

Matrix Forensics

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

RICHARD BARNES

Git Hash: bd21e6118f0a1c2020da7affec3203644ed3c2b3
Compiled on: 2018/11/06 at 09:45:47

github.com/r-barnes/MatrixForensics

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 5 |
| 2 | Nomenclature | 6 |
| 3 | Basics | 7 |
| 3.1 | Fundamental Theorem of Linear Algebra | 8 |
| 3.2 | Matrix Properties | 10 |
| 3.3 | Rank | 10 |
| 3.4 | Identities | 11 |
| 3.5 | Matrix Multiplication | 11 |
| 3.6 | Time Complexities | 11 |
| 4 | Derivatives | 13 |
| 4.1 | Useful Rules for Derivatives | 13 |
| 4.2 | Derivatives of vector norms | 13 |
| 5 | Matrix Rogue Gallery | 14 |
| 5.1 | Non-Singular vs. Singular Matrices | 14 |
| 5.2 | Diagonal Matrix | 14 |
| 5.3 | Dyads | 15 |
| 5.4 | Hermitian Matrix | 15 |
| 5.5 | Idempotent Matrix | 16 |
| 5.6 | Orthogonal Matrix | 17 |
| 5.7 | Permutation Matrix | 18 |
| 5.8 | Positive Definite | 18 |
| 5.9 | Positive Semi-Definite | 18 |
| 5.9.1 | Loewner order | 19 |
| 5.10 | Projection Matrix | 19 |
| 5.11 | Singular Matrix | 19 |
| 5.12 | Symmetric Matrix | 19 |
| 5.13 | Skew-Hermitian | 20 |
| 5.14 | Toeplitz Matrix, General Form | 20 |
| 5.15 | Toeplitz Matrix, Discrete Convolution | 21 |
| 5.16 | Triangular Matrix | 22 |
| 5.17 | Vandermonde Matrix | 22 |
| 6 | Matrix Decompositions | 24 |
| 6.1 | LLT/UTU: Cholesky Decomposition | 24 |
| 6.2 | LDLT | 24 |
| 6.3 | PCA: Principle Components Analysis | 24 |
| 6.4 | QR: Orthogonal-triangular | 25 |
| 6.5 | SVD: Singular Value Decomposition | 25 |
| 6.6 | Eigenvalue Decomposition for Diagonalizable Matrices | 27 |
| 6.7 | Eigenvalue (Spectral) Decomposition for Symmetric Matrices | 27 |
| 6.8 | Schur Complements | 28 |

| | |
|--|-----------|
| <i>CONTENTS</i> | 3 |
| 7 Transpose Properties | 29 |
| 8 Determinant Properties | 30 |
| 9 Trace Properties | 31 |
| 10 Inverse Properties | 32 |
| 11 Pseudo-Inverse Properties | 33 |
| 11.1 Moore-Penrose Pseudoinverse | 33 |
| 12 Hadamard Identities | 34 |
| 13 Eigenvalue Properties | 35 |
| 13.0.1 Weyl's Inequality | 35 |
| 14 Norms | 36 |
| 14.1 General Properties | 36 |
| 14.2 Matrices | 36 |
| 14.2.1 Frobenius norm | 36 |
| 14.2.2 Operator Norms | 36 |
| 14.2.3 Spectral Radius | 37 |
| 14.3 Vectors | 38 |
| 14.3.1 Identities | 38 |
| 14.3.2 Bounds | 38 |
| 15 Bounds | 39 |
| 15.1 Matrix Gain | 39 |
| 15.2 Rayleigh quotients | 39 |
| 16 Linear Equations | 40 |
| 16.1 Least-Squares | 40 |
| 16.1.1 Regularized least-squares with low-rank data | 40 |
| 16.2 Minimum Norm Solutions | 41 |
| 17 Updates | 42 |
| 17.1 Removing a row from $\mathbf{A}^T \mathbf{A}$ ($\mathbf{A}^T \mathbf{A} \rightarrow \mathbf{A}_{\setminus i}^T \mathbf{A}_{\setminus i}$) | 42 |
| 17.2 $\mathbf{1}_r^T \mathbf{A} \mathbf{1}_c$ | 42 |
| 17.3 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ | 42 |
| 18 Optimization | 44 |
| 18.1 Standard Forms | 44 |
| 18.2 Transformations | 45 |
| 18.2.1 Linear-Fractional to Linear | 45 |
| 18.2.2 LP as SOCP | 45 |
| 18.2.3 QCQP as SOCP | 46 |
| 18.2.4 QP as SOCP | 46 |
| 18.2.5 Sum of L2 Norms to SOCP | 46 |
| 18.2.6 Minimax of L2 Norms to SOCP | 47 |
| 18.2.7 Hyperbolic Constraints to SOCP | 47 |
| 18.2.8 Matrix Fractional to SOCP | 47 |
| 18.2.9 Fractional Objective to SOCP | 48 |
| 18.2.10 Chance-Constrained LP to SOCP | 48 |
| 18.2.11 Robust LP with Box Uncertainty as LP | 48 |

| | |
|--|-----------|
| 18.2.12 Robust LP with Ellipsoidal Uncertainty as SOCP | 49 |
| 18.3 Useful Problems | 49 |
| 19 Algorithms | 50 |
| 19.1 Gram-Schmidt | 50 |

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 | Basics

3.1 Fundamental Theorem of Linear Algebra

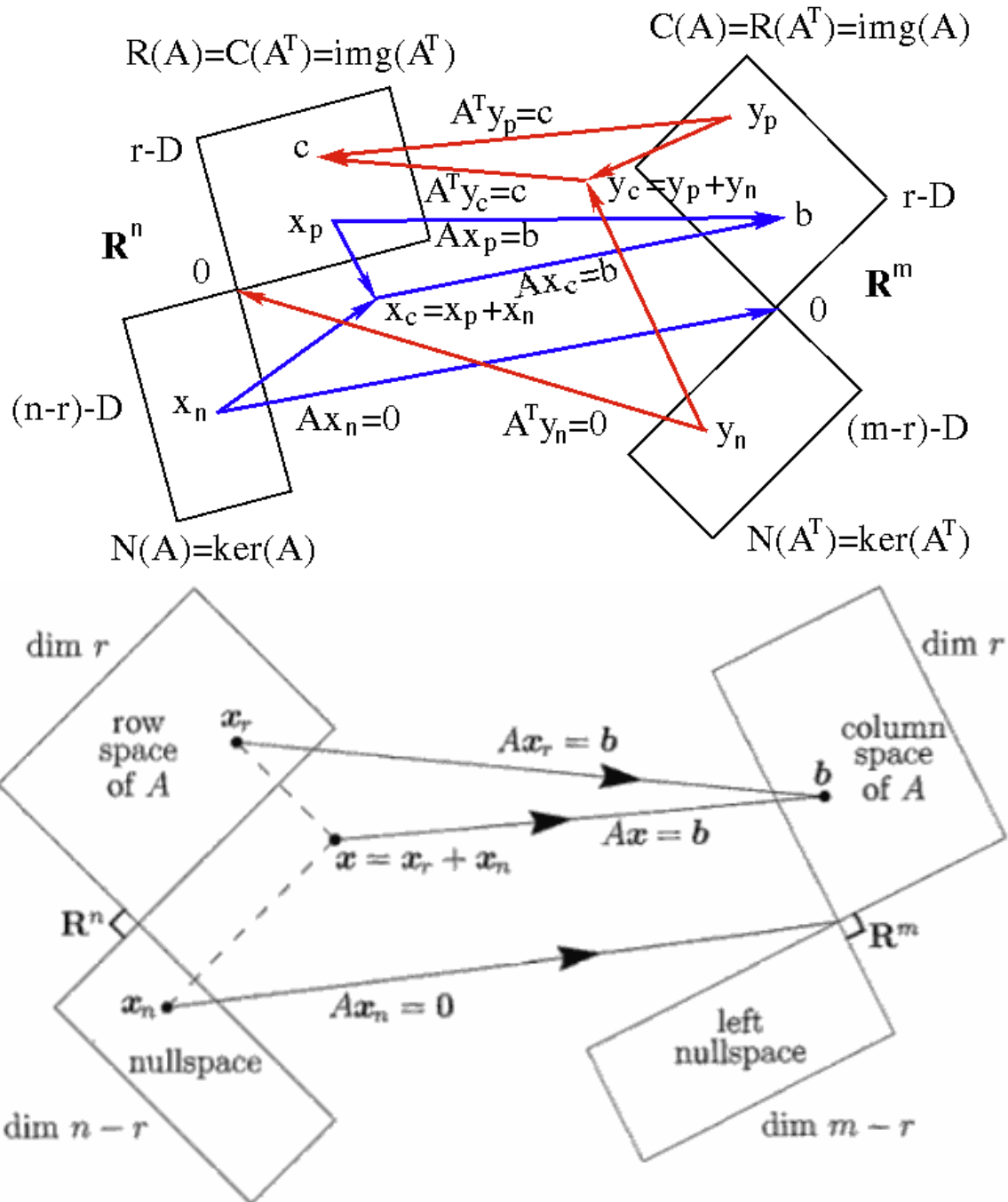
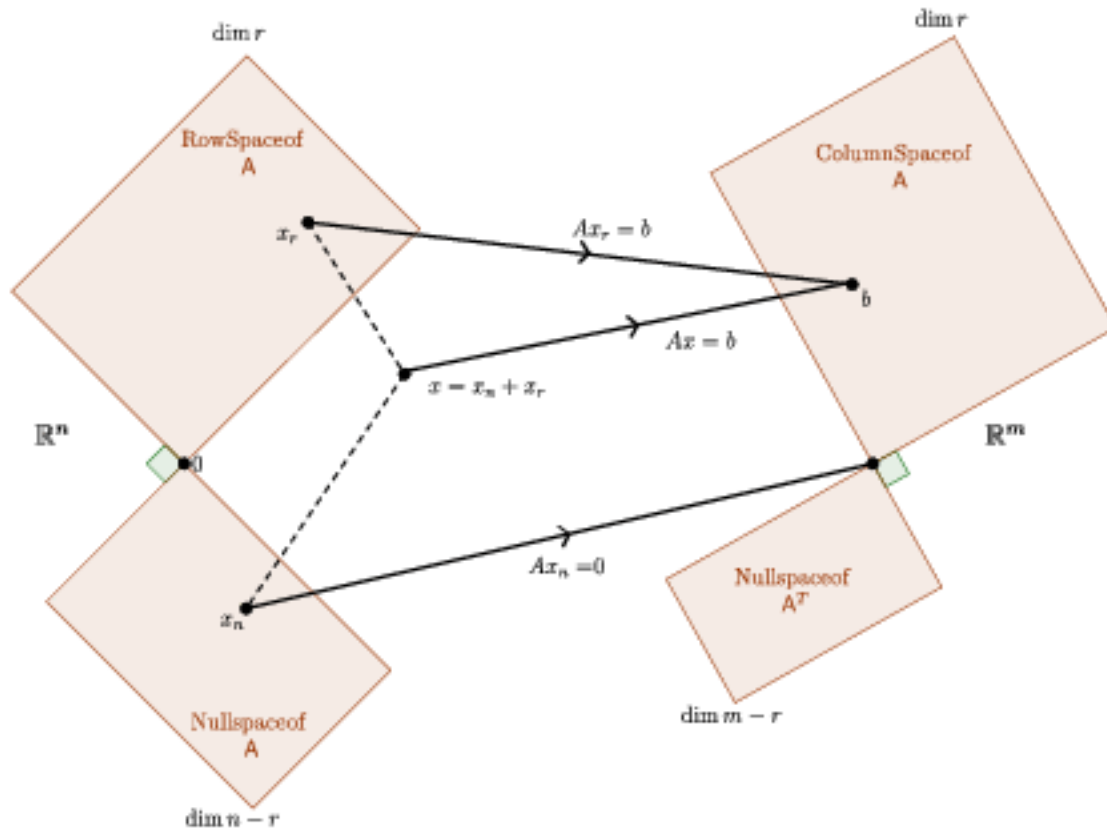
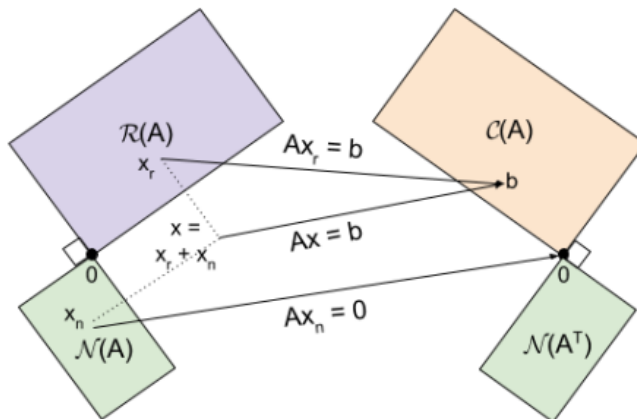


Figure 3.4 The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any m by n matrix.



Matrix A converts n -tuples into m -tuples $\mathbb{R}^n \rightarrow \mathbb{R}^m$.
That is, linear transformation T_A is a map between rows and columns



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

$\dim(\mathcal{C}) \equiv \text{rank}(A)$
 $\dim(\mathcal{N}) \equiv \text{nullity}(A)$

Theorems

$\dim(\mathcal{C}) + \dim(\mathcal{N}) = n$
 $\dim(\mathcal{R}) = \dim(\mathcal{C})$

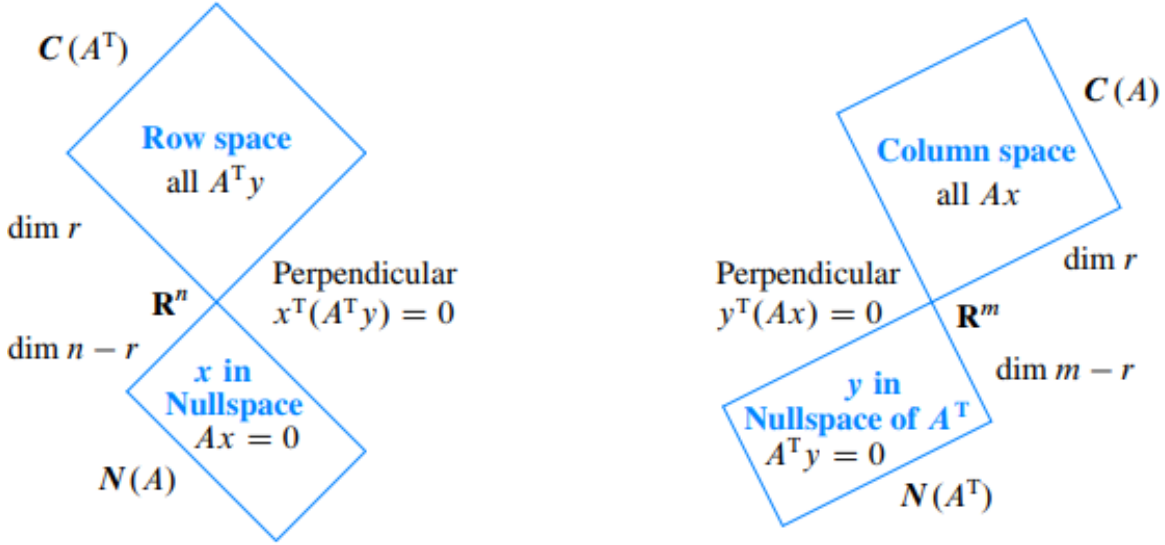


Figure 1: Dimensions and orthogonality for any m by n matrix A of rank r .

3.2 Matrix Properties

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

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

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

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

3.3 Rank

If $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{n,r}$, then

$$[1] \quad \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n \leq \text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})) \quad \text{Sylvester's Inequality} \quad (5)$$

If \mathbf{AB} , \mathbf{ABC} , \mathbf{BC} are defined, then

$$[1] \quad \text{rank}(\mathbf{AB}) + \text{rank}(\mathbf{BC}) \leq \text{rank}(\mathbf{B}) + \text{rank}(\mathbf{ABC}) \quad \text{Frobenius's inequality} \quad (6)$$

If $\dim(\mathbf{A}) = \dim(\mathbf{B})$, then

$$\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \quad \text{Subadditivity} \quad (7)$$

If $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_l$ have n_1, n_2, \dots, n_l columns, so that $\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_l$ is well-defined, then

$$[1] \quad \text{rank}(\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_l) \geq \sum_{i=1}^{l-1} \text{rank}(\mathbf{A}_i\mathbf{A}_{i+1}) - \sum_{i=2}^{l-1} \text{rank}(\mathbf{A}_i) \geq \sum_{i=1}^l \text{rank}(\mathbf{A}_i) - \sum_{i=1}^{l-1} n_i \quad (8)$$

3.4 Identities

$$\left(\sum_{i=1}^n \mathbf{z}_i\right)^2 = \mathbf{z}^T \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \mathbf{z} \quad (9)$$

3.5 Matrix Multiplication

$$(\mathbf{AB})_{kl} = \sum_m \mathbf{A}_{km} \mathbf{B}_{ml} \quad \mathbf{A} \in \mathbb{R}^{k,m}, \mathbf{B} \in \mathbb{R}^{m,l} \quad (10)$$

3.6 Time Complexities

| Operation | Input | Output | Algorithm | Time |
|-------------------|--------------------------------------|--|--------------------------|---------------------------|
| Matmult | $A, B \in n \times n$ | $n \times n$ | Schoolbook | $O(n^3)$ |
| | | | Strassen [2] | $O(n^{2.807})$ |
| | | | 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 [2] | $O(n^{2.807})$ |
| | | | Best | $O(n^\omega)$ |
| SVD | $A \in m \times n$ | $m \times m, m \times n, n \times n$ $m \times r, r \times r, n \times r$ | | $O(mn^2)$ $(m \geq n)$ |
| Determinant | $A \in n \times n$ | Scalar | Laplace expansion | $O(n!)$ |
| | | | Division-free [3] | $O(n!)$ |
| | | | LU decomposition | $O(n^3)$ |
| | | | Integer preserving [4] | $O(n^3)$ |
| Back substitution | A triangular | n solutions | Back substitution | $O(n^2)$ |

A comment on ω

The lower bound on matmult time complexity is $O(n^\omega)$, where ω is an unknown constant bounded by $2 \leq \omega \leq 2.373$. Algorithms achieving lower values of ω 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 | ω |
|-------------------------------|-------------|-----------|
| Standard | - | 3 |
| Strassen [2] | 1969 | 2.807 |
| Pan [5] | 1978 | 2.796 |
| Bini et al. [6] | 1979 | 2.78 |
| Schönhage [7] | 1981 | 2.548 |
| Schönhage [7] | 1981 | 2.522 |
| Romani [8] | 1982 | 2.517 |
| Coppersmith and Winograd [9] | 1982 | 2.496 |
| Strassen [10] | 1986 | 2.479 |
| Coppersmith and Winograd [11] | 1990 | 2.376 |
| Williams [12] | 2012 | 2.37294 |
| Le Gall [13] | 2014 | 2.3728639 |
| Williams [12] | 2012 | 2.3727 |

4 | Derivatives

4.1 Useful Rules for Derivatives

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

$$\partial \mathbf{A} = 0 \quad \text{where } \mathbf{A} \text{ is a constant} \quad (11)$$

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

$$\partial(\mathbf{X} + \mathbf{Y}) = \partial\mathbf{X} + \partial\mathbf{Y} \quad (13)$$

$$\partial(\text{tr}(\mathbf{X})) = \text{tr}(\partial\mathbf{X}) \quad (14)$$

$$\partial(\mathbf{X}\mathbf{Y}) = (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y}) \quad (15)$$

$$\partial(\mathbf{X} \circ \mathbf{Y}) = (\partial\mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial\mathbf{Y}) \quad (16)$$

$$\partial(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} \quad (17)$$

$$\partial(\det(\mathbf{X})) = \text{tr}(\text{adj}(\mathbf{X})\partial\mathbf{X}) \quad (18)$$

$$\partial(\det(\mathbf{X})) = \det(\mathbf{X}) \text{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \quad (19)$$

$$\partial(\ln(\det(\mathbf{X}))) = \text{tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \quad (20)$$

$$\partial(\mathbf{X}^T) = (\partial\mathbf{X})^T \quad (21)$$

$$\partial(\mathbf{X}^H) = (\partial\mathbf{X})^H \quad (22)$$

4.2 Derivatives of vector norms

$$\frac{\partial}{\partial \mathbf{x}} \|\mathbf{x} - \mathbf{a}\|_2 = \frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|_2} \quad (23)$$

$$\frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x} - \mathbf{a}\|_2} - \frac{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^T}{\|\mathbf{x} - \mathbf{a}\|_2^3} \quad (24)$$

$$\frac{\partial \|\mathbf{x}\|_2^2}{\partial \mathbf{x}} = \frac{\partial \|\mathbf{x}^T \mathbf{x}\|_2}{\partial \mathbf{x}} = 2\mathbf{x} \quad (25)$$

5 | Matrix Rogue Gallery

5.1 Non-Singular vs. Singular Matrices

For $\mathbf{A} \in \mathbb{R}^{n,n}$ (initially drawn from [14, p. 574]):

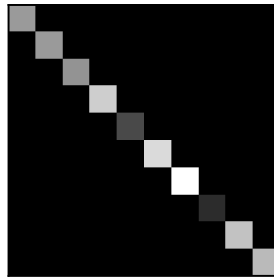
Non-Singular

\mathbf{A} is invertible
 The columns are independent
 The rows are independent
 $\det(\mathbf{A}) \neq 0$
 $\mathbf{A}\mathbf{x} = \mathbf{0}$ has one solution: $\mathbf{x} = \mathbf{0}$
 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 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
 $\det(\mathbf{A}) = 0$
 $\mathbf{A}\mathbf{x} = \mathbf{0}$ has infinitely many solutions
 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 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

5.2 Diagonal Matrix



$$A = \text{diag}(a_1, \dots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix} \quad (26)$$

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

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

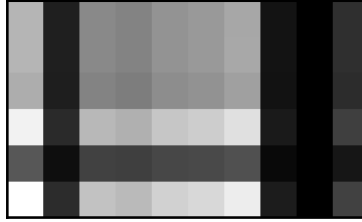
Special Properties

$$\text{eig}(A) = a_1, \dots, a_n \quad (27)$$

$$\det(A) = \prod_i a_i \quad (28)$$

$$A^{-1} = \begin{bmatrix} \frac{1}{a_1} & & \\ & \ddots & \\ & & \frac{1}{a_n} \end{bmatrix} \quad (29)$$

5.3 Dyads



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

$$\mathbf{A} = \mathbf{u}\mathbf{v}^T \quad \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n \quad (30)$$

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

5.4 Hermitian Matrix

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

$$\mathbf{H} = \mathbf{H}^H \quad (31)$$

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

$$\mathbf{H}_{ii} \in \mathbb{R} \quad (32)$$

$$\mathbf{H}\mathbf{H}^H = \mathbf{H}^H\mathbf{H} \quad (33)$$

$$\mathbf{x}^H\mathbf{H}\mathbf{x} \in \mathbb{R} \quad \forall \mathbf{x} \in \mathbb{C} \quad (34)$$

$$\mathbf{H}_1 + \mathbf{H}_2 = \text{Hermitian} \quad (35)$$

$$\mathbf{H}^{-1} = \text{Hermitian} \quad (36)$$

$$\mathbf{A} + \mathbf{A}^H = \text{Hermitian} \quad (37)$$

$$\mathbf{A} - \mathbf{A}^H = \text{Skew-Hermitian} \quad (38)$$

$$\mathbf{A}\mathbf{B} = \text{Hermitian iff } \mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} \quad (39)$$

$$\det(\mathbf{H}) \in \mathbb{R} \quad (40)$$

$$\text{eig}(\mathbf{H}) \in \mathbb{R} \quad (41)$$

5.5 Idempotent Matrix

A matrix \mathbf{A} is idempotent iff

$$\mathbf{A}\mathbf{A} = \mathbf{A} \quad (42)$$

Special Properties

$$\mathbf{A}^n = \mathbf{A} \quad \forall n \quad (43)$$

$$\mathbf{I} - \mathbf{A} \text{ is idempotent} \quad (44)$$

$$\mathbf{A}^H \text{ is idempotent} \quad (45)$$

$$\mathbf{I} - \mathbf{A}^H \text{ is idempotent} \quad (46)$$

$$\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) \quad (47)$$

$$\mathbf{A}(\mathbf{I} - \mathbf{A}) = 0 \quad (48)$$

$$\mathbf{A}^+ = \mathbf{A} \quad (49)$$

$$f(s\mathbf{I} + t\mathbf{A}) = (\mathbf{I} - \mathbf{A})f(s) + \mathbf{A}f(s + t) \quad (50)$$

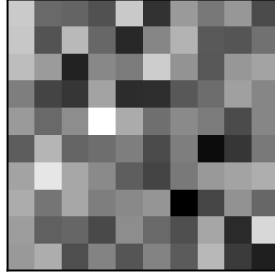
$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} \implies \mathbf{A}\mathbf{B} \text{ is idempotent} \quad (51)$$

$$\text{eig}(\mathbf{A})_i \in \{0, 1\} \quad (52)$$

$$\mathbf{A} \text{ is always diagonalizable} \quad (53)$$

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

5.6 Orthogonal Matrix



(Not much visible structure)

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (54)$$

A matrix \mathbf{U} is orthogonal iff:

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I} \quad (55)$$

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.

$$\mathbf{U}^T = \mathbf{U}^{-1} \quad (56)$$

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

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (58)$$

$$\mathbf{U} \mathbf{U}^T = \mathbf{I} \quad (59)$$

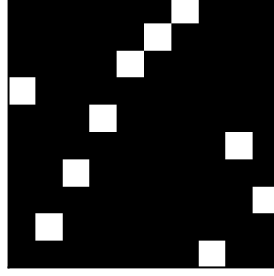
$$\det(\mathbf{U}) = \pm 1 \quad (60)$$

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.

$$\|\mathbf{U}\mathbf{x}\|_2^2 = (\mathbf{U}\mathbf{x})^T (\mathbf{U}\mathbf{x}) = \mathbf{x}^T \mathbf{U}^T \mathbf{U} \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \quad (61)$$

$$\|\mathbf{U}\mathbf{A}\mathbf{V}\|_F = \|\mathbf{A}\|_F \quad \forall \mathbf{A}, \mathbf{U}, \mathbf{V} \text{ with } \mathbf{U}, \mathbf{V} \text{ orthogonal} \quad (62)$$

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

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

$$\mathbf{x}^T (\alpha \mathbf{A} + (1 - \alpha) \mathbf{B}) \mathbf{x} = \alpha \mathbf{x}^T \mathbf{A} \mathbf{x} + (1 - \alpha) \mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0 \quad \forall \mathbf{x} \quad (63)$$

$$\alpha \mathbf{A} + (1 - \alpha) \mathbf{B} \in \mathbb{S}_+^n \quad (64)$$

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

5.9.1 Loewner order

If $\mathbf{A} - \mathbf{B} \succeq 0$, then we say $\mathbf{A} \succeq \mathbf{B}$. A sufficient condition for this is that $\lambda_n(\mathbf{A}) \geq \lambda_1(\mathbf{B})$.

5.10 Projection Matrix

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

$$\mathbf{P} \text{ is idempotent} \quad (65)$$

$$\mathbf{P} \mathbf{x} \in \mathcal{S} \quad \forall \mathbf{x} \quad (66)$$

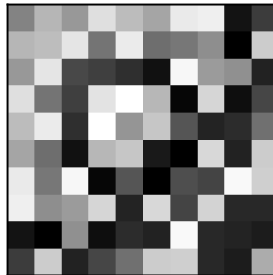
$$\mathbf{P} \mathbf{z} = \mathbf{z} \quad \forall \mathbf{z} \in \mathcal{S} \quad (67)$$

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 $\mathcal{N}(\mathbf{A}) \neq \{0\}$.

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).

$$\begin{bmatrix} a & b & c & d & e & f \\ b & g & l & m & o & p \\ c & l & h & n & q & r \\ d & m & n & i & s & t \\ e & o & q & s & j & u \\ f & p & r & t & u & k \end{bmatrix} \quad (68)$$

Special Properties

$$\mathbf{A} = \mathbf{A}^T \quad (69)$$

$$\text{eig}(A) \in \mathbb{R}^n \quad (70)$$

$$\text{Number of "free entries"} = \frac{n(n+1)}{2} \quad (71)$$

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} .

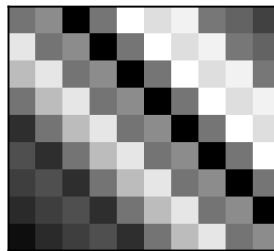
For a real, symmetric matrix with non-negative eigenvalues, the eigenvalues and singular values coincide.

5.13 Skew-Hermitian

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

$$\mathbf{H} = -\mathbf{H}^H \quad (72)$$

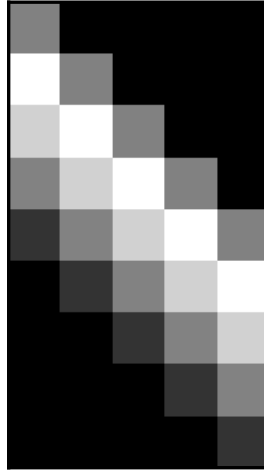
5.14 Toeplitz Matrix, General Form



Constant values on descending diagonals.

$$\begin{bmatrix}
 a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{-(n-1)} \\
 a_1 & a_0 & a_{-1} & \ddots & & \vdots \\
 a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\
 \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\
 \vdots & & \ddots & a_1 & a_0 & a_{-1} \\
 a_{n-1} & \dots & \dots & a_2 & a_1 & a_0
 \end{bmatrix} \quad (73)$$

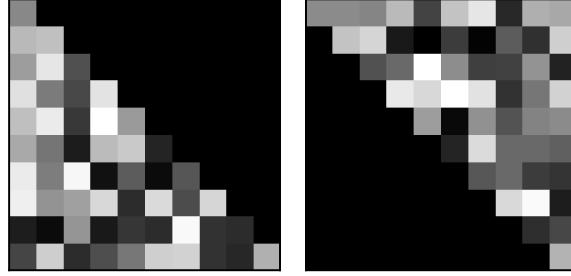
5.15 Toeplitz Matrix, Discrete Convolution



Constant values on main and subdiagonals.

$$\begin{bmatrix}
 h_m & 0 & 0 & \dots & 0 & 0 \\
 \vdots & h_m & 0 & \dots & 0 & 0 \\
 h_1 & \vdots & h_m & \dots & 0 & 0 \\
 0 & h_1 & \ddots & \ddots & 0 & 0 \\
 0 & 0 & h_1 & \ddots & h_m & 0 \\
 0 & 0 & 0 & \ddots & \vdots & h_m \\
 0 & 0 & 0 & \dots & h_1 & \vdots \\
 0 & 0 & 0 & \dots & 0 & h_1
 \end{bmatrix} \quad (74)$$

5.16 Triangular Matrix



$$\begin{bmatrix} a & b & c & d & e & f \\ & g & h & i & j & k \\ & & l & m & n & o \\ & & & p & q & r \\ & & & & s & t \\ & & & & & u \end{bmatrix} \quad \begin{bmatrix} a & & & & & \\ b & g & & & & \\ c & h & l & & & \\ d & i & m & p & & \\ e & j & n & q & s & \\ f & k & o & r & t & u \end{bmatrix} \quad (75)$$

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

$$\text{eig}(A) = \text{diag}(A) \quad (76)$$

$$\det(A) = \prod_i \text{diag}(A)_i \quad (77)$$

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

$$V = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \dots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \dots & \alpha_m^{n-1} \end{bmatrix} \quad (78)$$

Alternatively,

$$V_{i,j} = \alpha_i^{j-1} \quad (79)$$

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

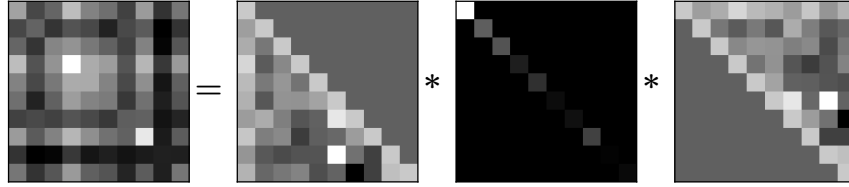


If \mathbf{A} is symmetric, positive definite, square, then

$$\mathbf{A} = \mathbf{U}^T \mathbf{U} = \mathbf{L} \mathbf{L}^T \quad (80)$$

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

$$\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T = \mathbf{L}^T \mathbf{D} \mathbf{L} \quad (81)$$

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

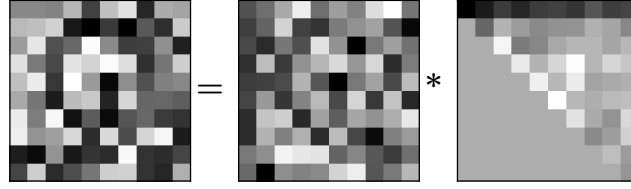
6.3 PCA: Principle Components Analysis

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 $\mathbf{x}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \mathbf{x}$.

$$\max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \mathbf{x}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \mathbf{x} \quad (82)$$

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

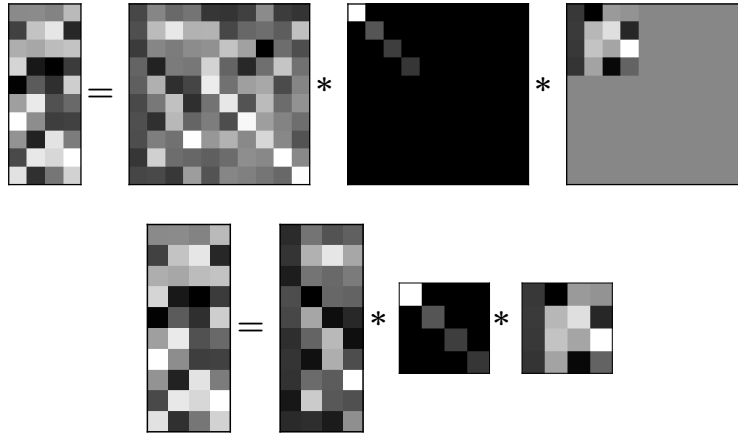


For $\mathbf{A} \in \mathbb{R}^{n,n}$, $\mathbf{A} = \mathbf{QR}$ where \mathbf{Q} is orthogonal and \mathbf{R} is an upper triangular matrix. If \mathbf{A} is non-singular, then \mathbf{Q} and \mathbf{R} are uniquely defined if $\text{diag}(\mathbf{R})$ are imposed to be positive.

Algorithms

Gram-Schmidt.

6.5 SVD: Singular Value Decomposition



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

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^r \sigma_i u_i v_i^T \quad (83)$$

where

$$U = \text{eigenvectors of } \mathbf{A}\mathbf{A}^T \quad \mathbb{R}^{m,m} \quad (84)$$

$$D = \text{diag}(\sigma_i) = \sqrt{\text{diag}(\text{eig}(\mathbf{A}\mathbf{A}^T))} \quad \mathbb{R}^{n,n} \quad (85)$$

$$V = \text{eigenvectors of } \mathbf{A}^T \mathbf{A} \quad \mathbb{R}^{n,n} \quad (86)$$

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

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (87)$$

$$\mathbf{A}^T \mathbf{u}_i = \sigma_i \mathbf{v}_i \quad (88)$$

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (89)$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (90)$$

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

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \quad (91)$$

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} .

$$\|\mathbf{A}\|_F^2 = \text{Frobenius norm} = \text{tr}(\mathbf{A}^T \mathbf{A}) = \sum_{i=1}^r \sigma_i^2 \quad (92)$$

$$\|\mathbf{A}\|_2^2 = \sigma_1^2 \quad (93)$$

$$\|\mathbf{A}\|_* = \text{nuclear norm} = \sum_{i=1}^r \sigma_i \quad (94)$$

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.

$$\kappa(\mathbf{A}) = \frac{\sigma_1}{\sigma_n} = \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^{-1}\|_2 \quad (95)$$

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

$$\min_{\mathbf{A}_k \in \mathbb{R}^{m,n}} \|\mathbf{A} - \mathbf{A}_k\|_F^2 : \text{rank } \mathbf{A}_k = k, 1 \leq k \leq \text{rank}(\mathbf{A}) \quad (96)$$

The optimal solution of this problem is given by

$$\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (97)$$

where

$$\frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_r^2} \quad (98)$$

$$1 - \frac{\|\mathbf{A}_k\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sigma_{k+1}^2 + \dots + \sigma_r^2}{\sigma_1^2 + \dots + \sigma_r^2} \quad (99)$$

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

Range and Nullspace

$$\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{V}_{nr}) \quad (100)$$

$$\mathcal{N}(\mathbf{A})^\perp \equiv \mathcal{R}(\mathbf{A}^T) = \mathcal{R}(\mathbf{V}_r) \quad (101)$$

$$\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U}_r) \quad (102)$$

$$\mathcal{R}(\mathbf{A})^\perp \equiv \mathcal{N}(\mathbf{A}^T) = \mathcal{R}(\mathbf{U}_{nr}) \quad (103)$$

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 $\epsilon \geq 0$.

6.6 Eigenvalue Decomposition for Diagonalizable Matrices

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \quad (104)$$

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

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad i = 1, \dots, n \quad (105)$$

6.7 Eigenvalue (Spectral) Decomposition for Symmetric Matrices

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \sum_i^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T \quad (106)$$

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

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad i = 1, \dots, n \quad (107)$$

6.8 Schur Complements

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

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{B} \end{bmatrix} \quad (108)$$

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

$$S = \mathbf{A} - \mathbf{X}\mathbf{B}^{-1}\mathbf{X}^T \quad (109)$$

Then

$$\mathbf{M} \succeq 0 \iff S \succeq 0 \quad (110)$$

$$\mathbf{M} \succ 0 \iff S \succ 0 \quad (111)$$

7 | Transpose Properties

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \quad (112)$$

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \quad (113)$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \quad (114)$$

8 | Determinant Properties

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

$$\det(I_n) = 1 \quad (115)$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A}) \quad (116)$$

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})} = \det(\mathbf{A})^{-1} \quad (117)$$

$$\det(AB) = \det(BA) \quad (118)$$

$$\det(AB) = \det(A) \det(B) \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n,n} \quad (119)$$

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}) \quad \mathbf{A} \in \mathbb{R}^{n,n} \quad (120)$$

$$\det(\mathbf{A}) = \prod \text{eig}(\mathbf{A}) \quad (121)$$

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

$$[15] \quad \det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA}) \quad \text{Sylvester's determinant identity} \quad (122)$$

9 | Trace Properties

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

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \mathbf{A}_{ii} \quad (123)$$

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) \quad (124)$$

$$\text{tr}(c\mathbf{A}) = c \text{tr}(\mathbf{A}) \quad (125)$$

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T) \quad (126)$$

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

$$\text{tr}(\mathbf{A}^T \mathbf{B}) = \text{tr}(\mathbf{A} \mathbf{B}^T) = \text{tr}(\mathbf{B}^T \mathbf{A}) = \text{tr}(\mathbf{B} \mathbf{A}^T) \quad (127)$$

$$\text{tr}(\mathbf{ABCD}) = \text{tr}(\mathbf{BCDA}) = \text{tr}(\mathbf{CDAB}) = \text{tr}(\mathbf{DABC}) \quad (128)$$

(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

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \quad (129)$$

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

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (130)$$

more generally

$$(\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_n)^{-1} = \mathbf{A}_n^{-1} \dots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1} \quad (131)$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} \quad (132)$$

11 | Pseudo-Inverse Properties

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

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \quad (133)$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \quad (134)$$

$$(\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+ \quad (135)$$

$$(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A} \quad (136)$$

11.1 Moore-Penrose Pseudoinverse

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T \quad (137)$$

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

$$\begin{aligned}
 & (\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij}B_{ij} \quad \forall i, j & (138) \\
 [16] \quad & \mathbf{A} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{A} & (139) \\
 & \mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C} & (140) \\
 [16] \quad & \mathbf{A} \circ (\mathbf{B} + \mathbf{C}) = \mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ \mathbf{C} & (141) \\
 [16] \quad & a(\mathbf{A} \circ \mathbf{B}) = (a\mathbf{A}) \circ \mathbf{B} = \mathbf{A} \circ (a\mathbf{B}) & (142) \\
 & (\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T & (143) \\
 & (\mathbf{A}^T \circ \mathbf{B}^T) = (\mathbf{A} \circ \mathbf{B})^T & (144) \\
 & (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \sum_{i,j} ((\mathbf{x} \mathbf{x}^T) \circ \mathbf{A}) & (145) \\
 [17] \quad & \mathbf{x}^T (\mathbf{A} \circ \mathbf{B}) \mathbf{y} = \text{tr}((\text{diag}(\mathbf{x}) \mathbf{A})^T \mathbf{B} \text{diag}(\mathbf{y})) \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m,n} & (146) \\
 & \text{tr}(\mathbf{A}^T \mathbf{B}) = \mathbf{1}^T (\mathbf{A} \circ \mathbf{B}) \mathbf{1} & (147)
 \end{aligned}$$

13 | Eigenvalue Properties

$\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n,n}$ and $\mathbf{u} \in \mathbb{C}^n$ 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.

$$\text{eig}(\mathbf{A}\mathbf{A}^T) = \text{eig}(\mathbf{A}^T\mathbf{A}) \quad (148)$$

(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.)

$$\text{eig}(\mathbf{A}^T\mathbf{A}) \geq 0 \quad (149)$$

$$\lambda_{\min}(\mathbf{A}) \leq \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \leq \lambda_{\max}(\mathbf{A}) \quad \mathbf{x} \neq 0 \quad (150)$$

13.0.1 Weyl's Inequality

If $\mathbf{M}, \mathbf{H}, \mathbf{P} \in \mathbb{R}^{n,n}$ are Hermitian matrices and $\mathbf{M} = \mathbf{H} + \mathbf{P}$ (\mathbf{H} is perturbed by \mathbf{P}) and \mathbf{M} has eigenvalues $\mu_1 \geq \dots \geq \mu_n$, \mathbf{H} has eigenvalues $\nu_1 \geq \dots \geq \nu_n$, and \mathbf{P} has eigenvalues $\rho_1 \geq \dots \geq \rho_n$, then $\nu_i + \rho_n \leq \mu_i \leq \nu_i + \rho_1 \quad \forall i$. If $j + k - n \geq i \geq r + s - 1$, then $\nu_j + \rho_k \leq \mu_i \leq \nu_r + \rho_s$. If $\mathbf{P} \succeq 0$, then $\mu_i \geq \nu_i \quad \forall i$.

Computation

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

14 | Norms

14.1 General Properties

Matrix norms satisfy some properties:

$$f(\mathbf{A}) \geq 0 \quad (151)$$

$$f(\mathbf{A}) = 0 \iff \mathbf{A} = 0 \quad (152)$$

$$f(c\mathbf{A}) = |c|f(\mathbf{A}) \quad (153)$$

$$f(\mathbf{A} + \mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B}) \quad (154)$$

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

14.2 Matrices

14.2.1 Frobenius norm

$$\|\mathbf{A}\|_F = \sqrt{\text{tr } \mathbf{A}\mathbf{A}^H} \quad (155)$$

$$= \sqrt{\sum_{i=1}^m \sum_{j=1}^n |\mathbf{A}_{ij}|^2} \quad (156)$$

$$= \sqrt{\sum_{i=1}^m \text{eig}(\mathbf{A}^H \mathbf{A})_i} \quad (157)$$

Special Properties

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_F \|\mathbf{x}\|_2 \quad \mathbf{x} \in \mathbb{R}^n \quad (158)$$

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F \quad (159)$$

$$\left\| \mathbf{C} - \mathbf{xx}^T \right\|_F^2 = \|\mathbf{C}\|_F^2 + \|\mathbf{x}\|_2^4 - 2\mathbf{x}^T \mathbf{Cx} \quad (160)$$

14.2.2 Operator Norms

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

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{u}\|_p=1} \|\mathbf{Au}\|_p \quad (161)$$

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{u}\|_1=1} \|\mathbf{A}\mathbf{u}\|_1 \quad (162)$$

$$= \max_{j=1,\dots,n} \sum_{i=1}^m |\mathbf{A}_{ij}| \quad (163)$$

$$= \text{Largest absolute column sum} \quad (164)$$

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{u}\|_\infty=1} \|\mathbf{A}\mathbf{u}\|_\infty \quad (165)$$

$$= \max_{j=1,\dots,m} \sum_{i=1}^n |\mathbf{A}_{ij}| \quad (166)$$

$$= \text{Largest absolute row sum} \quad (167)$$

$$\|\mathbf{A}\|_2 = \text{“spectral norm”} \quad (168)$$

$$= \max_{\|\mathbf{u}\|_2=1} \|\mathbf{A}\mathbf{u}\|_2 \quad (169)$$

$$= \sqrt{\max(\text{eig}(\mathbf{A}^T \mathbf{A}))} \quad (170)$$

$$= \text{Square root of largest eigenvalue of } \mathbf{A}^T \mathbf{A} \quad (171)$$

Special Properties

$$\|\mathbf{A}\mathbf{u}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{u}\|_p \quad (172)$$

$$\|\mathbf{A}\mathbf{B}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p \quad (173)$$

$$(174)$$

14.2.3 Spectral Radius

Not a proper norm.

$$\rho(\mathbf{A}) = \text{spectral radius}(\mathbf{A}) = \max_{i=1,\dots,n} |\text{eig}(\mathbf{A})_i| \quad (175)$$

Special Properties

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|_p \quad (176)$$

$$\rho(\mathbf{A}) \leq \min(\|\mathbf{A}\|_1, \|\mathbf{A}\|_\infty) \quad (177)$$

$$(178)$$

14.3 Vectors

$$\|\mathbf{x}\|_1 = \sum_i |\mathbf{x}_i| \quad \text{L1-norm} \quad (179)$$

$$\|\mathbf{x}\|_p = \left(\sum_i |\mathbf{x}_i|^p \right)^{1/p} \quad \text{P-norm} \quad (180)$$

$$\|\mathbf{x}\|_\infty = \max_i |\mathbf{x}_i| \quad \text{L}\infty\text{-norm, L-infinity norm} \quad (181)$$

14.3.1 Identities

$$2\|\mathbf{u}\|_2^2 + 2\|\mathbf{v}\|_2^2 = \|\mathbf{u} + \mathbf{v}\|_2^2 + \|\mathbf{u} - \mathbf{v}\|_2^2 \quad \text{Polarization Identity} \quad (182)$$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4} (\|\mathbf{x} + \mathbf{y}\|_2^2 - \|\mathbf{x} - \mathbf{y}\|_2^2) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{V} \quad \text{Polarization Identity} \quad (183)$$

14.3.2 Bounds

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \quad \text{Cauchy-Schwartz Inequality} \quad (184)$$

$$|\mathbf{x}^T \mathbf{y}| \leq \sum_{k=1}^n |\mathbf{x}_k \mathbf{y}_k| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q \quad \forall p, q \geq 1 : 1/p + 1/q = 1 \quad \text{Hölder Inequality} \quad (185)$$

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

$$\frac{1}{\sqrt{n}} \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{\text{card}(\mathbf{x})} \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_2 \leq n \|\mathbf{x}\|_\infty \quad (186)$$

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

15 | Bounds

15.1 Matrix Gain

$$\lambda_{\min}(\mathbf{A}^T \mathbf{A}) \leq \frac{\|\mathbf{Ax}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \lambda_{\max}(\mathbf{A}^T \mathbf{A}) \quad (187)$$

$$\max_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})} \implies \mathbf{x} = u_1 \quad (188)$$

$$\min_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \sqrt{\lambda_{\min}(\mathbf{A}^T \mathbf{A})} \implies \mathbf{x} = u_n \quad (189)$$

15.2 Rayleigh quotients

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

$$\frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}} \quad \mathbf{x} \neq 0 \quad (190)$$

$$\lambda_{\min}(\mathbf{A}) \leq \frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_{\max}(\mathbf{A}) \quad \mathbf{x} \neq 0 \quad (191)$$

$$\lambda_{\max}(A) = \max_{\mathbf{x}: \|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{Ax} = u_1 \quad (192)$$

$$\lambda_{\min}(A) = \min_{\mathbf{x}: \|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{Ax} = u_n \quad (193)$$

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. See section 16.1.

$\mathbf{Ax} = \mathbf{y}$ is *underdetermined* if it is short/wide ($n > m$); that is, if it 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:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 \quad (194)$$

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 [18, p. 161]

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (195)$$

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

$$\text{argmin}_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \quad (196)$$

has a closed form solution

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y} \quad (197)$$

However, if \mathbf{A} has a rank $r \ll \min(n, m)$ and a known low-rank decomposition $\mathbf{A} = \mathbf{L}\mathbf{R}^T$ with $\mathbf{L} \in \mathbb{R}^{m,r}$ and $\mathbf{R} \in \mathbb{R}^{n,r}$, then we can rewrite Equation 197 as

$$\mathbf{x} = (\mathbf{R}^T \mathbf{R} \mathbf{L}^T \mathbf{L} + \lambda \mathbf{I})^{-1} \mathbf{L}^T \mathbf{y} \quad (198)$$

This decreases the time complexity from $O(mn^2 + n^3)$ 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

$$\min_{\mathbf{x}: \mathbf{Ax}=\mathbf{y}} \|\mathbf{x}\|_2 \quad (199)$$

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 [18, p. 162]:

$$\mathbf{x}^* = \mathbf{A}^T (\mathbf{AA}^T)^{-1} \mathbf{y} \quad (200)$$

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

$$(\mathbf{AB})_{kl} = \sum_m \mathbf{A}_{km} \mathbf{B}_{ml} \quad \mathbf{A} \in \mathbb{R}^{k,m}, \mathbf{B} \in \mathbb{R}^{m,l} \quad (201)$$

then we have that

$$(\mathbf{A}^T \mathbf{A})_{kl} = \sum_m \mathbf{A}_{mk} \mathbf{A}_{ml} = \sum_{m \neq i} \mathbf{A}_{mk} \mathbf{A}_{ml} + \mathbf{A}_{jk} \mathbf{A}_{jl} = \sum_{m \neq i} \mathbf{A}_{mk} \mathbf{A}_{ml} + (\mathbf{A}_{k*})_j (\mathbf{A}_{l*})_j \quad (202)$$

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

Thus,

$$\mathbf{A}_{\setminus i}^T \mathbf{A}_{\setminus i} = \mathbf{A}^T \mathbf{A} - \mathbf{A}_{*j} \mathbf{A}_{*j}^T \quad (203)$$

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 $(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \sum_{i,j} ((\mathbf{x} \mathbf{x}^T) \circ \mathbf{A})$. If an entry $\mathbf{A}_{i,j}$ in the matrix changes subtract its old value $\mathbf{x}_i \mathbf{x}_j \mathbf{A}_{i,j}$ and add the new value $\mathbf{x}_i \mathbf{x}_j \mathbf{A}'_{i,j}$. If an entry \mathbf{x}_i changes TODO.

18 | Optimization

18.1 Standard Forms

Least Squares

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{Ax}\|_2 \quad (204)$$

LASSO

$$\min_{\mathbf{b} \in \mathbb{R}^n} \left(\frac{1}{N} \|\mathbf{y} - \mathbf{Xb}\|_2^2 + \lambda \|\mathbf{b}\|_1 \right) \quad (205)$$

LP: Linear program

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (206a)$$

$$\text{subject to} \quad \mathbf{A}_{\text{eq}} \mathbf{x} = \mathbf{b}_{\text{eq}}, \quad (206b)$$

$$\mathbf{Ax} \leq \mathbf{b} \quad (206c)$$

Linear Fractional Program

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

$$\text{subject to} \quad \mathbf{Ax} \leq \mathbf{b} \quad (207b)$$

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

QCQP: Quadratic Constrained Quadratic Programs

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{x}^T \mathbf{H}_0 \mathbf{x} + 2\mathbf{c}_0^T \mathbf{x} + \mathbf{d}_0 \quad (208a)$$

$$\text{subject to} \quad \mathbf{x}^T \mathbf{H}_i \mathbf{x} + 2\mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i \leq 0 \quad i \in \mathcal{I}, \quad (208b)$$

$$\mathbf{x}^T \mathbf{H}_j \mathbf{x} + 2\mathbf{c}_j^T \mathbf{x} + \mathbf{d}_j = 0 \quad j \in \mathcal{E} \quad (208c)$$

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

QP: Quadratic Program

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \mathbf{x}^T \mathbf{H}_0 \mathbf{x} + \mathbf{c}_0^T \mathbf{x} \quad (209a)$$

$$\text{subject to} \quad \mathbf{A}_{\text{eq}} \mathbf{x} = \mathbf{b}_{\text{eq}}, \quad (209b)$$

$$\mathbf{Ax} \leq \mathbf{b} \quad (209c)$$

SOCP: Second Order Cone Program (Standard Form)

$$\min_{\mathbf{x}} \quad \mathbf{c}^T \mathbf{x} \quad (210)$$

$$\text{s.t.} \quad \|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i, \quad i = 1, \dots, m \quad (211)$$

SOCP: Second Order Cone Program (Conic Standard Form)

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \quad (212)$$

$$\text{s.t. } (\mathbf{A}_i \mathbf{x} + \mathbf{b}_i, \mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i) \in \mathcal{K}_{m_i} \quad i = 1, \dots, m \quad (213)$$

18.2 Transformations**18.2.1 Linear-Fractional to Linear**

We transform a Linear-Fractional Program

$$\begin{aligned} & \underset{\mathbf{x}}{\text{maximize}} && \frac{\mathbf{c}^T \mathbf{x} + a}{\mathbf{d}^T \mathbf{x} + b} \\ & \text{subject to} && \mathbf{A} \mathbf{x} \leq \mathbf{b} \end{aligned} \quad (214a)$$

$$\quad (214b)$$

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 [19] by defining

$$\mathbf{y} = \frac{1}{\mathbf{d}^T \mathbf{x} + b} \cdot \mathbf{x} \quad (215)$$

$$t = \frac{1}{\mathbf{d}^T \mathbf{x} + b} \quad (216)$$

to form the equivalent program

$$\begin{aligned} & \underset{\mathbf{y}, t}{\text{maximize}} && \mathbf{c}^T \mathbf{y} + at \\ & \text{subject to} && \mathbf{A} \mathbf{y} \leq \mathbf{b}t, \end{aligned} \quad (217a)$$

$$\quad (217b)$$

$$\mathbf{d}^T \mathbf{y} + bt = 1, \quad (217c)$$

$$t \geq 0 \quad (217d)$$

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

18.2.2 LP as SOCP

The linear program

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (218a)$$

$$\text{subject to} \quad \mathbf{A} \mathbf{x} \leq \mathbf{b} \quad (218b)$$

becomes can be cast as an SOCP:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (219a)$$

$$\text{subject to} \quad \|\mathbf{C}_i \mathbf{x} + \mathbf{d}_i\|_2 \leq \mathbf{b}_i - \mathbf{a}_i^T \mathbf{x} \forall i \quad (219b)$$

where $C_i = 0, d_i = 0 \forall i$.

18.2.3 QCQP as SOCP

The quadratic constrained quadratic program

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{x}^T \mathbf{Q}_0 \mathbf{x} + \mathbf{a}_0^T \mathbf{x} \quad (220a)$$

$$\text{subject to} \quad \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + \mathbf{a}_i^T \mathbf{x} \leq b_i \quad i = 1, \dots, m \quad (220b)$$

with $\mathbf{Q}_i = \mathbf{Q}_i^T \succeq 0$, $i = 0, \dots, m$ can be cast as an SOCP:

$$\underset{\mathbf{x}, t}{\text{minimize}} \quad \mathbf{a}_0^T \mathbf{x} + t \quad (221a)$$

$$\text{subject to} \quad \left\| \begin{bmatrix} 2\mathbf{Q}_0^{1/2} \mathbf{x} \\ t - 1 \end{bmatrix} \right\|_2 \leq t + 1, \quad (221b)$$

$$\left\| \begin{bmatrix} 2\mathbf{Q}_i^{1/2} \mathbf{x} \\ b_i - \mathbf{a}_i^T \mathbf{x} - 1 \end{bmatrix} \right\|_2 \leq b_i - \mathbf{a}_i^T \mathbf{x} + 1 \quad i = 1, \dots, m \quad (221c)$$

18.2.4 QP as SOCP

The quadratic program

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \quad (222a)$$

$$\text{subject to} \quad \mathbf{a}_i^T \mathbf{x} \leq \mathbf{b}_i \quad (222b)$$

with $\mathbf{Q} = \mathbf{Q}^T \succeq 0$ can be cast as an SOCP:

$$\underset{\mathbf{x}, y}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} + y \quad (223a)$$

$$\text{subject to} \quad \left\| \begin{bmatrix} 2\mathbf{Q}^{1/2} \mathbf{x} \\ y - 1 \end{bmatrix} \right\|_2 \leq y + 1, \quad (223b)$$

$$\mathbf{a}_i^T \mathbf{x} \leq \mathbf{b}_i \quad \forall i \quad (223c)$$

18.2.5 Sum of L2 Norms to SOCP

$$\underset{\mathbf{x}}{\text{minimize}} \quad \sum_{i=1}^p \|\mathbf{A}_i \mathbf{x} - \mathbf{b}_i\|_2 \quad (224a)$$

becomes

$$\underset{\mathbf{x}, y}{\text{minimize}} \quad \sum_{i=1}^p y_i \quad (225a)$$

$$\text{subject to} \quad \|\mathbf{A}_i \mathbf{x} - \mathbf{b}_i\|_2 \leq y_i \quad i = 1, \dots, p \quad (225b)$$

18.2.6 Minimax of L2 Norms to SOCP

$$\underset{\mathbf{x}}{\text{minimize}} \quad \max_{i=1,\dots,p} \|\mathbf{A}_i \mathbf{x} - \mathbf{b}_i\|_2 \quad (226a)$$

becomes

$$\underset{\mathbf{x}, y}{\text{minimize}} \quad y \quad (227a)$$

$$\text{subject to} \quad \|\mathbf{A}_i \mathbf{x} - \mathbf{b}_i\|_2 \leq y \quad i = 1, \dots, p \quad (227b)$$

18.2.7 Hyperbolic Constraints to SOCP

For scalar w , a constraint of the form

$$w^2 \leq xy, \quad x \geq 0, \quad y \geq 0 \quad (228)$$

can be transformed into the SOCP constraint

$$\left\| \begin{bmatrix} 2w \\ x - y \end{bmatrix} \right\|_2 \leq x + y \quad (229) \quad [20]$$

For vector \mathbf{w} , a constraint of the form

$$\mathbf{w}^T \mathbf{w} = \|\mathbf{w}\|_2^2 \leq xy, \quad x \geq 0, \quad y \geq 0 \quad (230)$$

can be transformed into the SOCP constraint

$$\left\| \begin{bmatrix} 2\mathbf{w} \\ x - y \end{bmatrix} \right\|_2 \leq x + y \quad (231) \quad [20]$$

18.2.8 Matrix Fractional to SOCP

The problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad (\mathbf{F}\mathbf{x} + \mathbf{g})^T (\mathbf{P}_0 + \mathbf{x}_1 \mathbf{P} + \dots + \mathbf{x}_p \mathbf{P}_P)^{-1} (\mathbf{F}\mathbf{x} + \mathbf{g}) \quad (232a)$$

$$\text{subject to} \quad \mathbf{P}_0 + \mathbf{x}_1 \mathbf{P} + \dots + \mathbf{x}_p \mathbf{P}_P > 0, \quad (232b)$$

$$\mathbf{x} \geq 0 \quad (232c)$$

where $\mathbf{P}_i = \mathbf{P}_i^T \in \mathbb{R}^{n,n}$, $\mathbf{F} \in \mathbb{R}^{n,p}$, $\mathbf{g} \in \mathbb{R}^n$, and $\mathbf{x} \in \mathbb{R}^p$ can be transformed into the SOCP where $t_i \in \mathbb{R}, \mathbf{y}_i \in \mathbb{R}^n$:

$$\underset{\mathbf{x}, t}{\text{minimize}} \quad t_0 + \dots + t_p \quad (233a)$$

$$\text{subject to} \quad \mathbf{P}_0^{1/2} \mathbf{y}_0 + \dots + \mathbf{P}_p^{1/2} \mathbf{y}_p = \mathbf{F}\mathbf{x} + \mathbf{g}, \quad (233b) \quad [20]$$

$$\left\| \begin{bmatrix} 2\mathbf{y}_0 \\ t_0 - 1 \end{bmatrix} \right\|_2 \leq t_0 + 1, \quad (233c)$$

$$\left\| \begin{bmatrix} 2\mathbf{y}_i \\ t_i - x_i \end{bmatrix} \right\|_2 \leq t_i + x_i \quad i = 1, \dots, p \quad (233d)$$

18.2.9 Fractional Objective to SOCP

Convert

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \frac{f(x)^2}{g(x)} \end{aligned} \quad (234a)$$

$$\text{subject to} \quad g(x) > 0 \quad (234b)$$

to

$$\begin{aligned} & \underset{\mathbf{x}, t}{\text{minimize}} && t \end{aligned} \quad (235a)$$

$$\text{subject to} \quad f(x)^2 \leq tg(y), \quad (235b)$$

$$g(y) > 0, \quad (235c)$$

$$t \geq 0 \quad (235d)$$

and apply Equation 231.

18.2.10 Chance-Constrained LP to SOCP

The problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (236a)$$

$$\text{subject to} \quad \text{Prob}\{\mathbf{a}_i^T \mathbf{x} \leq \mathbf{b}_i\} \geq p_i \quad i = 1, \dots, m \quad (236b)$$

where $p_i > 0.5$ and all \mathbf{a}_i are independent normal random vectors with expected values $\bar{\mathbf{a}}_i$ and covariance matrices $\Sigma_i \succ 0$, can be transformed into the SOCP:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (237a)$$

$$\text{subject to} \quad \bar{\mathbf{a}}_i^T \mathbf{x} \leq b_i - \Phi^{-1}(p_i) \left\| \Sigma_i^{1/2} \mathbf{x} \right\|_2 \quad i = 1, \dots, m \quad (237b)$$

where $\Phi^{-1}(p)$ is the inverse cumulative probability distribution of a standard normal variable.

18.2.11 Robust LP with Box Uncertainty as LP

The problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (238a)$$

$$\text{subject to} \quad \mathbf{a}_i^T \mathbf{x} \leq b_i \quad \forall \mathbf{a}_i \in \{\hat{\mathbf{a}}_i + \rho_i \mathbf{u} : \|\mathbf{u}\|_\infty \leq 1\} \quad i = 1, \dots, m \quad (238b)$$

is equivalent to

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (239a)$$

$$\text{subject to} \quad \hat{\mathbf{a}}_i^T \mathbf{x} + \rho_i \|\mathbf{x}\|_1 \leq b_i \quad i = 1, \dots, m \quad (239b)$$

which is equivalent to:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (240a)$$

$$\text{subject to} \quad \hat{\mathbf{a}}_i^T \mathbf{x} + \rho_i \sum_{j=1}^n \mathbf{u}_j \leq b_i \quad i = 1, \dots, m, \quad (240b)$$

$$-\mathbf{u}_j \leq \mathbf{x}_j \leq \mathbf{u}_j \quad j = 1, \dots, n \quad (240c)$$

18.2.12 Robust LP with Ellipsoidal Uncertainty as SOCP

The problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (241a)$$

$$\text{subject to} \quad \mathbf{a}_i^T \mathbf{x} \leq b_i \quad \forall \mathbf{a}_i \in \{\hat{\mathbf{a}}_i + \mathbf{R}_i \mathbf{u} : \|\mathbf{u}\|_2 \leq 1\} \quad i = 1, \dots, m \quad (241b)$$

is equivalent to

$$\underset{\mathbf{x}}{\text{minimize}} \quad \mathbf{c}^T \mathbf{x} \quad (242a)$$

$$\text{subject to} \quad \hat{\mathbf{a}}_i^T \mathbf{x} + \left\| \mathbf{R}_i^T \mathbf{x} \right\|_2 \leq b_i \quad i = 1, \dots, m \quad (242b)$$

18.3 Useful Problems

$$\text{average}(\mathbf{v}) = \min_{x \in \mathbb{R}} \|\mathbf{v} - x \mathbf{1}\|_2^2 \quad (243)$$

$$\text{median}(\mathbf{v}) = \min_{x \in \mathbb{R}} \|\mathbf{v} - x \mathbf{1}\|_1 \quad (244)$$

19 | Algorithms

19.1 Gram-Schmidt

TODO

Bibliography

- [1] Néstor Thome. Inequalities and equalities for $l=2$ (Sylvester), $l=3$ (Frobenius), and $l \geq 3$ matrices. *Aequationes mathematicae*, 90(5):951–960, 2016.
- [2] Volker Strassen. Gaussian elimination is not optimal. *Numerische mathematik*, 13(4):354–356, 1969.
- [3] Günter Rote. Division-free algorithms for the determinant and the pfaffian: algebraic and combinatorial approaches. In *Computational discrete mathematics*, pages 119–135. Springer, 2001.
- [4] 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>.
- [5] 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.
- [6] 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.
- [7] Arnold Schönhage. Partial and total matrix multiplication. *SIAM Journal on Computing*, 10(3):434–455, 1981.
- [8] Francesco Romani. Some properties of disjoint sums of tensors related to matrix multiplication. *SIAM Journal on Computing*, 11(2):263–267, 1982.
- [9] Don Coppersmith and Shmuel Winograd. On the asymptotic complexity of matrix multiplication. *SIAM Journal on Computing*, 11(3):472–492, 1982.
- [10] 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.
- [11] 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.
- [12] 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>.
- [13] 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>.
- [14] Gilbert Strang. *Introduction to Linear Algebra*. 2016.

- [15] James Joseph Sylvester. Xxxvii. on the relation between the minor determinants of linearly equivalent quadratic functions. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(4):295–305, 1851.
- [16] Elizabeth Million. The hadamard product. <http://buzzard.ups.edu/courses/2007spring/projects/million-paper.pdf>, 2007.
- [17] Thomas P Minka. Old and new matrix algebra useful for statistics. <https://tminka.github.io/papers/matrix/minka-matrix.pdf>, 2000.
- [18] Giuseppe C Calafiore and Laurent El Ghaoui. *Optimization models*. Cambridge University Press, 2014. ISBN 978-1-107-05087-7.
- [19] 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>.
- [20] Miguel Sousa Lobo, Lieven Vandenbergh, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear algebra and its applications*, 284(1-3):193–228, 1998.

Index

L^∞ -norm, 38

L^1 -norm, 38

p -norm, 38