

Matrix Forensics

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

RICHARD BARNES

Git Hash: db53fec66958ad54b3d27c37c14753e9cb3fdc55
Compiled on: 2018/09/25 at 18:18:24

github.com/r-barnes/MatrixForensics

Contents

1	Introduction	4
2	Nomenclature	5
3	Derivatives	6
3.1	Useful Rules for Derivatives	6
4	Matrix Rogue Gallery	7
4.1	Non-Singular vs. Singular Matrices	7
4.2	Diagonal Matrix	7
4.3	Dyads	8
4.4	Hermitian Matrix	9
4.5	Idempotent Matrix	9
4.6	Orthogonal Matrix	10
4.7	Permutation Matrix	11
4.8	Positive Definite	12
4.9	Positive Semi-Definite	12
4.10	Projection Matrix	13
4.11	Singular Matrix	13
4.12	Symmetric Matrix	13
4.13	Skew-Hermitian	14
4.14	Toeplitz Matrix, General Form	14
4.15	Toeplitz Matrix, Discrete Convolution	15
4.16	Triangular Matrix	16
4.17	Vandermonde Matrix	17
5	Matrix Decompositions	18
5.1	LLT/UTU: Cholesky Decomposition	18
5.2	LDLT	18
5.3	PCA: Principle Components Analysis	18
5.4	QR: Orthogonal-triangular	19
5.5	SVD: Singular Value Decomposition	19
5.6	Eigenvalue Decomposition for Diagonalizable Matrices	22
5.7	Eigenvalue (Spectral) Decomposition for Symmetric Matrices	22
5.8	Schur Complements	22
6	Matrix Properties	24
7	Transpose Properties	25

<i>CONTENTS</i>	3
8 Determinant Properties	26
9 Trace Properties	27
10 Inverse Properties	28
11 Pseudo-Inverse Properties	29
11.1 Moore-Penrose Pseudoinverse	29
12 Hadamard Identities	30
13 Eigenvalue Properties	31
14 Norms	32
14.1 Matrices	32
14.1.1 Frobenius norm	32
14.1.2 Operator Norms	32
14.1.3 Spectral Radius	33
14.2 Vectors	34
15 Bounds	35
15.1 Matrix Gain	35
15.2 Norms	35
15.3 Rayleigh quotients	35
16 Linear Equations	37
16.1 Least-Squares	37
16.2 Minimum Norm Solutions	37
17 $\mathbf{1}_r^T \mathbf{A} \mathbf{1}_c$	39
18 $\mathbf{x}^T \mathbf{A} \mathbf{x}$	40
19 Algorithms	41
19.1 Gram-Schmidt	41

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 \mathbf{A} and \mathbf{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} .
$\text{eig}(\mathbf{A})$	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 | Derivatives

3.1 Useful Rules for Derivatives

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

$$\partial \mathbf{A} = 0 \quad \text{where } \mathbf{A} \text{ is a constant} \quad (1)$$

$$\partial(c\mathbf{X}) = c\partial\mathbf{X} \quad (2)$$

$$\partial(\mathbf{X} + \mathbf{Y}) = \partial\mathbf{X} + \partial\mathbf{Y} \quad (3)$$

$$\partial(\text{tr}(\mathbf{X})) = \text{tr}(\partial(\mathbf{X})) \quad (4)$$

$$\partial(\mathbf{X}\mathbf{Y}) = (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y}) \quad (5)$$

$$\partial(\mathbf{X} \circ \mathbf{Y}) = (\partial\mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial\mathbf{Y}) \quad (6)$$

$$\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} \quad (7)$$

$$\partial(\det(\mathbf{X})) = \text{tr}(\text{adj}(\mathbf{X})\partial\mathbf{X}) \quad (8)$$

$$\partial(\det(\mathbf{X})) = \det(\mathbf{X}) \text{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \quad (9)$$

$$\partial(\ln(\det(\mathbf{X}))) = \text{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \quad (10)$$

$$\partial(\mathbf{X}^T) = (\partial\mathbf{X})^T \quad (11)$$

$$\partial(\mathbf{X}^H) = (\partial\mathbf{X})^H \quad (12)$$

4 | Matrix Rogue Gallery

4.1 Non-Singular vs. Singular Matrices

For $\mathbf{A} \in \mathbb{R}^{n,n}$ (initially drawn from [1, p. 574]):

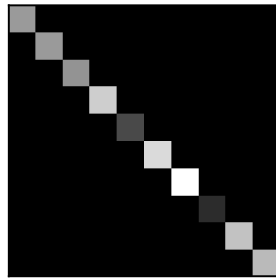
Non-Singular

\mathbf{A} is invertible
 The columns are independent
 The rows are independent
 $\det(\mathbf{A}) \neq 0$
 $\mathbf{Ax} = 0$ has one solution: $\mathbf{x} = 0$
 $\mathbf{Ax} = \mathbf{b}$ has one solution: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
 \mathbf{A} has n nonzero pivots
 \mathbf{A} has full rank $r = n$
 The reduced row echelon form is $\mathbf{R} = \mathbf{I}$
 The column space is all of \mathbb{R}^n
 The row space is all of \mathbb{R}^n
 All eigenvalues are nonzero
 $\mathbf{A}^T\mathbf{A}$ is symmetric positive definite
 \mathbf{A} has n positive singular values

Singular

\mathbf{A} is not invertible
 The columns are dependent
 The rows are dependent
 $\det(\mathbf{A}) = 0$
 $\mathbf{Ax} = 0$ has infinitely many solutions
 $\mathbf{Ax} = \mathbf{b}$ has either no or infinitely many solutions
 \mathbf{A} has $r < n$ pivots
 \mathbf{A} has rank $r < n$
 \mathbf{R} has at least one zero row
 The column space has dimension $r < n$
 The row space has dimension $r < n$
 Zero is an eigenvalue of \mathbf{A}
 $\mathbf{A}^T\mathbf{A}$ is only semidefinite
 \mathbf{A} has $r < n$ singular values

4.2 Diagonal Matrix



$$\mathbf{A} = \text{diag}(a_1, \dots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} \quad (13)$$

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

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

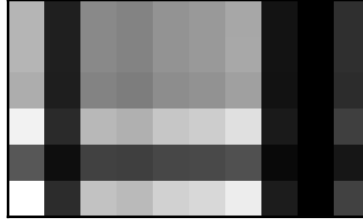
Special Properties

$$\text{eig}(A) = a_1, \dots, a_n \quad (14)$$

$$\det(A) = \prod_i a_i \quad (15)$$

$$A^{-1} = \begin{bmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{bmatrix} \quad (16)$$

4.3 Dyads



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

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T \quad \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n \quad (17)$$

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 $\mathbf{A}\mathbf{x} = (\mathbf{u}\mathbf{v}^T)\mathbf{x} = (\mathbf{v}^T\mathbf{u})\mathbf{x}$.
- $\mathbf{A}\mathbf{x} = c\mathbf{u}$ (\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 $\text{rank}(\mathbf{A}) = 1$.
- 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 $c\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$.

4.4 Hermitian Matrix

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

$$\mathbf{H} = \mathbf{H}^H \quad (18)$$

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

$$\mathbf{H}_{ii} \in \mathbb{R} \quad (19)$$

$$\mathbf{H}\mathbf{H}^H = \mathbf{H}^H\mathbf{H} \quad (20)$$

$$\mathbf{x}^H\mathbf{H}\mathbf{x} \in \mathbb{R} \quad \forall \mathbf{x} \in \mathbb{C} \quad (21)$$

$$\mathbf{H}_1 + \mathbf{H}_2 = \text{Hermitian} \quad (22)$$

$$\mathbf{H}^{-1} = \text{Hermitian} \quad (23)$$

$$\mathbf{A} + \mathbf{A}^H = \text{Hermitian} \quad (24)$$

$$\mathbf{A} - \mathbf{A}^H = \text{Skew-Hermitian} \quad (25)$$

$$\mathbf{A}\mathbf{B} = \text{Hermitian iff } \mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} \quad (26)$$

$$\det(\mathbf{H}) \in \mathbb{R} \quad (27)$$

$$\text{eig}(\mathbf{H}) \in \mathbb{R} \quad (28)$$

4.5 Idempotent Matrix

A matrix \mathbf{A} is idempotent iff

$$\mathbf{A}\mathbf{A} = \mathbf{A} \quad (29)$$

Special Properties

$$\mathbf{A}^n = \mathbf{A} \quad \forall n \quad (30)$$

$$\mathbf{I} - \mathbf{A} \text{ is idempotent} \quad (31)$$

$$\mathbf{A}^H \text{ is idempotent} \quad (32)$$

$$\mathbf{I} - \mathbf{A}^H \text{ is idempotent} \quad (33)$$

$$\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) \quad (34)$$

$$\mathbf{A}(\mathbf{I} - \mathbf{A}) = 0 \quad (35)$$

$$\mathbf{A}^+ = \mathbf{A} \quad (36)$$

$$f(s\mathbf{I} + t\mathbf{A}) = (\mathbf{I} - \mathbf{A})f(s) + \mathbf{A}f(s + t) \quad (37)$$

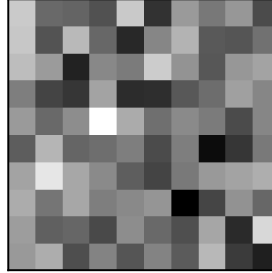
$$\mathbf{AB} = \mathbf{BA} \implies \mathbf{AB} \text{ is idempotent} \quad (38)$$

$$\text{eig}(\mathbf{A})_i \in \{0, 1\} \quad (39)$$

$$\mathbf{A} \text{ is always diagonalizable} \quad (40)$$

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

4.6 Orthogonal Matrix



(Not much visible structure)

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (41)$$

A matrix \mathbf{U} is orthogonal iff:

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I} \quad (42)$$

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.

$$\mathbf{U}^T = \mathbf{U}^{-1} \quad (43)$$

$$\mathbf{U}^{-T} = \mathbf{U} \quad (44)$$

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (45)$$

$$\mathbf{U} \mathbf{U}^T = \mathbf{I} \quad (46)$$

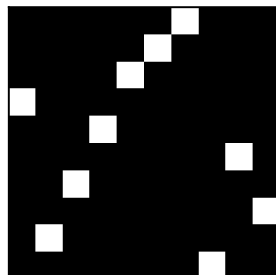
$$\det(\mathbf{U}) = \pm 1 \quad (47)$$

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.

$$\|\mathbf{U}\mathbf{x}\|_2^2 = (\mathbf{U}\mathbf{x})^T (\mathbf{U}\mathbf{x}) = \mathbf{x}^T \mathbf{U}^T \mathbf{U} \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \quad (48)$$

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\|_F = \|\mathbf{A}\|_F \quad \forall \mathbf{A}, \mathbf{U}, \mathbf{V} \text{ with } \mathbf{U}, \mathbf{V} \text{ orthogonal} \quad (49)$$

4.7 Permutation Matrix



TODO

4.8 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 $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} \leq 1\}$. The eigenvalues λ_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 $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{A} \mathbf{x}\|_2 \leq 1\}$.

4.9 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 $\alpha \in [0, 1]$:

$$\mathbf{x}^T (\alpha \mathbf{A} + (1 - \alpha) \mathbf{B}) \mathbf{x} = \alpha \mathbf{x}^T \mathbf{A} \mathbf{x} + (1 - \alpha) \mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0 \quad \forall \mathbf{x} \quad (50)$$

$$\alpha \mathbf{A} + (1 - \alpha) \mathbf{B} \in \mathbb{S}_+^n \quad (51)$$

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

4.10 Projection Matrix

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

$$\mathbf{P} \text{ is idempotent} \quad (52)$$

$$\mathbf{P}\mathbf{x} \in \mathcal{S} \quad \forall \mathbf{x} \quad (53)$$

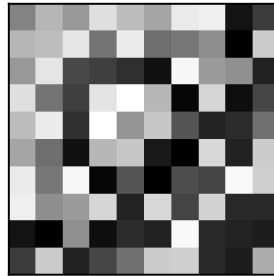
$$\mathbf{P}\mathbf{z} = \mathbf{z} \quad \forall \mathbf{z} \in \mathcal{S} \quad (54)$$

4.11 Singular Matrix

A square matrix that is not invertible.

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

4.12 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).

$$\begin{bmatrix} a & b & c & d & e & f \\ b & g & l & m & o & p \\ c & l & h & n & q & r \\ d & m & n & i & s & t \\ e & o & q & s & j & u \\ f & p & r & t & u & k \end{bmatrix} \quad (55)$$

Special Properties

$$\mathbf{A} = \mathbf{A}^T \quad (56)$$

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} .

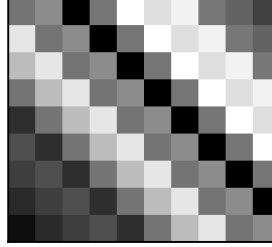
For a real, symmetric matrix with non-negative eigenvalues, the eigenvalues and singular values coincide.

4.13 Skew-Hermitian

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

$$\mathbf{H} = -\mathbf{H}^H \quad (57)$$

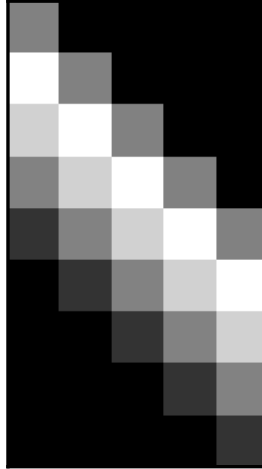
4.14 Toeplitz Matrix, General Form



Constant values on descending diagonals.

$$\begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \dots & \dots & a_2 & a_1 & a_0 \end{bmatrix} \quad (58)$$

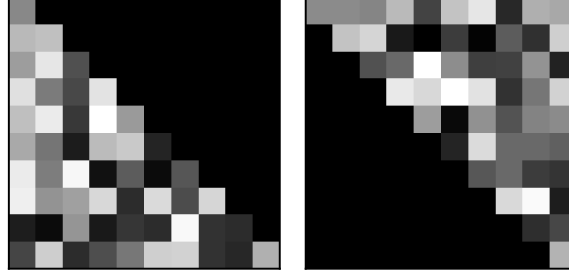
4.15 Toeplitz Matrix, Discrete Convolution



Constant values on main and subdiagonals.

$$\begin{bmatrix}
 h_m & 0 & 0 & \dots & 0 & 0 \\
 \vdots & h_m & 0 & \dots & 0 & 0 \\
 h_1 & \vdots & h_m & \dots & 0 & 0 \\
 0 & h_1 & \ddots & \ddots & 0 & 0 \\
 0 & 0 & h_1 & \ddots & h_m & 0 \\
 0 & 0 & 0 & \ddots & \vdots & h_m \\
 0 & 0 & 0 & \dots & h_1 & \vdots \\
 0 & 0 & 0 & \dots & 0 & h_1
 \end{bmatrix} \tag{59}$$

4.16 Triangular Matrix



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

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

$$\text{eig}(A) = \text{diag}(A) \quad (61)$$

$$\det(A) = \prod_i \text{diag}(A)_i \quad (62)$$

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.

4.17 Vandermonde Matrix

$$V = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \dots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \dots & \alpha_m^{n-1} \end{bmatrix} \quad (63)$$

Alternatively,

$$V_{i,j} = \alpha_i^{j-1} \quad (64)$$

Uses

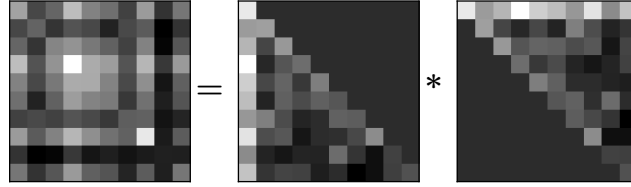
Polynomial interpolation of data.

Special Properties

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

5 | Matrix Decompositions

5.1 LLT/UTU: Cholesky Decomposition



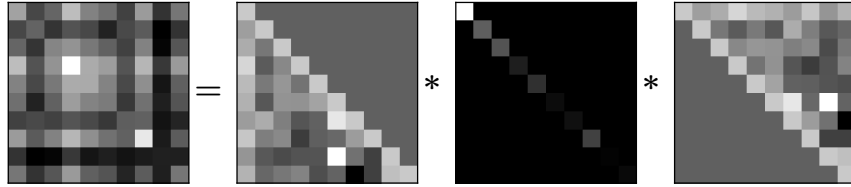
The diagram shows a square matrix A on the left, followed by an equals sign, then a lower triangular matrix L , followed by a multiplication sign $*$, and finally the transpose of L , L^T , which is an upper triangular matrix.

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

$$\mathbf{A} = \mathbf{U}^T \mathbf{U} = \mathbf{L} \mathbf{L}^T \quad (65)$$

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

5.2 LDLT



The diagram shows a square matrix A on the left, followed by an equals sign, then a lower triangular matrix L , followed by a multiplication sign $*$, then a diagonal matrix D , followed by another multiplication sign $*$, and finally the transpose of L , L^T , which is an upper triangular matrix.

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

$$\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T = \mathbf{L}^T \mathbf{D} \mathbf{L} \quad (66)$$

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

5.3 PCA: Principle Components Analysis

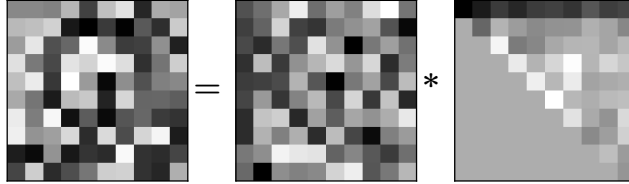
Find normalized directions in data space such that the variance of the projections of the centered data points is maximal. For centered data $\tilde{\mathbf{X}}$, the

mean-square variation of data along a vector \mathbf{x} is $\mathbf{x}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \mathbf{x}$.

$$\max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \mathbf{x}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \mathbf{x} \quad (67)$$

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.

5.4 QR: Orthogonal-triangular

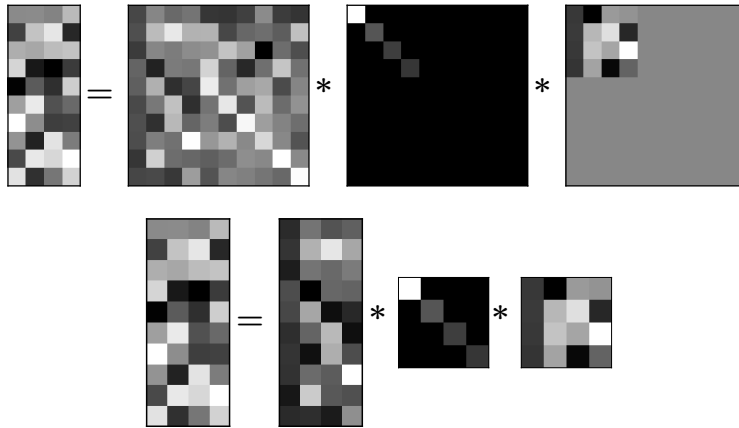


For $\mathbf{A} \in \mathbb{R}^{n,n}$, $\mathbf{A} = \mathbf{Q}\mathbf{R}$ where \mathbf{Q} is orthogonal and \mathbf{R} is an upper triangular matrix. If \mathbf{A} is non-singular, then \mathbf{Q} and \mathbf{R} are uniquely defined if $\text{diag}(\mathbf{R})$ are imposed to be positive.

Algorithms

Gram-Schmidt.

5.5 SVD: Singular Value Decomposition



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

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (68)$$

where

$$\mathbf{U} = \text{eigenvectors of } \mathbf{A}\mathbf{A}^T \quad \mathbb{R}^{m,m} \quad (69)$$

$$\mathbf{D} = \text{diag}(\sigma_i) = \sqrt{\text{diag}(\text{eig}(\mathbf{A}\mathbf{A}^T))} \quad \mathbb{R}^{n,m} \quad (70)$$

$$\mathbf{V} = \text{eigenvectors of } \mathbf{A}^T \mathbf{A} \quad \mathbb{R}^{n,n} \quad (71)$$

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

We also have that

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (72)$$

$$\mathbf{A}^T \mathbf{u}_i = \sigma_i \mathbf{v}_i \quad (73)$$

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (74)$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (75)$$

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

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \quad (76)$$

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} .

$$\|\mathbf{A}\|_F^2 = \text{Frobenius norm} = \text{tr}(\mathbf{A}^T \mathbf{A}) = \sum_{i=1}^r \sigma_i^2 \quad (77)$$

$$\|\mathbf{A}\|_2^2 = \sigma_1^2 \quad (78)$$

$$\|\mathbf{A}\|_* = \text{nuclear norm} = \sum_{i=1}^r \sigma_i \quad (79)$$

The **condition number** κ 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.

$$\kappa(\mathbf{A}) = \frac{\sigma_1}{\sigma_n} = \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^{-1}\|_2 \quad (80)$$

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 problem

$$\min_{\mathbf{A}_k \in \mathbb{R}^{m,n}} \|\mathbf{A} - \mathbf{A}_k\|_F^2 : \text{rank } \mathbf{A}_k = k, 1 \leq k \leq \text{rank}(\mathbf{A}) \quad (81)$$

The optimal solution of this problem is given by

$$\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (82)$$

where

$$\frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_r^2} \quad (83)$$

$$1 - \frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_{k+1}^2 + \dots + \sigma_r^2}{\sigma_1^2 + \dots + \sigma_r^2} \quad (84)$$

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

Range and Nullspace

$$\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{V}_{nr}) \quad (85)$$

$$\mathcal{N}(\mathbf{A})^\perp \equiv \mathcal{R}(\mathbf{A}^T) = \mathcal{R}(\mathbf{V}_r) \quad (86)$$

$$\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U}_r) \quad (87)$$

$$\mathcal{R}(\mathbf{A})^\perp \equiv \mathcal{N}(\mathbf{A}^T) = \mathcal{R}(\mathbf{U}_{nr}) \quad (88)$$

where \mathbf{V}_r is the first r columns of \mathbf{V} and \mathbf{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 $\epsilon \geq 0$.

5.6 Eigenvalue Decomposition for Diagonalizable Matrices

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

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1} \quad (89)$$

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

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

$$\mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad i = 1, \dots, n \quad (90)$$

5.7 Eigenvalue (Spectral) Decomposition for Symmetric Matrices

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

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T = \sum_i^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T \quad (91)$$

where $\mathbf{U} \in \mathbb{R}^{n,n}$ is an orthogonal matrix whose columns \mathbf{u}_i are the eigenvectors of \mathbf{A} and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$ 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 $\mathbf{u}_1, \dots, \mathbf{u}_n$ satisfy

$$\mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad i = 1, \dots, n \quad (92)$$

5.8 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

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{B} \end{bmatrix} \quad (93)$$

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

$$S = \mathbf{A} - \mathbf{XB}^{-1}\mathbf{X}^T \quad (94)$$

Then

$$\mathbf{M} \succeq 0 \iff S \succeq 0 \quad (95)$$

$$\mathbf{M} \succ 0 \iff S \succ 0 \quad (96)$$

6 | Matrix Properties

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad (\text{left distributivity}) \quad (97)$$

$$(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA} \quad (\text{right distributivity}) \quad (98)$$

$$\mathbf{AB} \neq \mathbf{BA} \quad (\text{in general}) \quad (99)$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \quad (\text{associativity}) \quad (100)$$

7 | Transpose Properties

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \quad (101)$$

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \quad (102)$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \quad (103)$$

8 | Determinant Properties

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

$$\det(I_n) = 1 \quad (104)$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A}) \quad (105)$$

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})} = \det(\mathbf{A})^{-1} \quad (106)$$

$$\det(AB) = \det(BA) \quad (107)$$

$$\det(AB) = \det(A) \det(B) \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n,n} \quad (108)$$

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}) \quad \mathbf{A} \in \mathbb{R}^{n,n} \quad (109)$$

$$\det(\mathbf{A}) = \prod \text{eig}(\mathbf{A}) \quad (110)$$

9 | Trace Properties

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

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \mathbf{A}_{ii} \quad (111)$$

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) \quad (112)$$

$$\text{tr}(c\mathbf{A}) = c \text{tr}(\mathbf{A}) \quad (113)$$

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T) \quad (114)$$

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

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = \text{tr}(\mathbf{A} \mathbf{B}^T) = \text{tr}(\mathbf{B}^T \mathbf{A}) = \text{tr}(\mathbf{B} \mathbf{A}^T) \quad (115)$$

$$\text{tr}(\mathbf{ABCD}) = \text{tr}(\mathbf{BCDA}) = \text{tr}(\mathbf{CDAB}) = \text{tr}(\mathbf{DABC}) \quad (116)$$

(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

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \quad (117)$$

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

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (118)$$

more generally

$$(\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_n)^{-1} = \mathbf{A}_n^{-1} \dots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1} \quad (119)$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \quad (120)$$

11 | Pseudo-Inverse Properties

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

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \quad (121)$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \quad (122)$$

$$(\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+ \quad (123)$$

$$(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A} \quad (124)$$

11.1 Moore-Penrose Pseudoinverse

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T \quad (125)$$

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

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

$$\begin{aligned}
 & (\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij}B_{ij} \quad \forall i, j & (126) \\
 [2] \quad & \mathbf{A} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{A} & (127) \\
 & \mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C} & (128) \\
 [2] \quad & \mathbf{A} \circ (\mathbf{B} + \mathbf{C}) = \mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ \mathbf{C} & (129) \\
 [2] \quad & a(\mathbf{A} \circ \mathbf{B}) = (a\mathbf{A}) \circ \mathbf{B} = \mathbf{A} \circ (a\mathbf{B}) & (130) \\
 & (\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T & (131) \\
 & (\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T & (132) \\
 & (x^T \mathbf{A} x) = \sum_{i,j} ((xx^T) \circ \mathbf{A}) & (133)
 \end{aligned}$$

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 $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ and $\mathbf{u} \neq 0$. Equivalently, $(\lambda\mathbf{I}_n - \mathbf{A})\mathbf{u} = 0$ 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. λ_1 is the largest eigenvalue and λ_n the smallest.

$$\text{eig}(\mathbf{A}\mathbf{A}^T) = \text{eig}(\mathbf{A}^T\mathbf{A}) \quad (134)$$

(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.)

$$\text{eig}(\mathbf{A}^T\mathbf{A}) \geq 0 \quad (135)$$

Computation

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

14 | Norms

14.1 Matrices

Matrix norms satisfy some properties:

$$f(\mathbf{A}) \geq 0 \quad (136)$$

$$f(\mathbf{A}) = 0 \iff \mathbf{A} = 0 \quad (137)$$

$$f(c\mathbf{A}) = |c|f(\mathbf{A}) \quad (138)$$

$$f(\mathbf{A} + \mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B}) \quad (139)$$

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

14.1.1 Frobenius norm

$$\|\mathbf{A}\|_F = \sqrt{\text{tr } \mathbf{A}\mathbf{A}^H} \quad (140)$$

$$= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |\mathbf{A}_{ij}|^2} \quad (141)$$

$$= \sqrt{\sum_{i=1}^m \text{eig}(\mathbf{A}^H \mathbf{A})_i} \quad (142)$$

Special Properties

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_F \|\mathbf{x}\|_2 \quad \mathbf{x} \in \mathbb{R}^n \quad (143)$$

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F \quad (144)$$

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.

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{u}\|_p=1} \|\mathbf{Au}\|_p \quad (145)$$

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{u}\|_1=1} \|\mathbf{A}\mathbf{u}\|_1 \quad (146)$$

$$= \max_{j=1,\dots,n} \sum_{i=1}^m |\mathbf{A}_{ij}| \quad (147)$$

$$= \text{Largest absolute column sum} \quad (148)$$

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{u}\|_\infty=1} \|\mathbf{A}\mathbf{u}\|_\infty \quad (149)$$

$$= \max_{j=1,\dots,m} \sum_{i=1}^n |\mathbf{A}_{ij}| \quad (150)$$

$$= \text{Largest absolute row sum} \quad (151)$$

$$\|\mathbf{A}\|_2 = \text{“spectral norm”} \quad (152)$$

$$= \max_{\|\mathbf{u}\|_2=1} \|\mathbf{A}\mathbf{u}\|_2 \quad (153)$$

$$= \sqrt{\max(\text{eig}(\mathbf{A}^T \mathbf{A}))} \quad (154)$$

$$= \text{Square root of largest eigenvalue of } \mathbf{A}^T \mathbf{A} \quad (155)$$

Special Properties

$$\|\mathbf{A}\mathbf{u}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{u}\|_p \quad (156)$$

$$\|\mathbf{A}\mathbf{B}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p \quad (157)$$

$$(158)$$

14.1.3 Spectral Radius

Not a proper norm.

$$\rho(\mathbf{A}) = \text{spectral radius}(\mathbf{A}) = \max_{i=1,\dots,n} |\text{eig}(\mathbf{A})_i| \quad (159)$$

Special Properties

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|_p \quad (160)$$

$$\rho(\mathbf{A}) \leq \min(\|\mathbf{A}\|_1, \|\mathbf{A}\|_\infty) \quad (161)$$

$$(162)$$

14.2 Vectors

P-norm:

$$\|\mathbf{x}\|_p = \left(\sum_i |\mathbf{x}_i|^p \right)^{1/p} \quad (163)$$

15 | Bounds

15.1 Matrix Gain

$$\lambda_{\min}(\mathbf{A}^T \mathbf{A}) \leq \frac{\|\mathbf{Ax}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \lambda_{\max}(\mathbf{A}^T \mathbf{A}) \quad (164)$$

$$\max_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})} \implies \mathbf{x} = u_1 \quad (165)$$

$$\min_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \sqrt{\lambda_{\min}(\mathbf{A}^T \mathbf{A})} \implies \mathbf{x} = u_n \quad (166)$$

15.2 Norms

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

$$\frac{1}{\sqrt{n}} \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{\text{card}(\mathbf{x})} \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_2 \leq n \|\mathbf{x}\|_\infty \quad (167)$$

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

$$\frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}} \quad \mathbf{x} \neq 0 \quad (168)$$

$$\lambda_{\min}(\mathbf{A}) \leq \frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_{\max}(\mathbf{A}) \quad \mathbf{x} \neq 0 \quad (169)$$

$$\lambda_{\max}(A) = \max_{\mathbf{x} \parallel \mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{Ax} = u_1 \quad (170)$$

$$\lambda_{\min}(A) = \min_{\mathbf{x} \parallel \mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{Ax} = u_n \quad (171)$$

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

16 | Linear Equations

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

$\mathbf{Ax} = \mathbf{y}$ 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 $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

16.1 Least-Squares

For an overdetermined system we wish to find:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 \quad (172)$$

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.

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (173)$$

16.2 Minimum Norm Solutions

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

$$\min_{\mathbf{x}: \mathbf{Ax}=\mathbf{y}} \|\mathbf{x}\|_2 \quad (174)$$

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 , so $\mathbf{A}\mathbf{A}^T c = \mathbf{y}$, therefore:

$$\mathbf{x}^* = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{y} \tag{175}$$

$$17 \quad | \quad \mathbf{1}_r^T \mathbf{A} \mathbf{1}_c$$

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 $\mathbb{R}^{c \times 1}$.

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

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.

18 | $\mathbf{x}^T \mathbf{A} \mathbf{x}$

Reduces to: Scalar

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

Plain English: TODO

Algorithm: TODO

Update Algorithm: We make use of the identity $(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \sum_{i,j} ((\mathbf{x} \mathbf{x}^T) \circ \mathbf{A})$. 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.

19 | Algorithms

19.1 Gram-Schmidt

TODO

Bibliography

- [1] Gilbert Strang. *Introduction to Linear Algebra*. 2016.
- [2] Elizabeth Million. The hadamard product. <http://buzzard.ups.edu/courses/2007spring/projects/million-paper.pdf>, 2007.