MATH50003 Numerical Analysis (2022–23)

Problem Sheet 3

This problem sheet explores the error in using divided differences and using dual numbers.

Please complete the problems using pen-and-paper, though some can be verified using Julia.

Problem 1 Suppose our floating point approximation $f^{\rm FP} : F \to F$ has *relative accuracy*:

$$f^{\rm FP}(x) = f(x)(1 + \delta_x^{\rm r})$$

where

$$|\delta_x^{\rm r}| \leq c \epsilon_{\rm m}.$$

Suppose further that $f(0) = f^{\rm FP}(0) = 0$ and assume that $f'(0) \neq 0$. Show that divided differences achieves relative accuracy:

$$\frac{f^{\rm FP}(h)}{h} = f'(0)(1 + \varepsilon_h)$$

where

$$|\varepsilon_h| \leq \frac{M}{2f'(0)} h (1 + c\epsilon_{\rm m}) + c\epsilon_{\rm m}$$

for $M = \sup_{0 \leq t \leq h} |f''(t)|$.

SOLUTION

$$\begin{aligned}
\frac{f^{\rm FP}(h)}{h} &= \frac{f(h)}{h}(1 + \delta_x^{\rm r}) = (f'(0) + f''(t)h/2)(1 + \delta_x^{\rm r}) = f'(0)(1 + f''(t)h/(2f'(0)) \\
&= f'(0)(1 + \underbrace{f''(t)h/(2f'(0)) + \delta_x^{\rm r} + f''(t)h/(2f'(0))\delta_x^{\rm r}}_{\varepsilon_h})
\end{aligned}$$

where $t \in [0,h]$. The bound follows.

END

Problem 2.1 For

$$\frac{f(x+h)-f(x-h)}{2h}=f'(x)+\delta_{x,h}^{\mathrm{T}},$$

bound the absolute error $|\delta_{x,h}^{\mathrm{T}}|$ in terms of

$$M=\max_{y\in[x-h,x+h]}\left|f'''(y)\right|.$$

SOLUTION

By Taylor's theorem, the approximation around $x+h$ is

$$f(x+h)=f(x)+f'(x)h+\frac{f''(x)}{2}h^2+\frac{f'''(z_1)}{6}h^3,$$

for some $z_1\in(x,x+h)$ and similarly

$$f(x-h)=f(x)+f'(x)(-h)+\frac{f''(x)}{2}h^2-\frac{f'''(z_2)}{6}h^3,$$

for some $z_2\in(x-h,x)$.

Subtracting the second expression from the first we obtain

$$f(x+h)-f(x-h)=f'(x)(2h)+\frac{f'''(z_1)+f'''(z_2)}{6}h^3.$$

Hence,

$$\frac{f(x+h)-f(x-h)}{2h}=f'(x)+\underbrace{\frac{f'''(z_1)+f'''(z_2)}{12}h^2}_{\delta_{x,h}^{\mathrm{T}}}.$$

Thus, the error can be bounded by

$$\left|\delta_{x,h}^{\mathrm{T}}\right|\leq\frac{M}{6}h^2.$$

END

Problem 2.2 Assume that

$$f^{\mathrm{FP}}(x)=f(x)+\delta_x^f$$

where $|\delta_x^f|\leq c\epsilon_{\mathrm{m}}$. For the *absolute error* $\delta_{x,h}^{\mathrm{CD}}$ satisfying

$$\frac{f^{\mathrm{FP}}(x+h)\ominus f^{\mathrm{FP}}(x-h)}{2h}=f'(x)+\delta_{x,h}^{\mathrm{CD}}$$

find a bound on $|\delta_{x,h}^{\mathrm{CD}}|$ in terms of M. You may assume all operations result in numbers in the normalised range, $h=2^{-n}$, $x\oplus h=x+h$ and $x\ominus h=x-h$.

SOLUTION

In floating point we have

$$
\begin{aligned}
\frac{f^{\mathrm{FP}}(x+2h) \ominus f^{\mathrm{FP}}(x-2h)}{2h} &= \frac{f(x+h) + \delta_{x+h} - f(x-h) - \delta_{x-h}}{2h}(1+\delta_1) \\
&= \frac{f(x+h) - f(x-h)}{2h}(1+\delta_1) + \frac{\delta_{x+h} - \delta_{x-h}}{2h}(1+\delta_1
\end{aligned}
$$

Applying Taylor's theorem we get

$$
(f^{\mathrm{FP}}(x+h) \ominus f^{\mathrm{FP}}(x-h)) \oslash (2h) = f'(x) + \underbrace{f'(x)\delta_1 + \delta^{\mathrm{T}}_{x,h}(1+\delta_1) + \frac{\delta_{x+h} - \delta_{x-h}}{2h}(}_{\delta^{\mathrm{CD}}_{x,h}}
$$

where

$$
|\delta^{\mathrm{CD}}_{x,h}| \leq \frac{|f'(x)|}{2}\epsilon_{\mathrm{m}} + \frac{M}{3}h^2 + \frac{2c\epsilon_{\mathrm{m}}}{h}
$$

END

Problem 3.1 For the second-order derivative approximation

$$
\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = f''(x) + \delta^{\mathrm{T}}_{x,h}
$$

bound the absolute error $|\delta^{\mathrm{T}}_{x,h}|$ in terms of

$$
M = \max_{y \in [x-h, x+h]} |f'''(y)|.
$$

SOLUTION Using the same two formulas as in 1.1 we have

$$
f(x+h) = f(x) + f'(x)h + \frac{f''(x)}{2}h^2 + \frac{f'''(z_1)}{6}h^3,
$$

for some $z_1 \in (x, x+h)$ and

$$
f(x-h) = f(x) + f'(x)(-h) + \frac{f''(x)}{2}h^2 - \frac{f'''(z_2)}{6}h^3,
$$

for some $z_2 \in (x-h, x)$.

Summing the two we obtain

$$
f(x+h) + f(x-h) = 2f(x) + f''(x)h^2 + \frac{f'''(z_1)}{6}h^3 - \frac{f'''(z_2)}{6}h^3.
$$

Thus,

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + \frac{f'''(z_2) - f'''(z_1)}{6}h.$$

Hence, the error is

$$|\delta^{\mathrm{T}}_{x,h}| = \left| f''(x) - \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \right| = \left| \frac{f'''(z_2) - f'''(z_1)}{6}h \right| \le \frac{Mh}{3}.$$

END

Problem 3.2 Assume that

$$f^{\mathrm{FP}}(x) = f(x) + \delta^f_x$$

where $|\delta^f_x| \le c\epsilon_{\mathrm{m}}$. For the *absolute error* $\delta^{\mathrm{2D}}_{x,h}$ satisfying

$$(f^{\mathrm{FP}}(x+h) \ominus 2f^{\mathrm{FP}}(x) \oplus f^{\mathrm{FP}}(x-h))/h = f''(x) + \delta^{\mathrm{2D}}_{x,h}$$

find a bound on $|\delta^{\mathrm{2D}}_{x,h}|$ in terms of M and $F = \sup_{x-h\le t\le x+h} |f(t)|$. You may assume all operations result in numbers in the normalised range, $h = 2^{-n}$, $x \oplus h = x + h$ and $x \ominus h = x - h$.

SOLUTION

We have

$$\begin{aligned} f^{\mathrm{FP}}(x+h) \ominus 2f^{\mathrm{FP}}(x) &= (f(x+h) + \delta_{x+h} - 2f(x) - 2\delta^f_x)(1+\delta_1) \\ &= f(x+h) - 2f(x) + \underbrace{(f(x+h) - 2f(x))\delta_1 + (\delta_{x+h} - 2\delta^f_x)(1}_{\delta_2} \end{aligned}$$

where $|\delta_2| \le (3/2F + 4c)\epsilon_m$. Therefore

$$\begin{aligned} (f^{\mathrm{FP}}(x+h) \ominus 2f^{\mathrm{FP}}(x)) \oplus f^{\mathrm{FP}}(x-h) &= ((f(x+h) - 2f(x) + \delta_2) + f(x-h) + \delta_{x-} \\ &= f(x+h) - 2f(x) + f(x-h) + \underbrace{(f(x+h)} \end{aligned}$$

where $|\delta_4| \le (5F + 10c)\epsilon_{\mathrm{m}}$. Putting everything together we have

$$\begin{aligned} \frac{(f^{\mathrm{FP}}(x+h) \ominus 2f^{\mathrm{FP}}(x)) \oplus f^{\mathrm{FP}}(x-h)}{h} &= \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + \frac{\delta_4}{h^2} \\ &= f''(x) + \frac{f'''(z_2) - f'''(z_1)}{6}h + \frac{\delta_4}{h^2} \end{aligned}$$

that is the error is bounded by

$$\frac{Mh}{3} + (5F + 10c)\frac{\epsilon_{\mathrm{m}}}{h^2}$$

END

Problem 4 Show that dual numbers $\mathbb{D}$ are a *commutative ring*, that is, for all $a, b, c \in \mathbb{D}$ the following are satisfied:

1. *additive associativity*: $(a + b) + c = a + (b + c)$
2. *additive commutativity*: $a + b = b + a$
3. *additive identity*: There exists $0 \in \mathbb{D}$ such that $a + 0 = a$.
4. *additive inverse*: There exists $-a$ such that $(-a) + a = 0$.
5. *multiplicative associativity*: $(ab)c = a(bc)$
6. *multiplictive commutativity*: $ab = ba$
7. *multiplictive identity*: There exists $1 \in \mathbb{D}$ such that $1a = a$.
8. *distributive*: $a(b + c) = ab + ac$

SOLUTION In what follows we write $a = a_r + a_d\epsilon$ and likewise for $b, c \in \mathbb{D}$.

Additive associativity and commutativity and existence of additive inverse are both immediate results of dual number addition reducing to element-wise real number addition. Furthermore, by definition of addition on $\mathbb{D}$ the dual number $0 + 0\epsilon$ acts as the additive identity since

$$(a_r + a_d\epsilon) + (0 + 0\epsilon) = (a_r + a_d\epsilon).$$

We explicitly prove multiplicative commutativity

$$ab = (a_r + a_d\epsilon)(b_r + b_d\epsilon) = a_rb_r + (a_rb_d + a_db_r)\epsilon = b_ra_r + (b_ra_d + b_da_r)\epsilon = ba.$$

We also explicitly prove multiplicative associativity:

$$\begin{gathered}(ab)c = ((a_rb_r + (a_rb_d + a_db_r)\epsilon)c = a_rb_rc_r + ((a_rb_d + a_db_r)c_r + a_rb_rc_d)\epsilon = a_rb_rc_r \\ + (a_rb_dc_r + a_db_rc_r + a_rb_rc_d)\epsilon\end{gathered}$$

and

$$a(bc) = a((b_rc_r + (b_rc_d + b_dc_r)\epsilon) = a_rb_rc_r + (a_rb_dc_r + a_db_rc_r + a_rb_rc_d)\epsilon.$$

The number $1 + 0\epsilon$ serves as the multiplicative identity. Note that for any dual number a, we have

$$(1 + 0\epsilon)(a_r + a_d\epsilon) = 1a_r + (a_r0 + 1a_d)\epsilon = a_r + a_d\epsilon = a.$$

Finally we show distributivity of multiplication:

$$a(b + c) = a(b_r + c_r + (b_d + c_d)\epsilon) = (a_rb_r + a_rc_r) + (a_rb_d + a_rc_d + a_db_r + a_dc_r)\epsilon,$$

$$\begin{gathered}ab + ac = a_rb_r + (a_db_r + a_rb_d)\epsilon + a_rc_r + (a_dc_r + a_rc_d)\epsilon = (a_rb_r + a_rc_r) \\ + (a_rb_d + a_rc_d + a_db_r + a_dc_r)\epsilon.\end{gathered}$$

END

Problem 5.1 What is the correct definition of division on dual numbers, i.e.,

$$(a + b\epsilon)/(c + d\epsilon) = s + t\epsilon$$

for what choice of s and t?

SOLUTION

As with complex numbers, division is easiest to understand by first multiplying with the conjugate, that is:

$$\frac{a + b\epsilon}{c + d\epsilon} = \frac{(a + b\epsilon)(c - d\epsilon)}{(c + d\epsilon)(c - d\epsilon)}.$$

Expanding the products and dropping terms with ϵ^2 then leaves us with the definition of division for dual numbers (where the denominator must have non-zero real part):

$$\frac{a}{c} + \frac{bc - ad}{c^2}\epsilon.$$

Thus we have $s = \frac{a}{c}$ and $t = \frac{bc-ad}{c^2}$.

END

Problem 5.2 A *field* is a commutative ring such that $0 \neq 1$ and all nonzero elements have a multiplicative inverse, i.e., there exists a^{-1} such that $aa^{-1} = 1$. Can we use Problem 5.1 to define $a^{-1} := 1/a$ to make $\mathbb{D}$ a field? Why or why not?

SOLUTION

Fields require that all nonzero elements have a unique multiplicative inverse. However, this is not the case for dual numbers. To give an explicit counter example, we show that there is no dual number z which is the inverse of $0 + \epsilon$, i.e. a dual number z such that

$$\frac{(0 + \epsilon)}{(z_r + z_d\epsilon)} = 1 + 0\epsilon.$$

By appropriate multiplication with the conjugate we show that

$$\frac{(0 + \epsilon)(z_r - z_d\epsilon)}{(z_r + z_d\epsilon)(z_r - z_d\epsilon)} = \frac{z_r\epsilon}{z_r^2} = \frac{\epsilon}{z_r}.$$

This proves that no choice of real part z_r can reach the multiplicative identity $1 + 0\epsilon$ when starting from the number $0 + \epsilon$. More general results for zero real part dual numbers can also be proved.

END

Problem 6 Use dual numbers to compute the derivative of the following functions at $x = 0.1$:

$$\exp(\exp x \cos x + \sin x), \prod_{k=1}^{3} \left(\frac{x}{k} - 1\right), \text{ and } f_2^{\mathrm{s}}(x) = 1 + \frac{x-1}{2 + \frac{x-1}{2}}$$

SOLUTION

We now compute the derivatives of the three functions by evaluating for $x = 0.1 + \epsilon$. For the first function we have:

$$\begin{aligned}\exp(\exp(0.1+\epsilon)\cos(0.1+\epsilon) + \sin(0.1+\epsilon)) &= \exp((\exp(0.1) + \epsilon\exp(0.1))(\cos(0.1\\ &= \exp(\exp(0.1)\cos(0.1) + \sin(0.1) + (\\ &= \exp(\exp(0.1)\cos(0.1) + \sin(0.1))\\ &(0.1)(\cos(0.1) -\end{aligned}$$

therefore the derivative is the dual part

$$\exp(\exp(0.1)\cos(0.1) + \sin(0.1))(\exp(0.1)(\cos(0.1) - \sin(0.1)) + \cos(0.1))$$

For the second function we have:

$$\begin{aligned}(0.1 + \epsilon - 1)\left(\frac{0.1+\epsilon}{2} - 1\right)\left(\frac{0.1+\epsilon}{3} - 1\right) &= (-0.9+\epsilon)(-0.95+\epsilon/2)(-29/30 +\\ &= (171/200 - 1.4\epsilon)(-29/30 + \epsilon/3)\\ &= -1653/2000 + 983\epsilon/600\end{aligned}$$

Thus the derivative is $983/600$.

For the third function we have:

$$\begin{aligned}1 + \frac{0.1+\epsilon-1}{2 + \frac{0.1+\epsilon-1}{2}} &= 1 + \frac{-0.9+\epsilon}{1.55 + \epsilon/2}\\ &= 1 - 18/31 + 2\epsilon/1.55^2\end{aligned}$$

Thus the derivative is $2/1.55^2$.

END