Lecture 2: Math background review II and Convexity I

Dabeen Lee

Industrial and Systems Engineering, KAIST

IE 539: Convex Optimization

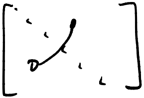
~~August 30~~, 2022

September 1

Outline

- Symmetric matrices, eigenvalue decomposition, positive semidefnite matricex
- Gradient, Hessian, Multivariate calculus.
- Convex sets.
- ~~(If time allows) Convex functions~~.
- (I~~f time allows) First-order, second-order characterizations of convex functions,~~

Symmetric matrices, eigenvalue-eigenvector pair



Symmetric matrices $M \in \mathbb{R}^{d \times d}$ is symmetric if $M = M^T$

$$M_{ij} = M_{ji}$$

eigenvalue-eigenvector pair If $\underset{\mathbb{R}^{d \times d}}{M} \cdot \overset{\mathbb{R}^{d}}{v} = \underset{\mathbb{R}}{\lambda} \cdot \underset{\mathbb{R}^{d}}{v}$, then

λ: eigen value

v : eigen vector

(λ, v) : eigen pair.

Eigenvalue decomposition

$v^1, \ldots, v^k$ are orthonormal if

① $\langle v^i, v^j \rangle = 0$

② $\|v^i\|_2 = 1$

Theorem

Let M be a symmetric matrix. Then M can be written as $M = Q\Lambda Q^\top$ where

1. *Q is an orthonormal matrix, i.e. $Q^\top Q = QQ^\top = I$, whose columns are the eigenvectors of M,*
2. *Λ is a diagonal matrix whose diagonal entries are the eigenvalues of M.*

$$M = \begin{bmatrix} v^1 & v^2 & v^3 & \cdots & v^d \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} \begin{bmatrix} v^{1T} \\ v^{2T} \\ \vdots \\ v^{dT} \end{bmatrix}$$

Express a symmetric matrix using its eigenvalues and eigvenvectors

$$M = \sum_{i=1}^{d} \lambda_i \cdot v^i \cdot v^{iT}$$

Positive semidefinite matices

square M. ($M = M^T$)

We say that a symmetric matrix is positive semidefinite (PSD) if

all of its eigenvalues are nonnegative.

$$M = \sum_{i=1}^{d} \lambda_i v^i \cdot v^{iT}$$

Theorem

Let M be an $d \times d$ symmetric matrix. Then M is PSD if and only if $x^\top M x \geq 0$ for all $x \in \mathbb{R}^d$.

Proof $Q Q^T = I$.

$\Downarrow$

invertible

$\Downarrow$

$\text{rank}(Q) = d$

$\Downarrow$

Columns $v^1, v^2, \ldots, v^d$ are linearly independent. $\longrightarrow$ $v^1, \ldots, v^d$ span $\mathbb{R}^d$

$\forall x \in \mathbb{R}^d$, $x = \sum_{i=1}^{d} \alpha_i \cdot v^i$

$x^T M x = \left(\sum_{i=1}^{d} \alpha_i v^{iT}\right) \cdot \left(\sum_{i=1}^{d} \lambda_i v^i v^{iT}\right) \cdot \left(\sum_{i=1}^{d} \alpha_i v^i\right)$

$= \sum_{i=1}^{d} \alpha_i^2 \lambda_i$

Let $f : \mathbb{R}^d \to \mathbb{R}$.

Lipschitz continuity f is Lipschitz continuous if

$$\underline{|f(x) - f(y)| \le L \cdot \|\underline{x} - \underline{y}\|}$$

for some L. $\forall x, y \in \mathbb{R}^d$

Partial derivatives

$$\frac{\partial f}{\partial x_i} = \lim_{t \to 0} \frac{f(x + t \cdot e^i) - f(x)}{t} \qquad e^i = (0, \dots 0, 1, 0 \cdots 0)$$

Gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_d} \right)^T$$

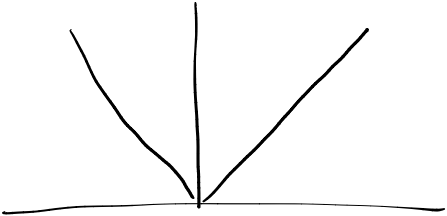
Lipschitz continuous but not differentiable function

Lipschitz continuous function are

- continuous.
- differentiable almost everywhere

$f(x) = |x| : \mathbb{R} \to \mathbb{R}$

$|f(x) - f(y)| \le |x-y|$



Second partial derivatives, Hessian, Schwarz's theorem

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

The Hessian

row i, column j $= \dfrac{\partial^2 f}{\partial x_i \partial x_j}$

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(x) & \frac{\partial^2 f}{\partial x_d \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_d^2}(x) \end{pmatrix}$$

row j column i $= \dfrac{\partial^2 f}{\partial x_j \partial x_i}$

Schwarz's theorem

If all second-order partial derivatives are "continuous", then $\nabla^2 f(x)$ is symmetric

Multivariate differentiation

Exercise: $f(t) = g(x + t \cdot y) : \mathbb{R} \to \mathbb{R}$

A

$f'(t) = ?$ $\quad y^T \nabla g(x+ty)$.

$f''(t) = y^T \nabla^2 g(x+ty) \cdot y$.

Let A be a $n \times d$ matrix and $b \in \mathbb{R}^n$, and let $f(x) = g(Ax - b) : \mathbb{R}^d \to \mathbb{R}$. Then

$A^T \nabla g(Ax-b)$

$\mathbb{R}^d$ $\quad \mathbb{R}^n$

- $\nabla f(x) = A^T \nabla g(Ax - b)$
- $\nabla^2 f(x) = A^T \nabla^2 g(Ax-b) A$

Example

Consider $f(x) = Ax - b$. Then $\nabla f(x) = A^T$

$= g(Ax-b)$

$g(y) = y$

Multivariate differentiation

Consider a quadratic function $f(x) = x^\top Q x + p^\top x = \sum_{i=1}^d \sum_{j=1}^{d} Q_{ij} x_i x_j + \sum_{i=1}^d p_i x_i$.
Then

- $\nabla f(x) = (Q + Q^T)x + p$
- $\nabla^2 f(x) = Q + Q^T$

Example

Consider $f(x) = \|Ax - b\|_2^2$. Then $\nabla f(x) =$, $\nabla^2 f(x) =$

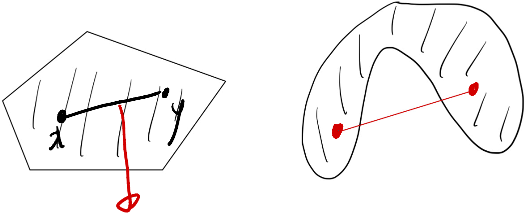
$$= (Ax-b)^T (Ax-b)$$

Convex set

A set $X \subseteq \mathbb{R}^d$ is <u>convex</u> if for any $x, y \in X$ and $\lambda \in [0,1]$,

$$\underline{\lambda x + (1-\lambda) y} \in X.$$

In words,



$\lambda \cdot x + (1-\lambda) y$

Convex combination, convex hull

Given $v^1, \ldots, v^k \in \mathbb{R}^d$, a convex combination of $v^1, \ldots, v^k$ is

a linear combination $\sum_{i=1}^{k} \alpha_i \cdot v^i$ where

$$\sum_{i=1}^{k} \alpha_i = 1, \quad \alpha_1, \ldots, \alpha_k \geq 0.$$

The convex hull of a set X, denoted $\mathrm{conv}(X)$, is

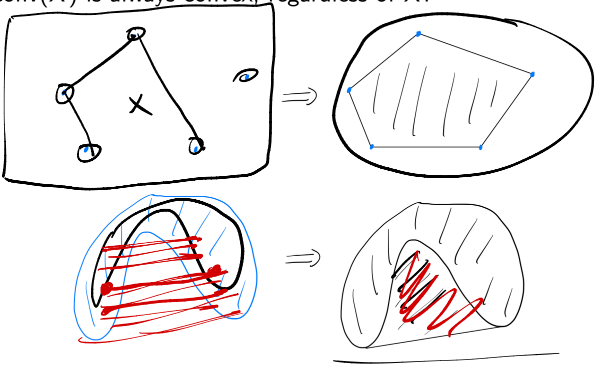
the set of all convex combinations of vectors in X.

Convex hull example

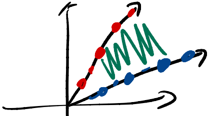
By definition,

$$\mathrm{conv}(X) = \left\{ \sum_{i=1}^{n} \lambda_i v^i : \begin{array}{c} n \in \mathbb{N},\ v^1, \ldots, v^n \in X, \\ \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_1, \ldots, \lambda_n \geq 0 \end{array} \right\}.$$

Exercise: conv(X) is always convex, regardless of X.



Cones, convex cones, conic combination



A set $C \subseteq \mathbb{R}^d$ is a <u>cone</u> if for any $v \in C$ and $\alpha \geq 0$,

$$\alpha \cdot v \in C.$$

A <u>convex cone</u> is a cone that is <u>convex</u>

Given $v^1, \dots, v^k \in \mathbb{R}^d$, a <u>conic combination</u> of $v^1, \dots, v^k$ is

a linear combination $\sum_{i=1}^{k} \alpha_i \cdot v^i$ where

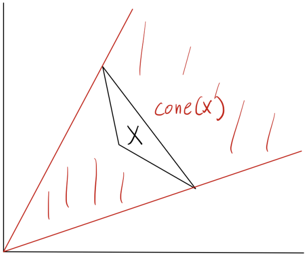
$$\alpha_1, \dots, \alpha_k \geq 0$$

Conic hull

The conic hull of a set X, denoted cone(X), is the set of all conic combinations of points in X. By definition,

$$\text{conv}(X) = \left\{ \sum_{i=1}^{n} \lambda_i v^i : \begin{array}{l} n \in \mathbb{N},\ v^1, \ldots, v^n \in X, \\ \lambda_1, \ldots, \lambda_n \geq 0 \end{array} \right\}.$$

Exercise: cone(X) is always a convex cone, regardless of X.



Affine combination, affine hull

Given $v^1, \ldots, v^k \in \mathbb{R}^d$, an affine combination of $v^1, \ldots, v^k$ is

a linear combination $\sum_{i=1}^{k} \alpha_i \cdot v^i$ where

$\sum_{i=1}^{k} \alpha_i = 1$,

The affine hull of a set X is

the set of all affine combinations of vectors in X ,

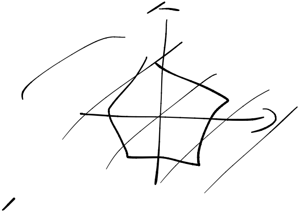
The affine hull of X = The affine suspace spanned by X

Affine subspace, linear hull

The affine suspace spanned by $X =$ affine hull of X

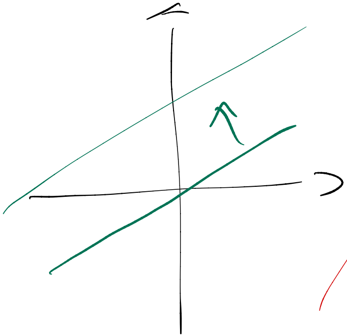
Linear hull of $X =$ the "linear" subspace spanned by X

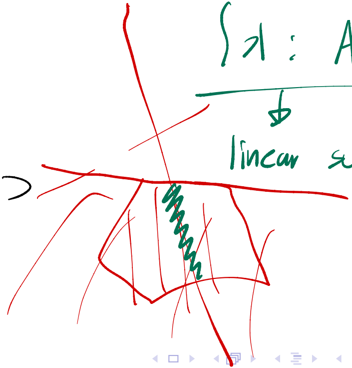
Affine subspace vs linear subspace



Theorem

An affine subspace is a translation of a linear subspace. For an affine subspace $V \subseteq \mathbb{R}^d$, there exist matrices A and b such that $V = \{x \in \mathbb{R}^d : Ax = b\}$.





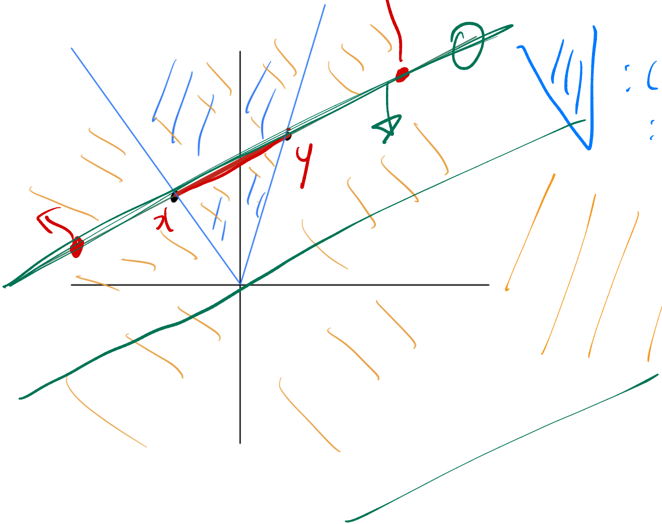
$\{x : Ax = 0\}$

linear subspace

Comparison

$-\alpha \cdot x + (2-\alpha) y$

: X.



: Cone(X)

: the conic hull of X

- linear subspace spanned by X

! affine subspace

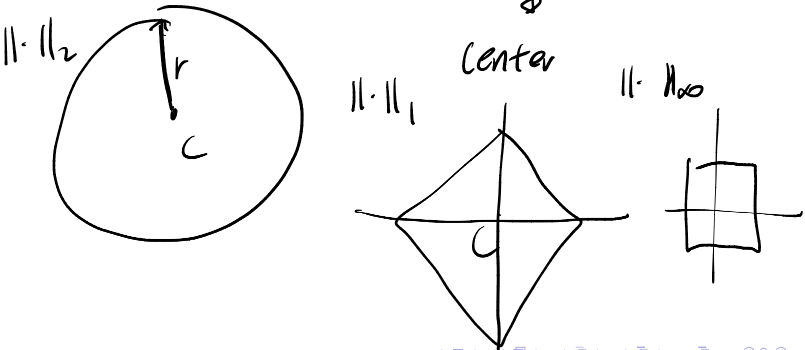
Examples

We saw that the convex hull and conic hull of a set are convex and that the linear subspace and affine subspace spanned by a set are convex. There are many more examples.

- $\emptyset$, $\{v\}$

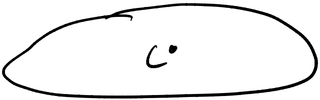
radius

- Norm ball: $\{x : \|x - c\| \le r\}$



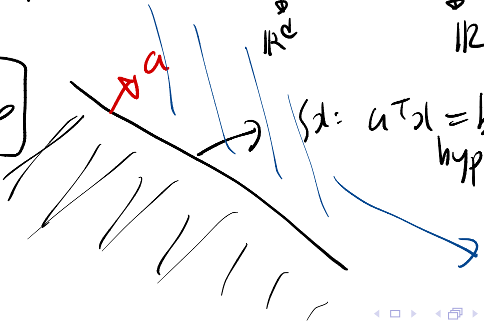
Examples

- Ellipsoid: $\{x : (x-c)^T \underbrace{Q}_{\text{PD}} (x-c) \le r\}$



- Half-space $\{x : \underbrace{a}_{\mathbb{R}^d}{}^T x \le \underbrace{b}_{\mathbb{R}}\}$

$\boxed{a^T x \le b}$



$\{x : a^T x = b\}$ hyperplane

$a^T x \geq b$

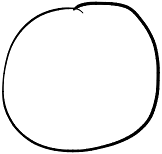
Examples

polyhedron

= finite intersection of half-spaces

$\{x : Ax \le b\}$

$A \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^{n}$



$\{\|x\|_2 \le 1\} = \bigcap_{d \in \mathbb{R}^n} \{d^T x \le \|d\|_2\}$

Examples

polytope = polyhedron & bounded.

$$\text{Simplex} = \{x \in \mathbb{R}^d : \sum_{i=1}^{d} x_i = 1, \; x \geq 0\}$$

$$\mathbb{R}_+^d = \{x \in \mathbb{R}^d : x \geq 0\}:$$

nonnegative orthant.

$$\mathbb{R}_{++}^d = \{x \in \mathbb{R}^d : x > 0\}$$

positive orthant.

Convex cone example

Norm cone

$$\{(x,t) \;:\; \|x\| \le t\}$$

Positive semidefinite cone

: the set of all positive semidefinite matrices