53}))$ er bit 53 en og alle bits til høyre for dette er 0. Bit 52 er en og derfor skal en legges til i bit 52. Svaret er derfor

$$(1 + (2^{-51} + 2^{-52} + 2^{-53})) - 1 \approx 2^{-50}$$

Fra Sauer 2E, avsnitt 0.4

1a Små verdier av x gjør at 1 og $\sec x$ er like. Vi skriver derfor om uttrykket:

$$\frac{1 - \sec x}{\tan^2 x} = \frac{1 - \sec x}{\sec^2 x - 1} = \frac{1 - \sec x}{(\sec x - 1)(\sec x + 1)} = \frac{-1}{\sec x + 1}$$

2 For å løse $x^2 + 3x - 8^{-14} = 0$ kan vi bruke abc-formelen eller fullføring av kvadratet. Begge deler gir

$$x = \frac{-3 \pm \sqrt{9 + 4 \cdot 8^{-14}}}{2}.$$

Tilfellet med minus foran rottegnet gir roten $r_1 = -3$. I det andre tilfellet derimot må vi skrive om utrykket for å hindre oversvømming.

$$r_2 = \frac{-3+\sqrt{9+4\cdot 8^{-14}}}{2} = \frac{(-3+\sqrt{9+4\cdot 8^{-14}})(-3-\sqrt{9+4\cdot 8^{-14}})}{2(-3-\sqrt{9+4\cdot 8^{-14}})}$$
$$= \frac{3^2-(\sqrt{9+4\cdot 8^{-14}})^2}{-12} = \frac{-4\cdot 8^{-14}}{-12} \approx 7.5791\cdot 10^{-14}..$$

Avrundet til 3 desimaler får vi $r_1 = -3.00$ og $r_2 = 7.58\cdot 10^{-14}$.

NB! Merk at abc-formelen gir $r_2 = 7.57\cdot 10^{-14}$ som ikke har 3 riktige desimaler.

CP1a Vi bruker svaret fra oppgave 1a. Følgende Python-kode besvarer spørsmålet.

```
import numpy as np
x = 10.0**np.array(range(-1,-15,-1))
res = np.ones( [14,3] )
res[:,0] = x
res[:,1] = (1-1/np.cos(x))/(np.tan(x)**2)
res[:,2] = 1.0/(-1-1/np.cos(x))
print(res)
```

Koden produserer:

```
>> cp0_4_1a
[[  1.00000000e-01  -4.98747914e-01  -4.98747914e-01]
 [  1.00000000e-02  -4.99987500e-01  -4.99987500e-01]
 [  1.00000000e-03  -4.99999875e-01  -4.99999875e-01]
 [  1.00000000e-04  -4.99999994e-01  -4.99999999e-01]
 [  1.00000000e-05  -5.00000041e-01  -5.00000000e-01]
 [  1.00000000e-06  -5.00044450e-01  -5.00000000e-01]
 [  1.00000000e-07  -5.10702591e-01  -5.00000000e-01]
 [  1.00000000e-08   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-09   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-10   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-11   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-12   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-13   0.00000000e+00  -5.00000000e-01]
 [  1.00000000e-14   0.00000000e+00  -5.00000000e-01]]
```

CP5 Hypotenusen har lengde $h = \sqrt{x^2+y^2}$ som er nesten lik x. Vi kan derfor ikke regne ut $h-x$ direkte. Da får vi null som svar. Vi må skrive om ved å omforme

$$h-x = \frac{h-x}{1} = \frac{(h-x)(h+x)}{h+x} = \frac{h^2-x^2}{h+x} = \frac{y^2}{h+x}$$

```
import math
x = 3344556600
y = 1.2222222
h = math.sqrt(x**2+y**2)
diff = y**2/(x+h)
print(diff)
```

Kall til koden (*python filnavn.py*) produserer:

```
2.23322144731e-10
```

Fra Sauer 2E, avsnitt 1.1

1a Vi regner ut forskjellige verdier av $f(x) = x^3 - 9$.

x	0	1	2	3
$f(x)$	-9	-8	-1	18

Funksjonen $f(x)$ skifter fortegn på intervallet $[2, 3]$ og fordi $f(x)$ er kontinuerlig kan vi bruke mellomverdisetningen til å konkludere at $f(x)$ tar verdien 0 i minst et punkt c på intervallet $[2, 3]$.

3a For at feilen skal være mindre enn $1/8$, så må c_k være midtpunktet i et intervall av lengde $1/4$. Dvs at vi gjør 2 halveringer.

k	a_k	$f(a_k)$	c_k	$f(c_k)$	b_k	$f(b_k)$
0	2	$-$	2,5	$+$	3	$+$
1	2	$-$	2,25	$+$	2,5	$+$
2	2	$-$	2,125		2,25	$+$

Svaret er $x = 2{,}125$.

CP1 Vi lager ét Python-script for alle tre deloppgavene.

```python
import numpy as np

def halvering(f,tol):
    a=0;b=1;
    for a in range(10):
        if (f(a)*f(a+1)<0):
            b=a+1
            break
        if (f(-a)*f(-a-1)<0):
            a=-a
            b=a-1
            break
    feil=0.5
    while(feil>tol):
        c=(a+b)/2
        if (f(a)*f(c)>0):
            a=c
        else:
            b=c
        feil /= 2
    c=(a+b)/2
    return c

TOL=.5*10**(-6) # smaller than necessary
print(halvering(lambda x:x**3-9,TOL))
print(halvering(lambda x:3*x**3+x**2-x-5,TOL))
print(halvering(lambda x:np.cos(x)**2+6-x,TOL))
```

Vi får svarene: a) 2.08008 b) 1.16973 c) 6.77609

Fra Sauer 2E, avsnitt 1.2

1 **a)** For å finne fikspunktene til $3/x$ må vi løse $3/x = x$. Multipliserer med x på begge sider: $3 = x^2$. Derfor er fikspunktene $x = \pm\sqrt{3}$.

b) For å finne fikspunktene til $x^2 - 2x + 2$ må vi løse $x^2 - 2x + 2 = x$. Trekker fra x på begge sider: $x^2 - 3x + 2$ og løser. Fikspunktene $x = 1$ og $x = 2$.

c) For å finne fikspunktene til $x^2 - 4x + 2$ må vi løse $x^2 - 4x + 2 = x$. Trekker fra x på begge sider: $x^2 - 5x + 2$ og løser. Det er to fikspunkt $x = \frac{5\pm\sqrt{17}}{2}$.

CP1 Vi lager ét Python-script for alle tre deloppgavene.

```python
import math
def fikspunkt( f , f_name , x , TOL ):
    i=0
    print("\\begin{eqnarray*}")
    while(True):
        resultat = f(x)
        print("x_{{{0}}}&=&{1}(x_{{{2}}})={3:.9f}\\\\".
           format(i+1,f_name,i,resultat));
        if(math.fabs(resultat-x) < TOL):
            break
        i=i+1;
        x=resultat
    print("\\end{eqnarray*}")

# a)
f = lambda x : (2*x+2)**(1.0/3)
fikspunkt(f,'f',1,0.5e-8)

# b)
g = lambda x : math.log(7-x)
fikspunkt(g,'g',1,0.5e-8)

# c)
h = lambda x: math.log(4-math.sin(x));
fikspunkt(h,'h',1,0.5e-8)
```

a) Vi skriver om til $x = \sqrt[3]{2x+2}$. Kjører vi scriptet får vi LaTeX-kode som gir resultatet

$$
\begin{aligned}
x_1 &= f(x_0) = 1.587401052 \\
x_2 &= f(x_1) = 1.729675293 \\
x_3 &= f(x_2) = 1.760814726 \\
x_4 &= f(x_3) = 1.767485065 \\
x_5 &= f(x_4) = 1.768907380 \\
x_6 &= f(x_5) = 1.769210364 \\
x_7 &= f(x_6) = 1.769274893 \\
x_8 &= f(x_7) = 1.769288636 \\
x_9 &= f(x_8) = 1.769291562 \\
x_{10} &= f(x_9) = 1.769292186 \\
x_{11} &= f(x_{10}) = 1.769292318 \\
x_{12} &= f(x_{11}) = 1.769292347 \\
x_{13} &= f(x_{12}) = 1.769292353 \\
x_{14} &= f(x_{13}) = 1.769292354
\end{aligned}
$$

b) Vi skriver om til $x = \ln(7 - x)$. Resultatet ble

$$
\begin{aligned}
x_1 &= g(x_0) = 1.791759469 \\
x_2 &= g(x_1) = 1.650242089 \\
x_3 &= g(x_2) = 1.677051310 \\
x_4 &= g(x_3) = 1.672027415 \\
x_5 &= g(x_4) = 1.672970788 \\
x_6 &= g(x_5) = 1.672793712 \\
x_7 &= g(x_6) = 1.672826952 \\
x_8 &= g(x_7) = 1.672820712 \\
x_9 &= g(x_8) = 1.672821884 \\
x_{10} &= g(x_9) = 1.672821664 \\
x_{11} &= g(x_{10}) = 1.672821705 \\
x_{12} &= g(x_{11}) = 1.672821697 \\
x_{13} &= g(x_{12}) = 1.672821699
\end{aligned}
$$

c) Vi skriver om til $x = \ln(4 - \sin x)$. Resultatet ble

$$
\begin{aligned}
x_1 &= h(x_0) = 1.150106418 \\
x_2 &= h(x_1) = 1.127262133 \\
x_3 &= h(x_2) = 1.130356190 \\
x_4 &= h(x_3) = 1.129928735 \\
x_5 &= h(x_4) = 1.129987634 \\
x_6 &= h(x_5) = 1.129979515 \\
x_7 &= h(x_6) = 1.129980634 \\
x_8 &= h(x_7) = 1.129980480 \\
x_9 &= h(x_8) = 1.129980501 \\
x_{10} &= h(x_9) = 1.129980498
\end{aligned}
$$