# Matrix Forensics

*A brief guide to matrix math*
*and its efficient implementation*

RICHARD BARNES

Git Hash: 0ce356ce3f3bef1b495bfa9e508dfef86380157a
Compiled on: 2018/11/01 at 16:29:50

github.com/r-barnes/MatrixForensics

# Contents

# 1 Introduction

**Goals:** TODO

**Contributing:** Please contribute on Github at `https://github.com/r-barnes/MatrixForensics` either by opening an issue or making a pull request. If you are not comfortable with this, please send your contribution to `rijard.barnes@gmail.com`.

**Contributors:** Richard Barnes

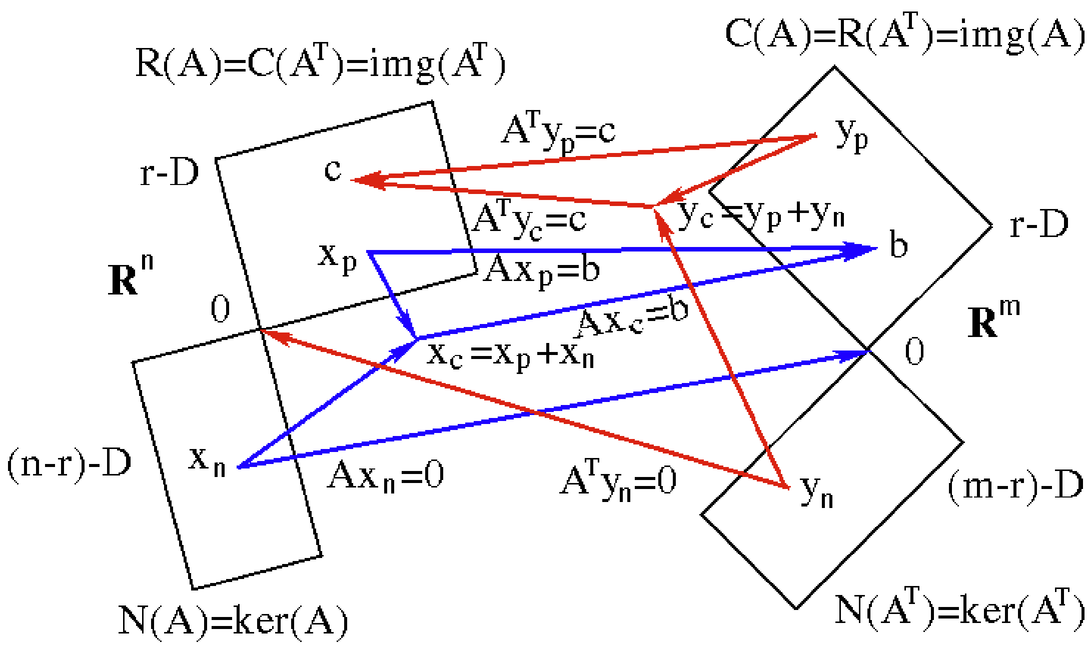
**Funding:** TODO

# 2 Nomenclature

| | |
|---|---|
| $\mathbf{A}$ | Matrix. |
| $\mathbf{a}$ | (Column) vector. |
| $a$ | Scalar. |
| $\mathbf{A}_{ij}$ | Matrix indexed. Returns $i$th row and $j$th column. |
| $\mathbf{A} \circ \mathbf{B}$ | Hadamard (element-wise) product of matrices A and B. |
| $\mathcal{N}(\mathbf{A})$ | Nullspace of the matrix $\mathbf{A}$. |
| $\mathcal{R}(\mathbf{A})$ | Range of the matrix $\mathbf{A}$. |
| $\det(\mathbf{A})$ | Determinant of the matrix $\mathbf{A}$. |
| $\mathrm{eig}(\mathbf{A})$ | Eigenvalues of the matrix $\mathbf{A}$. |
| $\mathbf{A}^H$ | Conjugate transpose of the matrix $\mathbf{A}$. |
| $\mathbf{A}^T$ | Transpose of the matrix $\mathbf{A}$. |
| $\mathbf{A}^+$ | Pseudoinverse of the matrix $\mathbf{A}$. |
| $\mathbf{x} \in \mathbb{R}^n$ | The entries of the $n$-vector $\mathbf{x}$ are all real numbers. |
| $\mathbf{A} \in \mathbb{R}^{m,n}$ | The entries of the matrix $\mathbf{A}$ with $m$ rows and $n$ columns are all real numbers. |
| $\mathbf{A} \in \mathbb{S}^n$ | The matrix $\mathbf{A}$ is symmetric and has $n$ rows and $n$ columns. |
| $\mathbf{I}_n$ | Identity matrix with $n$ rows and $n$ columns. |
| $\{0\}$ | The empty set |

# 3 Basics

## 3.1 Fundamental Theorem of Linear Algebra



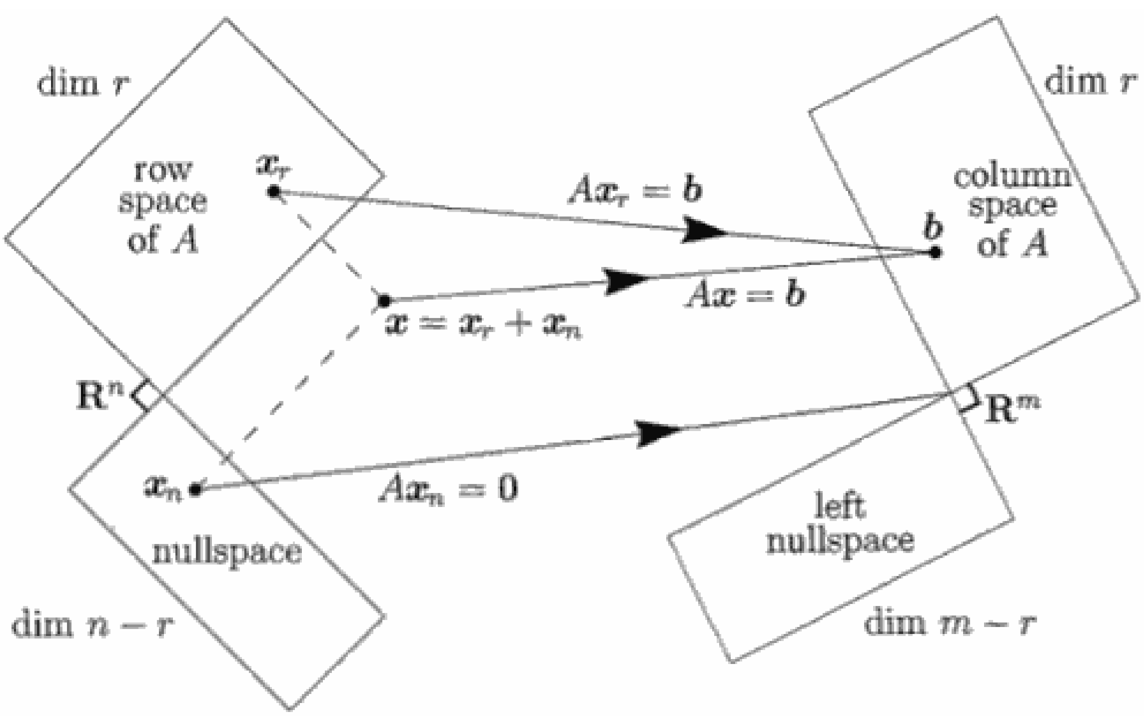
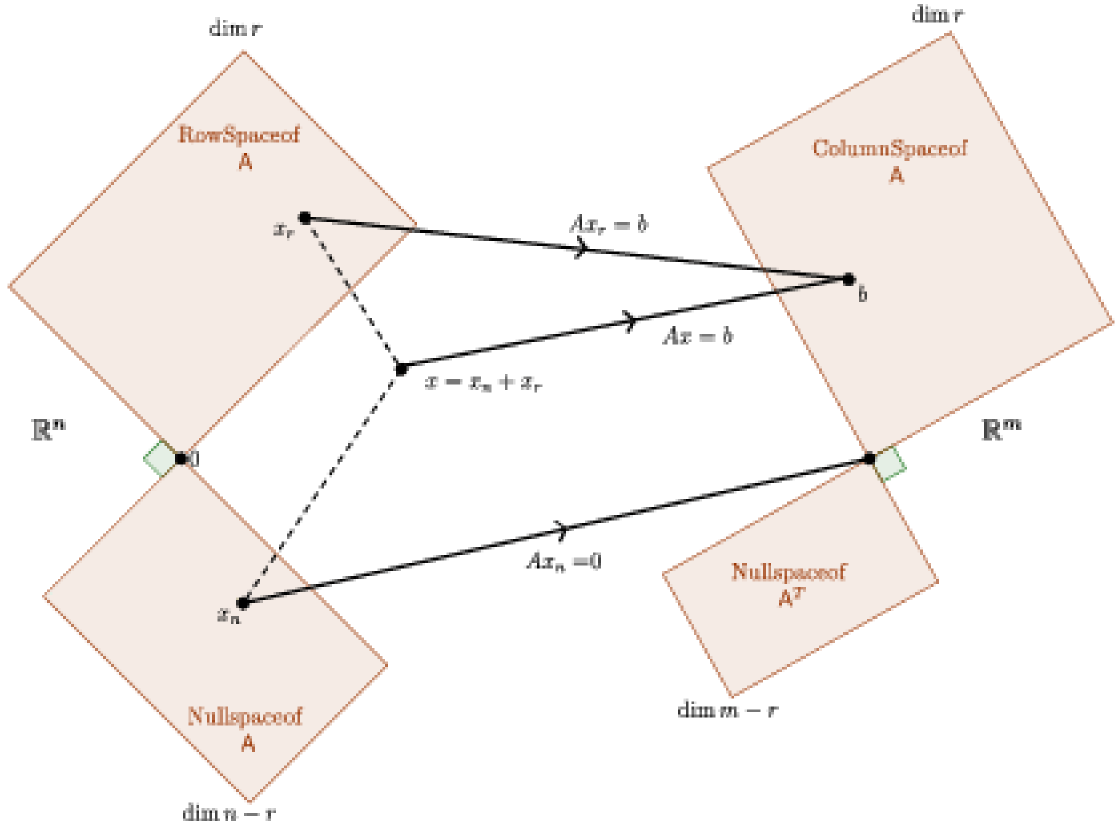
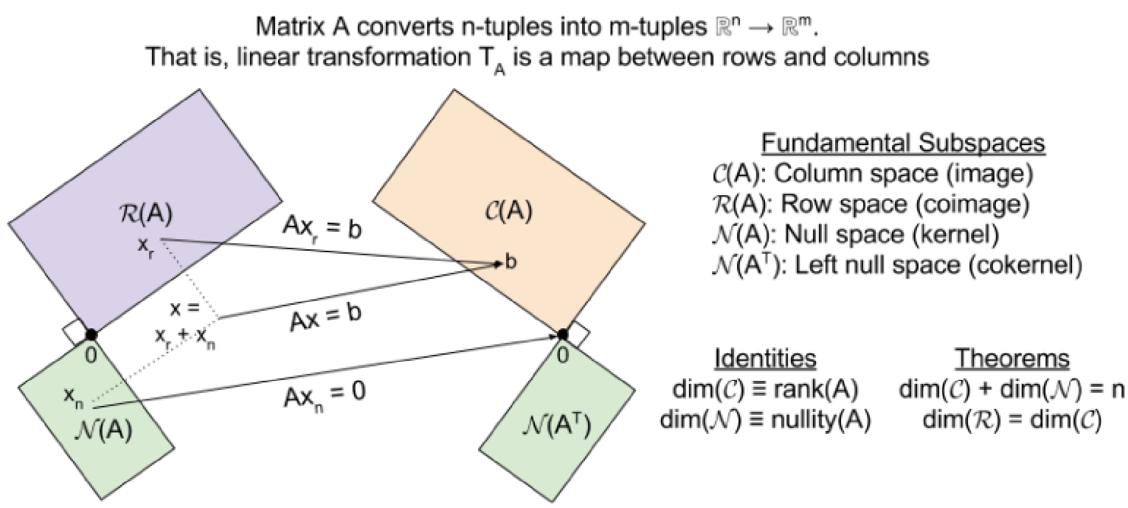




**Figure 3.4** The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any $m$ by $n$ matrix.


dim $r$
RowSpaceof
A
$x_r$
$Ax_r = b$
dim $r$
ColumnSpaceof
A
$b$
$Ax = b$
$x = x_n + x_r$
$\mathbb{R}^n$
$\mathbb{R}^m$
0
$Ax_n = 0$
Nullspaceof
$A^T$
$x_n$
dim $m - r$
Nullspaceof
A
dim $n - r$


Matrix A converts n-tuples into m-tuples $\mathbb{R}^n \to \mathbb{R}^m$.
That is, linear transformation $T_A$ is a map between rows and columns
$\mathcal{R}$(A)
$x_r$
$Ax_r = b$
$\mathcal{C}$(A)
b
x =
$x_r + x_n$
Ax = b
0
0
$x_n$
$Ax_n = 0$
$\mathcal{N}$(A)
$\mathcal{N}(A^T)$
Fundamental Subspaces
$\mathcal{C}$(A): Column space (image)
$\mathcal{R}$(A): Row space (coimage)
$\mathcal{N}$(A): Null space (kernel)
$\mathcal{N}(A^T)$: Left null space (cokernel)
Identities
dim($\mathcal{C}$) ≡ rank(A)
dim($\mathcal{N}$) ≡ nullity(A)
Theorems
dim($\mathcal{C}$) + dim($\mathcal{N}$) = n
dim($\mathcal{R}$) = dim($\mathcal{C}$)

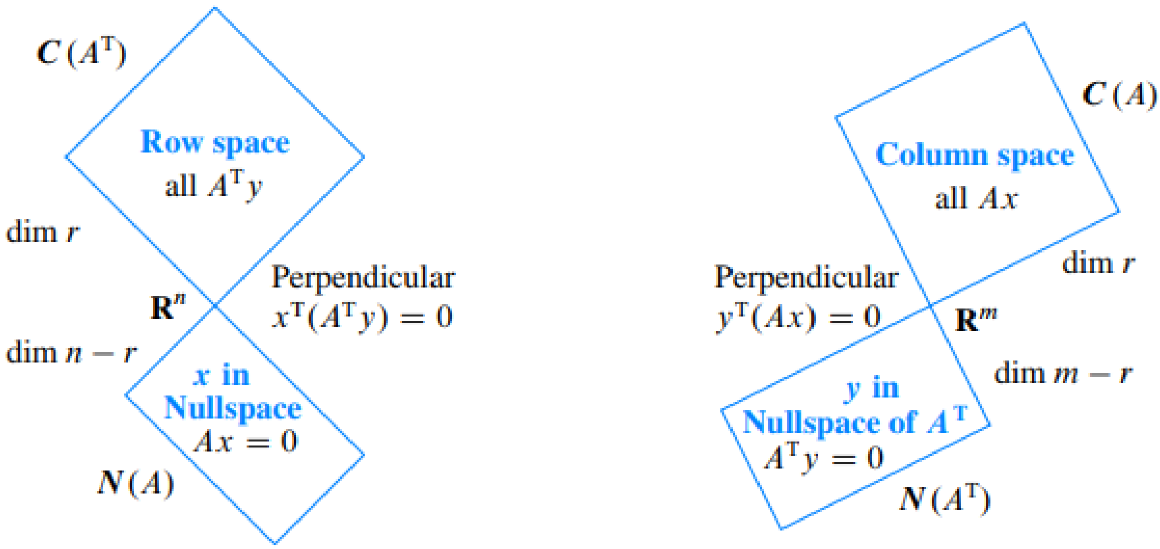



Figure 1: Dimensions and orthogonality for any $m$ by $n$ matrix $A$ of rank $r$.

## 3.2 Matrix Properties

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC} \quad \text{(left distributivity)} \tag{1}$$

$$(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA} \quad \text{(right distributivity)} \tag{2}$$

$$\mathbf{AB} \neq \mathbf{BA} \quad \text{(in general)} \tag{3}$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \quad \text{(associativity)} \tag{4}$$

## 3.3 Matrix Multiplication

$$(\mathbf{AB})_{kl} = \sum_m \mathbf{A}_{km}\mathbf{B}_{ml} \quad \mathbf{A} \in \mathbb{R}^{k,m}, \mathbf{B} \in \mathbb{R}^{m,l} \tag{5}$$

## 3.4 Time Complexities

| Operation | Input | Output | Algorithm | Time |
|---|---|---|---|---|
| Matmult | $A, B \in n \times n$ | $n \times n$ | Schoolbook | $O(n^3)$ |
| | | | Strassen [1] | $O(n^{2.807})$ |
| | | | Best | $O(n^\omega)$ |
| Matmult | $A \in n \times m, B \in m \times p$ | $n \times p$ | Schoolbook | $O(nmp)$ |
| Inversion | $A \in n \times n$ | $n \times n$ | Gauss–Jordan elimination | $O(n^3)$ |
| | | | Strassen [1] | $O(n^{2.807})$ |
| | | | Best | $O(n^\omega)$ |
| SVD | $A \in m \times n$ | $m \times m, m \times n, n \times n$<br>$m \times r, r \times r, n \times r$ | | $O(mn^2)$<br>$(m \geq n)$ |
| Determinant | $A \in n \times n$ | Scalar | Laplace expansion | $O(n!)$ |
| | | | Division-free [2] | $O(n!)$ |
| | | | LU decomposition | $O(n^3)$ |
| | | | Integer preserving [3] | $O(n^3)$ |
| Back substitution | $A$ triangular | $n$ solutions | Back substitution | $O(n^2)$ |

#### A comment on $\omega$

The lower bound on matmult time complexity is $O(n^\omega)$, where $\omega$ is an unknown constant bounded by $2 \leq \omega \leq 2.373$. Algorithms achieving lower values of $\omega$ tend to be less efficient in practice for all but the largest matrices. Of the algorithm with times of less than $O(n^3)$, only the Strassen algorithm has seen serious attempts at optimized implementation. Most matmult implementations use highly optimized variants of the standard $O(n^3)$ algorithm. At this point, memory and bus speeds dominate the performance of implementations, so simple Big-O notation cannot be used to reliably compare matmult performances.

| Name | Year | $\omega$ |
|---|---|---|
| Standard | - | 3 |
| Strassen [1] | 1969 | 2.807 |
| Pan [4] | 1978 | 2.796 |
| Bini et al. [5] | 1979 | 2.78 |
| Schönhage [6] | 1981 | 2.548 |
| Schönhage [6] | 1981 | 2.522 |
| Romani [7] | 1982 | 2.517 |
| Coppersmith and Winograd [8] | 1982 | 2.496 |
| Strassen [9] | 1986 | 2.479 |
| Coppersmith and Winograd [10] | 1990 | 2.376 |
| Williams [11] | 2012 | 2.37294 |
| Le Gall [12] | 2014 | 2.3728639 |
| Williams [11] | 2012 | 2.3727 |

# 4 Derivatives

## 4.1 Useful Rules for Derivatives

For general $\mathbf{A}$ and $\mathbf{X}$ (no special structure):

$$\partial\mathbf{A} = 0 \quad \text{where } \mathbf{A} \text{ is a constant} \tag{6}$$

$$\partial(c\mathbf{X}) = c\partial\mathbf{X} \tag{7}$$

$$\partial(\mathbf{X}+\mathbf{Y}) = \partial\mathbf{X} + \partial\mathbf{Y} \tag{8}$$

$$\partial(\mathrm{tr}(\mathbf{X})) = \mathrm{tr}(\partial(\mathbf{X})) \tag{9}$$

$$\partial(\mathbf{X}\mathbf{Y}) = (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y}) \tag{10}$$

$$\partial(\mathbf{X}\circ\mathbf{Y}) = (\partial\mathbf{X})\circ\mathbf{Y} + \mathbf{X}\circ(\partial\mathbf{Y}) \tag{11}$$

$$\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} \tag{12}$$

$$\partial(\det(\mathbf{X})) = \mathrm{tr}(\mathrm{adj}(\mathbf{X})\partial\mathbf{X}) \tag{13}$$

$$\partial(\det(\mathbf{X})) = \det(\mathbf{X})\,\mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \tag{14}$$

$$\partial(\ln(\det(\mathbf{X}))) = \mathrm{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \tag{15}$$

$$\partial(\mathbf{X}^T) = (\partial\mathbf{X})^T \tag{16}$$

$$\partial(\mathbf{X}^H) = (\partial\mathbf{X})^H \tag{17}$$

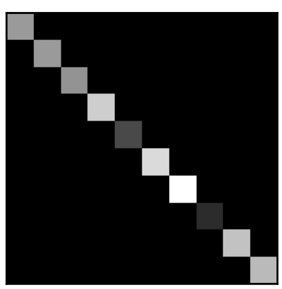
# 5 | Matrix Rogue Gallery

## 5.1 Non-Singular vs. Singular Matrices

For $\mathbf{A} \in \mathbb{R}^{n,n}$ (initially drawn from [13, p. 574]):

| **Non-Singular** | **Singular** |
|---|---|
| $\mathbf{A}$ is invertible | $\mathbf{A}$ is not invertible |
| The columns are independent | The columns are dependent |
| The rows are independent | The rows are dependent |
| $\det(\mathbf{A}) \neq 0$ | $\det(\mathbf{A}) = 0$ |
| $\mathbf{Ax} = 0$ has one solution: $\mathbf{x} = 0$ | $\mathbf{Ax} = 0$ has infinitely many solutions |
| $\mathbf{Ax} = \mathbf{b}$ has one solution: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ | $\mathbf{Ax} = \mathbf{b}$ has either no or infinitely many solutions |
| $\mathbf{A}$ has $n$ nonzero pivots | $\mathbf{A}$ has $r < n$ pivots |
| $\mathbf{A}$ has full rank $r = n$ | $\mathbf{A}$ has rank $r < n$ |
| The reduced row echelon form is $\mathbf{R} = \mathbf{I}$ | $\mathbf{R}$ has at least one zero row |
| The column space is all of $\mathbb{R}^n$ | The column space has dimension $r < n$ |
| The row space is all of $\mathbb{R}^n$ | The row space has dimension $r < n$ |
| All eigenvalues are nonzero | Zero is an eigenvalue of $\mathbf{A}$ |
| $\mathbf{A}^T\mathbf{A}$ is symmetric positive definite | $\mathbf{A}^T\mathbf{A}$ is only semidefinite |
| $\mathbf{A}$ has $n$ positive singular values | $\mathbf{A}$ has $r < n$ singular values |

## 5.2 Diagonal Matrix



$$A = \operatorname{diag}(a_1, \ldots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} \tag{18}$$

Square matrix. Entries above diagonal are equal to entries below diagonal.

Number of "free entries": $\frac{n(n+1)}{2}$.

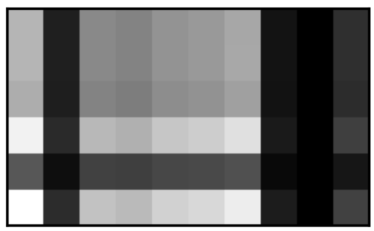
### Special Properties

$$\operatorname{eig}(A) = a_1, \ldots, a_n \tag{19}$$

$$\det(A) = \prod_i a_i \tag{20}$$

$$A^{-1} = \begin{bmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{bmatrix} \tag{21}$$

## 5.3 Dyads



$\mathbf{A} \in \mathbb{R}^{m,n}$ is a dyad if it can be written as

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T \quad \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n \tag{22}$$

### Special Properties

- The columns of $\mathbf{A}$ are copies of $\mathbf{u}$ scaled by the values of $\mathbf{v}$.
- The rows of $\mathbf{A}$ are copies of $\mathbf{u}^T$ scaled by the values of $\mathbf{v}$.
- If $\mathbf{A}$ is a dyad, it acts on a vector $\mathbf{x}$ as $\mathbf{Ax} = (\mathbf{u}\mathbf{v}^T)\mathbf{x} = (\mathbf{v}^T\mathbf{u})\mathbf{x}$.
- $\mathbf{Ax} = c\mathbf{u}$ ($\mathbf{A}$ scales $\mathbf{x}$ and points it along $\mathbf{u}$).
- $\mathbf{A}_{ij} = \mathbf{u}_i\mathbf{v}_j$.
- If $\mathbf{u}, \mathbf{v} \neq 0$, then $\operatorname{rank}(\mathbf{A}) = 1$.
- If $m = n$, $\mathbf{A}$ has one eigenvalue $\lambda = \mathbf{v}^T\mathbf{u}$ and eigenvector $\mathbf{u}$.
- A dyad can always be written in a normalized form $c\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$.

## 5.4 Hermitian Matrix

$\mathbf{H} \in \mathbb{C}^{m,n}$ is Hermitian iff

$$\mathbf{H} = \mathbf{H}^H \tag{23}$$

where $\mathbf{H}^H$ is the conjugate transpose of $\mathbf{H}$.

For $\mathbf{H} \in \mathbb{R}^{m,n}$, Hermitian and symmetric matrices are equivalent.

### Special Properties

$$\mathbf{H}_{ii} \in \mathbb{R} \tag{24}$$
$$\mathbf{H}\mathbf{H}^H = \mathbf{H}^H\mathbf{H} \tag{25}$$
$$\mathbf{x}^H\mathbf{H}\mathbf{x} \in \mathbb{R} \quad \forall \mathbf{x} \in \mathbb{C} \tag{26}$$
$$\mathbf{H}_1 + \mathbf{H}_2 = \text{Hermitian} \tag{27}$$
$$\mathbf{H}^{-1} = \text{Hermitian} \tag{28}$$
$$\mathbf{A} + \mathbf{A}^H = \text{Hermitian} \tag{29}$$
$$\mathbf{A} - \mathbf{A}^H = \text{Skew-Hermitian} \tag{30}$$
$$\mathbf{AB} = \text{Hermitian iff } \mathbf{AB} = \mathbf{BA} \tag{31}$$
$$\det(\mathbf{H}) \in \mathbb{R} \tag{32}$$
$$\text{eig}(\mathbf{H}) \in \mathbb{R} \tag{33}$$

## 5.5 Idempotent Matrix

A matrix $\mathbf{A}$ is idempotent iff

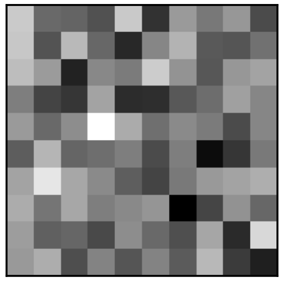
$$\mathbf{AA} = \mathbf{A} \tag{34}$$

### Special Properties

$$\mathbf{A}^n = A \quad \forall n \tag{35}$$
$$\mathbf{I} - \mathbf{A} \quad \text{is idempotent} \tag{36}$$
$$\mathbf{A}^H \quad \text{is idempotent} \tag{37}$$
$$\mathbf{I} - \mathbf{A}^H \quad \text{is idempotent} \tag{38}$$
$$\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) \tag{39}$$
$$\mathbf{A}(I - \mathbf{A}) = 0 \tag{40}$$
$$\mathbf{A}^+ = \mathbf{A} \tag{41}$$
$$f(s\mathbf{I} + t\mathbf{A}) = (\mathbf{I} - \mathbf{A})f(s) + \mathbf{A}f(s+t) \tag{42}$$
$$\mathbf{AB} = \mathbf{BA} \implies \mathbf{AB} \text{ is idempotent} \tag{43}$$
$$\text{eig}(\mathbf{A})_i \in \{0, 1\} \tag{44}$$
$$\mathbf{A} \text{ is always diagonalizable} \tag{45}$$

$\mathbf{A} - \mathbf{I}$ may not be idempotent.

## 5.6 Orthogonal Matrix



(Not much visible structure)

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{46}$$

A matrix $\mathbf{U}$ is orthogonal iff:

$$\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = I \tag{47}$$

Square matrix. The columns form an orthonormal basis of $\mathbb{R}^n$.

### Special Properties

- The eigenvalues of $\mathbf{U}$ are placed on the unit circle.
- The eigenvectors of $\mathbf{U}$ are unitary (have length one).
- $\mathbf{U}^{-1}$ is orthogonal.

$$\mathbf{U}^T = \mathbf{U}^{-1} \tag{48}$$

$$\mathbf{U}^{-T} = \mathbf{U} \tag{49}$$

$$\mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{50}$$

$$\mathbf{U}\mathbf{U}^T = \mathbf{I} \tag{51}$$

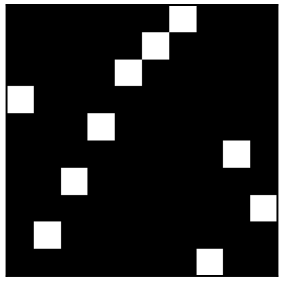
$$\det(\mathbf{U}) = \pm 1 \tag{52}$$

Orthogonal matrices preserve the lengths and angles of the vectors they operator on. The converse is true: any matrix which preserves lengths and angles is orthogonal.

$$\|\mathbf{U}\mathbf{x}\|_2^2 = (\mathbf{U}\mathbf{x})^T(\mathbf{U}\mathbf{x}) = \mathbf{x}^T\mathbf{U}^T\mathbf{U}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \tag{53}$$

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\|_F = \|\mathbf{A}\|_F \quad \forall \mathbf{A}, \mathbf{U}, \mathbf{V} \text{ with } U, V orthogonal \tag{54}$$

## 5.7 Permutation Matrix



TODO

## 5.8 Positive Definite

$\mathbf{A} \in \mathbb{S}^n$ is positive definite (denoted $\mathbf{A} \succ 0$) if any of the following are true:

- $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \forall \mathbf{x} \in \mathbb{R}^n$.
- $\text{eig}(\mathbf{A}) > 0$

### Special Properties

- If $\mathbf{A}$ is PD and invertible, $\mathbf{A}^{-1}$ is also PD.
- If $\mathbf{A}$ is PD and $c \in \mathbb{R}$ then $c\mathbf{A}$ is PD.
- The diagonal entries $\mathbf{A}_{ii}$ are real and non-negative, so $\text{tr}(\mathbf{A}) \geq 0$.
- $\det(\mathbf{A}) > 0$
- For $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{A}^T \mathbf{A} \succ 0 \iff \mathbf{A}$ is full-column rank ($\text{rank}(\mathbf{A}) = n$)
- For $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{A}\mathbf{A}^T \succ 0 \iff \mathbf{A}$ is full-row rank ($\text{rank}(\mathbf{A}) = m$)
- $\mathbf{P} \succ 0$ defines a full-dimensional, bounded ellipsoid centered at the origin and defined by the set $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : x^T \mathbf{P}^{-1} x \leq 1\}$. The eigenvalues $\lambda_i$ and eigenvectors $u_i$ of $\mathbf{P}$ define the orientation and shape of the ellipsoid. $u_i$ are the semi-axes while the lengths of the semi-axes are given by $\sqrt{\lambda_i}$. Using the Cholesky decomposition, $\mathbf{P}^{-1} = \mathbf{A}^T \mathbf{A}$, an equivalent definition of the ellipsoid is $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{A}\mathbf{x}\|_2 \leq 1\}$.

## 5.9 Positive Semi-Definite

$\mathbf{A}$ is positive semi-definite (denoted $\mathbf{A} \succeq 0$) if any of the following are true:

- $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$.
- $\text{eig}(\mathbf{A}) \geq 0$

### Special Properties

- For $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{A}^T\mathbf{A} \succeq 0$
- For $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{A}\mathbf{A}^T \succeq 0$
- The positive semi-definite matrices $\mathbb{S}^n_+$ form a convex cone. For any two PSD matrices $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n_+$ and some $\alpha \in [0, 1]$:

$$\mathbf{x}^T(\alpha\mathbf{A} + (1-\alpha)\mathbf{B})\mathbf{x} = \alpha\mathbf{x}^T\mathbf{A}\mathbf{x} + (1-\alpha)\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0 \quad \forall\mathbf{x} \tag{55}$$

$$\alpha\mathbf{A} + (1-\alpha)\mathbf{B} \in \mathbb{S}^n_+ \tag{56}$$

- For $\mathbf{A} \in \mathbb{S}^n_+$ and $\alpha \geq 0$, $\alpha\mathbf{A} \succeq 0$, so $\mathbb{S}^n_+$ is a cone.

## 5.10 Projection Matrix

A square matrix $\mathbf{P}$ is a projection matrix that projects onto a vector space $\mathcal{S}$ iff

$$\mathbf{P} \text{ is idempotent} \tag{57}$$

$$\mathbf{P}\mathbf{x} \in \mathcal{S} \quad \forall\mathbf{x} \tag{58}$$

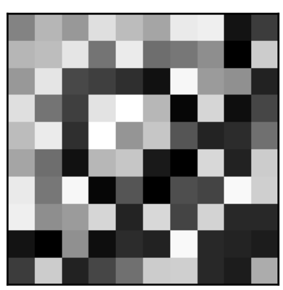
$$\mathbf{P}\mathbf{z} = \mathbf{z} \quad \forall\mathbf{z} \in \mathcal{S} \tag{59}$$

## 5.11 Singular Matrix

A square matrix that is not invertible.

$\mathbf{A} \in \mathbb{R}^{n,n}$ is singular iff $\det\mathbf{A} = 0$ iff $\mathcal{N}(A) \neq \{0\}$.

## 5.12 Symmetric Matrix



$\mathbf{A} \in \mathbb{S}^n$ is a symmetric matrix if $\mathbf{A} = \mathbf{A}^T$ (entries above diagonal are equal to entries below diagonal).

$$\begin{bmatrix} a & b & c & d & e & f \\ b & g & l & m & o & p \\ c & l & h & n & q & r \\ d & m & n & i & s & t \\ e & o & q & s & j & u \\ f & p & r & t & u & k \end{bmatrix} \tag{60}$$

### Special Properties

$$\mathbf{A} = \mathbf{A}^T \tag{61}$$
$$\operatorname{eig}(A) \in \mathbb{R}^n \tag{62}$$

Number of "free entries": $\frac{n(n+1)}{2}$.

If $\mathbf{A}$ is real, it can be decomposed into $\mathbf{A} = \mathbf{Q}^T\mathbf{D}\mathbf{Q}$ where $\mathbf{Q}$ is a real orthogonal matrix (the columns of which are eigenvectors of $\mathbf{A}$) and $\mathbf{D}$ is real and diagonal containing the eigenvalues of $\mathbf{A}$.

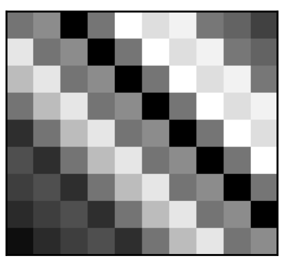
For a real, symmetric matrix with non-negative eignevalues, the eigenvalues and singular values coincide.

## 5.13 Skew-Hermitian

A matrix $\mathbf{H} \in \mathbb{C}^{m,n}$ is Skew-Hermitian iff

$$\mathbf{H} = -\mathbf{H}^H \tag{63}$$

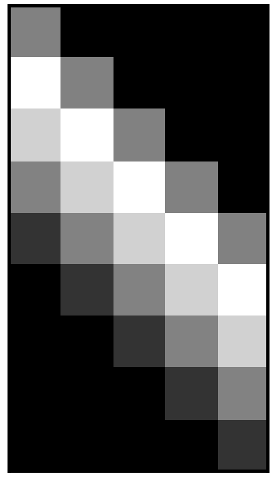
## 5.14 Toeplitz Matrix, General Form



Constant values on descending diagonals.

$$\begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix} \tag{64}$$

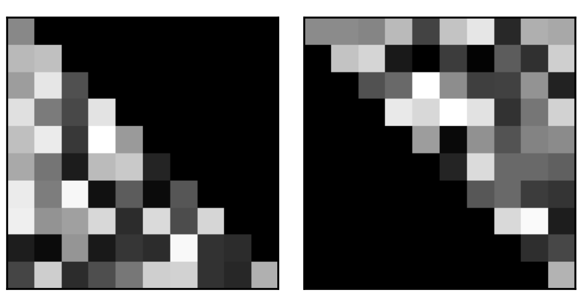
## 5.15 Toeplitz Matrix, Discrete Convolution



Constant values on main and subdiagonals.

$$\begin{bmatrix} h_m & 0 & 0 & \dots & 0 & 0 \\ \vdots & h_m & 0 & \dots & 0 & 0 \\ h_1 & \vdots & h_m & \dots & 0 & 0 \\ 0 & h_1 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & h_1 & \ddots & h_m & 0 \\ 0 & 0 & 0 & \ddots & \vdots & h_m \\ 0 & 0 & 0 & \dots & h_1 & \vdots \\ 0 & 0 & 0 & \dots & 0 & h_1 \end{bmatrix} \tag{65}$$

## 5.16 Triangular Matrix



$$\begin{bmatrix} a & b & c & d & e & f \\ & g & h & i & j & k \\ & & l & m & n & o \\ & & & p & q & r \\ & & & & s & t \\ & & & & & u \end{bmatrix} \quad \begin{bmatrix} a & & & & & \\ b & g & & & & \\ c & h & l & & & \\ d & i & m & p & & \\ e & j & n & q & s & \\ f & k & o & r & t & u \end{bmatrix} \tag{66}$$

Square matrices in which all elements either above or below the main diagonal are zero. An upper (left) and a lower (right) triangular matrix are shown above.

For an upper triangular matrix $A_{ij} = 0$ whenever $i > j$; for a lower triangular matrix $A_{ij} = 0$ whenever $i < j$.

### Special Properties

$$\text{eig}(A) = \text{diag}(A) \tag{67}$$

$$\det(A) = \prod_i \text{diag}(A)_i \tag{68}$$

The product of two upper (lower) triangular matrices is still upper (lower) triangular.

The inverse of a nonsingular upper (lower) triangular matrix is still upper (lower) triangular.

## 5.17 Vandermonde Matrix

$$V = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \dots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \dots & \alpha_m^{n-1} \end{bmatrix} \tag{69}$$

Alternatively,

$$V_{i,j} = \alpha_i^{j-1} \tag{70}$$

### Uses

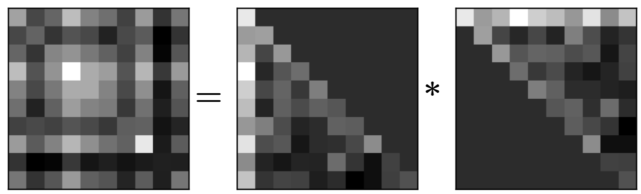
Polynomial interpolation of data.

### Special Properties

- $\det(V) = \prod_{1 \leq i < j \leq n} (x_j - x_i)$

# 6 | Matrix Decompositions

## 6.1 LLT/UTU: Cholesky Decomposition

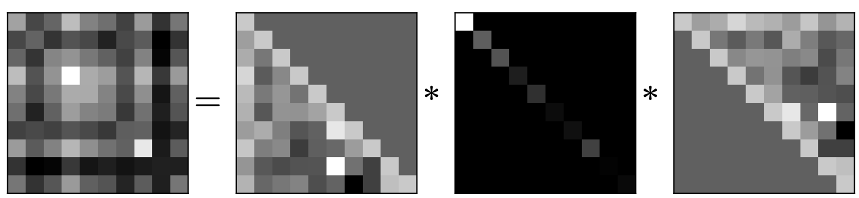


If $\mathbf{A}$ is symmetric, positive definite, square, then

$$\mathbf{A} = \mathbf{U}^T\mathbf{U} = \mathbf{L}\mathbf{L}^T \tag{71}$$

where $\mathbf{U}$ is a unique upper triangular matrix and $\mathbf{L}$ is a unique lower-triangular matrix.

## 6.2 LDLT



If $\mathbf{A}$ is a non-singular symmetric definite square matrix, then

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T = \mathbf{L}^T\mathbf{D}\mathbf{L} \tag{72}$$

where $\mathbf{L}$ is a unit lower triangular matrix and $\mathbf{D}$ is a diagonal matrix. If $\mathbf{A} \succ 0$, then $\mathbf{D}_{ii} > 0$.

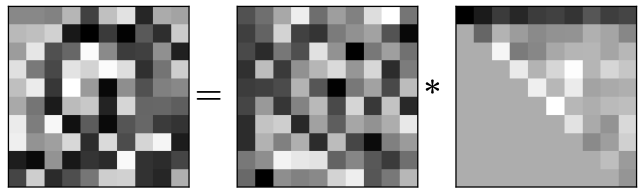
## 6.3 PCA: Principle Components Analysis

Find normalized directions in data space such that the variance of the projections of the centered data points is maximal. For centered data $\tilde{\mathbf{X}}$, the mean-square variation of data along a vector $\mathbf{x}$ is $\mathbf{x}^T\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\mathbf{x}$.

$$\max_{\mathbf{x}\in\mathbb{R}^n,\ \|\mathbf{x}\|_2=1} \mathbf{x}^T\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\mathbf{x} \tag{73}$$

Taking an SVD of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$ gives $H = \mathbf{U}_r\mathbf{D}^2\mathbf{U}^T$, which is maximized by taking $\mathbf{x} = \mathbf{u}_1$. By repeatedly removing the first principal components and recalculating, all the principal axes can be found.

## 6.4 QR: Orthogonal-triangular

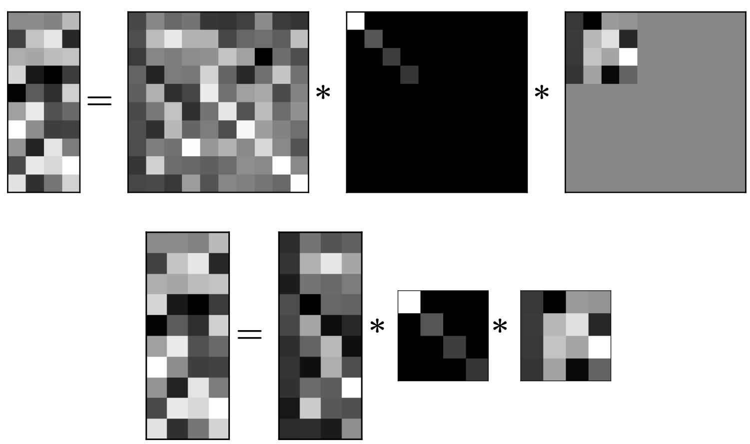


For $\mathbf{A} \in \mathbb{R}^{n,n}$, $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is an upper triangular matrix. If $\mathbf{A}$ is non-singular, then $\mathbf{Q}$ and $\mathbf{R}$ are uniquely defined if diag($\mathbf{R}$) are imposed to be positive.

### Algorithms

Gram-Schmidt.

## 6.5 SVD: Singular Value Decomposition



Any matrix $\mathbf{A} \in \mathbb{R}^{m,n}$ can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{74}$$

where

$$U = \text{eigenvectors of } \mathbf{A}\mathbf{A}^T \qquad \mathbb{R}^{m,m} \tag{75}$$

$$D = \text{diag}(\sigma_i) = \sqrt{\text{diag}(\text{eig}(\mathbf{A}\mathbf{A}^T))} \qquad \mathbb{R}^{n,m} \tag{76}$$

$$V = \text{eigenvectors of } \mathbf{A}^T\mathbf{A} \qquad \mathbb{R}^{n,n} \tag{77}$$

Let $\sigma_i$ be the non-zero singular values for $i = 1, \ldots, r$ where $r$ is the rank of $\mathbf{A}$; $\sigma_1 \geq \ldots \geq \sigma_r$.

We also have that

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \tag{78}$$
$$\mathbf{A}^T\mathbf{u}_i = \sigma_i \mathbf{v}_i \tag{79}$$
$$\mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{80}$$
$$\mathbf{V}^T\mathbf{V} = \mathbf{I} \tag{81}$$

$\mathbf{D}$ can be written in an expanded form:

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \tag{82}$$

The final $n-r$ columns of $\mathbf{V}$ give an orthonormal basis spanning $\mathcal{N}(\mathbf{A})$. An orthonormal basis spanning the range of $\mathbf{A}$ is given by the first $r$ columns of $\mathbf{U}$.

$$\|\mathbf{A}\|_F^2 = \text{Frobenius norm} = \operatorname{tr}(\mathbf{A}^T\mathbf{A}) = \sum_{i=1}^{r} \sigma_i^2 \tag{83}$$
$$\|\mathbf{A}\|_2^2 = \sigma_1^2 \tag{84}$$
$$\|\mathbf{A}\|_* = \text{nuclear norm} = \sum_{i=1}^{r} \sigma_i \tag{85}$$

The **condition number** $\kappa$ of an invertible matrix $\mathbf{A} \in \mathbb{R}^{n,n}$ is the ratio of the largest and smallest singular value. Matrices with large condition numbers are closer to being singular and more sensitive to changes.

$$\kappa(\mathbf{A}) = \frac{\sigma_1}{\sigma_n} = \|A\|_2 \cdot \left\|A^{-1}\right\|_2 \tag{86}$$

## Low-Rank Approximation

Approximating $\mathbf{A} \in \mathbb{R}^{m,n}$ by a matrix $\mathbf{A}_k$ of rank $k > 0$ can be formulated as the optimization probem

$$\min_{\mathbf{A}_k \in \mathbb{R}^{m,n}} \|\mathbf{A} - \mathbf{A}_k\|_F^2 : \operatorname{rank} \mathbf{A}_k = k, 1 \leq k \leq \operatorname{rank}(\mathbf{A}) \tag{87}$$

The optimal solution of this problem is given by

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{88}$$

where

$$\frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_1^2 + \ldots + \sigma_k^2}{\sigma_1^2 + \ldots + \sigma_r^2} \tag{89}$$
$$1 - \frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_{k+1}^2 + \ldots + \sigma_r^2}{\sigma_1^2 + \ldots + \sigma_r^2} \tag{90}$$

is the fraction of the total variance in $\mathbf{A}$ explained by the approximation $\mathbf{A}_k$.

### Range and Nullspace

$$\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{V}_{nr}) \tag{91}$$
$$\mathcal{N}(\mathbf{A})^\perp \equiv \mathcal{R}(\mathbf{A}^T) = \mathcal{R}(\mathbf{V}_r) \tag{92}$$
$$\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U}_r) \tag{93}$$
$$\mathcal{R}(\mathbf{A})^\perp \equiv \mathcal{N}(\mathbf{A}^T) = \mathcal{R}(\mathbf{U}_{nr}) \tag{94}$$

where $\mathbf{V}_r$ is the first $r$ columns of $V$ and $V_nr$ are the last $[r+1, n]$ columns; similarly for $\mathbf{U}$.

### Projectors

The projection of $\mathbf{x}$ onto $\mathcal{N}(\mathbf{A})$ is $(\mathbf{V}_{nr}\mathbf{V}_{nr}^T)\mathbf{x}$. Since $\mathbf{I}_n = \mathbf{V}_r\mathbf{V}_r^T + \mathbf{V}_{nr}\mathbf{V}_{nr}^T$, $(\mathbf{I}_n - \mathbf{V}_r\mathbf{V}_r^T)\mathbf{x}$ also works. The projection of $\mathbf{x}$ onto $\mathcal{R}(\mathbf{A})$ is $(\mathbf{U}_r\mathbf{U}_r^T)\mathbf{x}$.

If $\mathbf{A} \in \mathbb{R}^{m,n}$ is full row rank ($\mathbf{A}\mathbf{A}^T \succ 0$), then the minimum distance to an affine set $\{x : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \mathbf{b} \in \mathbb{R}^m$ is given by $\mathbf{x}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$.

If $\mathbf{A} \in \mathbb{R}^{m,n}$ is full column rank ($\mathbf{A}^T\mathbf{A} \succ 0$), then the minimum distance to an affine set $\{x : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \mathbf{b} \in \mathbb{R}^m$ is given by $\mathbf{x}^* = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.

### Computational Notes

Since $\sigma \approx 0$, a *numerical rank* can be estimated for the matrix as the largest $k$ such that $\sigma_k > \epsilon\sigma_1$ for $\epsilon \geq 0$.

## 6.6 Eigenvalue Decomposition for Diagonalizable Matrices

For a square, diagonalizable matrix $\mathbf{A} \in \mathbb{R}^{n,n}$

$$\mathbf{A} = U\Lambda U^{-1} \tag{95}$$

where $U \in \mathbb{C}^{n,n}$ is an invertible matrix whose columns are the eigenvectors of $\mathbf{A}$ and $\Lambda$ is a diagonal matrix containing the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ in the diagonal.

The columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ satisfy

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad i = 1, \ldots, n \tag{96}$$

## 6.7 Eigenvalue (Spectral) Decomposition for Symmetric Matrices

A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n,n}$ can be factored as

$$\mathbf{A} = U\Lambda U^T = \sum_i^n \lambda_i\mathbf{u}_i\mathbf{u}_i^T \tag{97}$$

where $U \in \mathbb{R}^{n,n}$ is an orthogonal matrix whose columns $\mathbf{u}_i$ are the eigenvectors of $\mathbf{A}$ and $\Lambda$ is a diagonal matrix containing the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ of $\mathbf{A}$ in the diagonal. These eigenvalues are always real. The eigenvectors can always be chosen to be real and to form an orthonormal basis.

The columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ satisfy

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad i = 1, \ldots, n \tag{98}$$

## 6.8 Schur Complements

For $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{B} \in \mathbb{S}^n$, $\mathbf{X} \in \mathbb{R}^{n,m}$ with $\mathbf{B} \succ 0$ and the block matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{B} \end{bmatrix} \tag{99}$$

and the Schur complement of $\mathbf{A}$ in $\mathbf{M}$

$$S = \mathbf{A} - \mathbf{X}\mathbf{B}^{-1}\mathbf{X}^T \tag{100}$$

Then

$$\mathbf{M} \succeq 0 \iff S \succeq 0 \tag{101}$$

$$\mathbf{M} \succ 0 \iff S \succ 0 \tag{102}$$

# 7 | Transpose Properties

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T \tag{103}$$

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \tag{104}$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \tag{105}$$

# 8 Determinant Properties

Geometrically, if a unit volume is acted on by $\mathbf{A}$, then $|\det(\mathbf{A})|$ indicates the volume after the transformation.

$$\det(I_n) = 1 \tag{106}$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A}) \tag{107}$$

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})} = \det(\mathbf{A})^{-1} \tag{108}$$

$$\det(AB) = \det(BA) \tag{109}$$

$$\det(AB) = \det(A)\det(B) \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n,n} \tag{110}$$

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}) \quad \mathbf{A} \in \mathbb{R}^{n,n} \tag{111}$$

$$\det(\mathbf{A}) = \prod \operatorname{eig}(\mathbf{A}) \tag{112}$$

# 9 Trace Properties

For $\mathbf{A} \in \mathbb{R}^{n,n}$

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} \mathbf{A}_{ii} \tag{113}$$

$$\operatorname{tr}(\mathbf{A}+\mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}) \tag{114}$$

$$\operatorname{tr}(c\mathbf{A}) = c\operatorname{tr}(\mathbf{A}) \tag{115}$$

$$\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A}^T) \tag{116}$$

For $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ of compatible dimensions,

$$\operatorname{tr}(\mathbf{A}^T\mathbf{B}) = \operatorname{tr}(\mathbf{A}\mathbf{B}^T) = \operatorname{tr}(\mathbf{B}^T\mathbf{A}) = \operatorname{tr}(\mathbf{B}\mathbf{A}^T) \tag{117}$$

$$\operatorname{tr}(\mathbf{ABCD}) = \operatorname{tr}(\mathbf{BCDA}) = \operatorname{tr}(\mathbf{CDAB}) = \operatorname{tr}(\mathbf{DABC}) \tag{118}$$

(Invariant under cyclic permutations)

# 10 Inverse Properties

The inverse of $\mathbf{A} \in \mathbb{C}^{n,n}$ is denoted $\mathbf{A}^{-1}$ and defined such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \tag{119}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\mathbf{A}$ is nonsingular if $\mathbf{A}^{-1}$ exists; otherwise, $\mathbf{A}$ is singular.

If individual inverses exist

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{120}$$

more generally

$$(\mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_n)^{-1} = \mathbf{A}_n^{-1} \ldots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1} \tag{121}$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \tag{122}$$

# 11 | Pseudo-Inverse Properties

For $\mathbf{A} \in \mathbb{R}^{m,n}$, a pseudoinverse satisfies:

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A} \tag{123}$$

$$\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+} \tag{124}$$

$$(\mathbf{A}\mathbf{A}^{+})^{T} = \mathbf{A}\mathbf{A}^{+} \tag{125}$$

$$(\mathbf{A}^{+}\mathbf{A})^{T} = \mathbf{A}^{+}\mathbf{A} \tag{126}$$

## 11.1 Moore-Penrose Pseudoinverse

$$\mathbf{A}^{+} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^{T} \tag{127}$$

where the foregoing comes from a singular-value decomposition and $\mathbf{D}^{-1} = \mathrm{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r})$

### Special Properties

- $\mathbf{A}^{+} = \mathbf{A}^{-1}$ if $\mathbf{A} \in \mathbb{R}^{n,n}$ and $\mathbf{A}$ is square and nonsingular.
- $\mathbf{A}^{+} = (\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}$, if $\mathbf{A} \in \mathbb{R}^{m,n}$ is full column rank $(r = n \leq m)$. $\mathbf{A}^{+}$ is a left inverse of $\mathbf{A}$, so $\mathbf{A}^{+}\mathbf{A} = \mathbf{V}_r\mathbf{V}_r^{T} = \mathbf{V}\mathbf{V}^{T} = \mathbf{I}_n$.
- $\mathbf{A}^{+} = \mathbf{A}^{T}(\mathbf{A}\mathbf{A}^{T})^{-1}$, if $\mathbf{A} \in \mathbb{R}^{m,n}$ is full row rank $(r = m \leq n)$. $\mathbf{A}^{+}$ is a right inverse of $\mathbf{A}$, so $\mathbf{A}\mathbf{A}^{+} = \mathbf{U}_r\mathbf{U}_r^{T} = \mathbf{U}\mathbf{U}^{T} = \mathbf{I}_m$.

# 12 Hadamard Identities

$$(\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij} B_{ij} \ \forall \ i, j \tag{128}$$

$$\mathbf{A} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{A} \tag{129}$$ [14]

$$\mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C} \tag{130}$$

$$\mathbf{A} \circ (\mathbf{B} + \mathbf{C}) = \mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ \mathbf{C} \tag{131}$$ [14]

$$a(\mathbf{A} \circ \mathbf{B}) = (a\mathbf{A}) \circ \mathbf{B} = \mathbf{A} \circ (a\mathbf{B}) \tag{132}$$ [14]

$$(\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T \tag{133}$$

$$(\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T \tag{134}$$

$$(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \sum_{i,j} \big( (\mathbf{x}\mathbf{x}^T) \circ \mathbf{A} \big) \tag{135}$$

$$\mathbf{x}^T (\mathbf{A} \circ \mathbf{B}) \mathbf{y} = \mathrm{tr}((\mathrm{diag}(\mathbf{x})\mathbf{A})^T \mathbf{B} \, \mathrm{diag}(\mathbf{y})) \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m,n} \tag{136}$$ [15]

$$\mathrm{tr}(\mathbf{A}^T \mathbf{B}) = \mathbf{1}^T (\mathbf{A} \circ \mathbf{B}) \mathbf{1} \tag{137}$$

# 13 | Eigenvalue Properties

$\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n,n}$ and $u \in \mathbb{C}^n$ is a corresponding eigenvector if $\mathbf{Au} = \lambda\mathbf{u}$ and $\mathbf{u} \neq 0$. Equivalantly, $(\lambda\mathbf{I}_n - \mathbf{A})\mathbf{u} = 0$ and $\mathbf{u} \neq 0$. Eigenvalues satisfy the equation $\det(\lambda\mathbf{I}_n - \mathbf{A}) = 0$.

Any matrix $\mathbf{A} \in \mathbb{R}^{n,n}$ has $n$ eigenvalues, though some may be repeated. $\lambda_1$ is the largest eigenvalue and $\lambda_n$ the smallest.

$$\mathrm{eig}(\mathbf{A}\mathbf{A}^T) = \mathrm{eig}(\mathbf{A}^T\mathbf{A}) \tag{138}$$

(Note that the number of entries in $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ may differ significantly leading to different compute times.)

$$\mathrm{eig}(\mathbf{A}^T\mathbf{A}) \geq 0 \tag{139}$$

## Computation

TODO: eigsh, small eigen value extraction, top-k

# 14 Norms

## 14.1 Matrices

Matrix norms satisfy some properties:

$$f(\mathbf{A}) \geq 0 \tag{140}$$
$$f(\mathbf{A}) = 0 \iff \mathbf{A} = 0 \tag{141}$$
$$f(c\mathbf{A}) = |c| f(\mathbf{A}) \tag{142}$$
$$f(\mathbf{A}+\mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B}) \tag{143}$$

Many popular matrix norms also satisfy "sub-multiplicativity": $f(\mathbf{AB}) \leq f(\mathbf{A})f(\mathbf{B})$.

### 14.1.1 Frobenius norm

$$\|\mathbf{A}\|_F = \sqrt{\operatorname{tr} \mathbf{A}\mathbf{A}^H} \tag{144}$$
$$= \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |\mathbf{A}_{ij}|^2} \tag{145}$$
$$= \sqrt{\sum_{i=1}^{m} \operatorname{eig}(A^H A)_i} \tag{146}$$

#### Special Properties

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_F \|\mathbf{x}\|_2 \quad \mathbf{x} \in \mathbb{R}^n \tag{147}$$
$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F \tag{148}$$
$$\left\|\mathbf{C} - \mathbf{x}\mathbf{x}^T\right\|_F^2 = \|\mathbf{C}\|_F^2 + \|\mathbf{x}\|_2^4 - 2\mathbf{x}^T\mathbf{C}\mathbf{x} \tag{149}$$

### 14.1.2 Operator Norms

For $p = 1, 2, \infty$ or other values, an operator norm indicates the maximum input-output gain of the matrix.

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{u}\|_p = 1} \|\mathbf{Au}\|_p \tag{150}$$

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{u}\|_1 = 1} \|\mathbf{Au}\|_1 \tag{151}$$
$$= \max_{j=1,\ldots,n} \sum_{i=1}^{m} |\mathbf{A}_{ij}| \tag{152}$$
$$= \text{Largest absolute column sum} \tag{153}$$

$$
\begin{aligned}
\|\mathbf{A}\|_\infty &= \max_{\|\mathbf{u}\|_\infty=1} \|\mathbf{Au}\|_\infty &(154)\\
&= \max_{j=1,\ldots,m} \sum_{i=1}^{n} |\mathbf{A}_{ij}| &(155)\\
&= \text{Largest absolute row sum} &(156)
\end{aligned}
$$

$$
\begin{aligned}
\|\mathbf{A}\|_2 &= \text{“spectral norm”} &(157)\\
&= \max_{\|\mathbf{u}\|_2=1} \|\mathbf{Au}\|_2 &(158)\\
&= \sqrt{\max(\operatorname{eig}(\mathbf{A}^T\mathbf{A}))} &(159)\\
&= \text{Square root of largest eigenvalue of } \mathbf{A}^T\mathbf{A} &(160)
\end{aligned}
$$

**Special Properties**

$$
\begin{aligned}
\|\mathbf{Au}\|_p &\le \|\mathbf{A}\|_p \|\mathbf{u}\|_p &(161)\\
\|\mathbf{AB}\|_p &\le \|\mathbf{A}\|_p \|\mathbf{B}\|_p &(162)\\
& &(163)
\end{aligned}
$$

### 14.1.3 Spectral Radius

Not a proper norm.

$$\rho(\mathbf{A}) = \text{spectral radius}(\mathbf{A}) = \max_{i=1,\ldots,n} |\operatorname{eig}(\mathbf{A})_i| \quad (164)$$

**Special Properties**

$$
\begin{aligned}
\rho(\mathbf{A}) &\le \|\mathbf{A}\|_p &(165)\\
\rho(\mathbf{A}) &\le \min(\ \|\mathbf{A}\|_1, \|\mathbf{A}\|_\infty) &(166)\\
& &(167)
\end{aligned}
$$

## 14.2 Vectors

$$
\begin{aligned}
\|\mathbf{x}\|_1 &= \sum_i |\mathbf{x}_i| & \text{(L1-norm)} \quad &(168)\\
\|\mathbf{x}\|_p &= (\sum_i |\mathbf{x}_i|^p)^{1/p} & \text{(P-norm)} \quad &(169)\\
\|\mathbf{x}\|_\infty &= \max_i |\mathbf{x}_i| & \text{(L}\infty\text{-norm, L-infinity norm)} \quad &(170)
\end{aligned}
$$

# 15 Bounds

## 15.1 Matrix Gain

$$\lambda_{\min}(\mathbf{A}^T\mathbf{A}) \leq \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \lambda_{\max}(\mathbf{A}^T\mathbf{A}) \tag{171}$$

$$\max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} \implies \mathbf{x} = u_1 \tag{172}$$

$$\min_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sqrt{\lambda_{\min}(\mathbf{A}^T\mathbf{A})} \implies \mathbf{x} = u_n \tag{173}$$

## 15.2 Norms

For $\mathbf{x} \in \mathbb{R}^n$

$$\frac{1}{\sqrt{n}}\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{\mathrm{card}(\mathbf{x})}\|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_2 \leq n\|\mathbf{x}\|_\infty \tag{174}$$

For any $0 < p < q$ we have that $\|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p$.

## 15.3 Rayleigh quotients

The Rayleigh quotient of $\mathbf{A} \in \mathbb{S}^n$ is given by

$$\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \quad \mathbf{x} \neq 0 \tag{175}$$

$$\lambda_{\min}(\mathbf{A}) \leq \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \leq \lambda_{\max}(\mathbf{A}) \quad \mathbf{x} \neq 0 \tag{176}$$

$$\lambda_{\max}(A) = \max_{\mathbf{x}:\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = u_1 \tag{177}$$

$$\lambda_{\min}(A) = \min_{\mathbf{x}:\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = u_n \tag{178}$$

where $u_1$ and $u_n$ are the eigenvectors associated with $\lambda_{\max}$ and $\lambda_{\min}$, respectively.

# 16 Linear Equations

The linear equation $\mathbf{Ax} = \mathbf{y}$ with $\mathbf{A} \in \mathbb{R}^{m,n}$ admits a solution iff $\text{rank}([\mathbf{Ay}]) = \text{rank}(\mathbf{A})$. If this is satisfied, the set of all solutions is an affine set $\mathcal{S} = \{\mathbf{x} = \bar{\mathbf{x}} + z : z \in \mathcal{N}(\mathbf{A})\}$ where $\bar{\mathbf{x}}$ is any vector such that $\mathbf{A}\bar{\mathbf{x}} = \mathbf{y}$. The solution is unique if $\mathcal{N}(\mathbf{A}) = \{0\}$.

$\mathbf{Ax} = \mathbf{y}$ is *overdetermined* if it is tall/skinny ($m > n$); that is, if there are more equations than unknowns. If $\text{rank}(\mathbf{A}) = n$ then $\dim \mathcal{N}(\mathbf{A}) = 0$, so there is either no solution or one solution. Overdetermined systems often have no solution ($\mathbf{y} \notin \mathcal{R}(\mathbf{A})$), so an approximate solution is necessary. See section 16.1.

$\mathbf{Ax} = \mathbf{y}$ is *underdetermined* if it is short/wide ($n > m$); that is, if has more unknowns than equations. If $\text{rank}(\mathbf{A}) = m$ then $\mathcal{R}(\mathbf{A}) = \mathbb{R}^m$, so $\dim \mathcal{N}(\mathbf{A}) = n - m > 0$, so the set of solutions is infinite. Therefore, finding a single solution that optimizes some quantity is of interest.

$\mathbf{Ax} = \mathbf{y}$ is *square* if $n = m$. If $\mathbf{A}$ is invertible, then the equations have the unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. See section 16.2.

## 16.1 Least-Squares

For an overdetermined system we wish to find:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 \tag{179}$$

Since $\mathbf{Ax} \in \mathcal{R}(\mathbf{A})$, we need a point $\tilde{\mathbf{y}} = \mathbf{Ax}^* \in \mathcal{R}(\mathbf{A})$ closest to $\mathbf{y}$. This point lies in the nullspace of $\mathbf{A}^T$, so we have $\mathbf{A}^T(\mathbf{y} - \mathbf{Ax}^*) = 0$. There is always a solution to this problem and, if $\text{rank}(\mathbf{A}) = n$, it is unique [16, p. 161]

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \tag{180}$$

### 16.1.1 Regularized least-squares with low-rank data

For $\mathbf{A} \in \mathbb{R}^{m,n}$, $\mathbf{y} \in \mathbb{R}^m$, $\lambda \geq 0$, the regularized least-squares problem

$$\text{argmin}_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \tag{181}$$

has a closed form solution

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y} \tag{182}$$

However, if $\mathbf{A}$ has a rank $r \ll \min(n, m)$ and a known low-rank decomposition $\mathbf{A} = \mathbf{LR}^T$ with $\mathbf{L} \in \mathbb{R}^{m,r}$ and $\mathbf{R} \in \mathbb{R}^{n,r}$, then we can rewrite Equation 182 as

$$\mathbf{x} = (\mathbf{R}^T\mathbf{R}\mathbf{L}^T\mathbf{L} + \lambda\mathbf{I})^{-1}\mathbf{L}^T\mathbf{y} \tag{183}$$

This decreases the time complexity from $O(mn^2 + n^\omega)$ to $O(nr^2 + mr^2)$.

## 16.2 Minimum Norm Solutions

For undertermined systems in which $\mathbf{A} \in \mathbb{R}^{m,n}$ with $m < n$. We wish to find

$$\min_{\mathbf{x}:\mathbf{Ax}=\mathbf{y}} \|\mathbf{x}\|_2 \tag{184}$$

The solution $\mathbf{x}^*$ must be orthogonal to $\mathcal{N}(\mathbf{A})$, so $\mathbf{x}^* \in \mathcal{R}(\mathbf{A}^T)$, so $\mathbf{x}^* = \mathbf{A}^T c$ for some $c$. Substituting into $\mathbf{Ax} = \mathbf{y}$ gives $\mathbf{AA}^T c = \mathbf{y}$, therefore [16, p. 162]:

$$\mathbf{x}^* = \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{y} \tag{185}$$

# 17 Updates

## 17.1 Removing a row from $\mathbf{A}^T\mathbf{A}$ ($\mathbf{A}^T\mathbf{A} \rightarrow \mathbf{A}_{\setminus i}^T\mathbf{A}_{\setminus i}$)

**Plain English:** Matrix times its transpose after eliminating row $i$ from the matrix

**Inputs:** $\mathbf{A} \in \mathbb{R}^{k,m}, \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$ and $i$, the row to remove from $\mathbf{A}$

**Reduces to:** $\mathbf{C} \in \mathbb{R}^{k,l}$

**Algorithm:**

Recall that

$$(\mathbf{AB})_{kl} = \sum_m \mathbf{A}_{km}\mathbf{B}_{ml} \quad \mathbf{A} \in \mathbb{R}^{k,m}, \mathbf{B} \in \mathbb{R}^{m,l} \tag{186}$$

then we have that

$$(\mathbf{A}^T\mathbf{A})_{kl} = \sum_m \mathbf{A}_{mk}\mathbf{A}_{ml} = \sum_{m \neq i} \mathbf{A}_{mk}\mathbf{A}_{ml} + \mathbf{A}_{jk}\mathbf{A}_{jl} = \sum_{m \neq i} \mathbf{A}_{mk}\mathbf{A}_{ml} + (\mathbf{A}_{k*})_j(\mathbf{A}_{l*})_j \tag{187}$$

where $(\mathbf{A}_k*)_j$ is the $j$th element of the $k$th column of $\mathbf{A}$.

Thus,

$$\mathbf{A}_{\setminus i}^T\mathbf{A}_{\setminus i} = \mathbf{A}^T\mathbf{A} - \mathbf{A}_{*j}\mathbf{A}_{*j}^T \tag{188}$$

## 17.2 $\mathbf{1}_r^T\mathbf{A}\mathbf{1}_c$

**Plain English:** The sum of the elements of the matrix.

**Reduces to:** Scalar

**Notation:** For $\mathbf{A} \in \mathbb{R}^{r \times c}$, $\mathbf{1}_r$ is in $\mathbb{R}^{r \times 1}$ and $\mathbf{1}_c$ is in $\mathbb{R}^{c \times 1}$.

**Algorithm:** Traverse all the element of the matrix in order keeping track of the sum. For applications where accuracy is important and the matrices have a large dynamic range, Kahan summation or a similar technique should be used.

**Update Algorithm:** If an entry changes, subtract its old value from the sum and add its new value to the sum.

## 17.3 $\mathbf{x}^T\mathbf{A}\mathbf{x}$

**Plain English:** TODO

**Reduces to**: Scalar

**Notation:** $\mathbf{A}$ must be in $\mathbb{R}^{i \times i}$. $\mathbf{x}$ is in $\mathbb{R}^{i \times 1}$.

**Algorithm:** TODO

**Update Algorithm:** We make use of the identity $(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \sum_{i,j}\left((\mathbf{x}\mathbf{x}^T)\circ\mathbf{A}\right)$. If an entry $\mathbf{A}_{i,j}$ in the matrix changes subtract its old value $\mathbf{x}_i\mathbf{x}_j\mathbf{A}_{ij}$ and add the new value $\mathbf{x}_i\mathbf{x}_j\mathbf{A}'_{ij}$. If an entry $\mathbf{x}_i$ changes TODO.

# 18 Optimization

## 18.1 Standard Forms

**Least Squares**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \tag{189}$$

**LASSO**

$$\min_{\mathbf{b}\in\mathbb{R}^n} \left(\frac{1}{N}\|\mathbf{y}-\mathbf{X}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1\right) \tag{190}$$

**LP: Linear program**

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^T\mathbf{x} && (191)\\ \text{s.t. } & \mathbf{A}_{\text{eq}}\mathbf{x} = \mathbf{b}_{\text{eq}} && (192)\\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} && (193) \end{aligned}$$

**Linear Fractional Program**

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \frac{\mathbf{c}^T\mathbf{x}+a}{\mathbf{d}^T\mathbf{x}+b} && (194a)\\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} && (194b) \end{aligned}$$

Additional constraints must ensure $\mathbf{d}^T\mathbf{x}+b$ has the same sign throughout the entire feasible region.

$$\begin{aligned} \max_{\mathbf{x}} \ & \frac{\mathbf{c}^T\mathbf{x}+a}{\mathbf{d}^T\mathbf{x}+b} && (195)\\ \text{s.t.} & && (196) \end{aligned}$$

**QCQP: Quadratic Constrainted Quadratic Programs**

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{x}^T\mathbf{H}_0\mathbf{x} + 2\mathbf{c}_0^T\mathbf{x} + \mathbf{d}_0 && (197)\\ \text{s.t. } & \mathbf{x}^T\mathbf{H}_i\mathbf{x} + 2\mathbf{c}_i^T\mathbf{x} + \mathbf{d}_i \leq 0, \quad i \in \mathcal{I} && (198)\\ & \mathbf{x}^T\mathbf{H}_j\mathbf{x} + 2\mathbf{c}_j^T\mathbf{x} + \mathbf{d}_j = 0, \quad j \in \mathcal{E} && (199) \end{aligned}$$

Note that, in general, QCQPs are NP-Hard.

**QP: Quadratic Program**

$$\begin{aligned} \min_{\mathbf{x}} \ & \frac{1}{2}\mathbf{x}^T\mathbf{H}_0\mathbf{x} + \mathbf{c}_0^T\mathbf{x} && (200)\\ \text{s.t. } & \mathbf{A}_{\text{eq}}\mathbf{x} = \mathbf{b}_{\text{eq}} && (201)\\ & \mathbf{A}\mathbf{x} \leq \mathbf{b} && (202) \end{aligned}$$

**SOCP: Second Order Cone Program (Standard Form)**

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \tag{203}$$

$$\text{s.t. } \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T\mathbf{x} + \mathbf{d}_i, \quad i = 1, \ldots, m \tag{204}$$

**SOCP: Second Order Cone Program (Conic Standard Form)**

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \tag{205}$$

$$\text{s.t. } (\mathbf{A}_i\mathbf{x} + \mathbf{b}_i, \mathbf{c}_i^T\mathbf{x} + \mathbf{d}_i) \in \mathcal{K}_{m_i} \quad i = 1, \ldots, m \tag{206}$$

## 18.2 Transformations

### 18.2.1 Linear-Fractional to Linear

We transform a Linear-Fractional Program

$$\underset{\mathbf{x}}{\text{maximize}} \quad \frac{\mathbf{c}^T\mathbf{x} + a}{\mathbf{d}^T\mathbf{x} + b} \tag{207a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{207b}$$

where $\mathbf{d}^T\mathbf{x} + b$ has the same sign throughout the entire feasible region to a linear program using the Charnes–Cooper transformation [17] by defining

$$\mathbf{y} = \frac{1}{\mathbf{d}^T\mathbf{x} + b} \cdot \mathbf{x} \tag{208}$$

$$t = \frac{1}{\mathbf{d}^T\mathbf{x} + b} \tag{209}$$

to form the equivalent program

$$\underset{\mathbf{y}, t}{\text{maximize}} \quad \mathbf{c}^T\mathbf{y} + at \tag{210a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{y} \leq \mathbf{b}t, \tag{210b}$$

$$\mathbf{d}^T\mathbf{y} + bt = 1, \tag{210c}$$

$$t \geq 0 \tag{210d}$$

We then have $\mathbf{x}^* = \frac{1}{t}\mathbf{y}$.

# 19 | Algorithms

## 19.1 Gram-Schmidt

TODO

# Bibliography


[1] Volker Strassen. Gaussian elimination is not optimal. *Numerische mathematik*, 13(4):354–356, 1969.

[2] Günter Rote. Division-free algorithms for the determinant and the pfaffian: algebraic and combinatorial approaches. In *Computational discrete mathematics*, pages 119–135. Springer, 2001.

[3] Erwin H. Bareiss. Sylvester's identity and multistep integer-preserving gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968. ISSN 00255718, 10886842. doi: 10.2307/2004533. URL `http://www.jstor.org/stable/2004533`.

[4] V Ya Pan. Strassen's algorithm is not optimal trilinear technique of aggregating, uniting and canceling for constructing fast algorithms for matrix operations. In *Foundations of Computer Science, 1978., 19th Annual Symposium on*, pages 166–176. IEEE, 1978. doi: 10.1109/SFCS.1978.34.

[5] DARIO ANDREA Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. $o(n^{2.7799})$ complexity for n by n approximate matrix multiplication. *Information processing letters*, 8(5): 234–235, 1979.

[6] Arnold Schönhage. Partial and total matrix multiplication. *SIAM Journal on Computing*, 10 (3):434–455, 1981.

[7] Francesco Romani. Some properties of disjoint sums of tensors related to matrix multiplication. *SIAM Journal on Computing*, 11(2):263–267, 1982.

[8] Don Coppersmith and Shmuel Winograd. On the asymptotic complexity of matrix multiplication. *SIAM Journal on Computing*, 11(3):472–492, 1982.

[9] Volker Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication. In *Foundations of Computer Science, 1986., 27th Annual Symposium on*, pages 49–54. IEEE, 1986.

[10] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 9(3):251 – 280, 1990. ISSN 0747-7171. doi: 10.1016/S0747-7171(08)80013-2. URL `http://www.sciencedirect.com/science/article/pii/S0747717108800132`. Computational algebraic complexity editorial.

[11] Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 887–898, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1245-5. doi: 10.1145/2213977.2214056. URL `http://doi.acm.org/10.1145/2213977.2214056`.

[12] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, ISSAC '14, pages 296–303, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2501-1. doi: 10.1145/2608628.2608664. URL `http://doi.acm.org/10.1145/2608628.2608664`.

[13] Gilbert Strang. *Introduction to Linear Algebra*. 2016.

[14] Elizabeth Million. The hadamard product. `http://buzzard.ups.edu/courses/2007spring/projects/million-paper.pdf`, 2007.

[15] Thomas P Minka. Old and new matrix algebra useful for statistics. https://tminka.github.io/papers/matrix/minka-matrix.pdf, 2000.

[16] Giuseppe C Calafiore and Laurent El Ghaoui. *Optimization models*. Cambridge University Press, 2014. ISBN 978-1-107-05087-7.

[17] A. Charnes and W. W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186, 1962. doi: 10.1002/nav.3800090303. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800090303.

# Index