1. Mathematics Preliminary

- limit, continuity, and derivative
- sets of continuous functions
- Taylor's Theorem
- orders of Convergence
- big $\mathcal{O}$ and little o
- vector and matrix norms

Limit, Continuity, Derivative

Limit of function f at c is defined as

$$\lim_{x \to c} f(x) = L$$

Definition: for any positive ϵ, there exists positive δ such that

$$|f(x) - L| < \epsilon \text{ whenever } |x - c| < \delta.$$

if there is no L with this property, the limit of f at c does not exist.

Continuity of function f at c is defined as

$$\lim_{x \to c} f(x) = f(c)$$

Derivative of function f at c is defined as

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}.$$

if $f'(c)$ exists, f is differentiable at c.

Fact: if f is differentiable at c, then f is continuous at c.

Continuous functions

$\mathbf{C}(\mathbf{R})$ = the set of all functions that are continuous on $\mathbf{R}$.

$\mathbf{C}^1(\mathbf{R})$ = the set of all functions for which f' is continuous on $\mathbf{R}$.

$\mathbf{C}^n(\mathbf{R})$ = the set of all functions for which $f^{(n)}$ is continuous on $\mathbf{R}$.

$\mathbf{C}^\infty(\mathbf{R})$ = the set of all functions each of whose derivatives is continuous on $\mathbf{R}$.

Fact:

$$\mathbf{C}^\infty(\mathbf{R}) \subset \cdots \subset \mathbf{C}^2(\mathbf{R}) \subset \mathbf{C}^1(\mathbf{R}) \subset \mathbf{C}(\mathbf{R}).$$

$\mathbf{C}^n([a,\ b])$ = the set of all functions of which $f^{(n)}$ is continuous on $[a,\ b]$.

Derivative and Gradient

Suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ and $x \in \mathbf{int\,dom}\, f$

the **derivative** (or **Jacobian**) of f at x is the matrix $Df(x) \in \mathbf{R}^{m\times n}$:

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

- when f is real-valued (*i.e.*, $f : \mathbf{R}^n \to \mathbf{R}$), the derivative $Df(x)$ is a row vector
- its transpose is called the **gradient** of the function:

$$\nabla f(x) = Df(x)^T, \qquad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \ldots, n$$

which is a column vector in $\mathbf{R}^n$

Second Derivative

suppose f is a real-valued function ($i.e.$, $f : \mathbf{R}^n \to \mathbf{R}$)

the second derivative or **Hessian matrix** of f at x, denoted $\nabla^2 f(x)$ is

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n$$

example: the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$

$$f(x) = (1/2)x^T P x + q^T x + r,$$

where $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$

- $\nabla f(x) = Px + q$
- $\nabla^2 f(x) = P$

Taylor's Theorem

if $f \in \mathbf{C}^n[a,\ b]$ and $f^{(n+1)}$ exists on $(a,\ b)$ then for any x and c in $[a,\ b]$

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k + E_n(x)$$

where E_n is Lagrange remainder, ξ is between x and c,

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1}.$$

- the first term is polynomial in x
- E_n is not a polynomial in x since ξ depends on x in a nonpolynomial way

Mean-Value Theorem

special case $n = 0$ of Taylor's Theorem

$$f(x) = f(c) + f'(\xi)(x - c)$$

where ξ is between c and x.

for example: $x = b$ and $c = a$

$$f(b) - f(a) = f'(\xi)(b - a), \quad a < \xi < b.$$

Rolle's Theorem: special case of Mean-value Theorem

if $f(a) = 0, f(b) = 0$, then $f'(\xi) = 0$ for some $\xi \in (a,\ b)$.

Intermediate Value Theorem

let $f \in \mathbf{C}([a,b])$ and assume p is a value between $f(a)$ and $f(b)$, that is

$$f(a) \leq p \leq f(b), \quad \text{or} \quad f(b) \leq p \leq f(a)$$

then there exists a point $c \in [a,b]$ for which $f(c) = p$

- if $f(a)$ and $f(b)$ have different signs, f must cross zero once in $[a,b]$
- useful for finding the roots of functions; see bisection algorithm

Extreme Value Theorem

let $f \in \mathbf{C}([a, b])$; then there exists a point $c \in [a, b]$ such that

$$f(c) \leq f(x) \qquad \forall x \in [a, b]$$

and a point $d \in [a, b]$ such that

$$f(d) \geq f(x) \qquad \forall x \in [a, b]$$

basically means

- one can maximize (minimize) a *continuous* function f over a closed and bounded interval
- the maximum and minimum values of f are attained (and finite)
- the point for which the extremum occurs is attained

Contractive Mapping

a mapping (or function) f is said to be *contractive* on $S \subseteq \mathbf{dom}\, f$ if there exists a scalar $\lambda < 1$ such that

$$|f(x) - f(y)| \leq \lambda |x - y|$$

for all $x, y \in S$

- loosely speaking, a contractive function is a *non-expansive* map
- every contractive mapping is Lipschitz continuous
- if f is continuously differentiable on $[a, b]$ with

$$\max_{x \in [a,b]} |f'(x)| < 1$$

  then f is contractive on $[a, b]$ (by Mean-Value Theorem)

examples: $f(x) = \|x\|, e^{-x}, \cos x$ on $[0, 1]$

Contractive Mapping Theorem

if $f : S \to S$ is contractive for all $x \in S$, then

- f has a *uniqued* fixed point in S
- this fixed point is a limit of every sequence

$$x_{n+1} = f(x_n) \quad \text{with} \quad x_0 \in S$$

Monotonic sequence

a sequence

$$a_1, a_2, \dots, a_n$$

of real numbers, denoted by $\{a_n\}$

is called **monotone increasing** if

$$a_{n+1} \geq a_n, \quad \forall n \in \mathbb{N}$$

and is called **monotone decreasing** if

$$a_{n+1} \leq a_n, \quad \forall n \in \mathbb{N}$$

$\{a_n\}$ is called *monotonic* or *monotone* if $\{a_n\}$ is monotone increasing or monotone decreasing

Convergent sequence

we say a sequence $\{a_n\}$ of real numbers has a limit $a^\star$, denoted by

$$a_n \to a^\star, \qquad n \to \infty$$

or by

$$\lim_{n\to} \; a_n \; = \; a^\star$$

if for all $\epsilon > 0$ there exists an integer N such that

$$|a_n - a^\star| < \epsilon, \qquad \forall n > N$$

sandwich theorem: let $\{x_n\}, \{y_n\}$, and $\{z_n\}$ be sequences in **R** s.t.

$$x_n \leq y_n \leq z_n$$

if $x_n \to a$ and $z_n \to a$, then $y_n \to a$

Bounded sequence

- a sequence $\{a_n\}$ is *bounded above* if there exists $M \in \mathbf{R}$ such that

$$a_n \leq M \qquad \forall n \in \mathbb{N}$$

- a sequence $\{a_n\}$ is *bounded below* if there exists $m \in \mathbf{R}$ such that

$$a_n \geq m \qquad \forall n \in \mathbb{N}$$

- $\{a_n\}$ is a *bounded sequence* if it is bounded below and bounded above:

$$\exists M > 0, \qquad |a_n| \leq M \qquad \forall n \in \mathbb{N}$$

Monotone Convergence Theorem:

a monotone sequence of real numbers has a finite limit if and only if the sequence is bounded

Order of Convergence for sequences

let $\{x_n\}$ be a sequence of real numbers converging to x^*, such that

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^p} = C$$

for some nonzero C and some p

then p is called the **order of convergence** for the sequence $\{x_n\}$

linear convergence: $p = 1$ and requires that $|C| < 1$

example:

$$x_n = \left(1 + \frac{1}{n}\right)^n; \quad \lim_{n\to\infty} x_n = e$$

we can show that

$$\frac{|x_{n+1} - e|}{|x_n - e|} \to 1$$

quadratic convergence: $p = 2$

example 1:

$$x_1 = 2; x_{n+1} = \frac{1}{2}x_n + \frac{1}{x_n}, n \geq 1$$

example 2: the Newton's method to find a root of a function $f(x)$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

if the convergence occurs in the Newton's method, one can show that

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^2} = \frac{f''(x^*)}{2f'(x^*)}$$

superlinear convergence: faster than linear but not as fast as quadratic

i.e., when the sequence satisfies

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|} = 0$$

but

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^2} = \infty$$

example:

$$x_{n+1} = x_n - (x_n^2 - 2)\frac{(x_n - x_{n-1})}{(x_n^2 - x_{n-1}^2)}$$

$x_n \to \sqrt{2}$ as $n \to \infty$ and we can show that

$$\frac{|x_{n+1} - \sqrt{2}|}{|x_n - \sqrt{2}|^{1.62}} \to 0.77$$

Big $\mathcal{O}$ notation

for a given function $g(x)$ and a point c, define $\mathcal{O}(g(x))$ as

$$\mathcal{O}(g(x)) = \{f(x) \mid \exists \delta, C > 0,\ |f(x)| \leq C|g(x)|,\ \text{ for } \ |x-c| < \delta\}$$

- $\mathcal{O}(g(x))$ is pronounced as 'big-oh of g'
- set of all functions that is bounded by $g(x)$ when $x \to c$
- set of all functions with a *smaller* or the *same* rate of growth as g
- the constant C is nonzero and independent of x

with an abuse of notation, the expression

$$f(x) = \mathcal{O}(g(x)), \quad x \to c$$

means that function f belongs to the set $\mathcal{O}(g(x))$

($=$ does not really mean 'equal' or 'symmetry')

examples

$$\begin{array}{rclrrclr} x^2 & = & \mathcal{O}(x), & x \to 0; & x & = & \mathcal{O}(x^2), & x \to \infty \\ e^{-x} & = & \mathcal{O}(1), & x \to \infty; & 2x + x\sin(x) & = & \mathcal{O}(x), & x \to \infty \end{array}$$

the formula

$$f(x) = h(x) + \mathcal{O}(g(x))$$

means

$$f(x) = h(x) + w(x) \qquad \text{where} \quad w(x) \in \mathcal{O}(g(x))$$

examples:

$$2x^2 + 3x + 4 = 2x^2 + \mathcal{O}(x), \quad x \to \infty; \qquad e^x = 1 + x + \mathcal{O}(x^2), \quad x \to 0$$

if g and h are two functions such that $g(x) = \mathcal{O}(h(x))$

if $f(x) = \mathcal{O}(g(x))$, then obviously $f(x) = \mathcal{O}(h(x))$

the upper bound provided by $\mathcal{O}$-notation may or may not be tight

as $x \to \infty$, $4x^3 = \mathcal{O}(x^3)$ (tight) $4x = \mathcal{O}(x^3)$ (not tight) in many computations, it is of interest to consider

computation time VS problem size

denote $t(n)$ a computation time as a function of problem size n

one would like to estimate the computational efficiency for large n

example: running time for an algorithm is $t(n) = 2n^2 + 6n + 1$

- $t(n) = \mathcal{O}(n^2)$
- we say the algorithm has the order of n^2 time complexity
- note that $t(n) = \mathcal{O}(n^3)$ is also correct but we would want to express the worst-case running time to be as smallest as possible

some common running time complexities

$$\mathcal{O}(1), \quad \mathcal{O}(\log(n)), \quad \mathcal{O}(n), \quad \mathcal{O}(n^2), \quad \mathcal{O}(2^n)$$

as n grows, which one refers to the fastest algorithm ?

Computational time versus n

n	$\log(n)$	n	n^2	2^n
10	1	10	100	1024
100	2	100	10000	1267650600228229401496703205376
1000	3	1000	1000000	1.0715e+301
10000	4	10000	100000000	`Inf`

$$1.0715e + 301 = 10715086071862673209484250490600018105614048117055336074437503883703510511249361224931983788156958581275946729175531468251871452856923140435984577574698574803934567774824230985421074605062371141877954182153046474983581941267398767559165543946077062914571196477686542167660429831652624386837205668069376$$

how many digits does it have ?

Little o notation

for a given function $g(x)$ and a point c, define $o(g(x))$ as

$$o(g(x)) = \{f(x) \mid \exists \delta > 0, \forall C > 0,\ |f(x)| \leq C|g(x)| \ \text{ for } \ |x - c| < \delta\}$$

equivalent condition: a function f is in $o(g(x)))$ if

$$\lim_{x \to c} \frac{|f(x)|}{|g(x)|} = 0$$

- $o(g(x))$ is pronounced as 'little-oh of g'
- set of all functions with a *smaller* rate of growth than g
- f becomes insignificant relative to g as $x \to c$

the expression $f(x) = o(g(x)), \quad x \to c$

means f belongs to the set $o(g(x))$

examples:

$$\cos(x) - 1 = o(x), \quad x \to 0, \qquad \frac{1}{n\log(n)} = o\left(\frac{1}{n}\right), \quad n \to \infty$$

- $o(g(x))$ excludes all functions that have the *same* rate of growth as g
- $e.g..$, $3n = o(n^2)$ but $3n^2 \neq o(n^2)$ as $n \to \infty$

Big $\mathcal{O}$ and Little o

- when $f(x) = \mathcal{O}(g(x))$, the bound $|f(x)| \leq C|g(x)|$ holds for *some* constant $C > 0$
- when $f(x) = o(g(x))$, the bound $|f(x)| \leq C|g(x)|$ holds for *all* constant $C > 0$
- hence, if $f(x) \in o(g(x))$ then $f(x) \in \mathcal{O}(g(x))$

Vector norms

a vector norm on $\mathbf{R}^n$ is a mapping $\|\cdot\| : \mathbf{R}^n \to [0, \infty)$ that satisfies

1. $\|\alpha x\| = |\alpha|\|x\|$ for any $\alpha \in \mathbf{R}$ (homogeneity)
2. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
3. $\|x\| = 0$ if and only if $x = 0$ (definiteness)

2-norm

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

1-norm

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

∞-norm

$$\|x\|_\infty = \max_k \{|x_1|, |x_2|, \ldots, |x_n|\}$$

Matrix norms

matrix norm of $A \in \mathbf{R}^{m\times n}$ is defined as

$$\|A\| = \max_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

also often called **operator norm** or **induced norm**

properties:

1. for any x, $\|Ax\| \leq \|A\|\|x\|$
2. $\|aA\| = |a|\|A\|$ (scaling)
3. $\|A+B\| \leq \|A\| + \|B\|$ (triangle inequality)
4. $\|A\| = 0$ if and only if $A = 0$ (positiveness)
5. $\|AB\| \leq \|A\|\|B\|$ (submultiplicative)

2-norm or **spectral norm**

$$\|A\|_2 \triangleq \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

1-norm

$$\|A\|_1 \triangleq \max_{\|x\|_1=1} \|Ax\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|$$

∞-norm

$$\|A\|_\infty \triangleq \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|$$

other definitions of matrix norm also exist

Frobenius norm:

$$\|A\|_F = \sqrt{\mathbf{tr}(A^T A)} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

References

Chapter 1 in

D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, 3rd edition, Brooks & Cole, 2002

Section 1.1 in

J. F. Epperson, *An Introduction to Numerical Methods and Analysis*, John Wiley & Sons, 2007

Chapter 1 in

N. G. De Bruijn, *Asymptotic Methods in Analysis*, Dover, 1958

Chapter 3 in

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd edition, MIT Press, 2001