CS 307-Optimization Algorithms and Techniques

Instructor : Chandresh Kumar Maurya
Homepage: chandreshiit.github.io/

Syllabus

- Part I--Introduction
  - Introduction to Optimization
  - Math Foundations
- Part II-- Linear optimization
  - Linear Optimization
- Part III-- Non-linear optimization

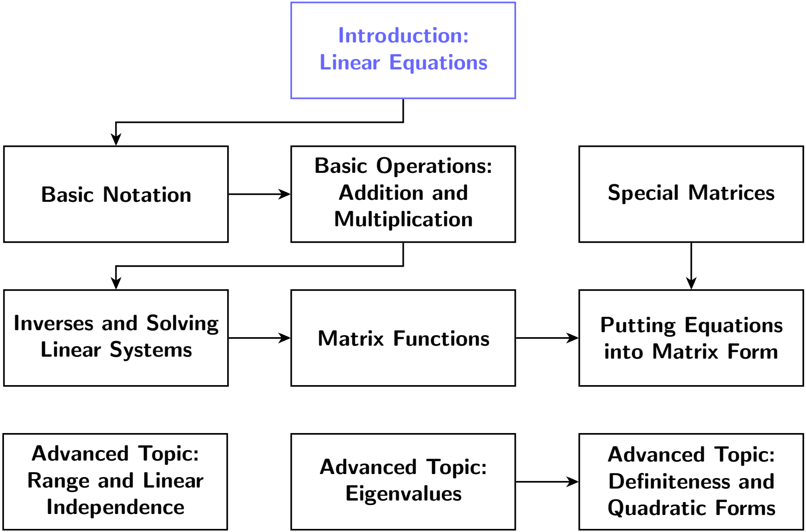
Books

- An Introduction to Optimization: Foundations and Fundamental Algorithms, Niclas Andr ´easson, Anton Evgrafov, and Michael Patriksson, 2nd and 3rd ed.
- Convex optimization, Stephen Boyd and Lieven Vandenberghe, 1st ed., 2004

What will be Covered?

- Mathematical foundation
  - Linear algebra
  - Real analysis

Linear Algebra Review



Linear equations

- Set of linear equations (two equations, two unknowns)

$$\begin{array}{rcrcr} 4x_1 & - & 5x_2 & = & -13 \\ -2x_1 & + & 3x_2 & = & 9 \end{array}$$

- Set of linear equations (two equations, two unknowns)

$$\begin{array}{rcrcr} 4x_1 & - & 5x_2 & = & -13 \\ -2x_1 & + & 3x_2 & = & 9 \end{array}$$

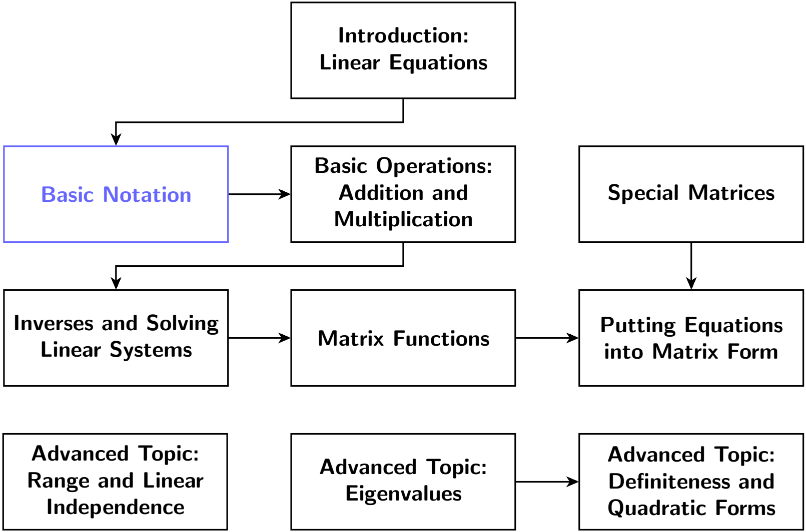
- Can represent compactly using matrix notation

$$Ax = b$$

with

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

Linear Algebra Review



Basic notation

- A matrix with real-valued entries, m rows, and n columns
$$A \in \mathbb{R}^{m \times n}$$

- A_{ij} denotes the entry in the ith row and jth column

- A (column) vector with n real-valued entries

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

- x_i denotes the ith entry

The Transpose

- The transpose operator A^T switches rows and columns of a matrix

$$A_{ij} = (A^T)_{ji}$$

- For a vector $x \in \mathbb{R}^n$, $x^T \in \mathbb{R}^{1 \times n}$ would represent a row vector

Elements of a Matrix

- Can write a matrix in terms of its columns

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

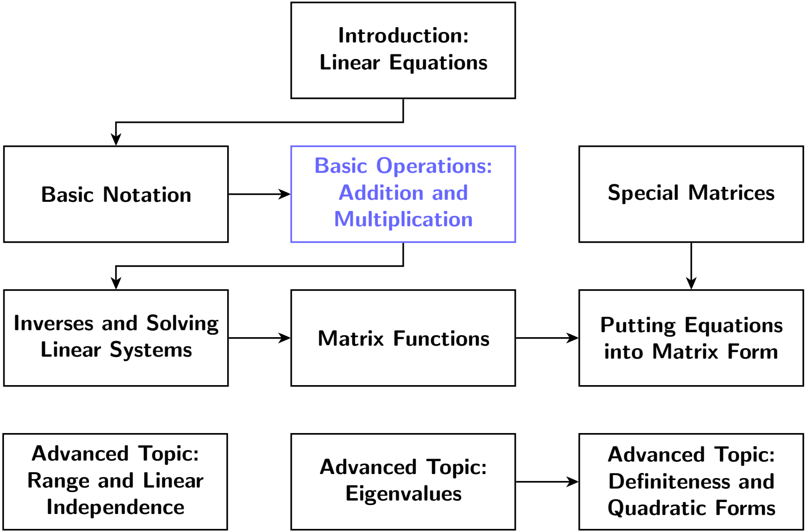
- Careful, a_i here corresponds to an entire *vector* $a_i \in \mathbb{R}^m$, not an element of a vector

- Similarly, can write a matrix in terms of rows

$$A = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}$$

- $a_1 \in \mathbb{R}^n$ here and $a_1 \in \mathbb{R}^m$ from previous slide are *not* the same vector

Linear Algebra Review



Matrix addition

- For two matrices *of the same size and type*, $A, B \in \mathbb{R}^{m\times n}$ addition is just sum of corresponding elements

$$A + B = C \in \mathbb{R}^{m\times n} \iff C_{ij} = A_{ij} + B_{ij}$$

- Addition is *undefined* for matrices of different sizes $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$

Matrix multiplication

- For two matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, their product is

$$AB = C \in \mathbb{R}^{m \times p} \iff C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

- Multiplication is undefined when number of columns in A doesn't equal number or rows in B (one exception: cA for $c \in \mathbb{R}$ taken to mean scaling A by c)

- Some imporant properties
  - Associative: ($A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{p \times q}$)

$$A(BC) = (AB)C$$

  - Distributive: ($A \in \mathbb{R}^{m \times n}$, $B, C \in \mathbb{R}^{n \times p}$)

$$A(B + C) = AB + AC$$

  - *NOT* commutative: (the dimensions might not even make sense, but this doesn't hold even when the dimensions are correct)

$$AB \neq BA$$

Vector-vector Products

- Inner product: $x, y \in \mathbb{R}^n$

$$x^T y \in \mathbb{R} = \sum_{i=1}^{n} x_i y_i$$

- Outer product: $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

$$xy^T \in \mathbb{R}^{n \times m} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{bmatrix}$$

Matrix-vector Products

- $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n \iff Ax \in \mathbb{R}^m$

- Writing A by rows, each entry of Ax is an inner product between x and a row of A

$$A = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}, \quad Ax \in \mathbb{R}^m = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

- Writing A by columns, Ax is a *linear combination* of the columns of A, with coefficients given by x

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}, \quad Ax \in \mathbb{R}^m = \sum_{i=1}^{n} a_i x_i$$

Matrix-matrix Products

- Write $A \in \mathbb{R}^{m \times n}$ by rows, $B \in \mathbb{R}^{n \times p}$ by columns: entries of AB are inner products of the rows of A and the columns of B

$$A = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix}, \quad B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix}$$

$$(AB)_{ij} = a_i^T b_j$$

- Write $A \in \mathbb{R}^{m \times n}$ by columns, $B \in \mathbb{R}^{n \times p}$ by rows: AB is a sum of outer products of columns of A and rows of B

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}, \quad B = \begin{bmatrix} - & b_1^T & - \\ - & b_2^T & - \\ & \vdots & \\ - & b_n^T & - \end{bmatrix}$$

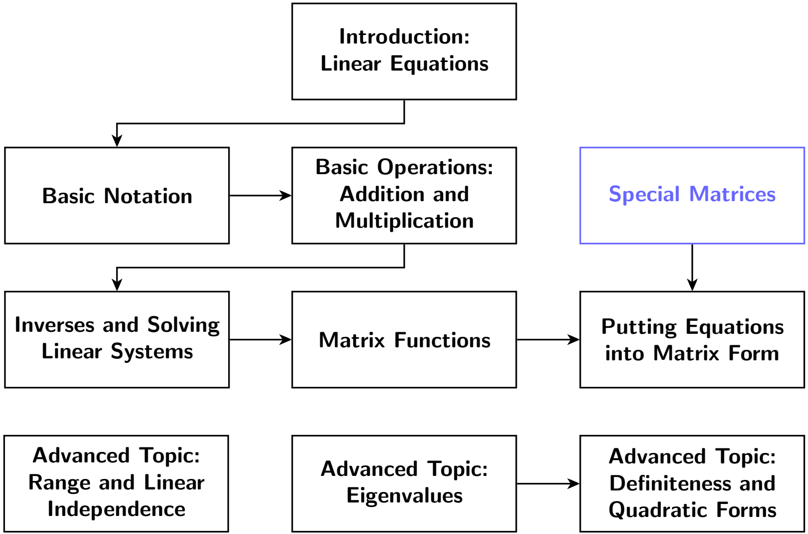
$$AB \in \mathbb{R}^{m \times p} = \sum_{i=1}^{n} a_i b_i^T$$

- Leave $A \in \mathbb{R}^{m \times n}$ as complete matrix, write $B \in \mathbb{R}^{n \times p}$ by columns: columns of AB are matrix-vector products between A and columns of B

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix}$$

$$AB \in \mathbb{R}^{m \times p} = \left[\begin{array}{cccc} Ab_1 & Ab_2 & \cdots & Ab_p \end{array} \right]$$

Linear Algebra Review



The Identity Matrix

$$I \in \mathbb{R}^{n\times n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Has the property that for any $A \in \mathbb{R}^{m\times n}$

$$AI = A = IA$$

(note that the identity matrices on the left and right are *different sizes*, $n \times n$ versus $m \times m$)

The Zero Matrix

$$0 \in \mathbb{R}^{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

- Useful in defining block forms for matrices; e.g. $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times q}$

$$C \in \mathbb{R}^{(m+p) \times (n+q)} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

The All-ones Vector

$$\mathbf{1} \in \mathbb{R}^n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

- Useful, for example, in compactly representing sums

$$a \in \mathbb{R}^n, \quad \mathbf{1}^T a = \sum_{i=1}^{n} a_i$$

The Standard Basis Vector

$$e_i \in \mathbb{R}^n = \begin{bmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \leftarrow i\text{th row}$$

- Can be used to extract entries of a vector $x \in \mathbb{R}^n$

$$x^T e_i = x_i$$

Symmetric Matrices

- Symmetric matrix: $A \in \mathbb{R}^{n \times n}$ with $A = A^T$
- Arise naturally in many settings
  - For $A \in \mathbb{R}^{m \times n}$, $A^T A \in \mathbb{R}^{n \times n}$ is symmetric

Diagonal Matrices

- For $d \in \mathbb{R}^n$

$$\operatorname{diag}(d) \in \mathbb{R}^{n \times n} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

- For example, the identity is given by $I = \operatorname{diag}(1)$

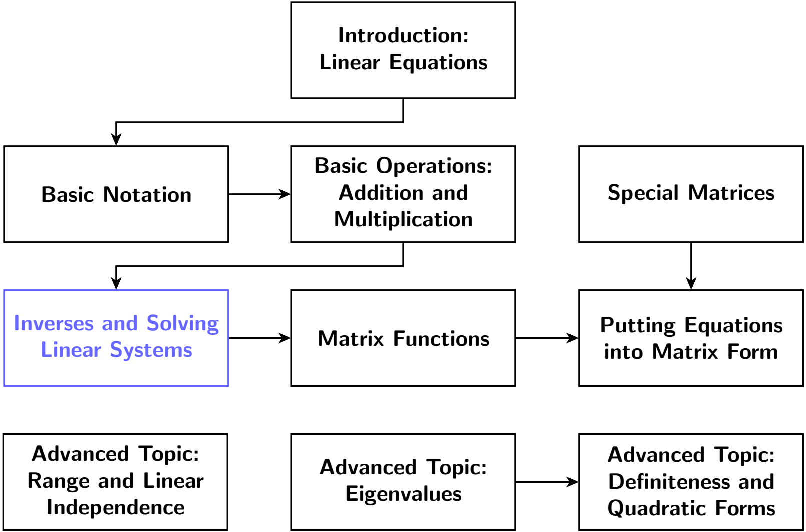
- Multiplying $A \in \mathbb{R}^{m \times n}$ by a diagonal matrix $D \in \mathbb{R}^{n \times n}$ on the right scales the *columns* of A

$$AD = \begin{bmatrix} | & | & & | \\ d_1 a_1 & d_2 a_2 & \cdots & d_n a_n \\ | & | & & | \end{bmatrix}$$

- Multiplying by a diagonal matrix $D \in \mathbb{R}^{m \times m}$ on the left scales the *rows* of A

$$DA = \begin{bmatrix} - & d_1 a_1^T & - \\ - & d_2 a_2^T & - \\ & \vdots & \\ - & d_m a_m^T & - \end{bmatrix}$$

Linear Algebra Review



The Matrix Inverse

- Inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ denoted A^{-1}

$$AA^{-1} = I = A^{-1}A$$

- May not exist (*non-singular* matrix has inverse, *singular* matrix does not)

$$A^{-1} \text{ exists } \iff Ax \neq 0 \text{ for all } x \neq 0$$

- Some important properties for $A, B \in \mathbb{R}^{n\times n}$ non-singular
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^T)^{-1} = (A^{-1})^T$

Solving Linear Equations

- Two linear equations

$$\begin{array}{rcrcr} 4x_1 & - & 5x_2 & = & -13 \\ -2x_1 & + & 3x_2 & = & 9 \end{array}$$

- In vector form, $Ax = b$, with

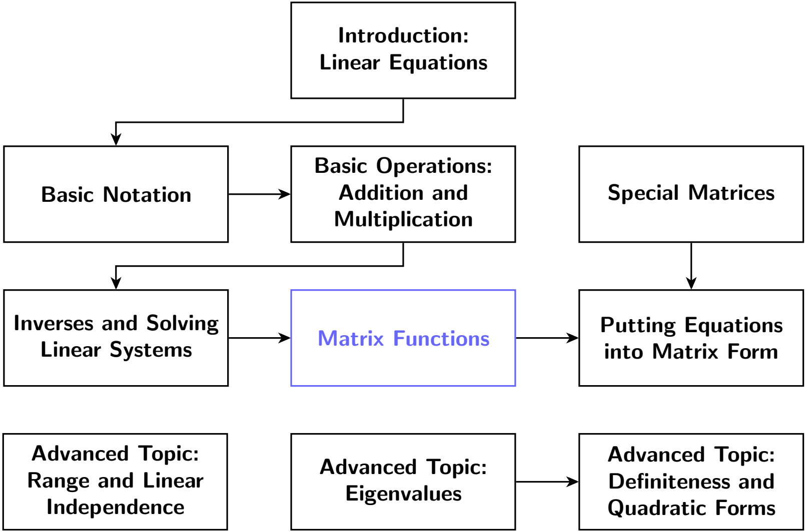
$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

- Solution using inverse

$$\begin{aligned} Ax &= b \\ A^{-1}Ax &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

- Won't worry here about how to compute inverse, but it's very similar to the standard method for solving linear equations

Linear Algebra Review



Notation for Functions

- $f(x) = x^2$, $f : \mathbb{R} \to \mathbb{R}$

- Function with matrix inputs/outputs

$$f : \mathbb{R}^{m \times n} \to \mathbb{R}^{p \times q}$$

Some Examples

- Transpose: $f(A) = A^T$

$$f : \mathbb{R}^{m \times n} \to \mathbb{R}^{n \times m}$$

- Inverse: $f(A) = A^{-1}$

$$f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$$

- Multiplication: $f(x) = Ax$ for $A \in \mathbb{R}^{m \times n}$

$$f : \mathbb{R}^{n} \to \mathbb{R}^{m}$$

The Trace

- $\mathrm{tr} : \mathbb{R}^{n \times n} \to \mathbb{R}$

$$\mathrm{tr}\, A = \sum_{i=1}^{n} A_{ii}$$

- Some properties
  - $\mathrm{tr}\, A = \mathrm{tr}\, A^T$, $A \in \mathbb{R}^{n \times n}$
  - $\mathrm{tr}(A + B) = \mathrm{tr}\, A + \mathrm{tr}\, B$, $A, B \in \mathbb{R}^{n \times n}$
  - $\mathrm{tr}\, AB = \mathrm{tr}\, BA$, $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$

Norms

- A vector norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ with

  1. $f(x) \geq 0$ and $f(x) = 0 \Leftrightarrow x = 0$

  2. $f(ax) = |a| f(x)$ for $a \in \mathbb{R}$

  3. $f(x + y) \leq f(x) + f(y)$

- ℓ_2 norm

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- ℓ_1 norm

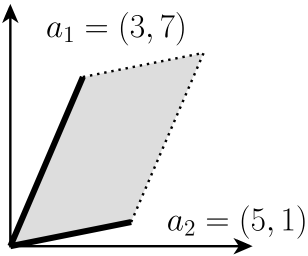
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

- ℓ_∞ norm

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

Determinant

- $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$ (sometimes denoted $|\cdot|$)



$$A = \begin{bmatrix} 3 & 7 \\ 5 & 1 \end{bmatrix}$$

- $|\det A|$ is the area of the parallelogram

- Can be formally defined by three properties

  1. Determinant of identity is one: $\det I = 1$

  2. Multiplying a row by scalar $t \in \mathbb{R}$ scales determinant:

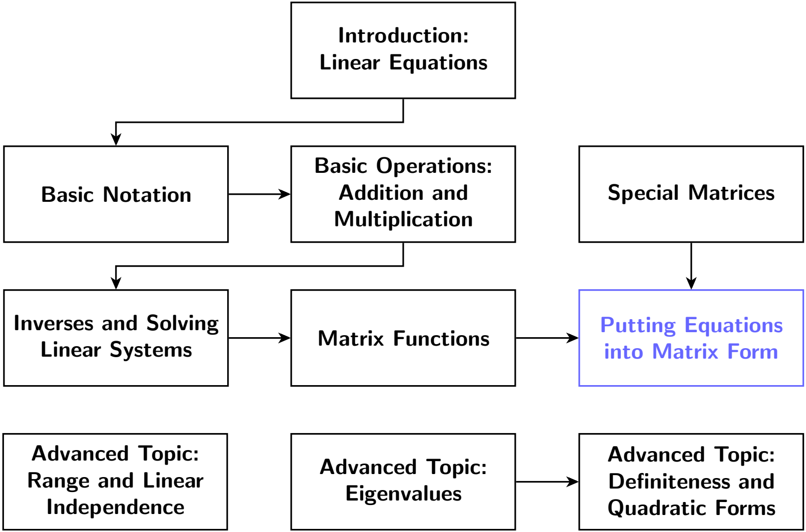
  $$\det \begin{bmatrix} — & t a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix} = t \det A$$

  3. Swapping rows negates determinant:

  $$\det \begin{bmatrix} — & a_2^T & — \\ — & a_1^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix} = -\det A$$

- Important properties
  - $\det A = \det A^T$
  - $\det AB = \det A \det B$
  - $\det A = 0 \Leftrightarrow$ A singular (non-invertible)
  - $\det A^{-1} = 1/\det A$

Linear Algebra Review



- Given $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$, $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = \sum_{i=1}^{m} (a_i^T x - b_i)^2$$

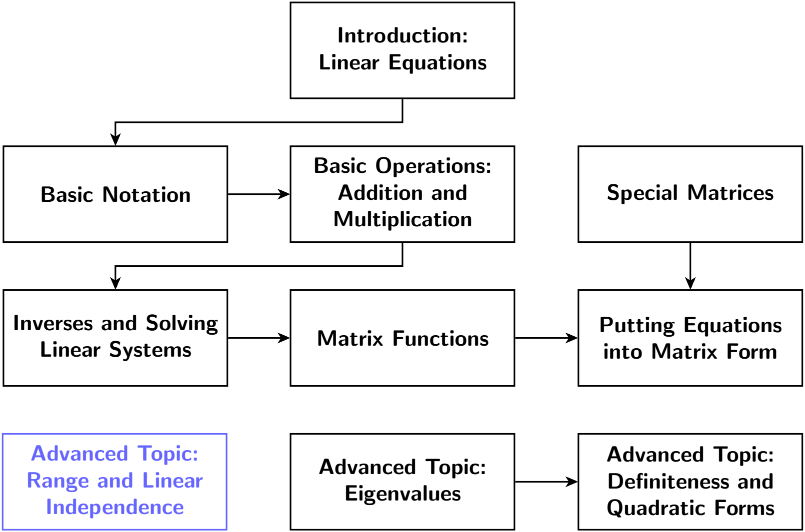
- $f : \mathbb{R}^{m \times n} \to \mathbb{R}$

$$f(A) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2$$

- Given $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, construct $A \in \mathbb{R}^{m \times n}$ such that

$$A_{ij} = (x_i - y_j)^2$$

Linear Algebra Review



Range

- For $A \in \mathbb{R}^{m \times n}$, *range* of A is the set of all vectors that can be written Ax for some $x \in \mathbb{R}^n$

$$\mathcal{R}(A) \subseteq \mathbb{R}^m = \{y : y = Ax, x \in \mathbb{R}^n\}$$

- The columns of A are *linearly independent* if no column is in the range of the remaining columns

$$a_i \notin \mathcal{R}(A_{-i}), \forall i = 1, \ldots, n$$

Rank

- *Rank* of $A \in \mathbb{R}^{m \times n}$ is the number of linearly indpendent columns

- Some important properties

  – $\operatorname{rank}(A) = \operatorname{rank}(A^T)$

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

$$\operatorname{rank}(A) = n \;\Leftrightarrow\; \mathcal{R}(A) = \mathbb{R}^n \;\Leftrightarrow\; A \text{ non-singular}$$

Orthogonality

- Two vectors $x, y \in \mathbb{R}^n$ are *orthogonal* if

$$x^T y = 0$$

- They are *orthonormal* if, in addition,

$$\|x\|_2 = \|y\|_2 = 1$$

- A matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if all it's columns are orthonormal, i.e.,

$$U^T U = I = U U^T$$

- Columns of an orthogonal matrix are linearly independent

Nullspace

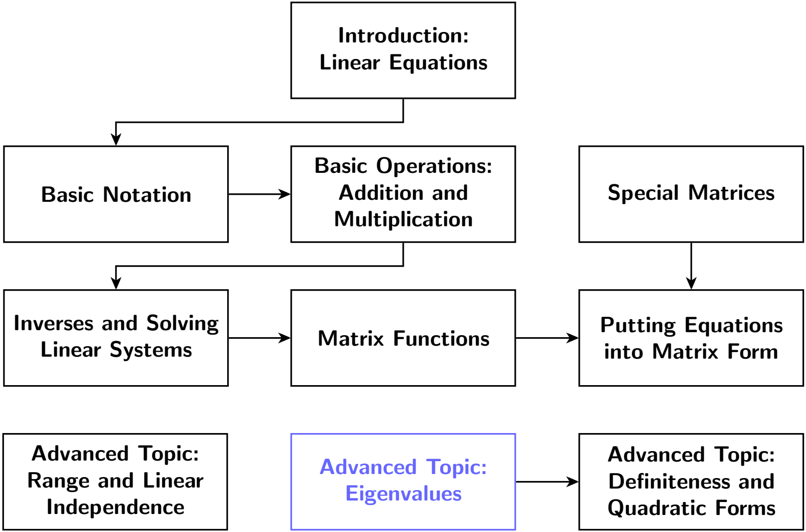
- for $A \in \mathbb{R}^{m \times n}$, *nullspace* of A is set of all vectors x s.t. $Ax = 0$

$$\mathcal{N}(A) \subseteq \mathbb{R}^n = \{x : Ax = 0\}$$

- $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$ are *orthogonal complements*

$$\mathcal{R}(A) \cup \mathcal{N}(A^T) = \mathbb{R}^m, \;\; \mathcal{R}(A) \cap \mathcal{N}(A^T) = \{0\}$$

Linear Algebra Review



Eigenvalues and Eigenvectors

- For $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an *eigenvalue* and $x \in \mathbb{C}^n \neq 0$ an *eigenvector* if

$$Ax = \lambda x$$

- Satisfied if $(\lambda I - A)x = 0$, which we know exists if and only if $\det(\lambda I - A) = 0$

- $\det(\lambda I - A)$ is a polynomial (of degree n) in λ, its n roots are the n eigenvalues of A

Diagonalization

- Write equations for all n eigenvalues as

$$A \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

- Write as $AX = X\Lambda$, which implies

$$A = X\Lambda X^{-1}$$

if X is invertible (A *diagonalizable*)

- Important properties of eigenvectors/eigenvalues
  - $\operatorname{tr} A = \sum_{i=1}^{n} \lambda_i$
  - $\det A = \prod_{i=1}^{n} \lambda_i$
  - $\operatorname{rank}(A) =$ number of non-zero eigenvalues
  - Eigenvalues of A^{-1} are $1/\lambda_i$, $i = 1, \ldots, n$, eigenvectors are the same

- An example: Given $A \in \mathbb{R}^{n \times n}$, what can we say about A^k as $k \to \infty$?

Symmetric Matrices

- For a symmetric matrix $A \in \mathbb{R}^{n\times n})$ $(A = A^T)$, we have the following properties

  1. All eigenvalues/eigenvectors of A are real (more correctly, eigenvectors can be chosen to be real)

  2. The eigenvectors of A are orthogonal (can be chosen to be orthogonal)

- Implies that A can be diagonalized as

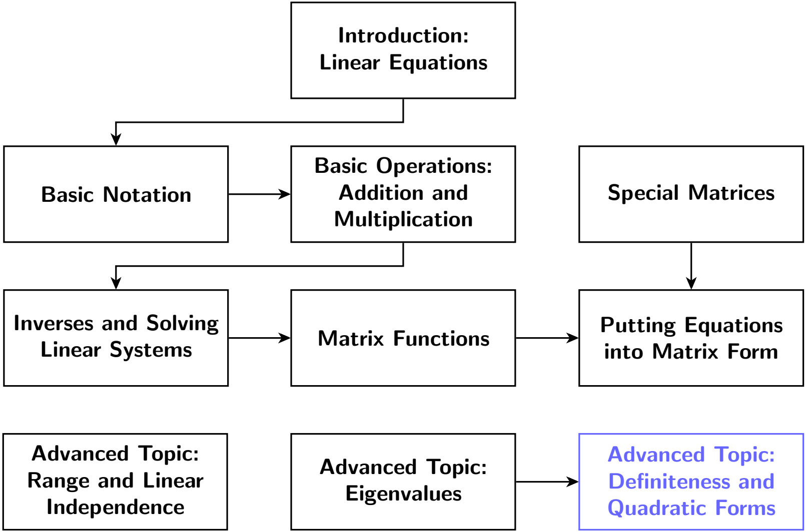
$$A = U\Lambda U^T$$

- Eigenvalues of symmetric matrix are real

- Eigenvectors of symmetric matrix can be chosen to be real

- Eigenvectors of symmetric matrix can be chosen to be orthogonal

Linear Algebra Review



Quadratic Forms

- A quadratic form is a function $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = x^T A x$$

  for some $A \in \mathbb{R}^{n \times n}$

- Can take A to be symmetric, since

$$x^T A x = (x^T A x)^T = x^T A^T x = x^T \frac{1}{2}(A + A^T) x$$

- $A \in \mathbb{R}^{n \times n}$ is positive definite (positive semidefinite) if $x^T A x > 0$ ($x^T A x \geq 0$) for all $x \in \mathbb{R}^n \neq 0$

- $A \in \mathbb{R}^{n \times n}$ is negative definite (negative semidefinite) if $x^T A x < 0$ ($x^T A x \leq 0$) for all $x \in \mathbb{R}^n \neq 0$

- A is indefinite if neither positive nor negative semidefinite

- Definiteness is characterized by eigenvalues of A
  - A positive definite $\Leftrightarrow \lambda_i > 0, \;\; \forall i$
  - A positive semidefinite $\Leftrightarrow \lambda_i \geq 0, \;\; \forall i$
  - A negative definite $\Leftrightarrow \lambda_i < 0, \;\; \forall i$
  - A negative semidefinite $\Leftrightarrow \lambda_i \leq 0, \;\; \forall i$