Note Title 2020-03-20

Derivatives)

1) Computer algebra system

2) $\frac{f(x+h)-f(x)}{h}$ (error $\approx 10^{-8}$) (numerical derivatives)

3) Write "matrix-friendly" code + evaluate in $\begin{bmatrix} x & 1 & & \\ & x & \ddots & \\ & & \ddots & 1 \\ 0 & & & x \end{bmatrix}$

to get $\begin{bmatrix} f(x) & f'(x) & \cdots & \vdots \\ & \ddots & \ddots & f'(x) \\ 0 & & & f(x) \end{bmatrix}$ (use dimension n to get $f(x), \ldots f^{(n-1)}(x)$)

variables, e.g. $a \iff A = \begin{bmatrix} a & a' & a''/2 & \cdots & a^{(n-1)}/(n-1)! \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & a'' \\ & & & \ddots & a' \\ 0 & & & & a \end{bmatrix}$ (Triangular Toeplitz)

This could be described even without matrices:

$c = a * b$

$[a\ a'\ a''] \quad [b\ b'\ b'']$

$c = [a\ a'\ a''] * [b\ b'\ b''] =$

$= [ab \quad ab'+a'b \quad ab''+2a'b'+a''b]$

Special case: $n=2$ (first derivative only)

"dual numbers": replace a with $a+\varepsilon a'$

Use algebra + $\varepsilon^2 = 0$

$c = ab$

$(a+\varepsilon a')(b+\varepsilon b')$

$c + \varepsilon c' = ab + \varepsilon(ab'+a'b) + \cancel{\varepsilon^2 a'b'}$ ($\varepsilon^2 = 0$)

(The input variable x becomes $x+\varepsilon \cdot 1$)

Ways to see this:

- operations in $\mathbb{R}[\varepsilon]/(\varepsilon^2)$
- $\varepsilon = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$
- "ε is infinitesimal"

"Complex step differentiation":

Suppose you have f holomorphic / analytic

You are interested in $f(x) \in \mathbb{R}$ $x \in \mathbb{R}$ and $f'(x)$

You have code that computes f and works also with complex input

$$f(x+ih) = f(x) + ih f'(x) - h^2 \frac{f''(x)}{2} + O(h^3)$$

$$\frac{\mathrm{Im}\left(f(x+ih)\right)}{h} = f'(x) + O(h^2)$$

(one order of precision better than $\frac{f(x+h)-f(x)}{h} = f'(x) + O(h)$)

In the numerator: error approx. $u \cdot f'(x)$

In total: error $O\left(\frac{u \cdot f'(x)}{h} + h^2\right)$

$h \approx u^{1/3}$ gives $O\left(f'(x) u^{2/3} + u^{2/3}\right)$ $\quad u \sim 10^{-16}$

$\sim 10^{-\frac{2}{3}\cdot 16} \approx 10^{-11}$

$a \gg b$

$f(x) = (a+ib)^3 = a^3 + 3ia^2b - 3ab^2 - ib^3$

$x = a+ib$

error in computing $\mathrm{Im}\, f(x)$ $3a^2b - b^3$ is proportional to $\mathrm{Im}\, f(x) \ll f(x)$

Previous (exact) techniques: automatic differentiation.

What we described (extending operations to derivatives) is forward mode of automatic differentiation

There is also reverse mode ("back-propagation" in machine learning)

General idea: first you run your code to compute $f(x) = y$, then you go through instruction backwards and compute for each variable a $\frac{\partial y}{\partial a}$.

(We will not see details here)

Remark: In case you have to compute $Jf \in \mathbb{R}^{m \times n}$ for a function $f: \mathbb{R}^n \to \mathbb{R}^m$

forward mode is faster if $n \ll m$,

reverse mode is faster for $n \gg m$. (machine learning)

Methods for specific matrix functions

First one: $\exp m(A) = I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \dots$

The solution of the vector-valued ODE

$v: \mathbb{R} \to \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$

$$\frac{d}{dt} v(t) = A v(t) \qquad v(0) = v_0$$

is $v(t) = \exp m(At) \cdot v_0$

(Proof: we can differentiate term by term

$$v(t) = \left(I + tA + t^2\frac{A^2}{2} + t^3\frac{A^3}{3!} + \dots\right) v_0$$

$$\frac{d}{dt} v(t) = \left(0 + A + 2t\frac{A}{2} + 3t^2\frac{A^3}{3!} + \dots\right) = A\left(I + tA + t^2\frac{A^2}{2} + \dots\right) v_0$$

$$= A v(t).$$

We have seen a few problematic methods:

$A = VJV^{-1}$ (Jordan form) $\rightsquigarrow$ $k(V)$ big, not so good

Series $I + A + \frac{A^2}{2} + \dots$ $\rightsquigarrow$ problems, intermediate growth...

Moler, Van Loan '78 e '03 "Nineteen dubious ways to compute the matrix exponential".

Problem with Maclaurin series: there is large intermediate growth in matrix powers.

$$\|A^n\| \sim \rho(A)^n \qquad \|A^n\|^{1/n} \to \rho(A)$$

Even if $\rho(A) < 1$ and so $\|A^n\| \to 0$, the intermediate powers may grow a lot.

Ex:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A^4 = 0$$

$\rho(A) = 0$

Typical behavior for a non-normal matrix

(normal means $A = QDQ^*$, with Q unitary and D diagonal)

(for a normal matrix, $A^n = QD^nQ^*$ $\|A^n\| = \rho(A)^n$)

For wrong matrices, the summands in $\sum \frac{A^k}{k!} = \exp m(A)$ have a "hump"

intermediate growth

and so does $\exp m(tA)$

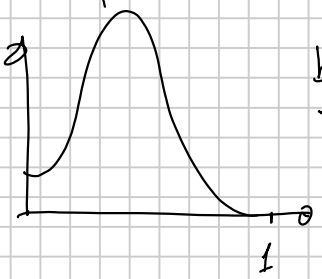
(For matrices that are far from normal)

Other dubious methods: use a method to discretize ODEs to solve

$$\frac{d}{dt}V(t) = AV \qquad V(0) = I.$$

$$V(t) = \exp(At)$$

Running an ODE solver up to $t=1$ should get $V(1) = \exp(A)$



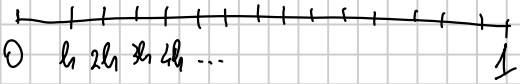
but one might have intermediate growth $\Rightarrow$ $\Rightarrow$ large errors.

Explicit Euler method

$$AV = \frac{d}{dt}V(t) \approx \frac{V(t+h)-V(t)}{h}$$

Discretize $[0,1]$ into n points with spacing $h = 1/n$

$V(0) = V_0$



$0 \quad h \quad 2h \quad 3h \quad 4h \quad \dots \quad 1$

$$\frac{V(t+h)-V(t)}{h} \approx AV(t)$$

$$V(h) = V(0) + hAV(0) =$$

$$V(2h) = V(h) + hAV(h)$$

$$\vdots$$

$$\hookrightarrow V(h) = (I + hA)V_0$$

$$\hookrightarrow V(2h) = (I+hA)V(h) = (I+hA)^2 V_0$$

$$V(nh) = (I+hA)^n V_0 \qquad V(1) \approx \left(I + \frac{1}{n}A\right)^n$$

Padé approximants: are a variant of Maclaurin series

Idea: fix degrees p, q

We would like to find

$$\exp(x) \approx \frac{N(x)}{D(x)} \quad \begin{matrix} \leftarrow \text{degree } p \\ \leftarrow \text{degree } q \end{matrix}$$

in the sense that $\exp(x) = \frac{N(x)}{D(x)} + O(x^m)$ with m as large as possible

e.g. $p = q = 2$

$$\exp(x) \approx \frac{ax^2+bx+c}{dx^2+ex+1} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + O(x^5)$$

take $f = 1$ as normalization

$p = 2,\ q = 1$

$p+1$ unknowns, $q+1$ unknowns

$$\frac{ax^2+bx+c}{1+ex} = (ax^2+bx+c)(1 - ex + e^2x^2 - e^3x^3 + \dots)$$

$$= c + (b-e)x + (a+ce^2-be)x^2 + (-ae+be^2-ce^3)x^3 + O(x^4)$$

$$\stackrel{!}{=} 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + O(x^4)$$

$p+q+1$ equations; $O(x^{p+q+1})$

but you normalize by setting one to 1

$$(*) \begin{cases} 1 = c \\ 1 = b - e \\ \frac{1}{2} = a + ce^2 - be \\ \frac{1}{3!} = -ae + be^2 - ce^3 \end{cases}$$

solve in a, b, c, e

(Generalization of Maclaurin expansion, which has degree 0 in the denominator.)

Padé approximants approximate the exponential better than Maclaurin expansion for small x

---

(There is a trick to convert $(*)$ into a linear system — can you see it?)