Useful Facts, Identities, Inequalities

Linear Algebra

Notation

Notation	Name	Comment
$\mathbf{A}$	a matrix	indicated by capitalization and bold font
$[\mathbf{A}]_{ij}$	the element from the i-th row and j-th column of $\mathbf{A}$	Sometimes denoted A_{ij} for brevity
$\mathcal{M}_n$	the set of complex-valued matrices with m rows and n columns	e.g., $\mathbf{A} \in \mathcal{M}_{m,n}$
$\mathcal{M}_n$	the set of complex-valued square matrices	e.g., $\mathbf{A} \in \mathcal{M}_n$
$\mathbf{I}$	the identity matrix	e.g., $\mathbf{I} = \mathsf{diag}\,[1, 1, \dots, 1]$
$\mathbf{e}$	vector of ones	
$\sigma_i(\mathbf{A})$	i-th singular value of $\mathbf{A}$	Typically, we assume sorted order: $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \dots$
$\lambda_i(\mathbf{A})$	the i-th eigenvalue of $\mathbf{A}$	may also write $\lambda(\mathbf{A}) = \{\lambda_i\}_i$ for the set of eigenvalues.
$\rho(\mathbf{A})$	the spectral radius of $\mathbf{A} \in \mathcal{M}_n$	$\rho(\mathbf{A}) = \max_{1 \leq i \leq n} \{\lvert\lambda_i\rvert\}$
$\bar{\mathbf{A}}$	the complex conjugate of $\mathbf{A}$	operates entrywise, e.g. $A_{ij} = \bar{A}_{ij}$
$\mathbf{A}^\top$	the matrix transpose of $\mathbf{A}$	
$\mathbf{A}^\dagger$	the conjugate transpose of $\mathbf{A}$	e.g., $\mathbf{A}^\dagger \overset{\text{def}}{=} (\bar{\mathbf{A}})^\top$
$\mathbf{A}^{-1}$	the matrix inverse of $\mathbf{A}$	e.g., $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
$\mathbf{A}^+$	the Moore-Penrose pseudoinverse of $\mathbf{A}$	sometimes may just use $\mathbf{A}^{-1}$ due to sloppy notation

Definitions

Definition 1 (Complex Conjugate). *Let $\mathbf{A} \in \mathcal{M}_n$, then $\mathbf{A}^\dagger = (\bar{\mathbf{A}})^\top$.*
Definition 2 (Symmetric Matrix). *A matrix $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}^\top$*
Definition 3 (Hermitian Matrix). *A matrix $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}^\dagger = (\bar{\mathbf{A}})^\top$.*
Definition 4 (Normal Matrix). *$\mathbf{A}$ normal $\Leftrightarrow \mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$*
That is, a normal matrix is one that commutes with its conjugate transpose.
Definition 5 (Unitary Matrix). *$\mathbf{A}$ unitary $\Leftrightarrow \mathbf{I} = \mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$.*
Definition 6 (Orthogonal Matrix). *A square matrix $\mathbf{A}$ whose columns and rows are orthogonal unit vectors, that is $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top\mathbf{A}$.*
Definition 7 (Idempotent matrix). *A matrix such that $\mathbf{A}\mathbf{A} = \mathbf{A}$.*
Definition 8 (Nilpotent Matrix). *A matrix such that $\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbb{0}$.*
Definition 9 (Unipotent Matrix). *A matrix $\mathbf{A}$ such that $\mathbf{A}^2 = \mathbf{I}$.*
Definition 10 (Stochastic Matrix). *Consider a matrix, say $\mathbf{P}$ with no negative entries. $\mathbf{P}$ is* row stochastic *matrix if each row sums to one; it is* column stochastic *if each column sums to one. By convention, usually mean row stochastic when using "stochastic" without qualification. A matrix is* doubly stochastic *if it is both row and column stochastic.*
Definition 11 (Matrix congruence). *Two square matrices $\mathbf{A}$ and $\mathbf{B}$ over a field are called* congruent *if there exists an invertible matrix $\mathbf{P}$ over the same field such that $\mathbf{P}^\top\mathbf{A}\mathbf{P} = \mathbf{B}$.*
Definition 12 (Matrix similarity). *Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are called* similar *if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.*
A transformation $\mathbf{A} \mapsto \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is called a similarity transformation *or* conjugation *of the matrix $\mathbf{A}$.*

Special Matrices and Vectors

The vector of all ones, $\mathbf{1} = (1, 1, \dots, 1)^\top$.

$$\mathbf{v}\mathbf{1}^\top = \mathbf{v} \otimes \mathbf{1} = \begin{pmatrix} v_1 & v_1 & \cdots & v_1 \\ v_2 & v_2 & \cdots & v_n \\ \vdots & \vdots & \ddots & \vdots \\ v_n & v_n & \cdots & v_n \end{pmatrix} \tag{1}$$

Standard Basis

The standard (euclidean) basis for $\mathbb{R}^n$ is denoted $\{\mathbf{e}_i\}_{i=1}^n$, with $\mathbf{e}_i \in \mathbb{R}^n$, and

$$[\mathbf{e}_i] \overset{\text{def}}{=} \begin{cases} 1 & \text{if } i = j \\ 0 & i \neq j \end{cases} \tag{2}$$

General Information

Symmetric Matrices

- Any matrix congruent to a symmetric matrix is itself symmetric, i.e. if $\mathbf{S}$ is symmetric, then so is $\mathbf{A}\mathbf{S}\mathbf{A}^\top$ for any matrix $\mathbf{A}$.
- Any symmetric real matrix can be diagonalized by an orthogonal matrix.

Determinants

Lemma 1. *[Sylvester's Identity]*
Let $\mathbf{A} \in \mathcal{M}_{m,n}$ and $\mathbf{B} \in \mathcal{M}_{n,m}$. Then:

$$\det(\mathbf{I}_m + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B}\mathbf{A}) \tag{3}$$

Determinant Facts

For $\mathbf{A} \in \mathcal{M}_n$ with eigenvalues $\lambda(\mathbf{A}) = \{\lambda_1, \dots \lambda_n\}$,

- $\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$
- $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$
- Also, $\det(\exp\{\mathbf{A}\}) = \exp\{\operatorname{tr}(\mathbf{A})\}$.

Lemma 2 (Block Matrix Determinants). *Let $\mathbf{M} \in \mathcal{M}_{n+m}$ be a square matrix partitioned as:*

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \tag{4}$$

where $\mathbf{A} \in \mathcal{M}_n$, $\mathbf{D} \in \mathcal{M}_m$, with B and C being $n \times m$ and $m \times n$ respectively.
If $\mathbf{B} = \mathbb{0}_{n,m}$ or $\mathbf{C} = \mathbb{0}_{m,n}$, then

$$\det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{D}) \tag{5}$$

Furthermore, if $\mathbf{A}$ is invertible, then

$$\det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) \tag{6}$$

A similar identity holds for $\mathbf{D}$ invertible.
Lemma 3 (Block Triangular Matrix Determinants). *Consider a matrix $\mathbf{M} \in \mathcal{M}_{m,n}$ of the form:*

$$\mathbf{M} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1n} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2n} \\ \vdots & & \ddots & \vdots \\ \mathbf{S}_{n1} & \mathbf{S}_{n2} & \cdots & \mathbf{S}_{nn} \end{pmatrix} \tag{7}$$

where each $\mathbf{S}_{ij}$ is an $m \times m$ matrix.
If $\mathbf{M}$ is in block triangular form (that is, $\mathbf{S}_{ij} = \mathbb{0}$ for $j > i$), then

$$\det(\mathbf{M}) = \prod_{i=1}^n \det(\mathbf{S}_{ii}) = \det(\mathbf{S}_{11})\det(\mathbf{S}_{22})\cdots\det(\mathbf{S}_{nn}) \tag{8}$$

Definiteness

Definition 13: Definiteness

A square matrix $\mathbf{A}$ is
- *positive definite* if $\mathbf{x}^\dagger\mathbf{A}\mathbf{x} > 0 \;\forall \mathbf{x} \neq \mathbf{0}$.
- *positive semi-definite* if $\mathbf{x}^\dagger\mathbf{A}\mathbf{x} \geq 0 \;\forall \mathbf{x}$.

Similar definitions apply for negative (semi-)definiteness and indefiniteness.

Some sources argue that only Hermitian matrices should count as positive definite, while others note that (real) matrices can be positive definite according to definition 13, although this is ultimately due to the symmetric part of the matrix.

Remark 1: non-symmetric positive definite matrices

If $\mathbf{A} \in \mathcal{M}_n(\mathbb{R})$ is a (not necessarily symmetric) real matrix, then $\mathbf{x}^\top\mathbf{A}\mathbf{x} > 0$ for all non-zero real $\mathbf{x}$ if the symmetric part of $\mathbf{A}$, defined as $\mathbf{A}_+ = (\mathbf{A} + \mathbf{A}^\top)/2$, is positive definite.
In fact, $\mathbf{x}^\top\mathbf{A}\mathbf{x} = \mathbf{x}^\top\mathbf{A}_+\mathbf{x} \;\forall \mathbf{x}$ because

$$\mathbf{x}^\top\mathbf{A}\mathbf{x} = (\mathbf{x}^\top\mathbf{A}\mathbf{x})^\top = \mathbf{x}^\top\mathbf{A}^\top\mathbf{x} = \frac{1}{2}\mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)\mathbf{x} \tag{9}$$

Facts about definite matrices

Assume $\mathbf{A}$ is positive definite. Then:

- $\mathbf{A}$ is always full rank.
- $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is also positive definite.
- $\mathbf{A}$ is positive definite $\Leftrightarrow$ there exists an invertible matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}\mathbf{B}^\top$.
- $A_{ii} > 0$
- $\operatorname{rank}(\mathbf{B}\mathbf{A}\mathbf{B}^\top) = \operatorname{rank}(\mathbf{B})$.
- $\det(\mathbf{A}) \leq \prod_i A_{ii}$
- If $\mathbf{X} \in \mathcal{M}_{n,r}$, with $n \leq r$ and $\operatorname{rank}(\mathbf{X}) = n$, then $\mathbf{X}\mathbf{X}^\top$ is positive definite.
- $\mathbf{A}$ positive definite $\Leftrightarrow \mathsf{eig}(\frac{\mathbf{A}+\mathbf{A}^\dagger}{2}) > 0$ (and ≥ 0 for $\mathbf{A}$ positive-semidefinite).
- If $\mathbf{B}$ is symmetric, then $\mathbf{A} - t\mathbf{B}$ is positive definite for sufficiently small t.

Projections

Definition 14

Let $\mathcal{V}$ be a vector space, and $\mathcal{U} \subseteq \mathcal{V}$ be a subspace. The projection of $\mathbf{v} \in \mathcal{V}$ onto some subspace with respect to some norm $\|\cdot\|_q$ is:

$$\mathbf{\Pi}\mathbf{v} \overset{\text{def}}{=} \operatorname*{argmin}_{\mathbf{u} \in \mathcal{U}} \|\mathbf{v} - \mathbf{u}\|_q \tag{10}$$

For the (weighted) Euclidean norm, $\mathbf{\Pi}$ can itself be expressed as a matrix.

$$\mathbf{\Pi}_\mathbf{D} = \mathbf{X}(\mathbf{X}^\top\mathbf{D}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{D} \tag{11}$$

Although sometimes we just write $\mathbf{\Pi}$ instead of $\mathbf{\Pi}_\mathbf{D}$.

Projection Facts

- Projections are idempotent, i.e., $\mathbf{\Pi}^2 = \mathbf{\Pi}$.
- A square matrix $\mathbf{\Pi} \in \mathcal{M}_n$ is called an *orthogonal projection matrix* if $\mathbf{\Pi}^2 = \mathbf{\Pi} = \mathbf{\Pi}^\dagger$.
- A non-orthogonal projection is called an *oblique projection*, usually expressed as $\mathbf{\Pi} = \mathbf{A}(\mathbf{B}^\top\mathbf{A})^{-1}\mathbf{B}^\top$.
- The triangle inequality applies, $\|\mathbf{x}\| = \|(\mathbf{I} - \mathbf{\Pi})\mathbf{x} + \mathbf{\Pi}\mathbf{x}\| \leq \|(\mathbf{I} - \mathbf{\Pi})\mathbf{x}\| + \|\mathbf{\Pi}\mathbf{x}\|$.

Singular Values

Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $r = \min\{m, n\}$, with $\sigma(\mathbf{A}) = \{\sigma_k(\mathbf{A})\}_{k=1}^r$, and U, V are subspaces with $U \subseteq \mathbb{C}^n$, $V \subseteq \mathbb{C}^m$.

$$\begin{aligned} \sigma_k(\mathbf{A}) &= \max_{\dim(U)=k} \min_{\substack{x \in U \\ \|\mathbf{x}\|=1}} \|\mathbf{A}\mathbf{x}\|_2 = \min_{\dim(U)=n-k+1} \max_{\substack{x \in U \\ \|\mathbf{x}\|=1}} \|\mathbf{A}\mathbf{x}\|_2 \\ &= \max_{\substack{\dim(U)=k \\ \dim(V)=k}} \min_{\substack{\mathbf{x} \in U \\ \mathbf{x} \in V}} \frac{\mathbf{y}^\top\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \max_{\dim(U)=k} \min_{\mathbf{x} \in U} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|_2} \end{aligned} \tag{12}$$

- $\sigma_1(\mathbf{A}) = \|\!|\mathbf{A}|\!\|_2$
- $\sigma_i(\mathbf{A}) = \sigma_i(\mathbf{A}^\top) = \sigma_i(\mathbf{A}^\dagger) = \sigma_i(\bar{\mathbf{A}})$.
- $\sigma_i^2(\mathbf{A}) = \lambda_i(\mathbf{A}\mathbf{A}^\dagger) = \lambda_i(\mathbf{A}^\dagger\mathbf{A})$.
- For $\mathbf{U}$ and $\mathbf{V}$ (square) unitary matrices of appropriate dimension, $\sigma_i(\mathbf{A}) = \sigma_i(\mathbf{U}\mathbf{A}\mathbf{V})$.
  - This also connects to why $\|\mathbf{A}\|_2 = \sigma_1$, because $\|\cdot\|_2$ is unitarily invariant, $\|\mathsf{diag}(\mathbf{d})\| = \max_i \lvert d_i\rvert$, so $\|\mathbf{A}\|_2 = \|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger\|_2 = \|\mathbf{\Sigma}\|_2 = \sigma_1(\mathbf{A})$.

Matrix Inverse

Identities

- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^\top\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{A}^{-1}$
-
-

Moore-Penrose Pseudoinverse

Decompositions

Symmetric Decomposition

A square matrix $\mathbf{A}$ can always be written as the sum of a symmetric matrix $\mathbf{A}_+$ and an antisymmetric matrix $\mathbf{A}_-$, such that $\mathbf{A} = \mathbf{A}_+ + \mathbf{A}_-$.

QR Decomposition

A decomposition of a matrix $\mathbf{A}$ into the product of an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{QR}$.

Singular Value Decomposition

Rewriting Matrices

We can rewrite some common expressions using the standard basis $\mathbf{e}_i$ and the single-entry matrix $\mathbf{J}^{ij} = \mathbf{e}_i\mathbf{e}_j^\top$:

$$\begin{aligned} A_{ij}B_{k\ell} &= [\mathbf{A}\mathbf{e}_j\mathbf{e}_k^\top\mathbf{B}]_{i\ell} && = [\mathbf{A}\mathbf{J}^{jk}\mathbf{B}]_{i\ell} \\ &= [\mathbf{A}^\top\mathbf{e}_i\mathbf{e}_k^\top\mathbf{B}]_{j\ell} && = [\mathbf{A}\mathbf{J}^{ik}\mathbf{B}]_{j\ell} \\ &= [\mathbf{A}\mathbf{e}_j\mathbf{e}_\ell^\top\mathbf{B}^\top]_{ik} && = [\mathbf{A}\mathbf{J}^{j\ell}\mathbf{B}]_{ik} \\ &= [\mathbf{A}^\top\mathbf{e}_i\mathbf{e}_\ell^\top\mathbf{B}^\top]_{jk} && = [\mathbf{A}^\top\mathbf{J}^{i\ell}\mathbf{B}^\top]_{jk} \end{aligned} \tag{13}$$

Note that the above identities can be derived from each other via appropriate transpositions. The single-entry matrix can also be used to extract rows and columns of a matrix. Let $\mathbf{A}$ be $n \times m$ and $\mathbf{J}^{ij}$ be $m \times p$. Then $\mathbf{AJ}^{ij}$ is a $p \times m$ matrix of zeros except for the j-th column, which is the i-th column of $\mathbf{A}$.

$$\mathbf{AJ}^{ij} = \begin{bmatrix} 0 & 0 & \cdots & A_{1i} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & A_{2i} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A_{ni} & 0 & \cdots & 0 \end{bmatrix} \tag{14}$$

For appropriately sized $\mathbf{J}^{ij}$, we can use the Kronecker delta to express:

$$[\mathbf{AJ}^{ij}]_{k,\ell} = \delta_{j\ell}A_{ki} \qquad [\mathbf{J}^{ij}\mathbf{A}]_{k,\ell} = \delta_{ik}A_{j\ell} \tag{15}$$

Quadratic Forms

Definition 15: Quadratic Form

Given $\mathbf{A} \in \mathcal{M}_n$, $\mathbf{x} \in \mathbb{C}^n$, the scalar value $\mathbf{x}^\dagger\mathbf{A}\mathbf{x}$ is called a quadratic form. Explicitly,

$$\mathbf{x}^\dagger\mathbf{A}\mathbf{x} = \sum_i^n \bar{x}_i \sum_j^n A_{ij}x_j = \sum_i^n \sum_j^n A_{ij}\bar{x}_i x_j \tag{16}$$

For $\mathbf{x} \in \mathbb{R}^n$, this becomes $\mathbf{x}^\top\mathbf{A}\mathbf{x} = \sum_{i,j}^n A_{ij}x_ix_j$.

The definiteness of quadratic forms maps to the definiteness of $\mathbf{A}$ in a pretty natural way (see remark 1).

Norms

Vector Norms

1. $\|\mathbf{x}\| \geq 0$.
2. $\|\mathbf{0}\| = 0$.
3. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
4. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ for any scalar a.

Definition 16 (p-norms)**.**

$$\|\mathbf{x}\|_p \overset{\text{def}}{=} \left(\sum_i |x_i|^p\right)^{1/p} \tag{17}$$

For $p = \infty$ this becomes the max norm *or* infinity norm*,*

$$\|\mathbf{x}\|_\infty \overset{\text{def}}{=} \max_i |x_i| \tag{18}$$

Vector Norm Inequalities

$|\langle\mathbf{x},\mathbf{y}\rangle| \leq \|\mathbf{x}\|_p\|\mathbf{q}\|_q$ for $\frac{1}{p}+\frac{1}{q}=1$. (Holder)
$|\langle\mathbf{u},\mathbf{v}\rangle| \leq \|\mathbf{u}\|_2 \leq \|\mathbf{v}\|_2$ (Cauchy)
$\|\mathbf{u}+\mathbf{v}\|_p \leq \|\mathbf{u}\|_p + \|\mathbf{v}\|_p$ for $p > 1$ (Minkowski)
$\|\mathbf{x}\|_q \geq \|\mathbf{x}\|_p$ for $p > q > 0$ (Generalized Mean)

Miscellaneous facts

- $\|\mathbf{Av}\|_2 = (\mathbf{v}^\top\mathbf{A}^\top\mathbf{A}\mathbf{v})^{1/2} \leq \|\mathbf{A}^\top\mathbf{A}\|_2^{1/2}(\mathbf{v}^\top\mathbf{v})^{1/2}$.
- $|\det(\mathbf{A})| = \sigma_1\sigma_2\cdots\sigma_n$, where σ_i is the i-th singular value of $\mathbf{A}$.
- All eigenvalues of a matrix are smaller in magnitude than the largest singular value, i.e., $\sigma_1(\mathbf{A}) > |\lambda_i(\mathbf{A})|$. In particular, $|\lambda_1(\mathbf{A})| = \rho(\mathbf{A}) \leq \sigma_1(\mathbf{A}) = \|\mathbf{A}\|_2$.

Matrix Norms

Properties

$|||\alpha\mathbf{A}||| = |\alpha||||\mathbf{A}|||$ (absolutely homogeneous)
$|||\mathbf{A}+\mathbf{B}||| \leq |||\mathbf{A}||| + |||\mathbf{B}|||$ (sub-additive)
$|||\mathbf{A}||| \geq 0$ (positive valued)
$|||\mathbf{A}||| = 0 \Rightarrow \quad \mathbf{A} = \mathbb{0}$ (definiteness)
$|||\mathbf{AB}||| \leq |||\mathbf{A}||||||\mathbf{B}|||$ (sub-multiplicative)

Some vector norms, when generalized to matrices, are also matrix norms.
We use $|||\cdot|||$ when denoting a norm satisfying all five of the above properties, and $\|\cdot\|$ to denote the "generalized norms" that are not sub-multiplicative.

Common Matrix Norms

Let $\mathbf{A} \in \mathcal{M}_{m,n}$.

Notation	Name	Formula
$|||\mathbf{A}|||_p$	L_p-norm	$\max_{\|\mathbf{x}\|_p=1} \|\mathbf{Ax}\|_p$
$|||\mathbf{A}|||_1$	L_1-norm	$|||\mathbf{A}|||_1 = \max_{1\leq j\leq n}\sum_{i=1}^m |A_{ij}|$
$|||\mathbf{A}|||_2$	L_2-norm	$|||\mathbf{A}|||_2 = \sigma_1(\mathbf{A})$ $= \sqrt{\lambda_{\max}(\mathbf{A}^\dagger\mathbf{A})}$
$|||\mathbf{A}|||_{\mathrm{F}}$	Frobenius norm	$|||\mathbf{A}|||_{\mathrm{F}} = \sqrt{\sum_{i=1}^m\sum_{j=1}^n |A_{ij}|^2}$ $= \sqrt{\mathrm{tr}(\mathbf{A}^\dagger\mathbf{A})}$ $= \sqrt{\sum_{i=1}^{\min m,n} \sigma_i^2}$
$|||\mathbf{A}|||_\infty$	L_∞-norm	$|||\mathbf{A}|||_\infty = \max_{1\leq i\leq m}\sum_{j=1}^n |A_{ij}|$

Miscellaneous Matrix Norm Facts

- $\|\mathbf{A}^+\|_2 = \sigma_{\min}(\mathbf{A})$

Other Norms

Matrix Norm Inequalities

- $|||\mathbf{A}|||_2 \leq \sqrt{|||\mathbf{A}|||_1|||\mathbf{A}|||_\infty}$
- $|||\cdot|||$ is an induced norm $\Rightarrow \rho(\mathbf{A}) \leq |||\mathbf{A}^r|||^{1/r}$
  - For $|||\cdot|||_2$ and $\mathbf{A}$ Hermitian, we have $|||\mathbf{A}|||_2 = \rho(\mathbf{A})$
- $|||\mathbf{A}|||_2 \leq |||\mathbf{A}|||_F \leq \sqrt{\min\{m,n\}}|||\mathbf{A}|||_2$
- Weyl's inequality (TODO)

Inequality	**Conditions**	**Comments**
$|||\mathbf{A}^{1/2}\mathbf{X}\mathbf{B}^{1/2}||| \leq \frac{1}{2}|||\mathbf{A}^v\mathbf{X}\mathbf{B}^{1-v} + \mathbf{A}^{1-v}\mathbf{X}\mathbf{B}^v||| \leq \frac{1}{2}|||\mathbf{AX}+\mathbf{XB}|||$	$0 < v \leq 1$, $\mathbf{A}$, $\mathbf{B},\mathbf{X} \in \mathcal{M}_n$, $\mathbf{A}$, $\mathbf{B}$ positive semidefinite	Heinz Inequality for Matrices
$|||\mathbf{Ab}||| \leq \frac{1}{4}|||(\mathbf{A}+\mathbf{B})^2||| \leq \frac{1}{2}|||\mathbf{A}^2+\mathbf{B}^2|||$	$\mathbf{A},\mathbf{B}$ positive semi-definite, $|||\cdot|||$ unitarily invariant.	Bhatia and Kittaneh
$\|\mathbf{ABx}\| \leq \|\mathbf{A}\|\|\mathbf{Bx}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|\|\mathbf{x}\|$	$\|\cdot\|$ is induced norm	...
$|||\mathbf{A}|||_p \leq |||\mathbf{A}|||_q$	...	...
$|||\mathbf{A}|||_p \leq |||\mathbf{A}|||_q$	...	...

Induced norms

Definition 17 (Induced Norm)**.** *A matrix norm $|||\cdot|||$ is said to be* induced *by the vector norm $\|\cdot\|$ if*

$$|||\mathbf{M}||| = \max_{\|\mathbf{x}\|=1}\|\mathbf{Mx}\| \tag{19}$$

For an induced norm $|||\cdot|||$ we have:

$$\begin{aligned} |||\mathbf{I}||| &= 1 \\ |||\mathbf{Ax}||| &\leq |||\mathbf{A}||||||\mathbf{x}||| && \text{for all matrices } \mathbf{A} \text{ and vectors } \mathbf{x} \\ |||\mathbf{AB}||| &\leq |||\mathbf{A}||||||\mathbf{B}||| && \text{for all } \mathbf{A},\mathbf{B} \end{aligned}$$

For a *weighted* norm with $\mathbf{W}$ a symmetric positive definite matrix, we have

$$\|\mathbf{x}\|_{\mathbf{W}} \overset{\text{def}}{=} \sqrt{\mathbf{x}^\top\mathbf{W}\mathbf{x}} = \left\|\mathbf{W}^{1/2}\mathbf{x}\right\|_2 \tag{20}$$

The corresponding induced matrix norm is

$$\begin{aligned} |||\mathbf{A}|||_{\mathbf{W}} &\overset{\text{def}}{=} \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_{\mathbf{W}}}{\|\mathbf{x}\|_W} = \max_{\mathbf{y}\neq\mathbf{0}} \frac{\left\|\mathbf{W}^{1/2}\mathbf{A}\mathbf{W}^{-1/2}\mathbf{y}\right\|_2}{\left\|\mathbf{W}^{1/2}\mathbf{y}\right\|_2} \\ &= |||\mathbf{W}^{1/2}\mathbf{A}\mathbf{W}^{-1/2}|||_2 \end{aligned} \tag{21}$$

Spectrum of a Matrix

Definition 18 (Spectral Radius)**.** *The spectral radius $\rho(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathcal{M}_n$ is*

$$\rho(\mathbf{A}) \overset{\text{def}}{=} \max\{|\lambda| : \lambda \textit{ is an eigenvalue of } \mathbf{A}\} \tag{22}$$

For square matrices, we have $\lim_{r\to\infty}|||\mathbf{A}^r|||^{1/r} = \rho(\mathbf{A})$.

Theorem 4. *If $|||\cdot|||$ is a submultiplicative matrix norm and $\mathbf{A} \in \mathcal{M}_n$, then*

$$\rho(\mathbf{A}) \leq |||\mathbf{A}||| \tag{23}$$

Proof. Let $\mathbf{v}$ be the eigenvector associated with λ, with $|\lambda| = \rho(\mathbf{A})$. Consider the matrix $\mathbf{V}$ with columns equal to $\mathbf{v}$, i.e. $[\mathbf{V}]_{ij} = v_j$. Note that $\mathbf{AV} = \lambda\mathbf{V}$. So we have,

$$|\lambda||||\mathbf{V}||| = |||\lambda\mathbf{V}||| = |||\mathbf{AV}||| \leq |||\mathbf{A}||||||\mathbf{V}||| \tag{24}$$

So $|\lambda| = \rho(\mathbf{A}) \leq |||\mathbf{A}|||$. □

Theorem 5 (Weyl's Inequality for Eigenvalues and Singular Values)**.** *Let $\mathbf{A} \in \mathcal{M}_n$ be a square matrix with singular values $sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ and eigenvalues $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Then*

$$|\lambda_1\cdots\lambda_k| \leq \sigma_1\cdots\sigma_k \tag{25}$$

for $k = 1,\ldots,n$, with equality for $k = n$.

Norm equivalence

Definition 19 (Unitarily Invariant Norm)**.** *...*

Facts

1. If $\mathbf{D} = \mathsf{diag}(d_1, d_2, \ldots, d_N)$, then $|||\mathbf{D}||| = \max_i |d_i|$.
2. $|||\mathbf{A}^\top\mathbf{A}|||_F = |||\mathbf{A}\mathbf{A}^\top|||_F \leq |||\mathbf{A}|||_F^2$
3. $|||\mathbf{A}+\mathbf{B}|||_F = |||\mathbf{A}|||_F + |||\mathbf{B}|||_F + 2\langle\mathbf{A},\mathbf{B}\rangle_F$

Hadamard Product

Definition 20. *Let $\mathbf{A},\mathbf{B} \in \mathcal{M}_{n,r}$. Then the Hadamard product (AKA Schur product, elementwise product) of $\mathbf{A}$ and $\mathbf{B}$ is denoted $\mathbf{A}\circ\mathbf{B}$ and defined such that*

$$[\mathbf{A}\circ\mathbf{B}]_{ij} \overset{\text{def}}{=} A_{ij}B_{ij} \tag{26}$$

Identities

- $\mathbf{A}\circ\mathbf{B} = \mathbf{B}\circ\mathbf{A}$
- $\mathbf{A}\circ(\mathbf{B}\circ\mathbf{C}) = (\mathbf{A}\circ\mathbf{B})\circ\mathbf{C}$
- $\mathbf{A}\circ(\mathbf{B}+\mathbf{C}) = (\mathbf{A}\circ\mathbf{B}) + (\mathbf{A}\circ\mathbf{C})$
- $(k\mathbf{A})\circ\mathbf{B} = \mathbf{A}\circ(k\mathbf{B}) = k(\mathbf{A}\circ\mathbf{B})$
- $(\mathbf{A}\circ\mathbf{B})^\top = \mathbf{A}^\top\circ\mathbf{B}^\top$
- $\mathbf{A}\circ\mathbb{1} = \mathbf{A}$ (identity operation, $\mathbb{1}$ is the matrix with all entries one).
- $\mathbf{A}\circ(\mathbf{x}\mathbf{y}^\dagger) = \mathbf{D}_\mathbf{x}\mathbf{A}\mathbf{D}_\mathbf{y} = \mathsf{diag}(\mathbf{x})\mathbf{A}\,\mathsf{diag}(\mathbf{y})$.
- $(\mathbf{X}\circ\mathbf{ab}^\top)(\mathbf{Y}\circ\mathbf{cd}^\top) = (\mathbf{ad}^\top)\circ(\mathbf{X}\,\mathsf{diag}(\mathbf{b}\circ\mathbf{c})\mathbf{Y})$
- $(\mathbf{A}\circ\mathbf{x}\mathbf{y}^\top)\mathbf{z} = \mathbf{x}\circ(\mathbf{A}(\mathbf{y}\circ\mathbf{z}))$

- For $\mathbf{D_x}$ and $\mathbf{D_y}$ diagonal matrices,

$$\begin{aligned}\mathbf{D_x}(\mathbf{A}\circ\mathbf{B})\mathbf{D_y} &= (\mathbf{D_x}\mathbf{A}\mathbf{D_y})\circ\mathbf{B} = \mathbf{A}\circ(\mathbf{D_x}\mathbf{B}\mathbf{D_y}) \\ &= (\mathbf{D_x}\mathbf{A})\circ(\mathbf{B}\mathbf{D_y}) = (\mathbf{A}\mathbf{D_y})\circ(\mathbf{D_x}\mathbf{B})\end{aligned} \tag{27}$$

- $\mathbf{a}\circ\mathbf{b} = \mathbf{D_a}\mathbf{b}$ (vector product)
- $(\mathbf{a}\circ\mathbf{b})(\mathbf{x}\circ\mathbf{y})^\top = (\mathbf{a}\mathbf{x}^\top)\circ(\mathbf{b}\mathbf{y}^\top) = (\mathbf{a}\mathbf{y}^\top)\circ(\mathbf{b}\mathbf{x}^\top)$
- $\mathbf{x}^\dagger(\mathbf{A}\circ\mathbf{B})\mathbf{y} = \mathrm{tr}(\mathbf{D}_\mathbf{x}^\dagger\mathbf{A}\mathbf{D_y}\mathbf{B}^\top)$
- $[(\mathbf{A}\circ\mathbf{B})\mathbf{x}]_i = [\mathbf{A}\mathbf{D_x}\mathbf{B}^\top]_{ii}$
- $\sum_j[\mathbf{A}\circ\mathbf{B}]_{ij} = [\mathbf{B}^\top\mathbf{A}]_{jj} = [\mathbf{A}\mathbf{B}^\top]_{ii}$. That is, the row sums of $\mathbf{A}\circ\mathbf{B}$ are the diagonal elements of $\mathbf{A}\mathbf{B}^\top$

Inequalities

For $\mathbf{A} = [A_{ij}] \in \mathcal{M}_{m,n}$, denote the decreasingly ordered Euclidean row and column lengths respectively by

$$\begin{aligned} r_1(\mathbf{A}) \geq r_2(\mathbf{A}) \geq \cdots \geq r_m(\mathbf{A}) \\ c_1(\mathbf{A}) \geq c_2(\mathbf{A}) \geq \cdots \geq c_n(\mathbf{A})\end{aligned} \tag{28}$$

where $r_k(\mathbf{A})$ is the k-th largest value of $\left(\sum_{j=1}^n |A_{ij}|^2\right)^{\frac{1}{2}}$ for $i = 1, 2, \ldots, m$ and similarly for $c_k(\mathbf{A})$. Then:

$$\sigma_1(\mathbf{A}\circ\mathbf{B}) \leq r_1(\mathbf{A})c_1(\mathbf{B}) \leq \begin{Bmatrix} r_1(\mathbf{A})\sigma_1(\mathbf{B}) \\ \sigma_1(\mathbf{A})c_1(\mathbf{B})\end{Bmatrix} \leq \sigma_1(\mathbf{A})\sigma_1(\mathbf{B}) \tag{29}$$

(from [HJ91], theorem 5.5.3, pg. 332)

			Conditions
$\lVert\lvert\mathbf{A}\circ\mathbf{B}\rvert\rVert_2$	$\leq$	$\left(\frac{1}{2}\lVert\lvert\mathbf{A}\rvert\rVert_2^2\lVert\lvert\mathbf{B}\rvert\rVert_2^2 + \lVert\lvert\mathbf{A}\mathbf{B}^\dagger\rvert\rVert_2\right)^{1/2} \leq \lVert\lvert\mathbf{A}\rvert\rVert_2\lVert\lvert\mathbf{B}\rvert\rVert_2$	$\mathbf{A},\mathbf{B}\in\mathcal{M}_n$
$\sigma_1(\mathbf{A}\circ\mathbf{B})$	$\leq$	$\sigma_1(\mathbf{A})\sigma_1(\mathbf{B})$	$\mathbf{A},\mathbf{B}\in\mathcal{M}_{m,n}$
$\det(\mathbf{A}\circ\mathbf{B})$	$\geq$	$\det(\mathbf{A})\det(\mathbf{B})$	$\mathbf{A},\mathbf{B}$ positive-semidefinite.

Stochastic Matrices

- $\rho(\mathbf{P}) = 1 \leq \sigma_1(\mathbf{P})$
- $\lVert\lvert\mathbf{P}\rvert\rVert_2 = \sigma_1(\mathbf{P}) = 1 \Leftrightarrow \mathbf{P}$ is doubly stochastic.
- If $\mathbf{P}$ is doubly stochastic, then so is $\mathbf{P}^\top\mathbf{P}$

Some Results on Stochastic Matrices

For $\mathbf{P}$ aperiodic and irreducible, with stationary distribution $\mathbf{d}$ and $\mathbf{D} \overset{\text{def}}{=} \mathsf{diag}(\mathbf{d})$, we have

$$\lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_2 \leq 1 \tag{30}$$

Proof. Note that by definition, $0 \leq P_{ij} < 1$, and $0 < d_i < 1$. Then,

$$\begin{aligned} \lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_F &= \left(\sum_i\sum_j[\mathbf{D}\mathbf{P}]_{ij}^2\right)^{1/2} \\ \sum_i\sum_j[\mathbf{D}\mathbf{P}]_{ij}^2 = \sum_i\sum_j d_i^2 P_{ij}^2 &\leq \sum_i d_i^2\sum_j P_{ij} = \sum_i d_i^2 \leq 1 \\ \Rightarrow \lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_2 &\leq \lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_F \leq 1^{1/2} = 1\end{aligned} \tag{31}$$

The bound is actually pretty tight, gets close to exact as the distribution becomes more concentrated in a single state. Generally $\lVert\lvert\mathbf{D}\mathbf{P}\rvert\rVert_2$ is pretty close to $\lVert\lvert\mathbf{d}\rvert\rVert_2^2$, which can be justified in a hand-wavy manner.
NB: We don't use strict inequality here, though we could, since I think this could be generalized to other kinds of stochastic matrices, and so I'm future-proofing. □

We can use a variation of the above to show that $\lVert\lvert\mathbf{P}\rvert\rVert_D$ is a non-expansion in the distribution-weighted Euclidean norm, as in lemma 6.

Lemma 6 (Stochastic Matrix is a Non-Expansion in Distribution Weighted Euclidean Norm)**.** *Let $\mathbf{P}\in\mathbb{R}^{N\times N}$ be an ergodic stochastic matrix with stationary distribution $\mathbf{d} = (d_1, d_2, \ldots, d_N$, with $\mathbf{D} = \mathsf{diag}(\mathbf{d})$.*
Then $\mathbf{P}$ is a non-expansion in the weighted Euclidean norm $\lVert\cdot\rVert_D$, that is,

$$\lVert\mathbf{P}\mathbf{z}\rVert_D \leq \lVert\mathbf{z}\rVert_D \tag{32}$$

Proof of lemma 6. (From Lemma 1 in [TV97])
Recall that $\mathbf{d}^\top\mathbf{P} = \mathbf{d}^\top$, and then expand:

$$\begin{aligned} \lVert\mathbf{P}\mathbf{z}\rVert_d^2 = (\mathbf{P}\mathbf{z})^\top\mathbf{D}\mathbf{P}\mathbf{z} &= \sum_{i=1}^N d(i)\left(\sum_j^N P_{ij}z_j\right)^2 \\ &\leq \sum_{i=1}^N d(i)\sum_j^N P_{ij}z_j^2 = \sum_{i=1}^N\sum_j^N d(i)P_{ij}z_j^2 \\ &= \sum_{j=1}^N\sum_i^N d(i)P_{ij}z_j^2 = \sum_{j=1}^N d(i)z_j^2 \\ &= \lVert\mathbf{z}\rVert_D^2\end{aligned} \tag{33}$$

Where we first applied Jensen's inequality as the function $f(x) = x^2$ is convex, and then interchange the order of summation via the Fubini-Tonelli theorem[1]. For the penultimate step, we recognize that $\sum_i^N P_{ij}d(i)$ corresponds to $\mathbf{d}^\top\mathbf{P}$, and then note that what remains is just $\lVert\mathbf{z}\rVert_D$. □

Lemma 7 (Bounded Features, Bounded Feature Matrix)**.** *Suppose we have $\lVert\mathbf{x}(i)\rVert_2 \leq K_b$ for some $K_b > 0$ for $i = 1, \ldots, N$, and let $\mathbf{X}\in\mathbb{R}^{N\times m}$ be the feature matrix with $[\mathbf{X}]_{ij} = x_j(i)$. Let $\mathbf{D}$ be a diagonal matrix $\mathbf{D} = \mathsf{diag}(\mathbf{d})$, with $d_i \geq 0$ and $\sum_i d_i = 1$. Then we have an L_2 norm bound of the form:*

$$\lVert\lvert\mathbf{D}\mathbf{X}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_2 \leq K_b \tag{34}$$

(This is a bit of a niche result, but even some experts have missed it)

lemma 7. First, note that

$$\lVert\lvert\mathbf{D}\mathbf{X}\rvert\rVert_2 = \lVert\lvert\mathbf{D}^{\frac{1}{2}}(\mathbf{D}^{\frac{1}{2}}\mathbf{X})\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\rvert\rVert_2\lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_2 \tag{35}$$

Then use the fact that $\lVert\lvert\cdot\rvert\rVert_2 \leq \lVert\lvert\cdot\rvert\rVert_F$ to get

$$\lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_F \tag{36}$$

and upon expanding

$$\begin{aligned}\lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_F^2 &= \sum_i\sum_j[\mathbf{D}^{\frac{1}{2}}\mathbf{X}]_{ij}^2 = \sum_i\sum_j d_i X_{ij}^2 = \sum_i d_i\sum_j X_{ij}^2 \\ &\leq \sum_i K_b^2 = K_b^2\end{aligned} \tag{37}$$

so we have

$$\lVert\lvert\mathbf{D}\mathbf{X}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_2 \leq \lVert\lvert\mathbf{D}^{\frac{1}{2}}\mathbf{X}\rvert\rVert_F \leq \left(K_b^2\right)^{\frac{1}{2}} = K_b \tag{38}$$

□

[1] Exchanging the order of summation can actually be ill-defined if the series contain subsequences diverging to *both* positive and negative infinity. Given that $d(\cdot)$, $P(\cdot,\cdot)$ and $z^2(\cdot)$ are positive-valued functions this does not happen– but since we are dealing with a *finite* state space, hence the sequence cannot diverge at all. Presumably the original proof invoked it for maximum generality, because this lemma would hold in the setting where the state space was continuous (requiring only changes to the notation). See [Coh80] or another book on analysis for further details.

Sequences, Series, and Products

Identities

Series identities

Identity	Conditions	Comments
$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$	$\|x\| < 1$	Infinite Power Series
$\left(\sum_{n=0}^{\infty} a_n x^n\right)^2 = \sum_{k=0}^{\infty} a_n^2 x^{2n} + 2 \sum_{\substack{n=1 \\ i+j=n \\ i<j}}^{\infty} a_i a_j x^n$		Square of Geometric Series
$\left(\sum_{k=1}^{n} a_k b_k\right)^2 = \sum_{k=1}^{n} a_k^2 \sum_{k=1}^{n} b_k^2 - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i b_j - a_j b_i)^2$		Lagrange's Identity
$\sum_{k=0}^{n} \log(a_k) = \log\left(\prod_{k=0}^{n} a_k\right)$	$0 < a_k \in \mathbb{R}$, $n \in \mathbb{N}$	Log-Sum Identity
$\sum_{k=0}^{n} \log(z_k) = \log\left(\prod_{k=0}^{n} z_k\right) - 2\pi i \lfloor \frac{\pi - \sum_{k=0}^{n} \arg(z_k)}{2\pi} \rfloor$	$n \in \mathbb{N}$	General Log-Sum Identity
$\sum_{k=0}^{m}\sum_{j=0}^{k} a_k b_j = \sum_{j=0}^{m}\sum_{k=j}^{m} a_k b_j$		Triangle-Sum Reordering
$\sum_{k=0}^{m}\sum_{j=k}^{p} a_k b_j = \sum_{j=0}^{m}\sum_{k=0}^{j} a_k b_j + \sum_{j=m+1}^{p}\sum_{k=0}^{m} a_k b_j$	$0 \le k \le m$, $0 \le k \le j \le p$	Quadrangle-Sum Reordering
$\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} a_k b_j = \sum_{j=0}^{\infty}\sum_{k=0}^{j} a_k b_{j-k}$		Infinite Double Sum Reordering
$\sum_{k=0}^{\infty}\sum_{j=0}^{k} a_k b_j = \sum_{j=0}^{\infty}\sum_{k=j}^{\infty} a_k b_j$		Infinite Double Sum Reordering

(Many taken from Wolfram/MathWorld)

Inequalities

Cauchy Schwarz

For $\{a_k\}$ and $\{b_k\}$ two sequences, we have

$$\left(\sum_{k}^{N} a_k b_k\right)^2 \le \left(\sum_{k}^{N} a_k^2\right)\left(\sum_{k}^{N} b_k^2\right) \tag{39}$$

Some direct implications:

$$(ac+bd)^2 \le (a^2+b^2)(c^2+d^2) \tag{40}$$

For $a_i, b_i > 0$, we have *Titu's Lemma*:

$$\frac{(a_1+a_2+\cdots+a_n)^2}{b_1+b_2+\cdots+b_n} \le \frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} + \cdots + \frac{a_n^2}{b_n} \tag{41}$$

For $0 \le x < 1$:

$$\sum_{k=0}^{\infty} a_k x^k \le \frac{1}{\sqrt{1-x^2}}\left(\sum_{k=0}^{\infty} a_k^2\right)^{\frac{1}{2}} \tag{42}$$

$$\sum_{k=1}^{n} \frac{a_k}{k} < \left(2\sum_{k=1}^{n} a_k^2\right)^{\frac{1}{2}} \tag{43}$$

Arithmetic and Geometric Mean Inequality

In general, the arithmetic mean is larger:

$$\left(\prod_{k=1}^{n} |a_k|\right)^{\frac{1}{n}} \le \frac{1}{n}\sum_{k=1}^{n} |a_k| \tag{44}$$

There are some particular implications of interest.
Let $\{a_i\}$ be a sequence of nonnegative real numbers, and let $\{p_i\}$ be a sequence of positive reals that sums to one. Then, via the exponential bound:

$$\prod_{k=1}^{n} a_k^{p_k} \le \sum_{k=1}^{n} p_k a_k \tag{45}$$

Power Mean Bound for Geometric Mean [Ste04, p. 122] For weights p_k, $k = 1, 2, \ldots, n$ with $p_k \ge 0$, $\sum_{k=1}^{n} p_k = 1$, and $x_k >\ge 0$, there is the bound:

$$\prod_{k=1}^{n} x_k^{p_k} \le \left[\sum_{k=1}^{n} p_k x_k^q\right]^{1/q} \tag{46}$$

for all $q > 0$.

Power Mean Inequality [Ste04, p. 123] For weights p_k, $k = 1, 2, \ldots, n$ with $p_k \ge 0$, $\sum_{k=1}^{n} p_k = 1$, and $x_k >\ge 0$, there is the bound:

$$\left[\sum_{k=1}^{n} p_k x_k^t\right]^{1/t} \le \left[\sum_{k=1}^{n} p_k x_k^q\right]^{1/q} \tag{47}$$

for all $-\infty < t < q < \infty$, with equality if and only if $x_1 = x_2 = \cdots = x_n$.

Sums of Squares

Product of two linear forms:

$$\sum_{j=1}^{n} u_j x_j \sum_{j=1}^{n} v_j x_j \le \frac{1}{2}\left[\sum_{j=1}^{n} u_j v_j + \left(\sum_{j=1}^{n} u_j^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{n} v_j^2\right)^{\frac{1}{2}}\right]\sum_{j=1}^{n} x_j^2 \tag{48}$$

for $\{u_i\}$, $\{v_i\}$, $\{x_i\}$ real valued.

General and Miscellaneous

(Facts/information that will be split into their own areas once enough are collected)

Algebraic Identities

$a^2 - b^2 = (a+b)(a-b)$

$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$ (Fibonacci-Brahmagupta)

Stirling's Approximation

Stirling's approximation: given as

$$\ln(n!) = n\ln(n) - n + \mathcal{O}(\ln n) \quad \text{or} \quad n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \approx n^n e^{-n} \tag{49}$$

With the following bounds that hold for $n \in \mathbb{N}_+$:

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \leq n! \leq e n^{n+\frac{1}{2}} e^{-n} \tag{50}$$

Inequalities

$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_p \|\mathbf{q}\|_q$ for $\frac{1}{p} + \frac{1}{q} = 1$. (Holder)

$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|_2 \leq \|\mathbf{v}\|_2$ (Cauchy-Schwarz)

$\|\mathbf{u} + \mathbf{v}\|_p \leq \|\mathbf{u}\|_p + \|\mathbf{v}\|_p$ for $p > 1$ (Minkowski)

$\|\mathbf{x}\|_q \geq \|\mathbf{x}\|_p$ for $p > q > 0$ (Generalized Mean)

(Jensen)

$\left(\prod_{k=1}^{n} |a_k|\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{k=1}^{n} |a_k|$ (AM-GM)

(Radon)

$1 + x \leq e^x$ for $x \in \mathbb{R}$ (Exponential Bound)

$1 + nx \leq (1+x)^n$ for $x \geq -1$, $n \geq 1$. (Bernoulli)