18-660: Numerical Methods for Engineering Design and Optimization

Xin Li
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213

Electrical & Computer ENGINEERING

Overview

- Convex Analysis
  - Convex function
  - Convex set
  - Convex optimization

Ordinary Least-Squares Regression

- Solve over-determined linear equation by optimization

M samples

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \cdot X = \begin{bmatrix} \\ B \\ \\ \end{bmatrix}$$

(M > N)

N coefficients

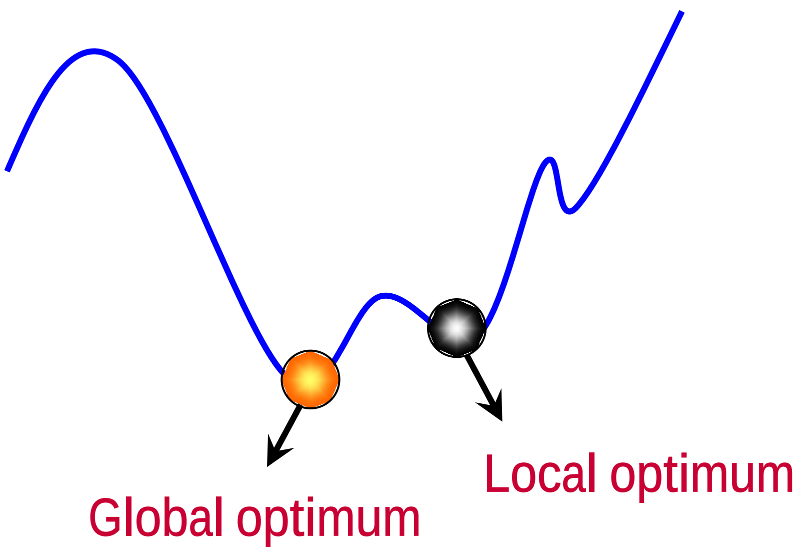
$$\min_{X} \quad \left\| A \cdot X - B \right\|_2^2$$

Unconstrained Nonlinear Programming

- Nonlinear cost function without constraints

$$\min_{X} \quad \left\| A \cdot X - B \right\|_2^2$$

- General nonlinear optimization is difficult to solve



Unconstrained Quadratic Programming

$$\min_{X} \quad \left\| A \cdot X - B \right\|_2^2$$

- However, ordinary least-squares regression is different from general nonlinear programming

- Optimization cost is a quadratic function of X and it is always non-negative for any given X
  - This is a unique property that enables us to solve least-squares regression efficiently

Positive Semi-Definite

- If a quadratic function $X^T A_Q X$ is always non-negative, the quadratic coefficient matrix A_Q is positive semi-definite
  - Assume that A_Q is symmetric so that its eigenvalues are real
  - Any asymmetric A_Q can be converted to a symmetric one

Row vector Matrix Column vector

$$X^T A_Q X = \left(X^T A_Q X\right)^T = X^T A_Q^T X$$

Scalar

$$X^T A_Q X = X^T \cdot \left[\frac{1}{2} \cdot \left(A_Q + A_Q^T\right)\right] \cdot X$$

Symmetric matrix

Positive Semi-Definite

- Simple example:

$$A_Q = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$X^T A_Q X = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 x_2$$

$$\frac{1}{2}\left(A_Q + A_Q^T\right) = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}$$

$$\frac{1}{2} X^T \left(A_Q + A_Q^T\right) X = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 x_2$$

Positive Semi-Definite

- A_Q is positive semi-definite if and only if all eigenvalues of A_Q are non-negative (necessary and sufficient condition)

- Eigenvalue decomposition

$$A_Q \cdot V_i = V_i \cdot \lambda_i$$

"Normalized" eigenvector: $||V_i||_2 = 1$

Eigenvalue

  - A_Q is symmetric $\rightarrow$ all eigenvalues are real
  - A_Q is symmetric $\rightarrow$ all eigenvectors are real and orthogonal

Positive Semi-Definite

- Eigenvalue decomposition

$$A_Q \cdot V_i = V_i \cdot \lambda_i \qquad V = \begin{bmatrix} V_1 & V_2 & \cdots \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \end{bmatrix}$$

$$A_Q \cdot V = V \cdot \Sigma$$

  - All eigenvectors are orthogonal

$$V^T V = I$$

Identity matrix

$$A_Q = V \cdot \Sigma \cdot V^{-1} = V \cdot \Sigma \cdot V^T$$

Positive Semi-Definite

- If one of the eigenvalues of A_Q is negative

$$A_Q = V \cdot \Sigma \cdot V^T \quad where \quad V^T V = I$$

$$\Downarrow$$

$$X^T A_Q X = \left(V^T X\right)^T \cdot \Sigma \cdot \left(V^T X\right) = \left(V^T X\right)^T \cdot \begin{bmatrix} \times & & & \\ & \times & & \\ & & \ddots & \\ & & & -\varepsilon \end{bmatrix} \cdot \left(V^T X\right)$$

Positive Semi-Definite

- If one of the eigenvalues of A_Q is negative

$$(V^T X)^T \cdot \begin{bmatrix} \times & & & \\ & \times & & \\ & & \ddots & \\ & & & -\varepsilon \end{bmatrix} \cdot (V^T X)$$

Select $V^T X = [0\ 0\ \ldots\ 0\ \boldsymbol{\Delta}]^T$

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ \Delta \end{bmatrix}^T \cdot \begin{bmatrix} \times & & & \\ & \times & & \\ & & \ddots & \\ & & & -\varepsilon \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \Delta \end{bmatrix} = -\varepsilon \cdot \Delta^2 < 0$$

$X^T A_Q X$ is negative

Positive Semi-Definite

- If a quadratic function $X^T A_Q X + B_Q^T X + C_Q$ is always non-negative (for any X), all eigenvalues of A_Q are non-negative
  - I.e., A_Q is positive semi-definite
  - Why? (You can prove this conclusion by following the steps of eigenvalue decomposition)
- The quadratic coefficient matrix for the least-squared error $err(X) = ||AX-B||_2^2$ is positive semi-definite

Positive Semi-Definite

- The reverse statement is NOT true

- Even if A_Q is positive semi-definite, $f(X) = X^T A_Q X + B_Q^T X + C_Q$ can be either positive or negative

$$f(x) = x^2 - 1$$

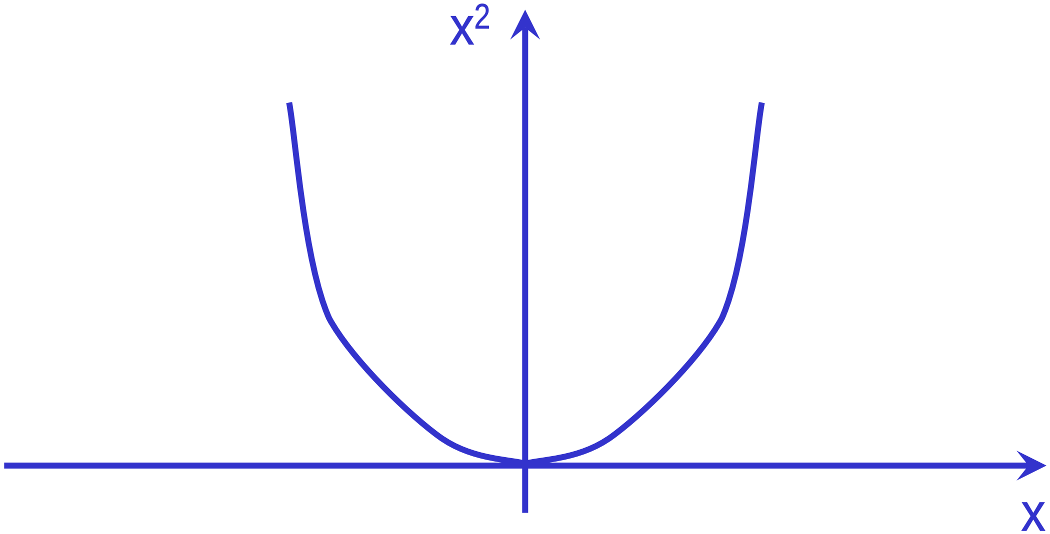
$$\Downarrow$$

$$f(0) = -1 < 0$$

Convex Function

- Positive semi-definite quadratic functions have special properties that can be utilized by nonlinear optimization

- A simple one-dimensional example $f(x) = x^2$

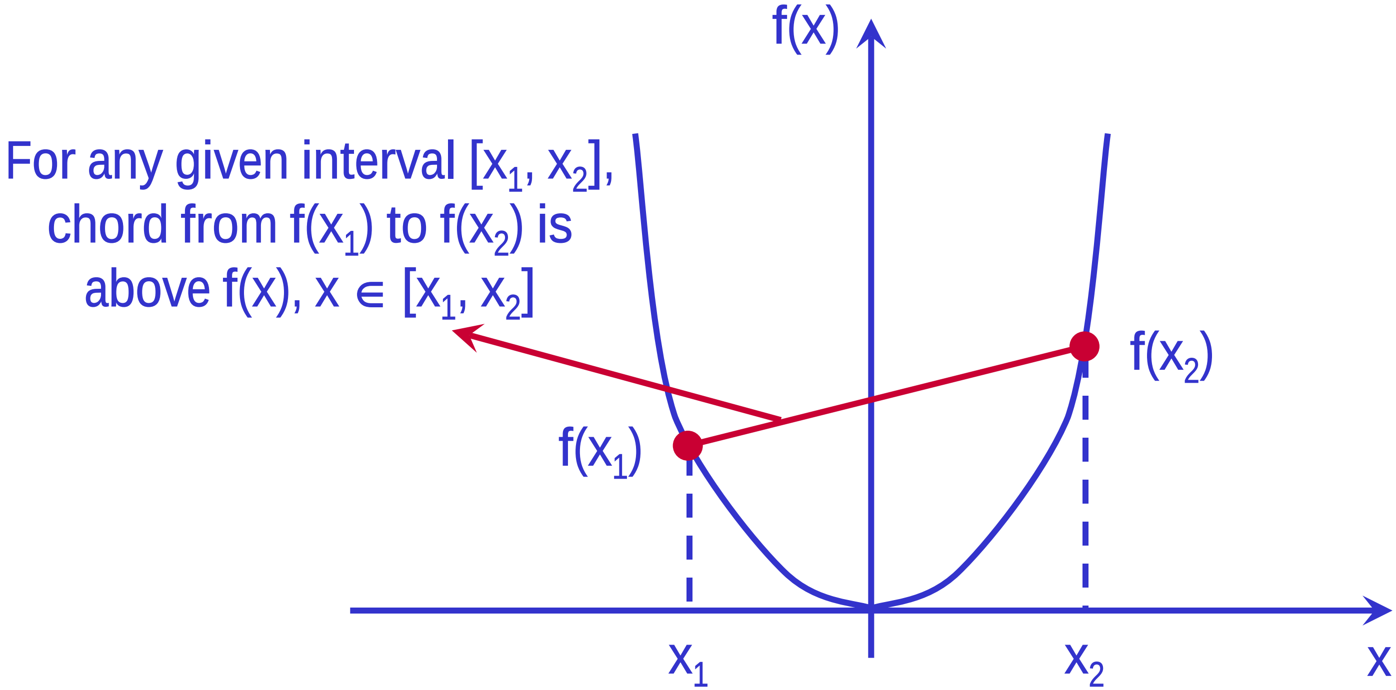


The function $f(X) = x^2$ is convex

Convex Function

- f(X) is convex, if for all *vectors* X_1, X_2 and $0 \le \boldsymbol{\alpha} \le 1$, we have

$$f\left[\alpha \cdot X_1 + (1-\alpha)\cdot X_2\right] \le \alpha \cdot f(X_1) + (1-\alpha)\cdot f(X_2)$$



A one-dimensional convex example

Convex Function

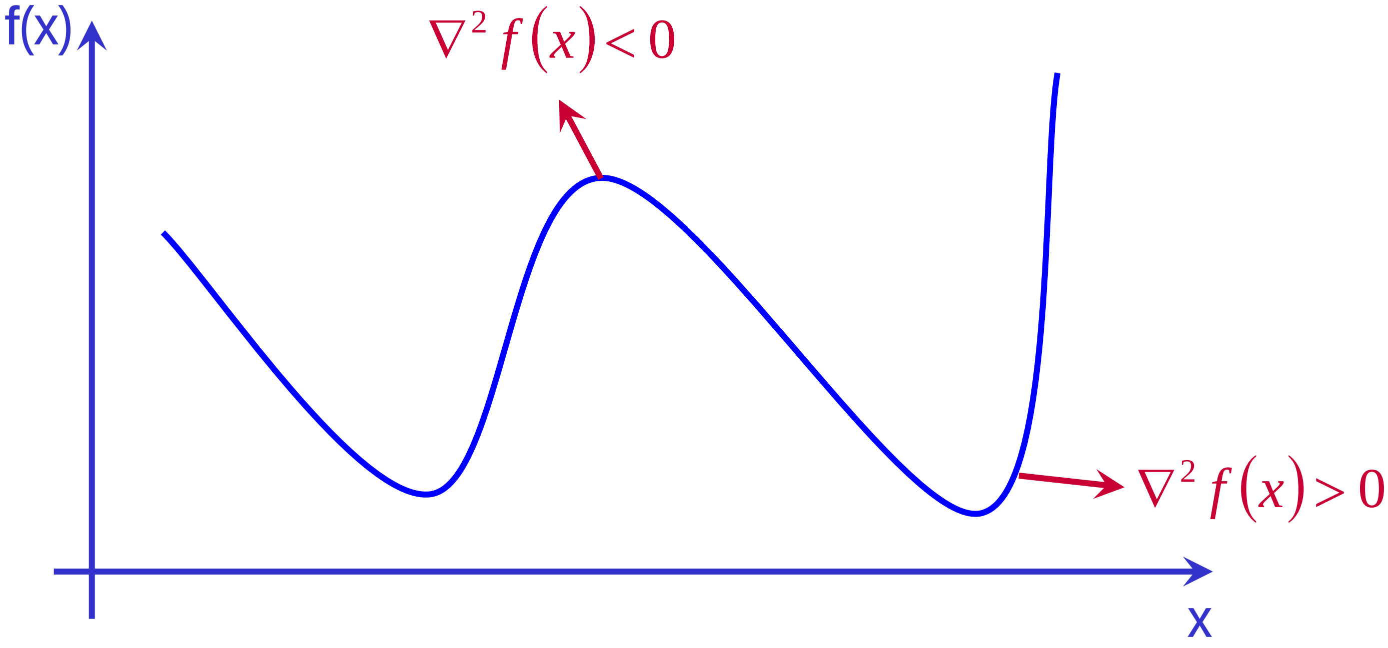
- Second-order sufficient condition for convexity
  - Not a necessary condition – convex function might not be smooth, and Hessian matrix might not exist

$$\nabla^2 f(X) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \end{bmatrix} \succ= 0$$

Hessian matrix is positive semi-definite for ALL X
(Hessian matrix depends on X)

Convex Function

- To guarantee convexity, Hessian matrix must be positive semi-definite for ALL X

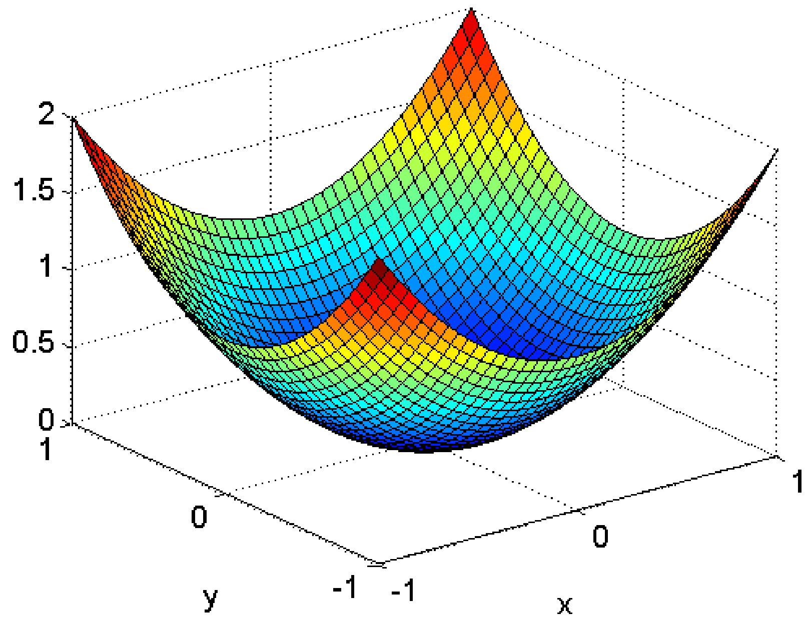


f(x) is NOT convex!

Convex Function

- A quadratic function $f(X) = X^T A_Q X + B_Q^T X + C_Q$ is convex if and only if A_Q is positive semi-definite

$$\nabla^2 f(X) = 2A_Q \quad \text{Constant}$$



$f(x, y) = x^2 + y^2$

Convex Function

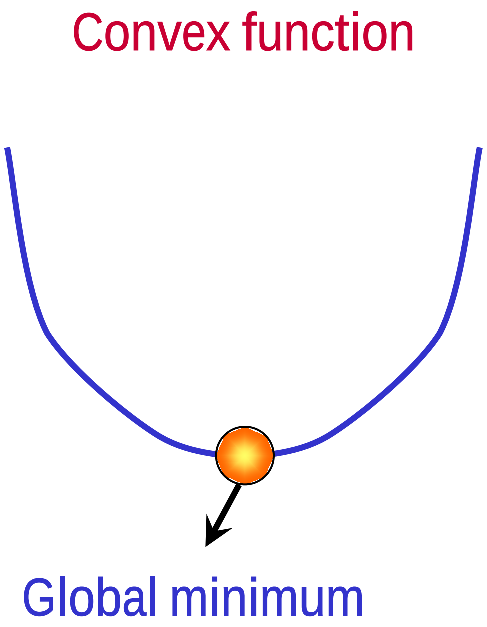
- Several popular examples of convex functions

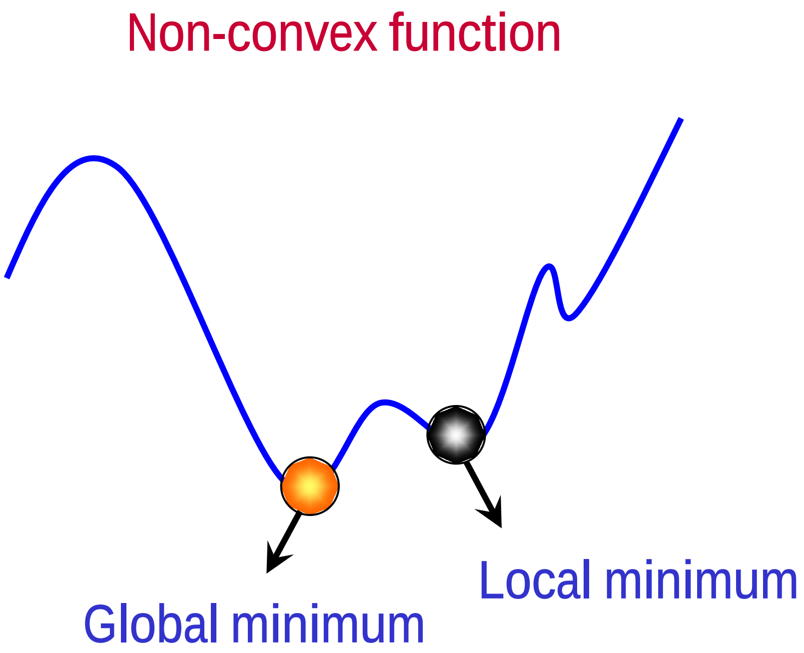
- One dimensional convex functions
  - Linear: $f(x) = bx + c$
  - Exponential: $f(x) = e^{ax}$
  - Power: $f(x) = x^a$ ($a < 0$ or $a > 1$, $x > 0$)

- N-dimensional convex functions
  - Linear: $f(X) = B^T X + C$
  - L_2-norm: $f(X) = ||X||_2$
  - Max: $f(X) = \max(x_1, x_2, \ldots, x_N)$
  - Log-sum-exp: $f(X) = \log(e^{x1} + e^{x2} + \ldots + e^{xN})$
  - Log-determinant: $f(X) = -\log[\det(X)]$ (X is positive definite)

Convex Function

- Minimizing a convex function is much easier than a general nonlinear programming
  - Convex functions do not contain local minima





Constrained Nonlinear Optimization

- The least-squares problem attempts to minimize a convex cost function without any constraints

- Many practical optimization problems contain both a cost function and a number of constraints
  - E.g., minimax optimization for regression

$$\min_{X,t} \quad t$$

Cost function

$$\text{S.T.} \quad \begin{cases} -t \le A(1,:)\cdot X - B_1 \le t \\ -t \le A(2,:)\cdot X - B_2 \le t \\ \vdots \\ -t \le A(M,:)\cdot X - B_M \le t \end{cases}$$

Constraints

Constrained Nonlinear Optimization

- A general nonlinear programming problem has the form of:

$$\min_{X} \quad f(X)$$

$$\text{S.T.} \quad \begin{cases} g_1(X) \le 0 \\ g_2(X) \le 0 \\ \vdots \end{cases}$$

- Equality constraints can be expressed in this general form

$$g(x) = 0 \quad \Rightarrow \quad \begin{cases} g(x) \le 0 \\ -g(x) \le 0 \end{cases}$$

Constrained Nonlinear Optimization

- A point X is feasible if it satisfies all constraints:

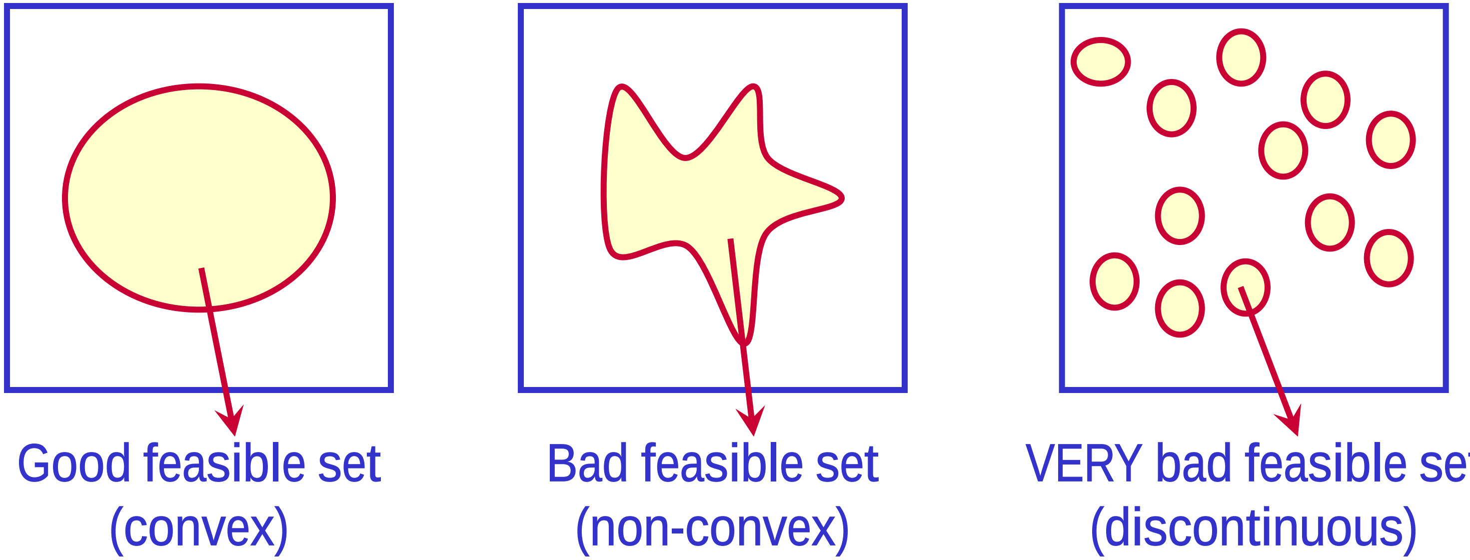
$$\begin{cases} g_1(X) \leq 0 \\ g_2(X) \leq 0 \\ \vdots \end{cases}$$

- The set of all feasible points is called the feasible set, or the constraint set

- An optimization is said to be feasible, if the corresponding feasible set is non-empty

Constrained Nonlinear Optimization

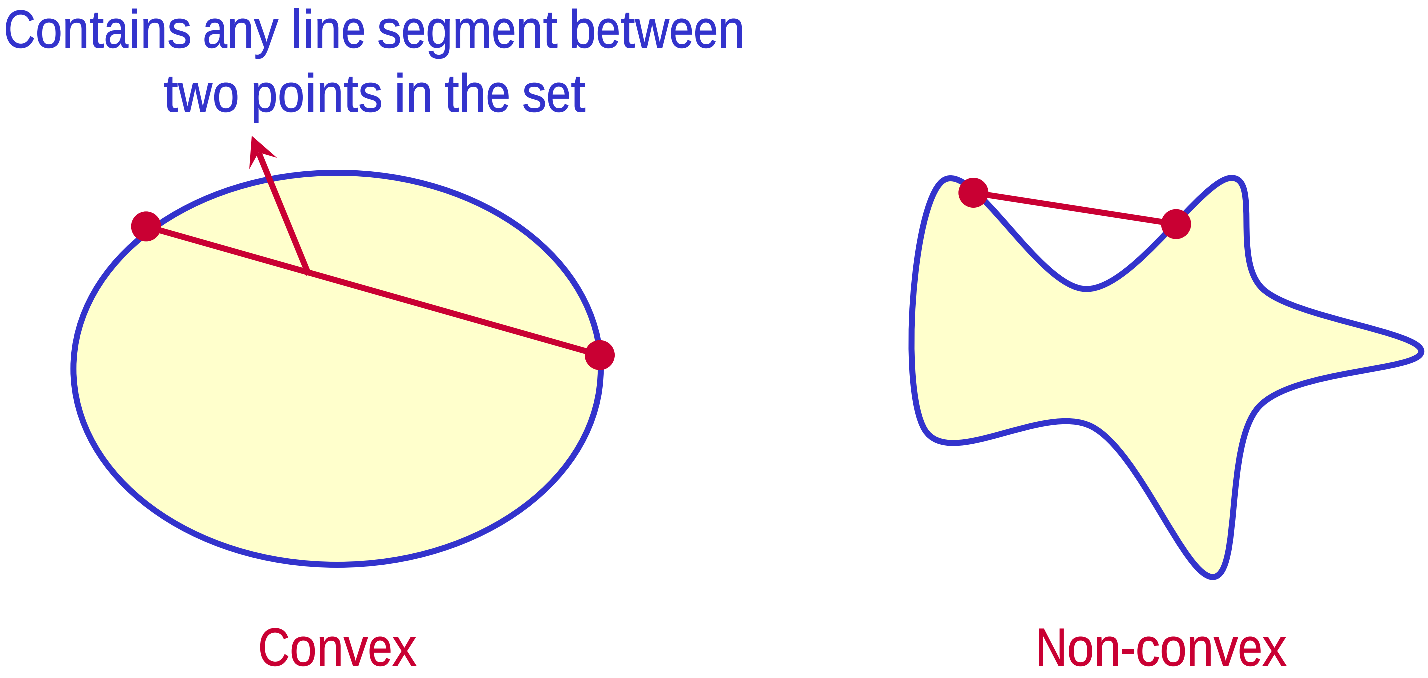
- Feasible set plays an important role in nonlinear optimization
- Even if the cost function is convex, nonlinear optimization can still be difficult given a “bad” feasible set



Convex Set

- A set D is convex, if for all X_1, $X_2 \in$ D and $0 \leq \boldsymbol{\alpha} \leq 1$, we have

$$\alpha \cdot X_1 + (1-\alpha) \cdot X_2 \in D$$



Convex Set

- Several popular examples of convex sets
  - Hyperplane: $\{X \mid B^T X = C\}$
  - Polytope: $\{X \mid B^T X \le C\}$
  - Ball: $\|X\|_2 \le C$
  - Positive semi-definite matrices (a non-trivial example):

$$\left\{X \middle| X \in R^{N\times N}, X = X^T, X \succ= 0\right\}$$

  - If X_1 and X_2 are positive semi-definite, their positive combination is also positive semi-definite

$$\alpha \cdot X_1 + (1-\alpha)\cdot X_2 \succ= 0$$

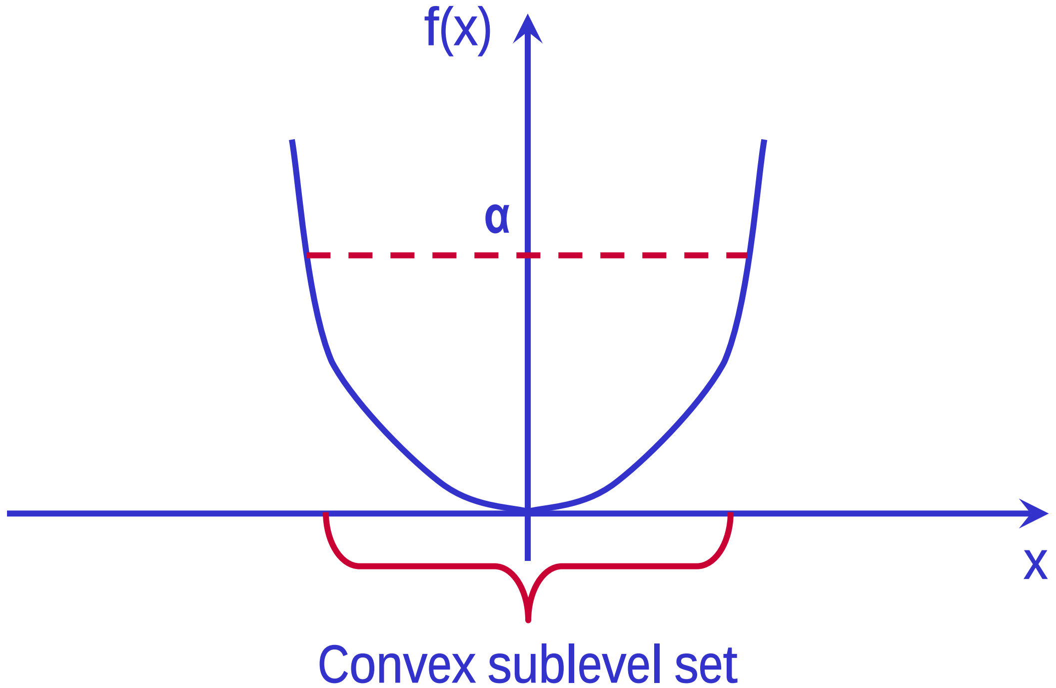
Positive coefficients

Convex Set

- Given a function f(X), the **α-sublevel set** is defined as:

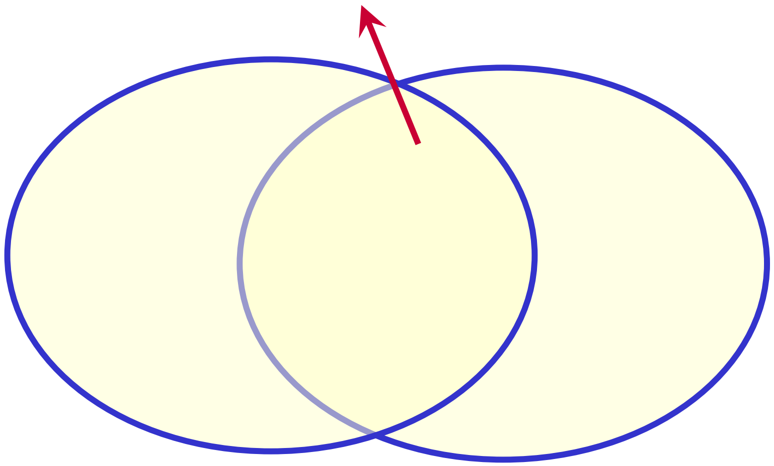
$$\{X \mid f(X) \le \alpha\}$$

- If f(X) is convex, its sublevel sets are convex



Convex Set

- Set convexity is preserved under intersection
  - If D_1 and D_2 are convex then $D_1 \cap D_2$ is convex



Convex Optimization

$$\min_{X} \quad f(X)$$

$$\text{S.T.} \quad \begin{cases} g_1(X) \leq 0 \\ g_2(X) \leq 0 \\ \quad \vdots \end{cases}$$

- If all $g_i(X)$'s are convex, the constraint set is convex
  - Constraint set is the intersection of all convex 0-sublevel sets

- The minimization of a convex cost function over a convex constraint set is called convex optimization

Convex Optimization

- The following optimizations are NOT convex, even if f(X) and g(X) are both convex

$$\begin{aligned} &\max_{X} \quad f(X) \\ &S.T. \quad g(X) \le 0 \end{aligned}$$

Maximizing a convex function is not a convex optimization

$$\begin{aligned} &\min_{X} \quad f(X) \\ &S.T. \quad g(X) \ge 0 \end{aligned}$$

Constraint set is not convex

Convex Optimization

- Linear programming is a special case of convex optimization

- Most convex optimization with smooth cost function and constraints can be efficiently and robustly solved
  - Decide if the optimization is feasible or infeasible
  - If feasible, provide the optimal solution

- Several good convex solvers
  - MOSEK (www.mosek.com)
  - CVX (www.stanford.edu/~boyd/cvx/)
  - More details on convex solver in future lectures...

Summary

- Convex analysis
  - Convex function
  - Convex set
  - Convex optimization