A

SUPPLEMENTAL MATHEMATICAL FACTS

A.1 Basic logic for proof

[1, p. 60]The negation of

for any $\epsilon > 0$, there exist $N > 0$, such that for all $n > N$, we have $|a_n - a| < \epsilon$

is

there exist an $\epsilon > 0$, such that for every $N > 0$, such that for all $n > N$, we have $|a_n - a| > \epsilon$.

[1, p. 60]The negation of

for any $\epsilon > 0$, there exist $\delta > 0$, such that for all $|x - x_0| < \delta$, we have $\left|f(x) - f(x_0)\right| < \epsilon$

is

there exist an $\epsilon > 0$, such that for every $\delta > 0$, such that for all $|x - x_0| < \delta$, we have $\left|f(x) - f(x_0)\right| > \epsilon$.

A.2 Some common limits

Lemma A.2.1 (Stirling approximation).
- *For positive integer n,*
$$\ln n! = n\ln n - n + O(\ln n).$$
-
$$\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq en^{n+1/2}e^{-n}, \forall n > 0.$$

Lemma A.2.2 (common limits summary).
-
$$\lim_{n\to\infty} \frac{\ln n}{n} = 0.$$
-
$$\lim_{n\to\infty} \frac{x^n}{n!} = 0, \forall x \in \mathbb{R}.$$
-
$$\lim_{n\to\infty} \frac{n!}{n^n} = 0.$$
-
$$\lim_{n\to\infty} M^{1/n} = 1$$
for any $M > 0$.
-
$$\lim_{n\to\infty} \frac{\ln n!}{n} = \infty, \lim_{n\to\infty} (n!)^{1/n} = \infty.$$

Proof. (2)see Lemma 3.6.4 and Lemma 1.4.3.(3)Lemma 1.4.3.(4) Lemma 1.5.1. (5) (a)Use Stirling approximation $\ln n! = n\ln n - n + O(\ln n)$ and $\ln n!/n = n - 1 + O(\ln n/n) \to \infty$.(b) Note that $(n!)^{1/n} = \exp(\ln(n!)^{1/n}) = \exp(\frac{\ln n!}{n})$. □

Note A.2.1. A helpful and general summary, as $n \to \infty$
$$\ln n \ll n^r(r > 0) \ll a^n(a > 1) \ll n! \ll n^n.$$

Lemma A.2.3 (property of e). *Define*
$$\lim_{n\to\infty} (1 + 1/n)^n = e$$

and then

$$\lim_{n\to\infty}(1 + x/n)^n = e^x$$

for any real x.

Proof.

$$\lim_{n\to\infty}((1 + x/n)^{n/x})^x = e^x$$

use the fact the $f(y) = y^x$ is continuous, such that function evaluation and limit can be exchanged. □

A.3 Common series summation

Lemma A.3.1. *[2, p. 1]*

- $$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$
- $$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$
- $$\sum_{k=1}^{n} k^3 = [\frac{n(n+1)}{2}]^2$$

Lemma A.3.2. *[2, p. 1]* *Assume $q \neq 1$.*

- $$\sum_{k=1}^{n} aq^{k-1} = a\frac{q^n - 1}{q-1}$$
- $$\sum_{k=0}^{n-1} kq^k = \frac{(n-1)q^n}{q-1} + \frac{(q-q^n)}{(q-1)^2}$$
- $$\sum_{k=0}^{n-1} (n-1-k)q^k = -(n-1)\frac{1}{q-1} - \frac{(q-q^n)}{(q-1)^2}$$

Proof. (3) use (1)(2), we have

$$\sum_{k=0}^{n-1} (n-1-k)q^k = (n-1)\frac{q^n-1}{q-1} - \frac{(n-1)q^n}{q-1} - \frac{(q-q^n)}{(q-1)^2}$$

□

A.4 Some common spaces

The metric space $(\mathbb{R}^n, d_2)$ is the set $\mathbb{R}^n$ with metric $d_2(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

[3, p. 122]The metric space l^2 is the set of all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that $\sum_{i=1}\infty x_i^2 < \infty$,i.e., $\sum_{i=1}\infty x_i^2$ converges. The metric is usually defined as

$$d_2(\{x_n\},\{y_n\}) = \sqrt{\sum_{k=1}^{\infty}(x_i - y_i)^2}$$

The metric space $l^p, 1 \leq p < \infty$,is the set of all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that $\sum_{i=1}\infty |x_i|^p < \infty$,i.e., $\sum_{i=1}\infty |x_i|^p$ converges. The metric is usually defined as

$$d_p(\{x_n\},\{y_n\}) = \sqrt[p]{\sum_{k=1}^{\infty}(x_i - y_i)^p}$$

The metric space l^∞, is the set all infinite sequence of real or complex numbers $\{x_1, x_2, ...\}$ such that every x_i is bounded. The metric is defined as

$$d_\infty(\{x_n\},\{y_n\}) = \sup_n |x_n - y_n|$$

[4, p. 75].The metric space $C[a,b] = (C[a,b], d_\infty)$ denote the set of real-valued(or complex valued) functions defined on the interval $[a,b]$. The metric d_∞ is given as

$$d_\infty(x,y) = \sup_t |x(t) - y(t)|$$

Remark A.4.1. **Caution!** Sometimes $C[a,b]$ refers to only continuous functions.[5, p. 23]

The metric space $(C[a,b], d_p)$ denote the set of real-valued(or complex valued) functions defined on the interval $[a,b]$. The metric d_p is given as

$$d_p(x,y) = [\int_a^b |x(t) - y(t)|^p \, dt]^{1/p}$$

where $1 \leq p < \infty$.[4, p. 75].

The vector space $\mathcal{L}(V,W)$ usually denotes the set of all linear operators from V into W.

A.4.1 Notations on continuously differentiable functions

- C^0 refers to continuous function

- C^1 refers to functions having continuous first derivatives, also called continuously differentiable functions.
- C^2 refers to functions having continuous second derivatives
- C^∞ refers to smooth functions

A.5 Different modes of continuity

Chain of inclusions for functions over a closed and bounded subset of the real line

$$continuouslyDifferntiable \subseteq LipschitzContinouous \subseteq UniformlyContinuous$$

Remark A.5.1.

- Continuously differentiable on a closed interval indicates the derivative is bounded $f' \leq M$, then we have
$$\left|f(x) - f(y)\right| = f'(s)\left|x - y\right| \leq M\left|x - y\right|$$
hence Lipschitz continuous.
- $f(x) = |x|$ is Lipschitz continuous but is not differentiable everywhere except at $x = 0$, therefore it is not continously differentiable.
- Lipschitz continuous $\rightarrow$ continous:
$$\left|f(x) - f(y)\right| \leq L\left|x - y\right| \rightarrow 0$$
as $\left|x - y\right| \rightarrow 0$

Lemma A.5.1 (differentiable implies continuous). *If f is differentiable on $[a, b]$, then it is continuous on $[a, b]$.*

Proof:

$$\lim_{y\rightarrow x} f(y) - f(x) = \lim_{y\rightarrow x}(y - x)(f(y) - f(x))/(y - x) =$$
$$\lim_{y\rightarrow x}(y - x) \lim_{y\rightarrow x}(f(y) - f(x))/(y - x) = 0$$

where we have use the property that if two limits exist then they can multiply.[3, p. 42].

Remark A.5.2. This lemma indicates that a function differentiable everywhere will be continous everywhere.

Lemma A.5.2 (differentiable everywhere NOT implies continuously differentiable). *A function is differentiable everywhere NOT implies it is continuously differentiable function.*

The standard example is

$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x}), x \neq 0 \\ 0, x = 0 \end{cases}$$

This function can be differentiated every where and $f'(0) = 0$, but $\lim_{x\rightarrow 0} f'(x)$ does not exist. See link.

A.5.1 continuity vs. uniform continuity

Definition A.5.1. *A function* $f : X \to Y$ *is* uniformly continuous *if for every* $\epsilon > 0$ *there exist a* $\delta > 0$ *such that for every* $x, x_0 \in X$*,*

$$\rho(x, x_0) \leq \delta \Rightarrow \rho(f(x), f(x_0)) < \epsilon$$

Theorem A.5.1. *[3, p. 154] If* f *is a continuous function from a compact metric space* M_1 *into a metric space* M_2*, then* f *is uniformly continuous on* M_1*.*

Corollary A.5.1.1. *[3, p. 154] If* f *is a continuous real-value function on a closed and bounded subset* X *of* $\mathbb{R}^n$*, then* f *is uniformly continuous on* X*.*

Example A.5.1. The function $f(x) = x^2$ is continuous but not uniformly continuous on the interval $(0, \infty)$.

Lemma A.5.3 (sufficient condition). *Let* $S = \mathbb{R}$*. if* f *is global Lipschitz continuous, i.e.*

$$\left|f(x_1) - f(x_2)\right| < M|x_1 - x_2|$$

$\forall x_1, x_2 \in S$*, then* f *is uniformly continuous.*

Proof: $\left|f(x_1) - f(x_2)\right| < M|x_1 - x_2| \to 0$

A.6 Exchanges of limits

A.6.1 Overall remark

Remark A.6.1.

- Usually, the necessary conditions for exchanging limits is difficult to find, therefore only sufficient conditions are given.
- Many operations are in nature taking limits, for example, summing infinite terms is taking limits on partial sums; integrals is taking limits on both summation and partitions; derivative is taking limits on quotient expressions.

A.6.2 exchange limits with infinite summations

Let $\lim_{m\to\infty}\sum_{n=1}^{\infty} f(m,n) = \sum_{n=1}^{\infty}\lim_{m\to\infty} f(m,n)$ Based on dominated convergence, if there is a $g(n)$ such that $f(m,n) < g(n), \forall m$ and $\sum_{n=1}^{\infty} g(n)$ exists, then we can exchange.

To use the dominated convergence theorem in Lebesgue integral, we can define a simple function $s_n$on $[0,\infty]$ take $f(m,n)$ on the interval $[m-1,m)$. Then the integral of s_n with respect to Legesbue measure on real lime will give the $\int_{[0,\infty)} s_n d\mu = \sum_{m=1}^{\infty} f(m,n)$

Theorem A.6.1. *[3, pp. 94, 373]Let $a_{m,n}$ be non-negative and $\sum_m^{\infty}\sum_n^{\infty} a_{m,n}$ exists, then*

$$\sum_m^{\infty}\sum_n^{\infty} a_{m,n} = \sum_n^{\infty}\sum_m^{\infty} a_{m,n}$$

Corollary A.6.1.1. *Let $a_{m,n}$ be increasing on both m,n and $\lim_{m\to\infty}\lim_{n\to\infty} a_{m,n}$ exists, then*

$$\lim_{m\to\infty}\lim_{n\to\infty} a_{m,n} = \lim_{n\to\infty}\lim_{m\to\infty} a_{m,n}$$

Proof: by constructing partial sums.

A.6.3 Exchange limits with integration and differentiation

Theorem A.6.2. *[3, p. 249] Let α be a function of bounded variation on $[a,b]$ and let f_n be a sequence of functions in $\mathcal{R}_\alpha[a,b]$ which converges uniformly to a function f. Then $f \in \mathcal{R}_\alpha[a,b]$ and*

$$\lim_{n\to\infty} \int_a^b f_n d\alpha = \int_a^b \lim_{n\to\infty} f d\alpha$$

Theorem A.6.3. *[3, p. 249] Let $\{f_n\}$ be a sequence of differentiable functions on (a,b). Suppose that*

- *f_n' is continuous on (a,b)*
- *$\{f_n\}$ converges pointwise to f*
- *$\{f_n'\}$ converges uniformly*

then f is differentiable on (a,b) and f_n' converges uniformly to f'.

A.6.4 Exchange differentiation with integration

Theorem A.6.4. *Let $f(x,y)$ be continuous on $[a,b]\times[c,d]$. Then*

$$\phi(y) = \int_a^b f(x,y)dx$$

defined above is continuous function on $[c,d]$

Proof: for any $\epsilon > 0$, there exist δ, such that

$$|\phi(y) - \phi(y')| \leq \int_a^b |f(x,y) - f(x,y')|\, dx \leq \epsilon(b-a) \forall |y-y'| < \delta$$

where we have the fact of $f(x,y) - f(x,y)$ is bounded (since continuous function on a compace set is uniformly continuous and will have maximum and minimum) which shows $\phi(y)$is uniformly continuous.

Theorem A.6.5. *Let f and f_y be continuous on $[a,b]\times[c,d]$. Then ϕ is differentiable and*

$$\phi_y = \int_a^b f_y(x,y)dx$$

Proof:

$$\frac{\phi(y+h) - \phi(y)}{h} = \frac{1}{h}\int_a^b f(x,y+h) - f(x,y)dx = \int_a^b f_y(x,z)$$

due to Taylor theorem, where $z \in [y, y+h]$. Then

$$\left|\frac{\phi(y+h)-\phi(y)}{h} - \int_a^b f_y(x,y)dx\right| \leq \int_a^b \left|f_y(x,z) - f_y(x,y)\right| dx$$

Because f_y is continuous on compact set, then it is uniformly continuous. Therefore given $\epsilon > 0$, there exists δ such that

$$\left|f_y(x,y') - f_y(x,y)\right| < \epsilon/(b-a), \forall \left|y - y'\right| < \delta$$

Taking $h < \delta$, we have

$$\left|\frac{\phi(y+h)-\phi(y)}{h} - \int_a^b f_y(x,y)dx\right| < \epsilon.$$

Take the limit on h and we get the result.

A.6.5 Exchange limit and function evaluations

Lemma A.6.1. *Let $\{x_n\}$ be a sequence with limit x, let f be a continuous function*

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x)$$

Proof: from the definition of continuous function.

A.7 Useful inequalities

Lemma A.7.1 (arithmetic-geometric mean inequality). *For $x_1, ..., X_n \geq 0$, we have*

$$(x_1 x_2 ... x_n)^{1/n} \leq \sum_{i=1}^{n} x_i / n.$$

Specifically,

$$\frac{x_1 + x_2}{x} \geq \sqrt{x_1 x_2}$$

Proof. use $y = \ln(x)$ and concavity of $\ln(x)$ □

A.7.1 Gronwall's inequality

see [6]

A.7.2 Inequality for norms

Lemma A.7.2. *[7]For L^p normed space, we have*

$$\|x\|_2 \leq \|x\|_1$$

where

$$\|x\|_2 = (\int_{-\infty}^{\infty} |f(x)|^2 \, d\mu(x))^{0.5}$$

and

$$\|x\|_1 = \int_{-\infty}^{\infty} |f(x)| \, d\mu(x)$$

Proof: for finite dimensional normed space cases: we need to prove

$$\sqrt{x_1^2 + x_2^2 + ... + x_n^2} \leq |x_1| + |x_2| + ... + |x_n|$$

By squaring both sides, we can get the result. For continuous case, TODO

Theorem A.7.1. *[7]For L^p normed space, we have*

$$\|x\|_q \leq \|x\|_q$$

whenever $p \leq q$ *where*

$$\|x\|_q = (\int_{-\infty}^{\infty} |f(x)|^q \, d\mu(x))^{1/q}$$

Proof: todo

Remark A.7.1. For complete description on L^p norms, see [7]

A.7.3 Young's inequality for product

Lemma A.7.3. *If* $a, b \geq 0$*, and* $p, q > 1, 1/p + 1/q = 1$*, then*

$$ab \leq a^p/p + b^q/q$$

Proof:

$$\log(a^p/p + b^q/q) \geq \log(a^p)/p + \log(a^q)/q = \log(a) + \log(b) = \log(ab)$$

where we use the fact of log is concave.

A.8 Useful properties of matrix

A.8.1 Matrix derivatives

Lemma A.8.1 (common matrix derivative in quadratic forms). *[8] For $A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n$, we have:*

$$\frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$$
$$\frac{\partial Ax}{\partial x} = A$$
$$\frac{\partial BAx}{\partial x} = BA$$
$$\frac{\partial x^T Ax}{\partial x} = (A + A^T)x$$
$$\frac{\partial x^T Ax}{\partial x} = 2A$$

Lemma A.8.2. *If $f(x) = g(Ax), A \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n$ for some differentiable function $g(y)$, then*

$$\nabla f = A^T \nabla g$$

In particularly, $a \in \mathbb{R}^n$, then

$$\nabla a^T Ax = A^T x$$

A.8.2 Matrix inversion lemma

Lemma A.8.3 (matrix inversion lemma). *[9, p. 120]*

- $(E - FH^{-1}G)^{-1} = E^{-1} + E^{-1}F(H - GE^{-1}F)^{-1}GE^{-1}$
- $(E - FH^{-1}G)^{-1}FH^{-1} = E^{-1}F(H - GE^{-1}F)^{-1}$

Proof. (1) can be verified (2) expand the parenthesis using (1). □

Corollary A.8.0.1 (matrix inversion of rank one update). *Let $H = I$, $F = \pm u \in \mathbb{R}^n$, and $G = \pm v \in \mathbb{R}^n$, we have*

- $$(E - uv)^{-1} = E^{-1} - \frac{E^{-1}uv^TE^{-1}}{1 + v^TE^{-1}u}$$
- $$(E - uv)^{-1} = E^{-1} + \frac{E^{-1}uv^TE^{-1}}{1 - v^TE^{-1}u}$$

A.8.3 Block matrix

Lemma A.8.4. *Given an $(m \times p)$ matrix A with q row partions and s colun partitions and a $(p \times n)$ matrix B with s row partions and r colun partitions,*

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q1} & A_{q2} & \cdots & A_{qs} \end{pmatrix}, B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1r} \\ B_{21} & B_{22} & \cdots & B_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{sr} \end{pmatrix},$$

then the matrix product

$$C = AB$$

can be formed blockwise, giving C as an $(m \times n)$ matrix with q row partitions and r column partitions. In particular,

$$C_{\alpha\beta} = \sum_{\gamma=1}^{s} A_{\alpha\gamma}B_{\gamma\beta}.$$

Lemma A.8.5 (sum of vector product to matrix product). *Consider column vectors $x_1, x_2, ..., x_N \in \mathbb{R}^d$ and column vectors $x_1, x_2, ..., x_N \in \mathbb{R}^d$. It follows that*

- $$\sum_{i=1}^{N} x_i^T y_i = X_C^T Y_C,$$

  where $X_C \in \mathbb{R}^{Nd}$ is a vector stacking all the $x_1, ..., x_N$ (similarly Y_C).
- $$\sum_{i=1}^{N} x_i y_i^T = X_R^T Y_R,$$

  where $X_R \in \mathbb{R}^{N \times d}$ is a matrix stacking all the $x_1^T, ..., x_N^T$ (similarly Y_R).

Lemma A.8.6 (block matrix inversion formula).

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} + B_{12}B_{22}^{-1}B_{21} & -B_{12}B_{22}^{-1} \\ -B_{22}^{-1}B_{21} & B_{22}^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} C_{11}^{-1} & -C_{11}^{-1}C_{12} \\ -C_{21}C_{11}^{-1} & A_{22}^{-1} + C_{21}C_{11}^{-1}C_{12} \end{pmatrix}$$

where

$$B_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12}, B_{12} = A_{11}^{-1}A_{12}, B_{21} = A_{21}A_{11}^{-1}$$

and

$$C_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21}, C_{12} = A_{12}A_{22}^{-1}, C_{21} = A_{22}^{-1}A_{21}$$

A.8.4 Matrix trace

Lemma A.8.7.
- $\|A\|_F^2 = Tr(AA^T)$

Lemma A.8.8 (matrix trace).

- *(linearity)*$Tr(aA + bB) = aTr(A) + bTr(B)$
- *(commutative)* $Tr(AB) = Tr(BA)$
- *(invariance under transposition)* $Tr(A) = Tr(A^T)$
- *(cyclic rule)* $Tr(ABCD) = Tr(DABC) = Tr(CDAB) = Tr(BCDA)$ *or* $Tr(ABC) = Tr(CAB) = Tr(BCA)$

Proof. (1)(2)(3) can be proved directly from definition. (4) We can group three elements together and commute with the fourth. For example, we can group (ABC) together and commute with D to prove the first equality. □

Corollary A.8.0.2.

- *(invariance under similar transformation)* $Tr(PAP^{-1}) = Tr(A)$
- $Tr(X^TY) = Tr(XY^T) = Tr(Y^TX) = Tr(XY^T)$

Proof. (1) Use cyclic rule. (2) Use invariance under transposition and commutative rule. □

Lemma A.8.9 (common matrix derivative involving matrix trace). *[8] Let $A, X, B \in \mathbb{R}^{m\times m}$. We have*

$$\frac{\partial Tr(X)}{\partial X} = I$$

$$\frac{\partial Tr(XA)}{\partial X} = \frac{\partial Tr(AX)}{X} = A^T$$

$$\frac{\partial Tr(X^T A)}{\partial X} = \frac{\partial Tr(AX^T)}{X} = A$$

$$\frac{\partial Tr(AXB)}{\partial X} = \frac{\partial Tr(BAX)}{X} = A^T B^T$$

$$\frac{\partial Tr(AX^T B)}{\partial X} = \frac{\partial Tr(BAX^T)}{X} = BA$$

$$\frac{\partial Tr(XX^T)}{\partial X} = 2X$$

$$\frac{\partial Tr(XX)}{\partial X} = 2X^T$$

Additional, we have chain rule given by

$$\frac{\partial Tr(X^T A^T AX)}{\partial X} = \frac{\partial Tr(XX^T A^T A)}{\partial X} = \frac{\partial Tr(XX^T A^T A)}{\partial XX^T}\frac{\partial XX^T}{\partial X} = 2A^T AX.$$

A.8.5 Matrix elementary operator

Lemma A.8.10 (elementary operator matrix). *Left multiply a matrix A by an elementary matrix is equivalent to performing row operation. The elementary operator matrix is given by*

- *(Interchange row i and j) For example, exchange row 2 and row 3:*

$$R_1 = \begin{bmatrix} 1 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 1 \end{bmatrix}$$

- *(Multiply row i by s) For example*

$$R_2 = \begin{bmatrix} 1 & & & \\ & s & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

- *(Add s times row i to row j) For example, add s times row 2 to row 3*

$$R_3 = \begin{bmatrix} 1 & & & \\ & 1 & 0 & \\ & s & 1 & \\ & & & 1 \end{bmatrix}.$$

Note that $R_3 = R_1 R_2 \neq R_2 R_1$.

Lemma A.8.11 (elementary operator matrix). *Right multiply a matrix A by an elementary matrix is equivalent to performing row operation. The elementary operator matrix is given by*

- *(Interchange column i and j) For example, exchange row 2 and row 3:*

$$C_1 = \begin{bmatrix} 1 & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & 1 \end{bmatrix}$$

- *(Multiply column i by s) For example*

$$C_2 = \begin{bmatrix} 1 & & & \\ & s & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

- *(Add s times column i to column j) For example, add s times column 2 to column 3*

$$C_3 = \begin{bmatrix} 1 & & & \\ & 1 & s & \\ & & 1 & \\ & & & 1 \end{bmatrix}.$$

Note that $C_3 = C_1C_2 \neq C_2C_1$.

A.8.6 Matrix determinant

Lemma A.8.12 (properties of determinant).

- *All elementary operator matrix has determinant 1.*
- *For matrix* $A \in \mathbb{R}^{n\times n}$,
$$det(kA) = k^n det(A).$$
- $det(AB) = det(A)det(B)$.
- *All elementary operation on a matrix will not change its determinant.*

A.9 Numerical integration

Definition A.9.1 (Newton-Cotes Formula). *Suppose we want to evaluate* $\int_a^b f(x)dx$*. We can evaluate* $f(x)$ *at* $n+1$ *equally spacing points* $x_i = a + i(b-a)/n$*, and then we approximate* $f(x)$ *by* n *degree of Lagrange polynomial and do the integral. Specifically, we have*

$$\int_a^b f(x)dx \approx \int_a^b L(x)dx = \int_a^b (\sum_{i=0}^{n} f(x_i)l_i(x)) = \sum_{i=1}^{n} f(x_i) \int_a^b l_i(x)dx = \sum_{i=1}^{n} f(x_i)w_i$$

where L is the Lagrange polynomial of degree n*, and* $l_i(x), i = 0, ..., n$ *is the (n+1) Lagrange polynomial basis, given as Lemma 3.5.2.*

Example A.9.1. Consider we use degree 1 Lagrange polynomial to approximate $f(x)$, then

$$L(x) = f(a)\frac{x-a}{b-a} + f(b)\frac{x-b}{a-b}$$

where $l_0(x) = \frac{x-a}{b-a}$ and $l_1(x) = \frac{x-b}{a-b}$. Then

$$w_0 = \int_a^b l_0(x)dx = \frac{1}{2}, w_1 = \int_a^b l_1(x)dx = \frac{1}{2}.$$

Table A.9.1: Closed Newton-Cotes Formula

Notation: $\int_a^b f(x)dx$, $f_i = f(x_i), x_i = a + i(b-a)/n$			
Degree	Name	Formula	Error term
1	Trapezoid rule	$\frac{b-a}{2}(f_0 + f_1)$	$-\frac{(b-a)^3}{12} f^{(2)}(\eta)$
2	Simpson's rule	$\frac{b-a}{6}(f_0 + 4f_1 + f_2)$	$-\frac{(b-a)^5}{2880} f^{(4)}(\eta)$
3	Simpson's 3/8 rule	$\frac{b-a}{8}(f_0 + 3f_1 + 3f_2 + f_3)$	$-\frac{(b-a)^3}{6480} f^{(5)}(\eta)$

Remark A.9.1 (Error analysis). For detailed error analysis, see [10, p. 252].

Remark A.9.2 (how to use). Usually, given the integral $\int_a^b f(x)dx$, we will first divide into smaller intervals and do the numerical integral on each interval and add them up. For example

$$\int_a^b f(x)dx = \int_a^{a+h} f(x)dx + \int_{a+h}^{a+2h} f(x)dx ... + \int_{b-h}^{b} f(x)dx.$$

Lemma A.9.1 (Trapezoid rule and the error bound). *Given the integral $\int_a^b f(x)dx$, we have*

$$\int_a^b f(x)dx \approx \frac{b-a}{n}\left(\frac{f(a)}{2} + \sum_{k=1}^{n-1}\left(f\left(a + k\frac{b-a}{n}\right)\right) + \frac{f(b)}{2}\right)$$

where we divide $b-a$ into n subintervals, and we do the Trapezoid rule on each subinterval. And the error bound is

$$\frac{(b-a)^3}{12n^2}\max_{x\in[a,b]} f^{(2)}(x)$$

Proof. Note that on each subinterval, the error is $-\frac{(b-a/n)^3}{12}f^{(2)}(\eta)$. Sum up n terms, and we have upper bound

$$\frac{(b-a)^3}{12n^3}n\max_{x\in[a,b]} f^{(2)}(x)$$

□

Lemma A.9.2 (Midpoint rule and the error bound). *Given the integral $\int_a^b f(x)dx$, we have*

$$\int_a^b f(x)dx \approx \frac{b-a}{n}\left(\sum_{k=1}^{n}\left(f\left(a + (k-0.5)\frac{b-a}{n}\right)\right)\right)$$

where we divide $b-a$ into n subintervals, and we do the Trapezoid rule on each subinterval. And the error bound is

$$\frac{(b-a)^3}{n^2}K$$

A.9.1 Gaussian quadrature

$$\int_a^b w(x)f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

which is exact when f is a polynomial.

Remark A.9.3. In Newton-Cotes formulas, we fix nodes and try to find suitable weights; in Gaussian quadrature, we use a weighted sum of function values at specified points within the domain of integration.

A.10 Vector calculus

Lemma A.10.1 (divergence theorem).

$$\iiint_V (\nabla \cdot \mathbf{F}) dV = \oiint_{S(V)} \mathbf{F} \cdot \hat{\mathbf{n}} dS$$

$$\iiint_V (\nabla \times \mathbf{F}) dV = \oiint_{S(V)} \hat{\mathbf{n}} \times \mathbf{F} dS$$

$$\iiint_V (\nabla f) dV = \oiint_{S(V)} \hat{\mathbf{n}} f dS$$

Lemma A.10.2 (Lapacian product rule). *Given functions $u : \mathbb{R}^n \to \mathbb{R}$ and $v : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\nabla^2(uv) = u\nabla v + 2\nabla u \cdot \nabla v + v\nabla^2 u.$$

Proof. Directly use product rule. □

A.11 Numerical linear algebra computation complexity

Note A.11.1. [11, p. 606]

- For a $m \times n$ matrix multiplying a n dimensional vector, mn.
- For a $n \times n$ matrix multiplying a $n \times n$ matrix, n^3(without optimization).
- For a $n \times n$ matrix, LU decomposition $2n^3/3$ (for symmetric matrix $n^3/3$).
- For a $m \times n$ matrix, Cholesky decomposition $4m^2n/3$ (for square matrix $4n^3/3$).
- For a $m \times n$ matrix, QR decomposition $4m^2n/3$ (for square matrix $4n^3/3$).

Note A.11.2 (solving triangular linear system). Let L be a $n \times n$ lower triangle matrix, the forward substitution algorithm for solving

$$Ly = d,$$

is given by

```
y(1) = d(1) / L(1,1);
for i=2:n
y(i) = (d(i) - L(i,1:i-1)* y(1:i-1))/L(i,i)
end
```

This algorithm has complexity of $O(n^2)$.

Let U be a $n \times n$ upper triangle matrix, the backward substitution algorithm for solving

$$Ux = d,$$

is given by

```
x(n) = d(n)/U(n,n);
for i = n - 1: -1 :1
x(i) = (d(i) - U(i,i + 1:n)*x(i + 1:n) )/U(i,i)
end
```

This algorithm has complexity of $O(n^2)$.

A.12 Distributions

Lemma A.12.1. *[12, p. 579] Let K be an externally given parameter. We have*

- $$\int_{-\infty}^{\infty} \delta(x)dx = 1, x\delta(x) = 0, \int_{-\infty}^{\infty} f(x)\delta(x-K)dx = f(K).$$
- $\delta(g(x)) = \sum_i \frac{\delta(x-x_i)}{|g'(x_i)|}$, *where* $x_i, i = 1, 2, ...$ *are the zeros of the function* $g(x)$.
- $\delta(\lambda x) = \frac{\delta(x)}{|\lambda|}, \delta(x-K) = \delta(K-x)$.
- *(step function definition)*
  $$H(x) \triangleq \frac{d}{dx}\max\{x, 0\}, H(x-A) \triangleq \frac{d}{dx}\max\{x-A, 0\}$$
- $H(x-K) + H(K-x) = 1$.
- $$\frac{dH(x-K)}{dx} = \delta(x-K), \frac{dH(K-x)}{dx} = -\delta(x-K).$$

Proof. Use $H(x-K) + H(K-x) = 1$ to prove $\frac{dH(K-x)}{dx}$. □

A.13 Common integrals

Lemma A.13.1.

- $$\int_0^\infty e^{-ax^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}, \int_{-\infty}^\infty e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$$
- $$\int_0^\infty xe^{-ax^2} dx = \frac{1}{2a}, \int_{-\infty}^\infty xe^{-ax^2} dx = 0$$
- $$\int_0^\infty x^2 e^{-ax^2} dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}}$$
- $$\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$$
- $$\int_0^\infty x^m e^{-ax^2} = \frac{\Gamma((m+1)/2)}{2a^{(m+1)/2}}$$

A.14 Nonlinear root finding

A.14.1 Bisection method

Methodology A.14.1.

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$.
- ***(Initial input)***: *Initial guess of* l_0 *and* r_0 *such that*

$$f(l_0) < 0, f(r_0) > 0; \text{ or } f(l_0) > 0, f(r_0) < 0.$$

- ***Repeat (i) is the iteration index***:
  - *Let* $m = \frac{r_i + l_i}{2}$.
  - *If* $f(l_i)f(m) < 0$, *then* $l_{i+1} = l_i, r_{i+1} = m$.
  - *If* $f(l_i)f(m) > 0$*(then we must have* $f(r_i)f(m) < 0$*), then* $l_{i+1} = m, r_{i+1} = r_i$.
  - *If* $f(l_i)f(m) = 0$, *then* m *is the root.*

A.14.2 Newton method

Methodology A.14.2.

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$. f *is assumed to be differentiable.*
- ***(Initial input)***: *Initial guess of* x_0.
- ***Repeat***(*i is the iteration index):*
  - *Let* $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$.
  - *If* $f(x_{i+1}) = 0$, *then* x_{i+1} *is the root.*

A.14.3 Secant method

Methodology A.14.3 (Secant method).

- ***(Goal)***: *We want to find a* $x^* \in \mathbb{R}$ *such that* $f(x^*) = 0$. f *is assumed to be differentiable.*
- ***(Initial input)***: *Initial guess of* x_0, x_1.
- ***Repeat***(*i is the iteration index):*

- *Let*
$$x_{i+1} = x_i - f(x_i)\frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})}$$
.
- *If $f(x_{i+1}) = 0$, then x_{i+1} is the root.*

Remark A.14.1 (derivation). Starting with initial guesses x_0, x_1, we construct a first order approximation of $f(x)$ via

$$y = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1).$$

And we solve the root for the first-order approximation problem via

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_1) + f(x_1) = 0 \implies x_2 = x_1 - f(x_1)\frac{x_1 - x_0}{f(x_1) - f(x_0)}.$$

Then we continue the process with x_1, x_2.

Remark A.14.2 (convergence property).

- There is no guarantee on the global convergence to the root of f.
- Only when the initial values x_0 and x_1 are sufficiently close to the root, the iterates x_n will converge to the root.

A.15 Interpolation

A.15.1 cubic interpolation

Definition A.15.1 (the cubic spine line functional form). *[13]*

- *Suppose $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$ are known.*
- *A cubic spine line is given by*

$$y(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, x_i \leq x \leq x_{i+1}, i = 1, 2, ..., n - 1.$$

- *There are $4n - 4$ unknowns.*
- *Note that*

$$\begin{aligned} y'(x) &= b_i + 2c_i\left(x - x_i\right) + 3d_i\left(x - x_i\right)^2, x_i < x < x_{i+1} \\ y''(x) &= 2c_i + 6d_i\left(x - x_i\right), x_i < x < x_{i+1} \\ y'''(x) &= 6d_{i,}x_i < x < x_{i+1} \end{aligned}$$

Definition A.15.2 (natural cubic spline condition). *[13]* *Let $h_i = x_{i+1} - x_i$*

- ***(spline line passing data points)**: for $i = 1, 2, ..., n - 1$, $a_i = y_i$; $a_{n-1} + b_{n-1}h_{n-1} + c_{n-1}h_{n-1}^2 + d_{n-1}h_{n-1}^3 = y_n$.*
- ***(interpolating function is continuous)**; that is,*

$$\lim_{x \to x_i-} y(x) = \lim_{x \to x_i+} y(x) \implies a_i + b_ih_i + c_ih_i^2 + d_ih_i^3 = a_{i+1}, \forall i = 1, 2, ..., n - 2.$$

- ***(interpolating function is differentiable)**; note that the interpolating function is differentiable on interval, therefore we require that,*

$$\lim_{x \to x_i-} y'(x) = \lim_{x \to x_i+} y'(x) \implies b_i + 2c_ih_i + 3d_ih_i^2 = b_{i+1}, \forall i = 1, 2, ..., n - 2$$

  :
- ***(interpolating function is twice differentiable and the second derivative at each endpoint is 0)**; that is,*

$$\lim_{x \to x_i-} y''(x) = \lim_{x \to x_i+} y''(x) \implies c_i + 3d_ih_i = c_{i+1}, \forall i = 1, 2, ..., n - 2,$$

  and $y''(x_1+) = y''(x_n-) = 0$.
- *these $4n - 4$ equations will solve the $4n - 4$ unknowns.*

BIBLIOGRAPHY

1. Hammack, R. H. *Book of Proof* (Richard Hammack, 2013).
2. Gradshteyn, I. S. & Ryzhik, I. M. *Table of integrals, series, and products* (Academic press, 2014).
3. Johnsonbaugh, R. & Pfaffenberger, W. Foundations of mathematical analysis (2010).
4. Moon, T. K. S. & Wynn, C. *Mathematical methods and algorithms for signal processing* **621.39: 51 MON** (2000).
5. Luenberger, D. G. *Optimization by vector space methods* (John Wiley & Sons, 1969).
6. Wikipedia. *Gronwall's inequality — Wikipedia, The Free Encyclopedia* [Online; accessed 10-July-2016]. 2016.
7. Wikipedia. *Lp space — Wikipedia, The Free Encyclopedia* [Online; accessed 15-August-2016]. 2016.
8. Petersen, K. & Pedersen, M. The matrix cookbook, nov 2012. *URL http://www2. imm. dtu. dk/pubdb/p. php* **3274.**
9. Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT press, 2012).
10. Atkinson, K. E. *An introduction to numerical analysis* (John Wiley & Sons, 1989).
11. Nocedal, J. & Wright, S. *Numerical optimization* (Springer Science & Business Media, 2006).
12. Prosperetti, A. Advanced mathematics for applications. *SIAM Review Vol. 55, Issue 2 (June 2013)* **55,** 403 (2013).
13. Hagan, P. S. & West, G. Interpolation methods for curve construction. *Applied Mathematical Finance* **13,** 89–129 (2006).