535520: Optimization Algorithms

Lecture 1 — Fundamentals

Ping-Chun Hsieh (謝秉均)

September 2, 2024

This Lecture

1. Optimization Problems: Formulation and Terminology

2. Optimality Conditions

3. Subgradients and Subdifferentials

Reading material:

- Chapters 1.1 and 2.1 of Dimitri Bertsekas's textbook "Nonlinear Programming"
- Chapters 2 and 3 of Stephen Boyd's textbook "Convex Optimization"

Basic Formulation of an Optimization Problem

Variable / parameter

Objective function

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

$$\text{subject to} \quad x \in A \subseteq X$$

Feasible set / constraints

Domain

- Nice properties of an objective function

  Continuity?

  Smoothness?

  Convexity?

  Differentiability?

  Separability?

- Nice properties of a feasible set

  Compactness?

  Convex sets?

  Unconstrained?

  Linear constraints?

  Discrete?

A Motivating Example: Portfolio Selection

- Portfolio selection (in hindsight)
  - Let $r_t = (r_{t,1}, \cdots, r_{t,n})$ denote "price ratio" of the n assets at each time t
  - Suppose initially we have total wealth $w > 0$
  - We want to choose an initial portfolio allocation vector a in hindsight such that the total wealth after T iterations is maximized

    Question: How to write down the optimization problem (e.g., objective function, constraints)?

Optimization in ML and Beyond

Machine Learning

Empirical Risk Minimization

Online Learning

Federated Learning

RL and Robotics

Policy Optimization

Adaptive Control

Learning From Demonstrations

Deep Learning

Seq2Seq

GANs

Representation Learning

Information Processing

Image Processing

Speech Signal Processing

Data compression

Computational Science

Physics

Chemistry

Bioinformatics

Network Optimization

Network Utility Maximization

Social Networks

Packet Scheduling

Optimization: 3 Questions to Answer

1. Characterization: Sufficient / necessary conditions of an optimal solution?

(Our focus today)

2. Algorithms: Iterative algorithms that find an optimal solution?

3. Convergence: Do the iterates converge to an optimum? How fast?

Optimality Conditions

(Structural information about optimal solutions)

Optimality Conditions (Necessary / Sufficient)

Unconstrained cases:

C1. FONC: First-order necessary conditions for local optimality

C2. SONC: Second-order necessary conditions for local optimality

C3. FOSC: First-order sufficient conditions for global optimality

C4. SOSC: Second-order sufficient conditions for local optimality

Constrained cases:

C5. FONC-C: First-order necessary conditions for constrained local optimality

C6. FOSC-C: First-order sufficient conditions for constrained global optimality

Non-differentiable cases:

C7. Fermat's Rule

Notation & Assumptions for This Lecture

Unless stated otherwise:

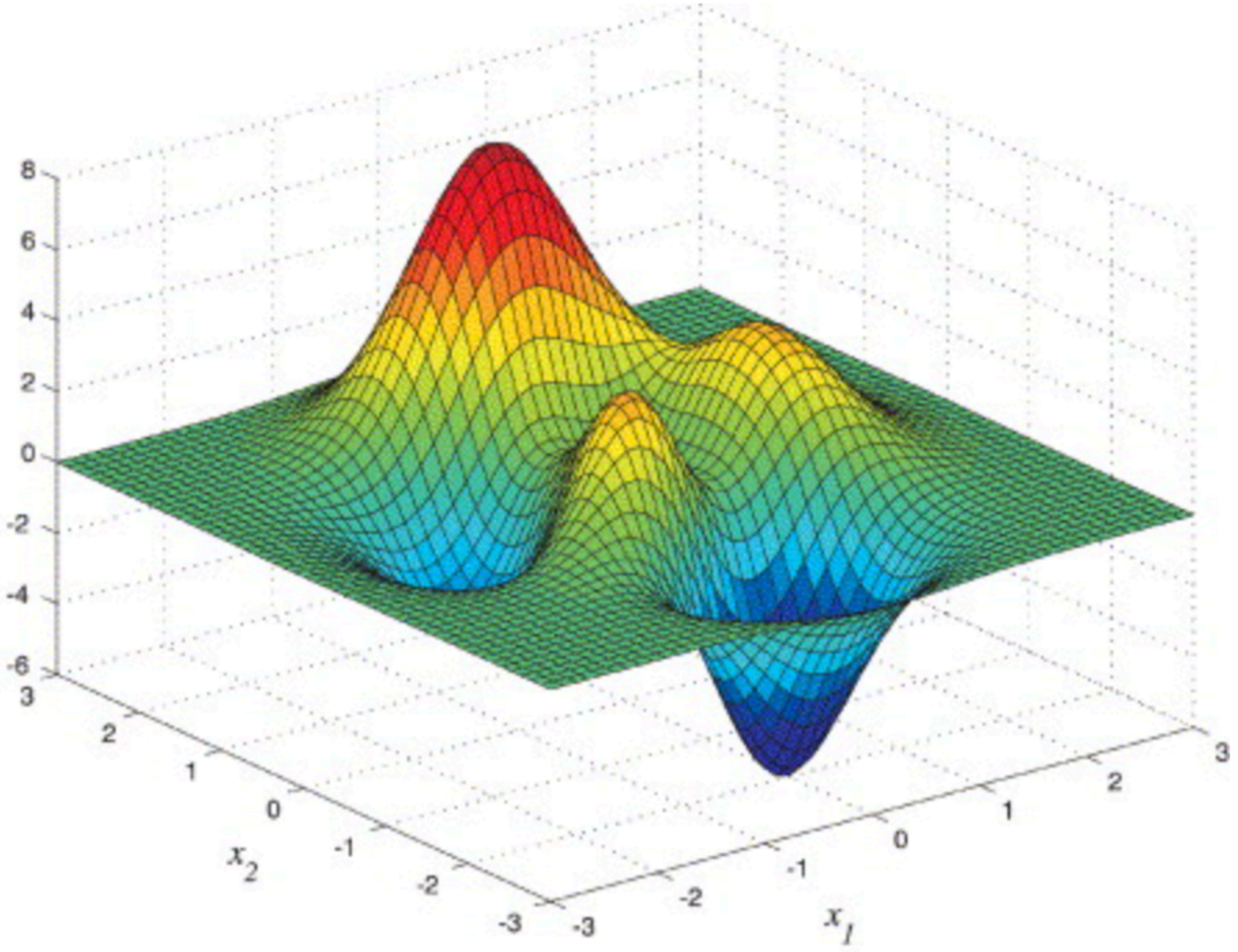
- $\|\cdot\|_p$ denotes the ℓ_p norm

- $\|\cdot\| \equiv \|\cdot\|_2$ denotes the Euclidean norm

- We focus on multivariate single-objective minimization problems (i.e., $f : \mathbb{R}^n \to \mathbb{R}$)

- The objective function is assumed differentiable

Local and Global Minima



Intuitively, let's make some observations:

- Where is the domain X?

- Where is the global minimizer?

- How about local minimizer(s)?

Local and Global Minima (Formally)

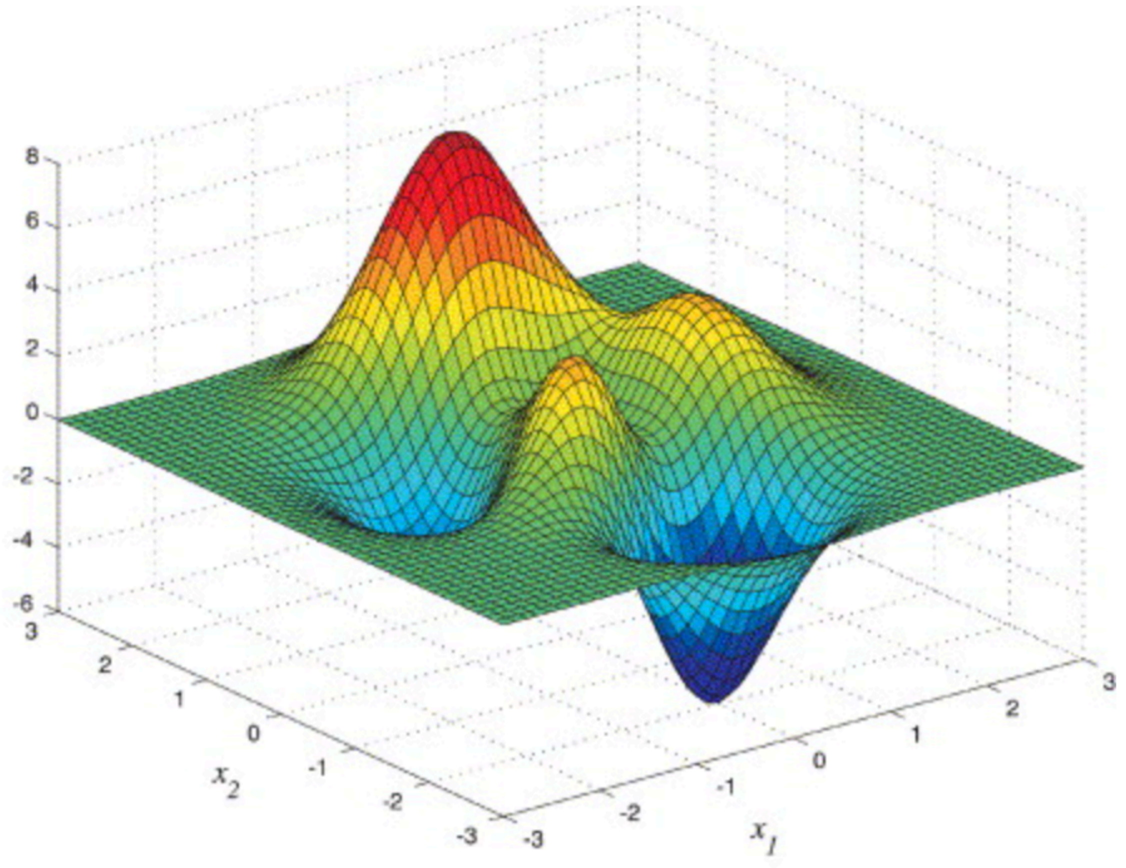
Definition: Given $f : X \to \mathbb{R}$, a vector $x^* \in X$ is a *local minimizer* if there exists some $\epsilon > 0$ such that

$$f(x^*) \leq f(x), \text{ for all } x \in X \text{ with } \|x - x^*\| < \epsilon$$

Definition: Given $f : X \to \mathbb{R}$, a vector $x^* \in X$ is a *global minimizer* if

$$f(x^*) \leq f(x), \text{ for all } x \in X$$

Remark: *Strict* local / global minimizers if "$\leq$" is replaced by "$<$" for all $x \neq x^*$



A Quick Recap of Notations, Calculus, and Linear Algebra (1/3)

Given $f : \mathbb{R}^n \to \mathbb{R}, x = (x_1, \cdots, x_n)$

- Gradient vector

$\nabla f(x)$

- Hessian matrix $\nabla^2 f(x)$

A real square matrix $H \in \mathbb{R}^{n \times n}$ is said to be

- Symmetric if:
- Positive semidefinite (psd), or $H \succcurlyeq 0$, if:
- Positive definite (pd), or $H \succ 0$ if:

- **Useful Properties:**

(1) $H \succcurlyeq 0$ if and only if all its eigenvalues are non-negative

(2) $H \succ 0$ if and only if all its eigenvalues are positive

(When discussing psd/pd, we can assume H is symmetric, without loss of generality)

A Quick Recap of Notations, Calculus, and Linear Algebra (2/3)

In this course, we will leverage Taylor's Theorem a lot! (for both intuition and analysis)

Taylor's Theorem (First-Order Version): Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be *continuously differentiable* on an open neighborhood S of a vector x.

(1) For all δ such that $x + \delta \in S$, we have

$$f(x + \delta) = f(x) + \delta^\top \nabla f(x) + o(\|\delta\|)$$

(2) For all δ such that $x + \delta \in S$, there exists $\alpha \in [0,1]$ such that

$$f(x + \delta) = f(x) + \delta^\top \nabla f(x + \alpha\delta)$$

Question: Why do we need "continuous differentiability"?

A Quick Recap of Notations, Calculus, and Linear Algebra (3/3)

Higher-order version of Taylor theorem:

Taylor's Theorem (Second-Order Version): Let $f : \mathbb{R}^n \to \mathbb{R}$ be *twice continuously differentiable* on an open neighborhood S of a vector x.

(1) For all δ such that $x + \delta \in S$, we have

$$f(x+\delta) = f(x) + \delta^\top \nabla f(x) + \frac{1}{2}\delta^\top \nabla^2 f(x)\delta + o(\|\delta\|^2)$$

(2) For all δ such that $x + \delta \in S$, there exists $\alpha \in [0,1]$ such that

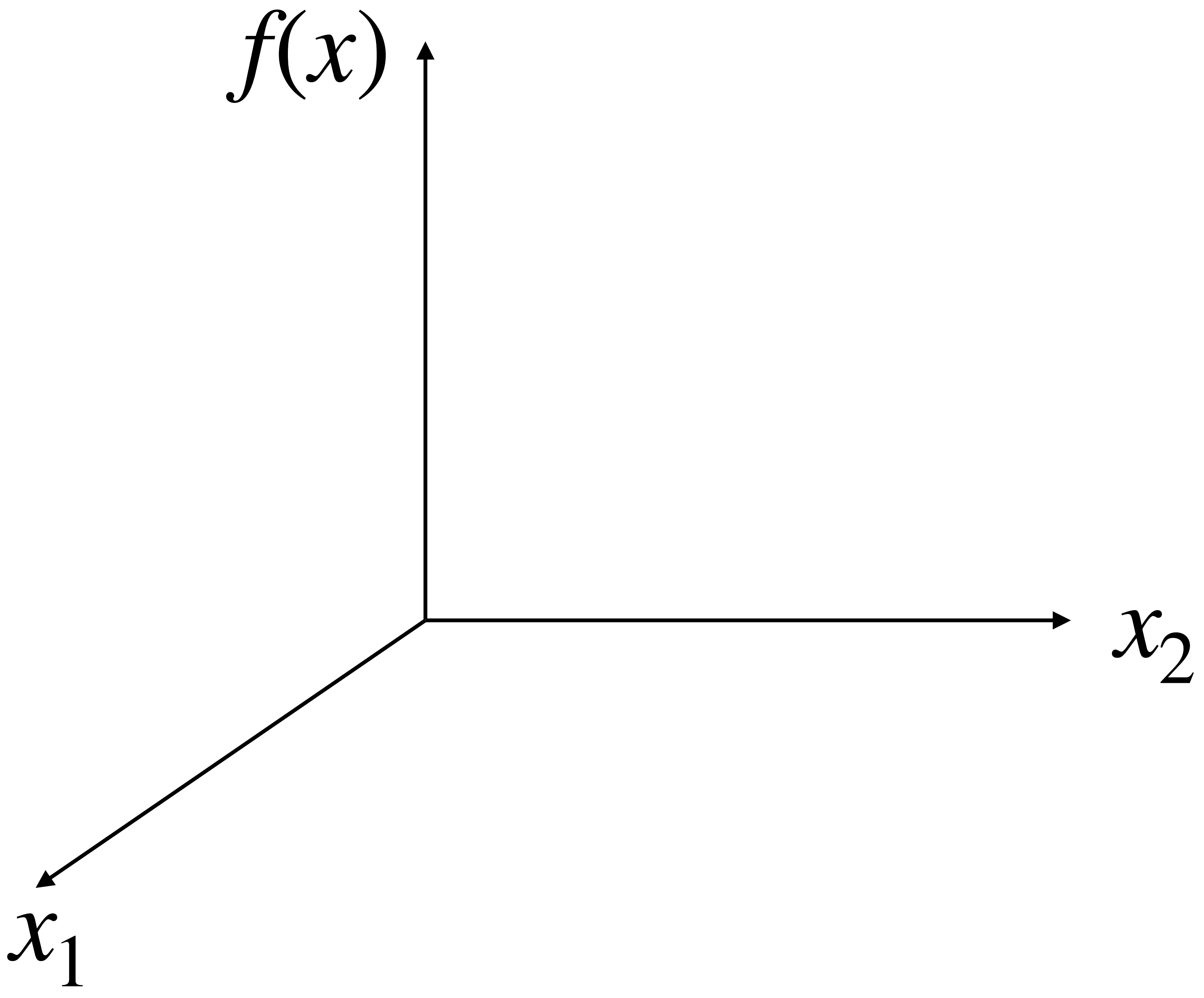
$$f(x+\delta) = f(x) + \delta^\top \nabla f(x) + \frac{1}{2}\delta^\top \nabla^2 f(x+\alpha\delta)\delta$$

C1. Necessary Conditions for Local Optimality: Unconstrained

Theorem (First-Order Necessary Condition, FONC): Let $f : \mathbb{R}^n \to \mathbb{R}$ be *continuously differentiable* on a neighborhood of x^*, which is a local minimizer. Then, we have $\nabla f(x^*) = 0$.

Intuition / Informal Proof:

$$f(x^* + \Delta x) - f(x^*) \approx \nabla f(x^*)^\top \Delta x =$$

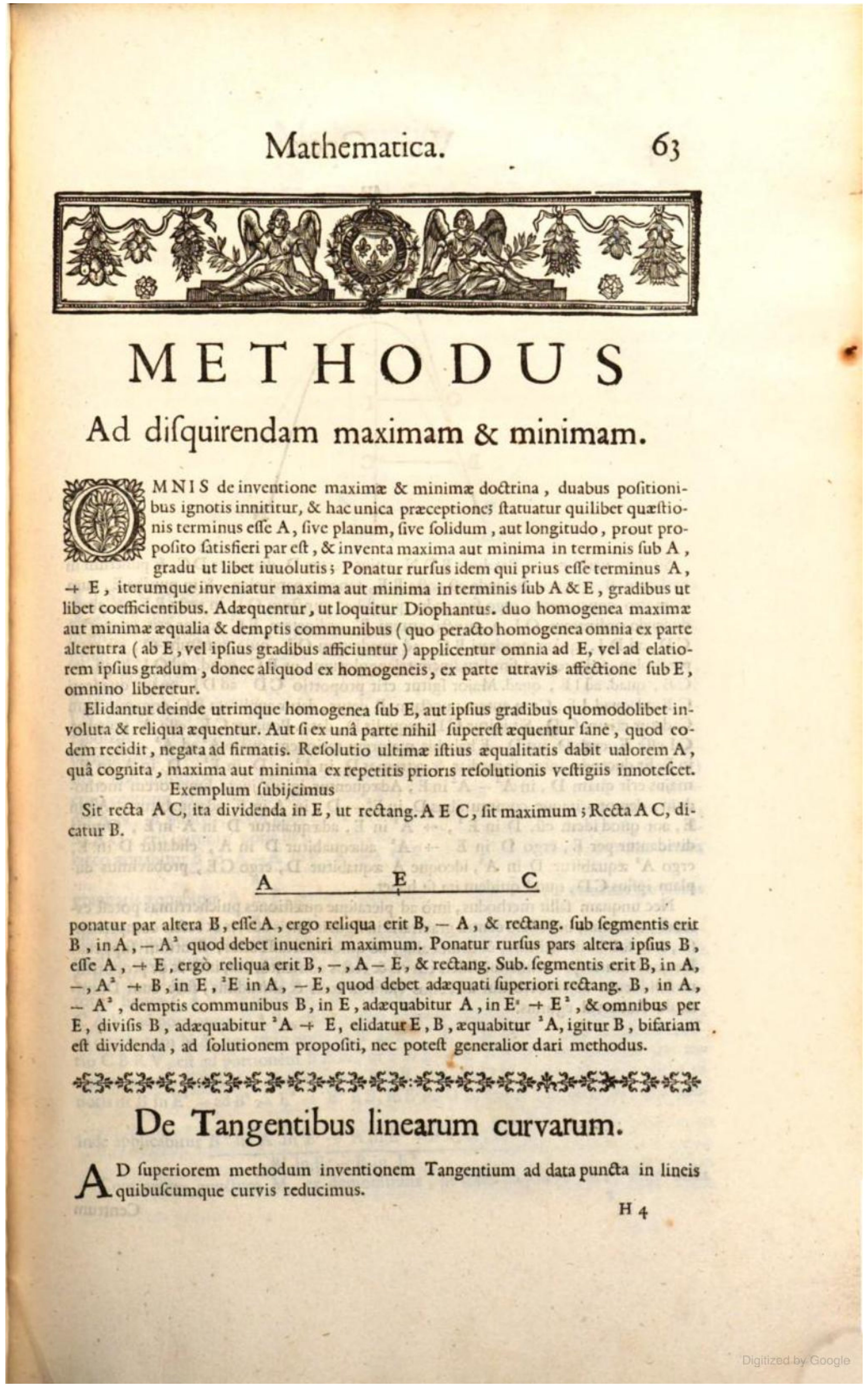


C1. Necessary Conditions for Local Optimality: Unconstrained

Theorem (First-Order Necessary Condition, FONC): Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on a neighborhood of x^*, which is a local minimizer. Then, we have $\nabla f(x^*) = 0$.

Proof: Construct a function $g(t) = f(x^* + td)$, where $d \in \mathbb{R}^n$ and $t > 0$

$$\lim_{t \downarrow 0} \frac{f(x^* + td) - f(x^*)}{t} =$$



METHODUS

Ad diſquirendam maximam & minimam.

OMNIS de inventione maximæ & minimæ doctrina, duabus poſitionibus ignotis innititur, & hac unica præceptione; ſtatuatur quilibet quæſtionis terminus eſſe A, ſive planum, ſive ſolidum, aut longitudo, prout propoſito ſatisfieri par eſt, & inventa maxima aut minima in terminis ſub A, gradu ut libet iuuolutis; Ponatur rurſus idem qui prius eſſe terminus A, + E, iterumque inveniatur maxima aut minima in terminis ſub A & E, gradibus ut libet coefficientibus. Adæquentur, ut loquitur Diophantus. duo homogenea maximæ aut minimæ æqualia & demptis communibus (quo peracto homogenea omnia ex parte alterutra (ab E, vel ipſius gradibus afficiuntur) applicentur omnia ad E, vel ad elatiorem ipſius gradum, donec aliquod ex homogeneis, ex parte utravis affectione ſub E, omnino liberetur.

Elidantur deinde utrimque homogenea ſub E, aut ipſius gradibus quomodolibet involuta & reliqua æquentur. Aut ſi ex unâ parte nihil ſupereſt æquentur ſane, quod eodem recidit, negata ad firmatis. Reſolutio ultimæ iſtius æqualitatis dabit ualorem A, quâ cognita, maxima aut minima ex repetitis prioris reſolutionis veſtigiis innoteſcet.

Exemplum ſubijcimus

Sit recta A C, ita dividenda in E, ut rectang. A E C, ſit maximum; Recta A C, dicatur B.

A E C

ponatur par altera B, eſſe A, ergo reliqua erit B, – A, & rectang. ſub ſegmentis erit B, in A, – A² quod debet inueniri maximum. Ponatur rurſus pars altera ipſius B, eſſe A, + E, ergò reliqua erit B, –, A – E, & rectang. Sub. ſegmentis erit B, in A, –, A² + B, in E, ²E in A, – E, quod debet adæquati ſuperiori rectang. B, in A, – A², demptis communibus B, in E, adæquabitur A, in E² + E², & omnibus per E, diviſis B, adæquabitur ²A + E, elidatur E, B, æquabitur ²A, igitur B, bifariam eſt dividenda, ad ſolutionem propoſiti, nec poteſt generalior dari methodus.

De Tangentibus linearum curvarum.

AD ſuperiorem methodum inventionem Tangentium ad data puncta in lineis quibuſcumque curvis reducimus.

An Interesting Fact:

- This necessary condition was originally formulated by Fermat in 1637, *without proof* (as expected!)

Pierre de Fermat
(1607-1665)

Remarks on FONC

- The condition $\nabla f(x^*) = 0$ is necessary but not sufficient for local optimality (any counterexample?)

- Despite this, the condition $\nabla f(x^*) = 0$ is still useful as it provides a candidate set of locally optimal solutions.

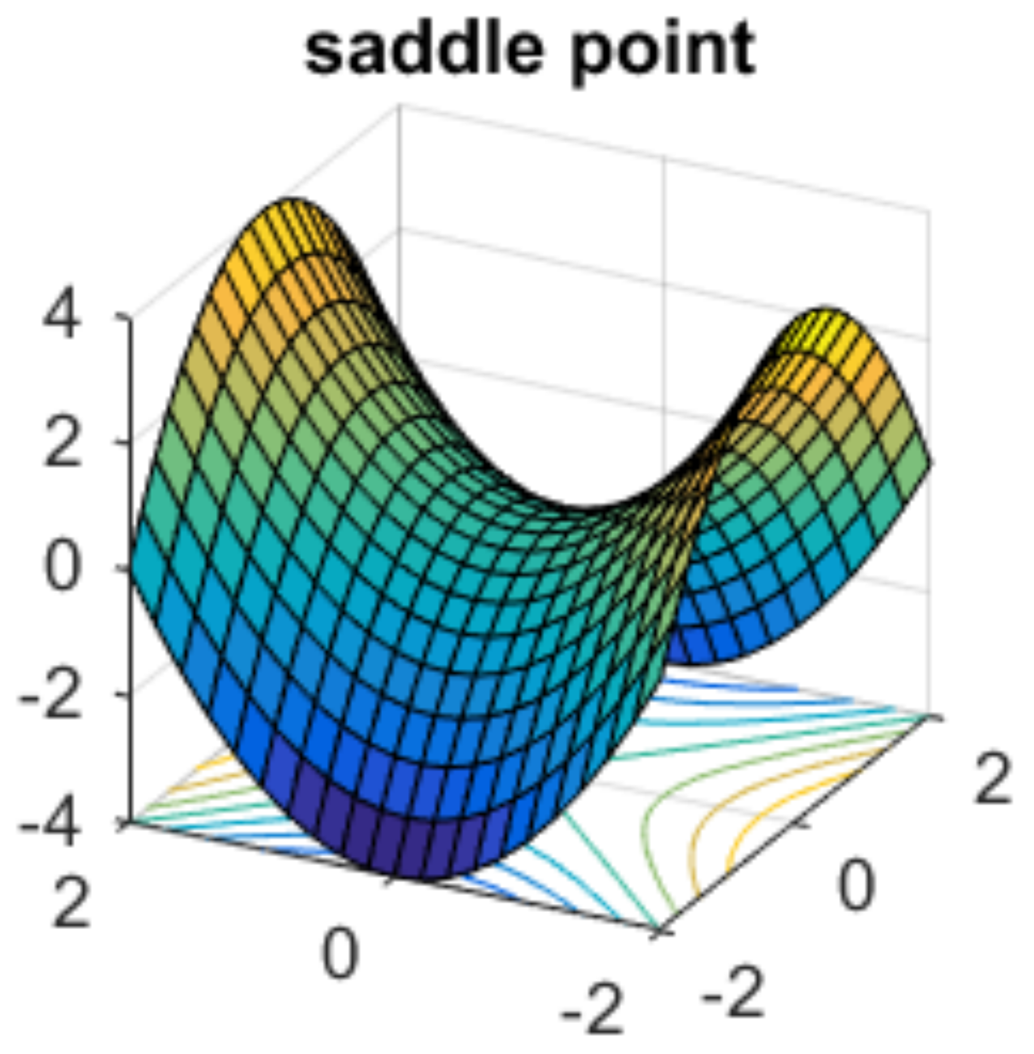
- **Terminology**: A point x with $\nabla f(x) = 0$ is termed as a "stationary point" or a "critical point" in the optimization literature.

Critical Points and Saddle Points

Given a differentiable $f : X \to \mathbb{R}$:

Definition: A vector $x_0 \in X$ is a *critical point* if $\nabla f(x)|_{x=x_0} = 0$

Definition: A vector $x_0 \in X$ is a *saddle point* if $\nabla f(x)|_{x=x_0} = 0$ and x_0 is not a local minimizer nor a local maximizer



- The existence of saddle point suggests that $\nabla f(x^*) = 0$ is not sufficient for local optimality

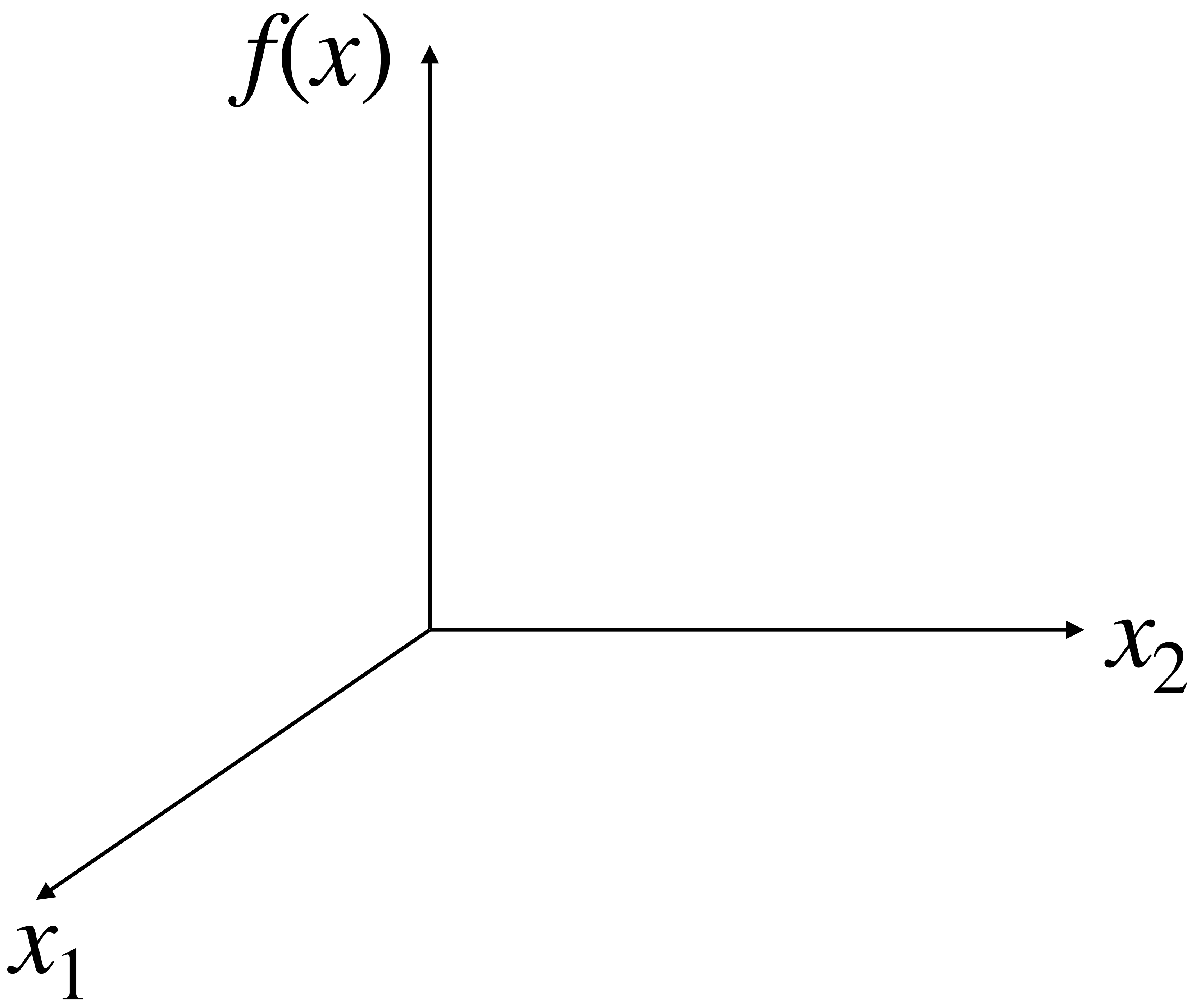
C2. Second-Order Necessary Condition: Unconstrained Cases

Theorem (Second-Order Necessary Condition, SONC): Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable on a neighborhood of x^*, and x^* is a local minimizer. Then, in addition to $\nabla f(x^*) = 0$, we must also have

$$\nabla^2 f(x^*) \succcurlyeq 0$$

Intuition / Informal Proof:

$$f(x^* + \Delta x) - f(x^*) \approx \nabla f(x^*)^\top \Delta x + \frac{1}{2} \Delta x^\top \nabla^2 f(x^*) \Delta x$$



Proof: You will be asked to prove this in HW0 :)

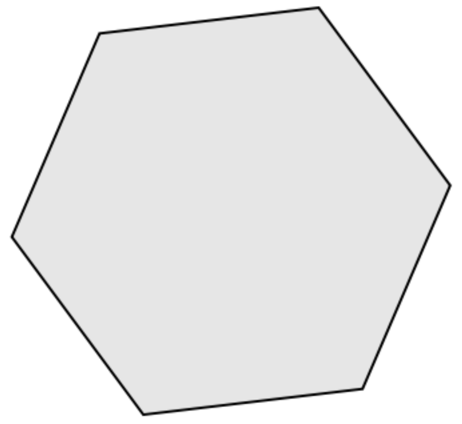
Next question: Are these conditions *sufficient*?

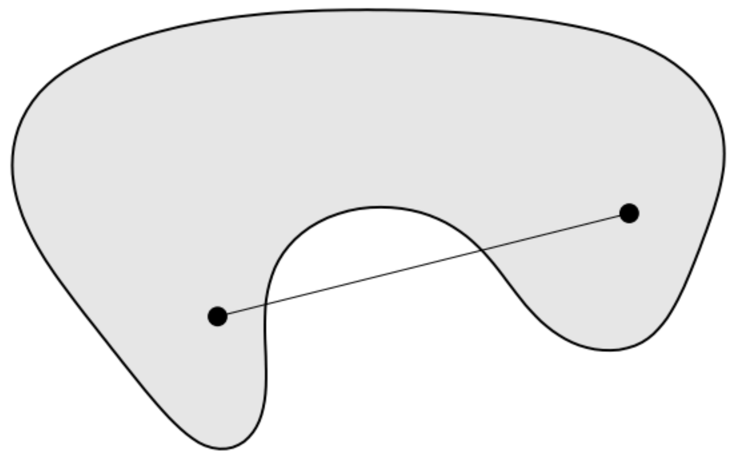
(If so, under what conditions?)

Convex Sets

Definition: A set $S \subset \mathbb{R}^n$ is called convex if for any $x, y \in S$, the line segment $\{\alpha x + (1-\alpha)y, \alpha \in [0,1]\}$ is also in S.

Examples:





- Convex hull: Let $x_1, \cdots, x_k \in \mathbb{R}^d$. The convex hull $\mathrm{CH}(x_1, \cdots, x_k) := \{\sum_i \alpha_i x_i : \alpha_i \geq 0, \sum_i \alpha_i = 1\}$
- Halfspace: $\{x : a^\top x \leq b\}$
- Hyperplane: $\{x : a^\top x = b\}$
- Ellipsoid: $\{x : (x-a)^\top A(x-a) \leq 1\}$
- Probability simplex: $\{x : x \geq 0, \sum_i x_i = 1\}$

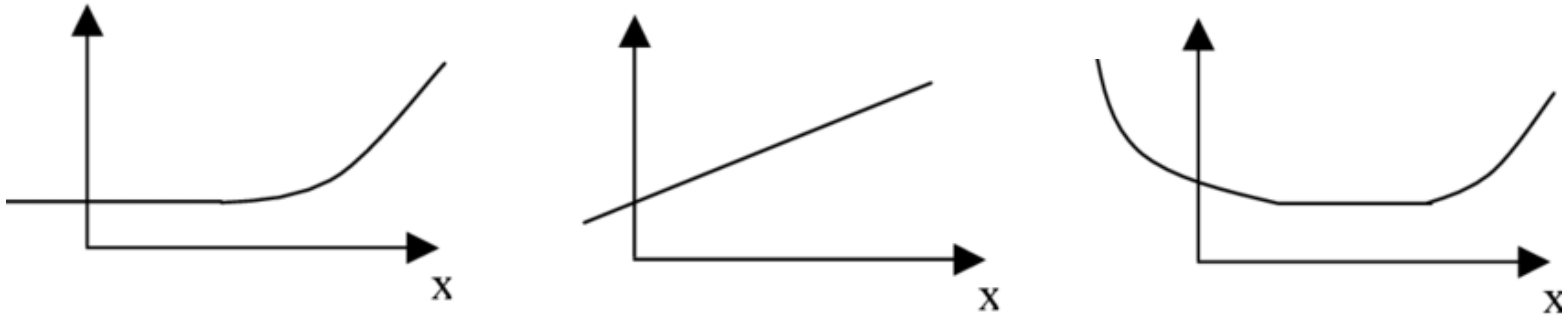
(For more properties of convex sets, please check Stephen Boyd's lecture slides)

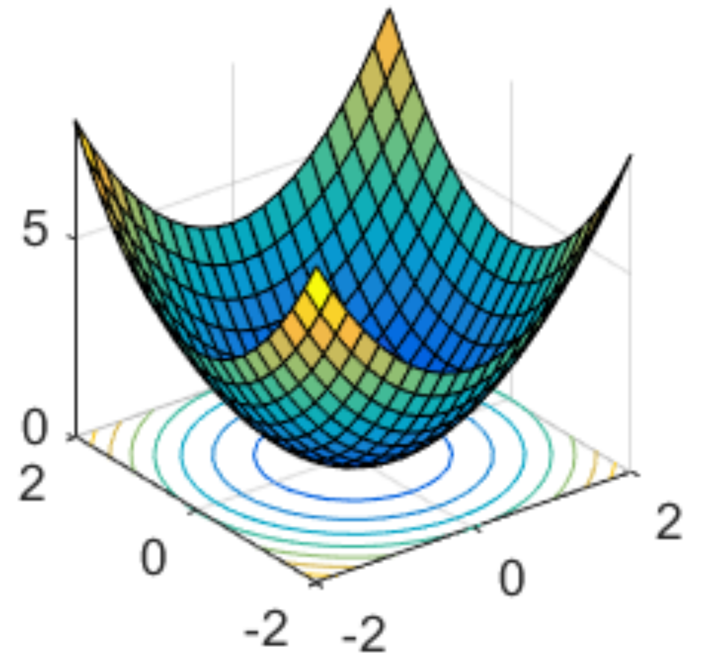
Convex and Concave Functions

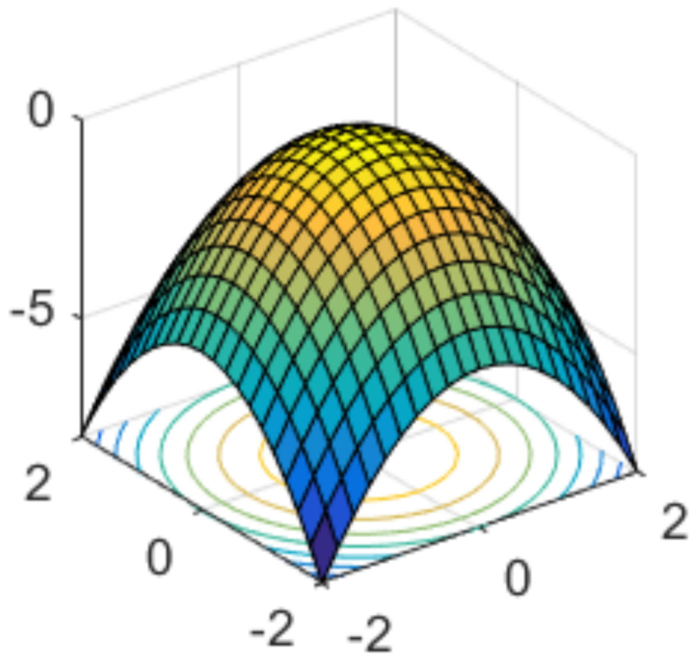
Definition: A function $f : X \to \mathbb{R}$ is called a *convex function* if its domain X is a convex set and for any $x, y \in X$ and any $\alpha \in [0,1]$, we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Intuition: "All the segments lie above the function"







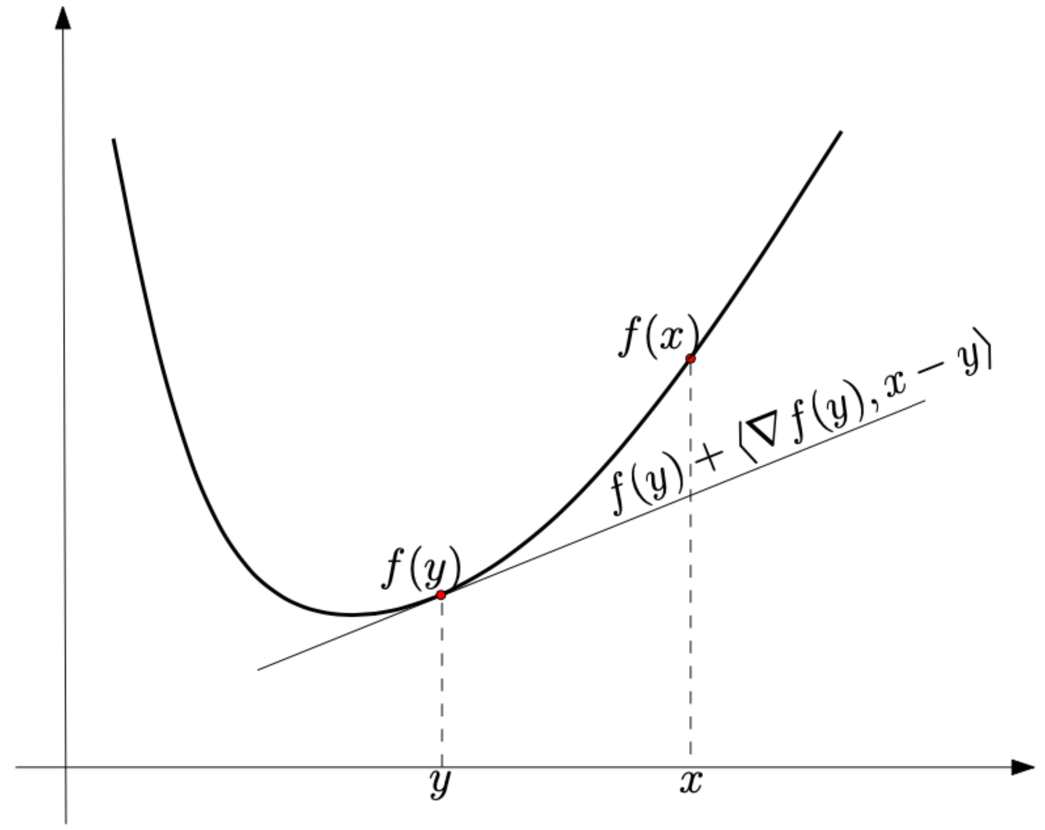
Remark: A function h is called *concave* if $-h$ is convex

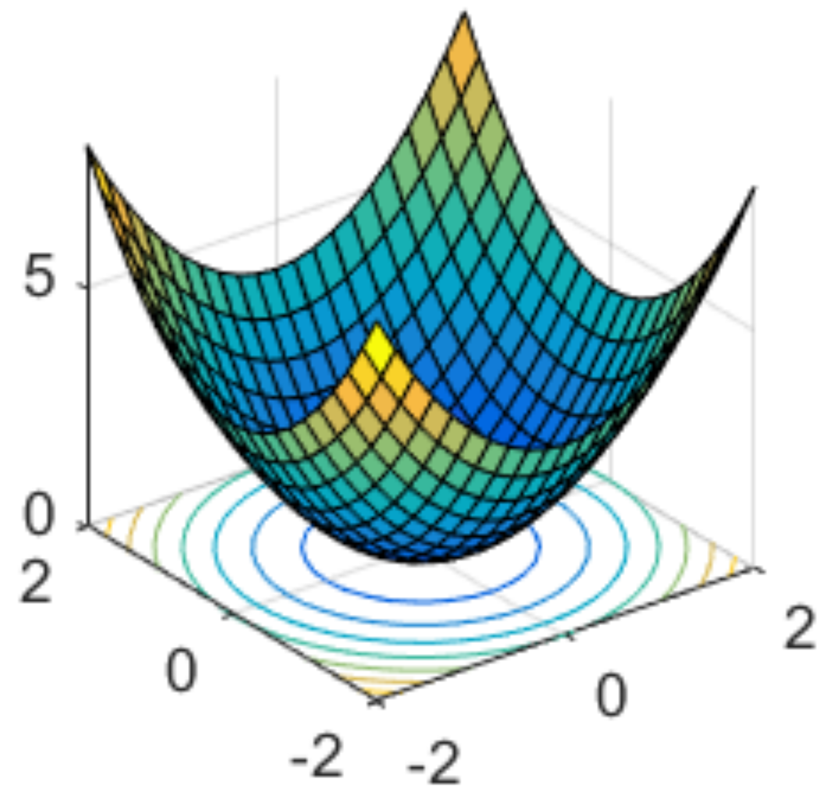
Question: Why do we need the domain X to be convex?

Characterizing Convex Functions Under Differentiability

- If $f : X \to \mathbb{R}$ is differentiable, then f is convex if and only if X is a convex set and $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$, for all $x, y \in X$. (a.k.a. first-order condition of convexity)

Intuition: "The function lies above the tangent"





- If $f : X \to \mathbb{R}$ is twice differentiable, then f is convex if and only if X is a convex open set and $\nabla^2 f(x) \succcurlyeq 0$, for all $x \in X$.

Exercise: Prove the above two properties

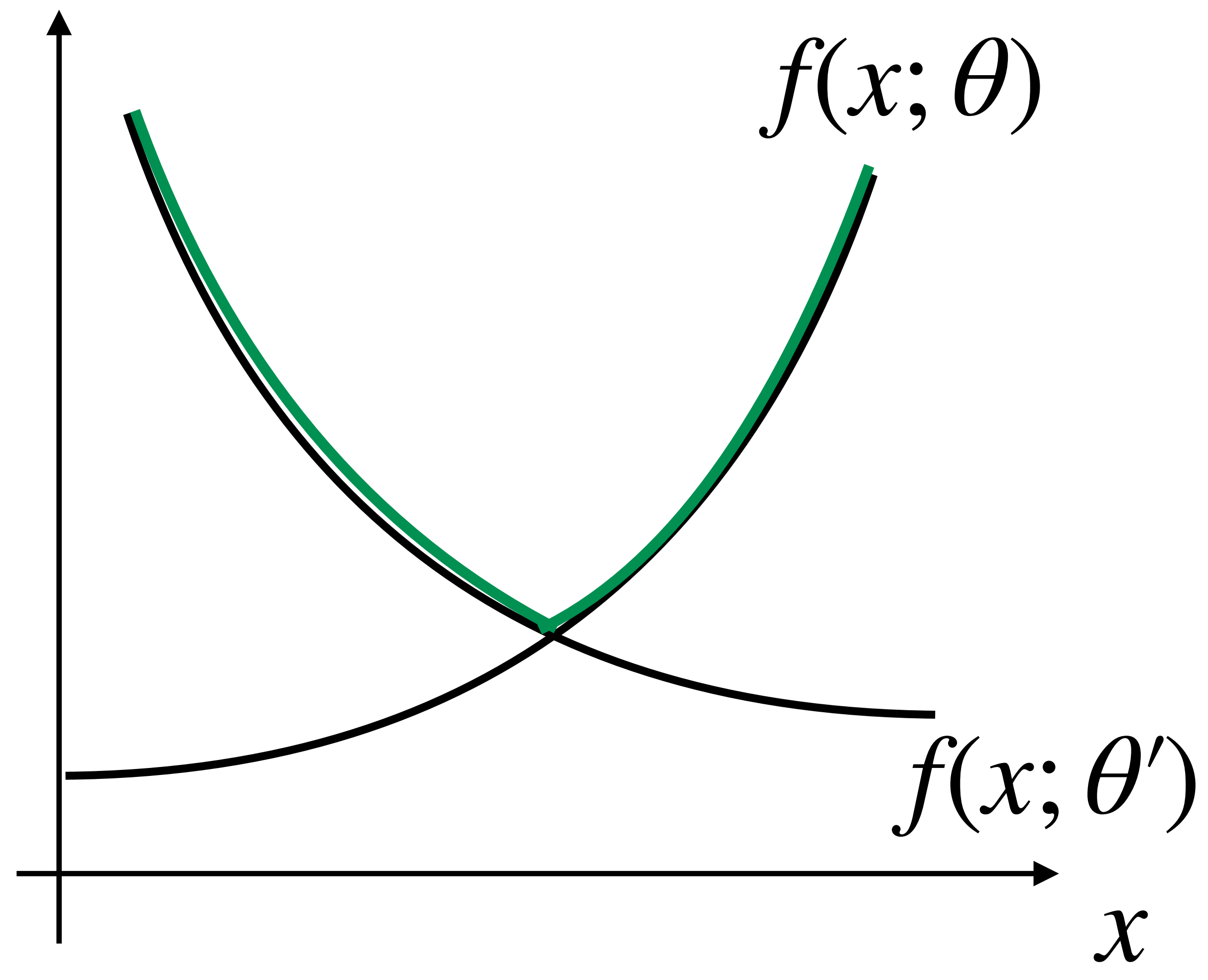
Example: Pointwise Maximum of Convex Functions

The pointwise maximum of a family of convex functions is still convex

- Let $f(x;\theta)$ be a convex function of x for every $\theta \in \Theta$, where Θ is an arbitrary index set.

Define $$F(x) := \max_{\theta \in \Theta} f(x;\theta)$$

- Question: Is $F(x)$ a convex function?



Popular Examples of Convex Functions

- Quadratic functions: $f(x) = \frac{1}{2} x^\top P x + q^\top x + r$, where $P \succcurlyeq 0$

- Negative entropy: $f(x) = \sum_{i=1}^{n} x_i \log x_i$, where x is a probability vector

- Log-sum-exp: $f(x) = \log\left(\exp(x_1) + \cdots + \exp(x_n)\right)$

- Log-determinant of pd matrices: $f(X) = \log \det(X)$

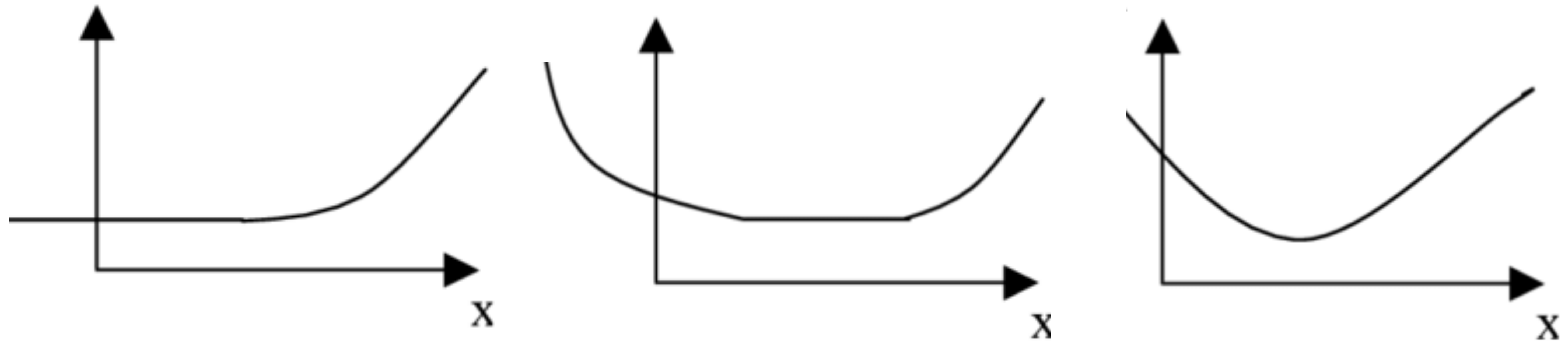
You will be asked to verify this in HW0 :)

Strictly Convex Functions

Definition: A function $f : X \to \mathbb{R}$ is called strictly convex if its domain X is a convex set and for any $x, y \in X$ with $x \neq y$ and any $\alpha \in (0,1)$, we have

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$$

Intuition: "All the line segments lie **strictly** above the function"



Characterizing Strictly Convex Functions Under Differentiability

- If $f : X \to \mathbb{R}$ is twice differentiable, then f is strictly convex if X is a convex set and $\nabla^2 f(x) \succ 0$, for all $x \in X$.

Remark: The condition $\nabla^2 f(x) \succ 0$ is only sufficient but not necessary for strict convexity (any counterexample?)

Remark: We will mention a related concept "strong convexity" when discussing gradient descent in Lecture 4

Nice Properties of Convex Functions

- Fact 1: If f is convex over X, then a local minimum is also a global minimum

(This makes convex optimization special!)

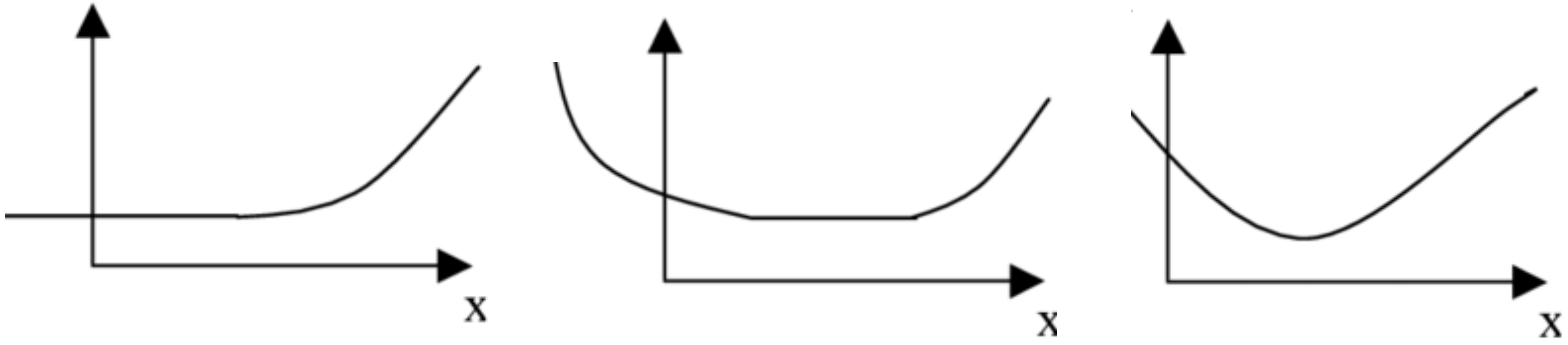
- Fact 2: If f is strictly convex over X, then there exists at most one global minimum

(Could you explain these by the definition of strict convexity?)

C3. First-Order Sufficient Condition for Global Optimality (Intuition)

Theorem (FOSC): If $f : X \to \mathbb{R}$ is *convex* and the set X is convex, then $\nabla f(x^*) = 0$ is sufficient for $x^* \in X$ to be a global minimizer.

Intuition:



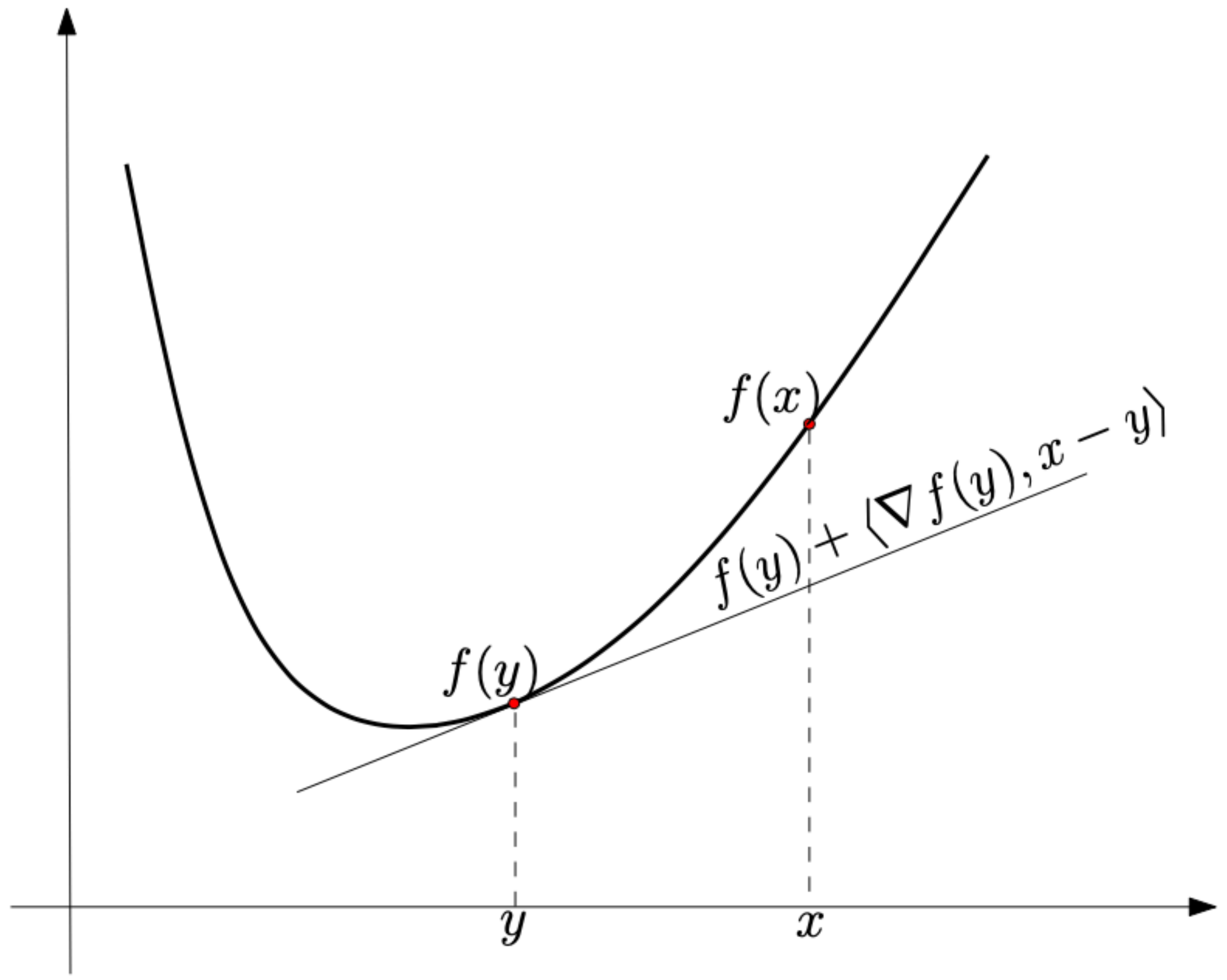
C3. First-Order Sufficient Conditions for Global Optimality (Formally)

Theorem (FOSC): If $f : X \to \mathbb{R}$ is *convex* and the set X is convex, then $\nabla f(x^*) = 0$ is sufficient for $x^* \in X$ to be a global minimizer.

Moreover, if X is also an open set, then $\nabla f(x^*) = 0$ is both necessary and sufficient for $x^* \in X$ to be a global minimizer.

Proof: Let $x^* \in X$ be a global minimizer. Then, by convexity, we have

$$f(x) \geq f(x^*) + \nabla f(x^*)^\top (x - x^*), \quad \text{for all } x \in X$$



Question: Why is "openness of domain X" needed?

C4. Second-Order Sufficient Condition (SOSC) for Local Optimality

Theorem (SOSC): Let $f : X \to \mathbb{R}$ be twice continuously differentiable. Then, $x^* \in X$ is a strict local minimizer of f if x^* satisfies:
(i) $\nabla f(x^*) = 0$ and (ii) $\nabla^2 f(x^*) \succ 0$.

Proof: $f(x^* + d) - f(x^*) = \nabla f(x^*)^\top d + \frac{1}{2} d^\top \nabla^2 f(x^*) d + o(\|d\|^2)$, for all $d \in \mathbb{R}^n$

Set of Optimal Solutions

- Fact 1: The set of optimal solutions (denoted by X^*) may be empty

Example: If the domain X is empty, then X^* is surely empty

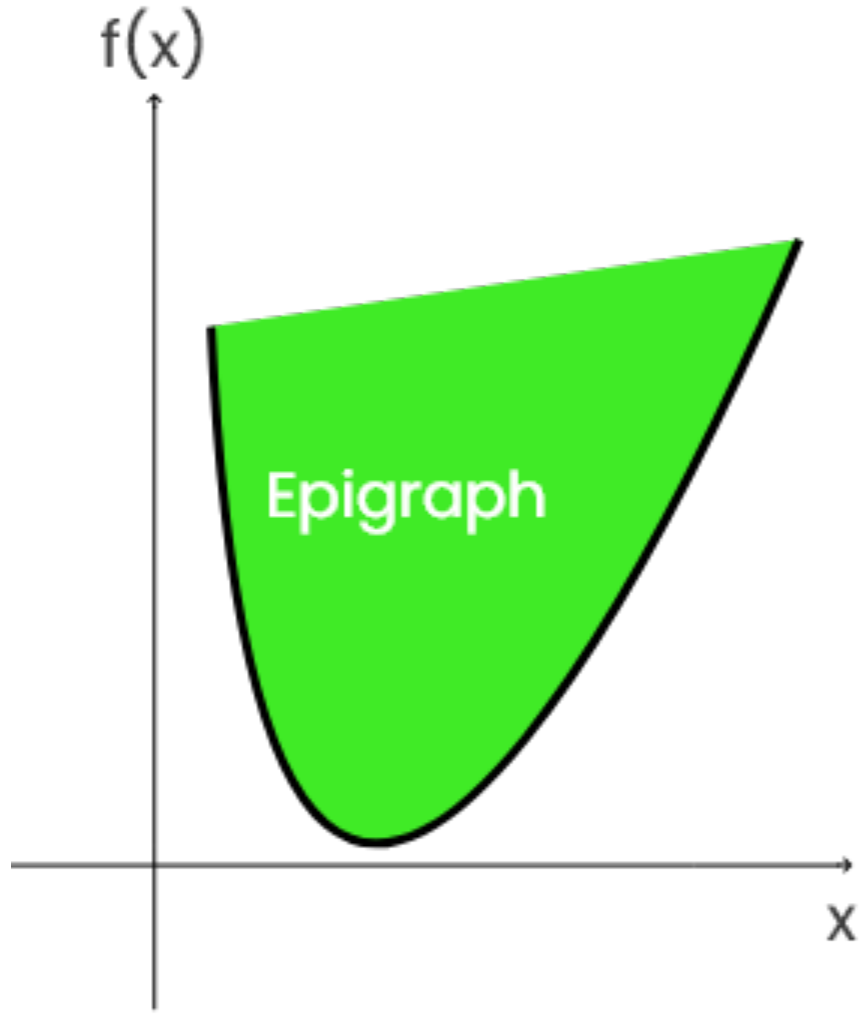
Example: When only the "$\inf$" exists but "$\min$" does not exist

$$f(x) = -\|x\|^2$$

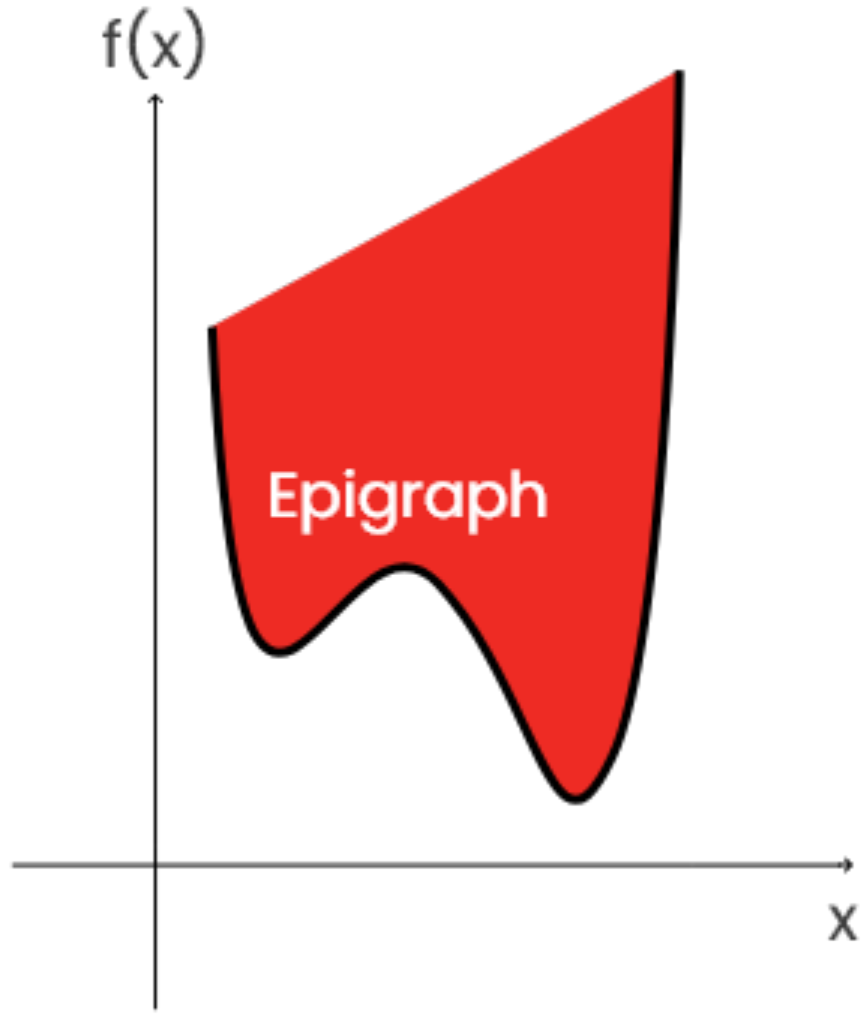
- Fact 2: Suppose the domain X is a convex set and f is a convex function.
  If X^* is not empty, then X^* must be a convex set (why?)

Remark: Definition of Convex Functions via *Epigraph*

Intuition: Can you find anything special in the regions in green and red?



A convex function

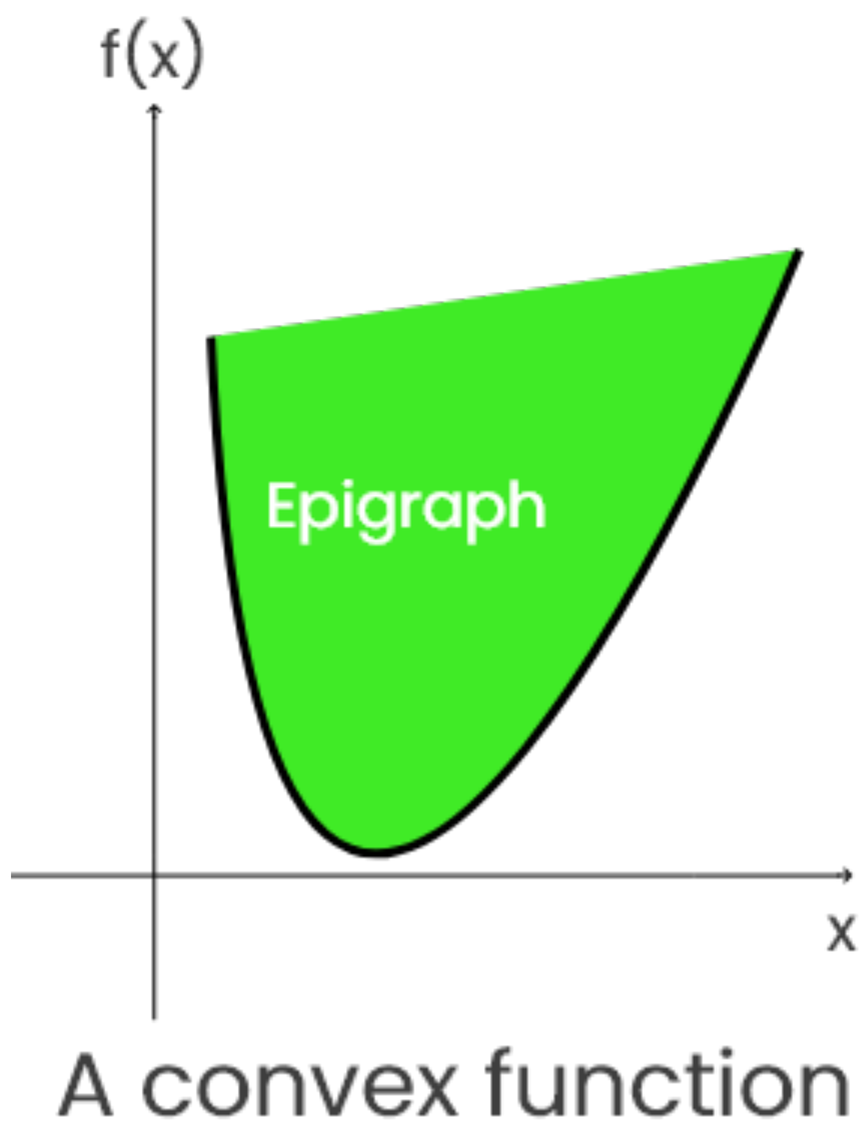


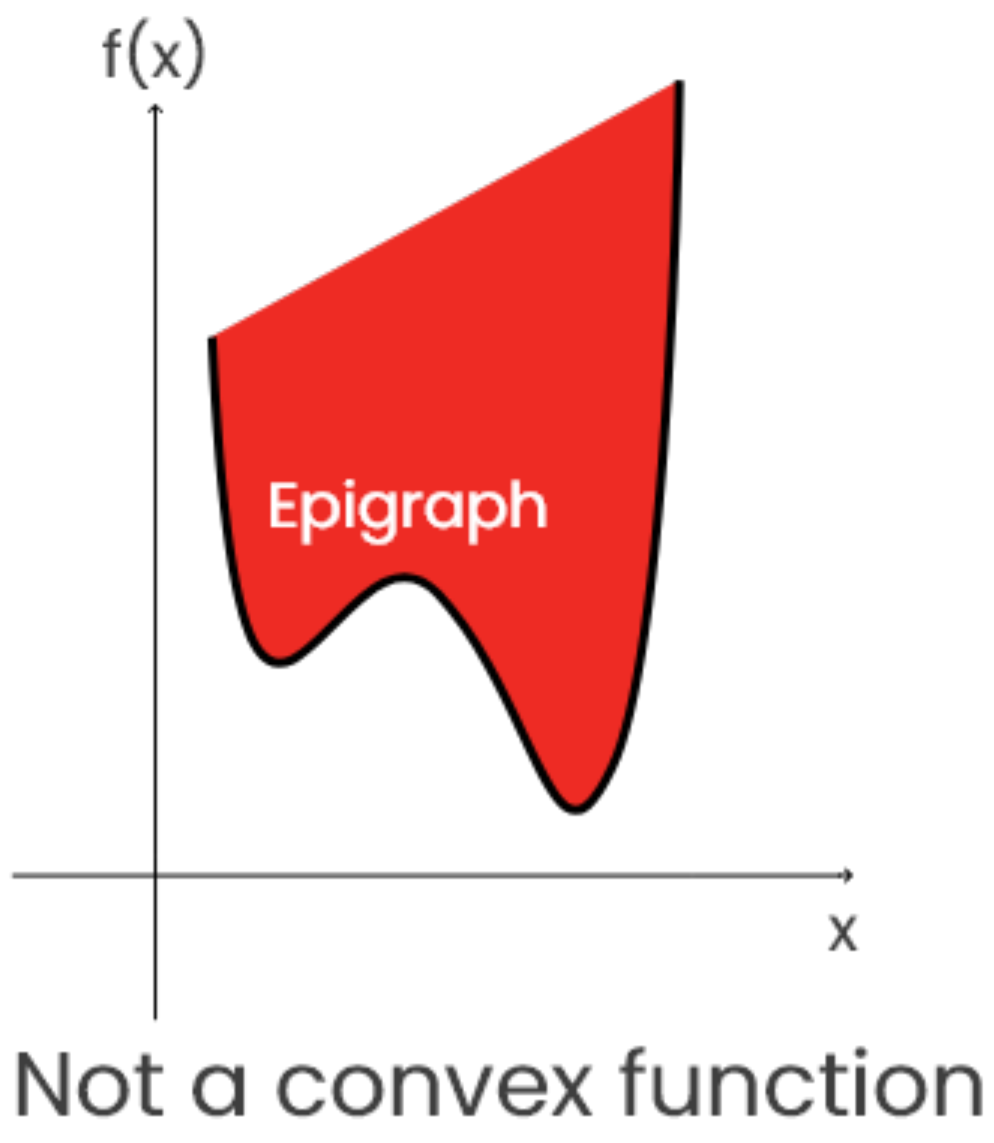
Not a convex function

Figure source: https://towardsdatascience.com/understand-convexity-in-optimization-db87653bf920

Remark: Definition of Convex Functions via *Epigraph*

Intuition: Can you find anything special in the regions in green and red?





Definition: The epigraph of a function $f : X \to \mathbb{R}$ is defined as

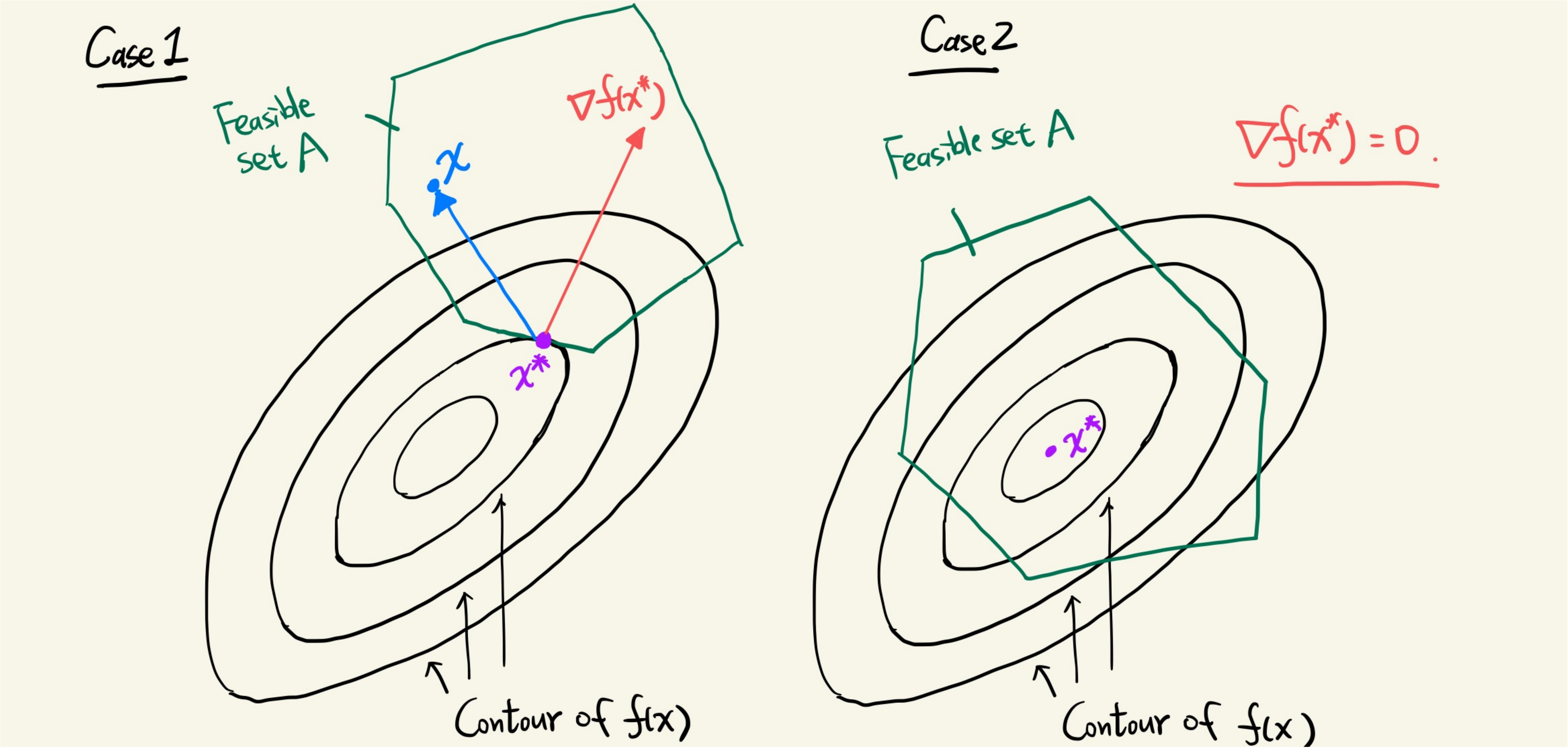
$$\textbf{epi}(f) := \{(x, \gamma) : x \in X, f(x) \leq \gamma\}$$

Property: A function f is *convex* if and only if its epigraph is a *convex set*.

Figure source: https://towardsdatascience.com/understand-convexity-in-optimization-db87653bf920

Optimality Conditions for Constrained Problems?

Let's start with some intuition! By "constrained": Feasible set $A \subseteq X$



C5 & C6. Optimality Conditions for Constrained Problems (Formally)

Theorem: Let $f : X \to \mathbb{R}$ be continuously differentiable and let $A \subseteq X$ be a convex feasible set.

(C5) If x^* is a local minimizer of f over A, then we have (Necessary)

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \quad \forall x \in A \quad \ldots\ldots \ (*)$$

(C6) If f is a convex function over A, then the condition (*) is also *sufficient* for x^* to be a global minimizer of f over A (Sufficient)

Remark: If $A = \mathbb{R}^n$ (i.e., unconstrained), then (*) reduces to $\nabla f(x^*) = 0$ (why?)

C5 & C6. Optimality Conditions for Constrained Problems (Formally)

(C5) If x^* is a local minimizer of f over A, then we have

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \quad \forall x \in A \quad \ldots\ldots (*)$$

Proof of (C5): Prove this by contradiction

Step 1: Suppose there exists some $x \in A$ such that $\nabla f(x)^\top (x - x^*) < 0$

Step 2: By Taylor's Theorem, for any $\varepsilon > 0$, we have

$$f(x^* + \varepsilon(x - x^*)) = f(x^*) + \varepsilon \nabla f(x')^\top (x - x^*)$$

where $x' = x^* + \alpha\varepsilon(x - x^*), \alpha \in [0,1]$

Step 3: Since $\nabla f(x)$ is continuous, we have that for all sufficiently small ε

(i) $\nabla f(x')^\top (x - x^*) < 0$ (ii) $x' \in A$ (why?)

These imply that $f(x^* + \varepsilon(x - x^*)) < f(x^*)$, for all sufficiently small $\varepsilon > 0$

This contradicts the fact that x^* is a local minimizer

C5 & C6. Optimality Conditions for Constrained Problems (Formally)

(C6) If f is a convex function over A, then the condition (*) is also *sufficient* for x^* to be a global minimizer of f over A

Proof of (C6):

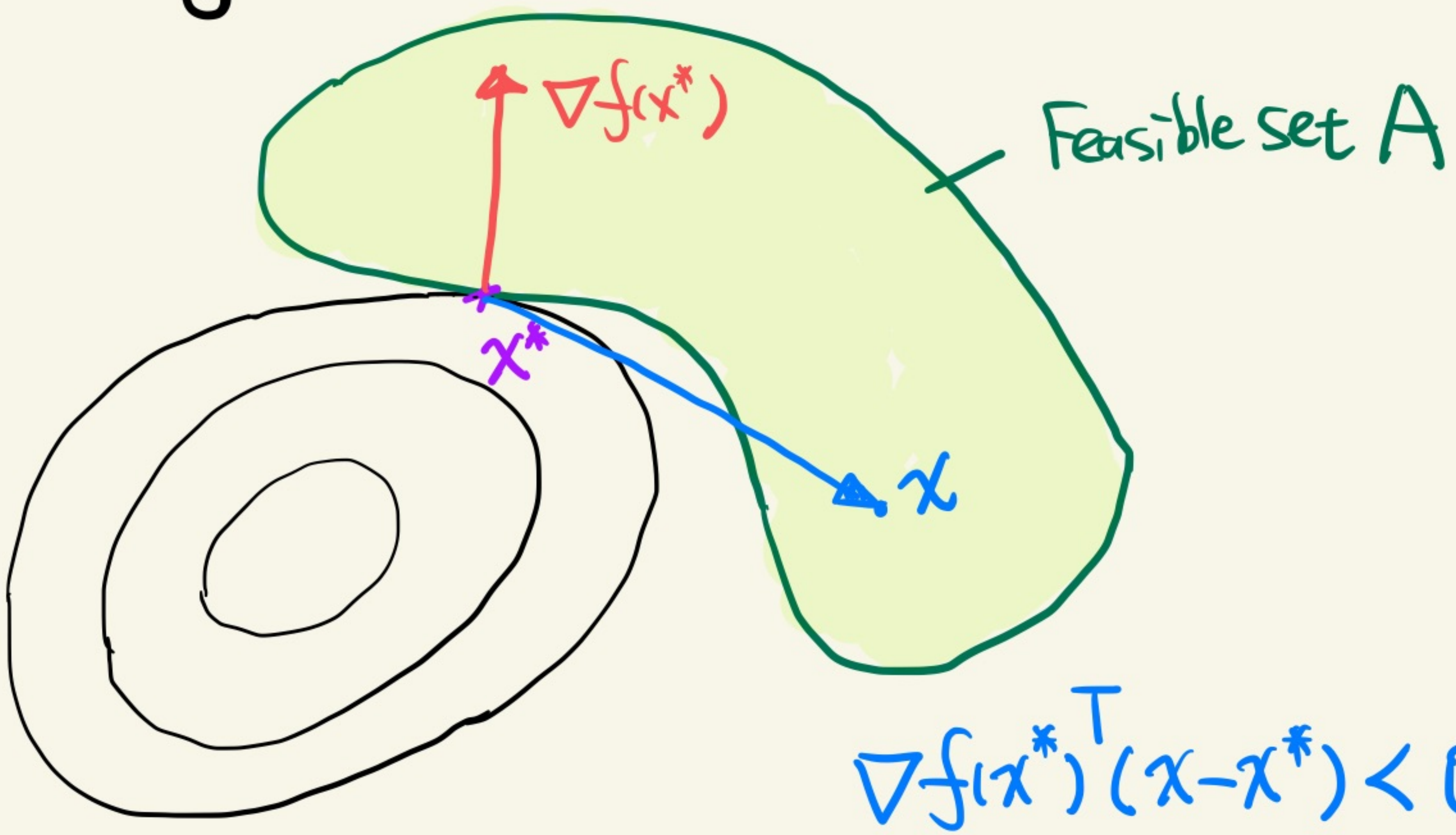
Step 1. By the convexity of f, we have

$$f(x) \geq f(x^*) + \underbrace{\nabla f(x^*)^\top (x - x^*)}_{\geq 0, \text{ for all } x \in A}, \text{ for all } x \in A$$

Step 2. Therefore, we have $f(x) \geq f(x^*)$, for all $x \in A$

Why "Convex" Feasible Sets?

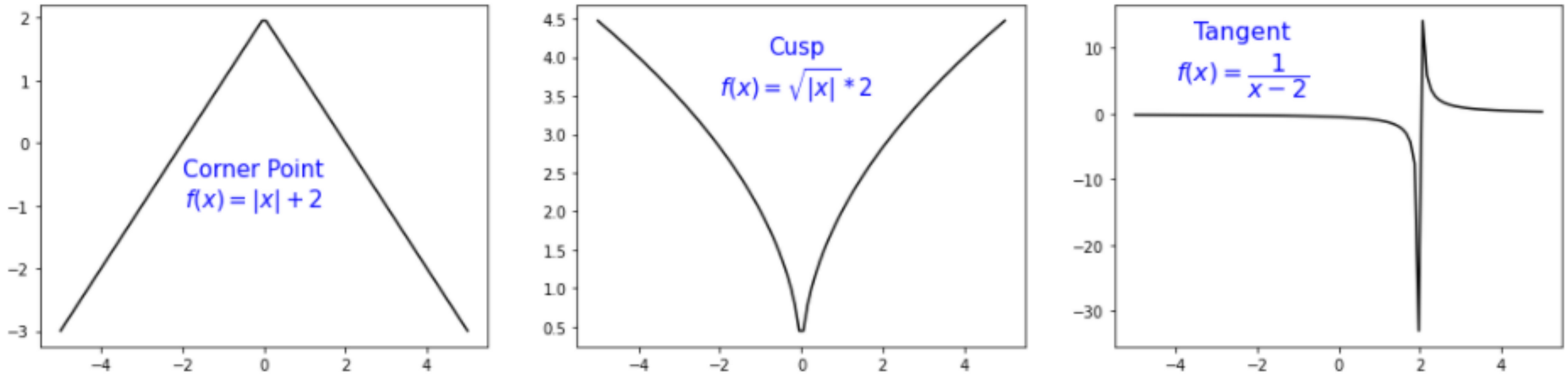
The necessary condition $\nabla f(x^*)^T(x-x^*) \geq 0$ may fail when A is <u>not convex</u>



$\nabla f(x^*)^T(x-x^*) < 0.$

Optimality Conditions Beyond Differentiability?

- What if there are some non-differentiable points in f?



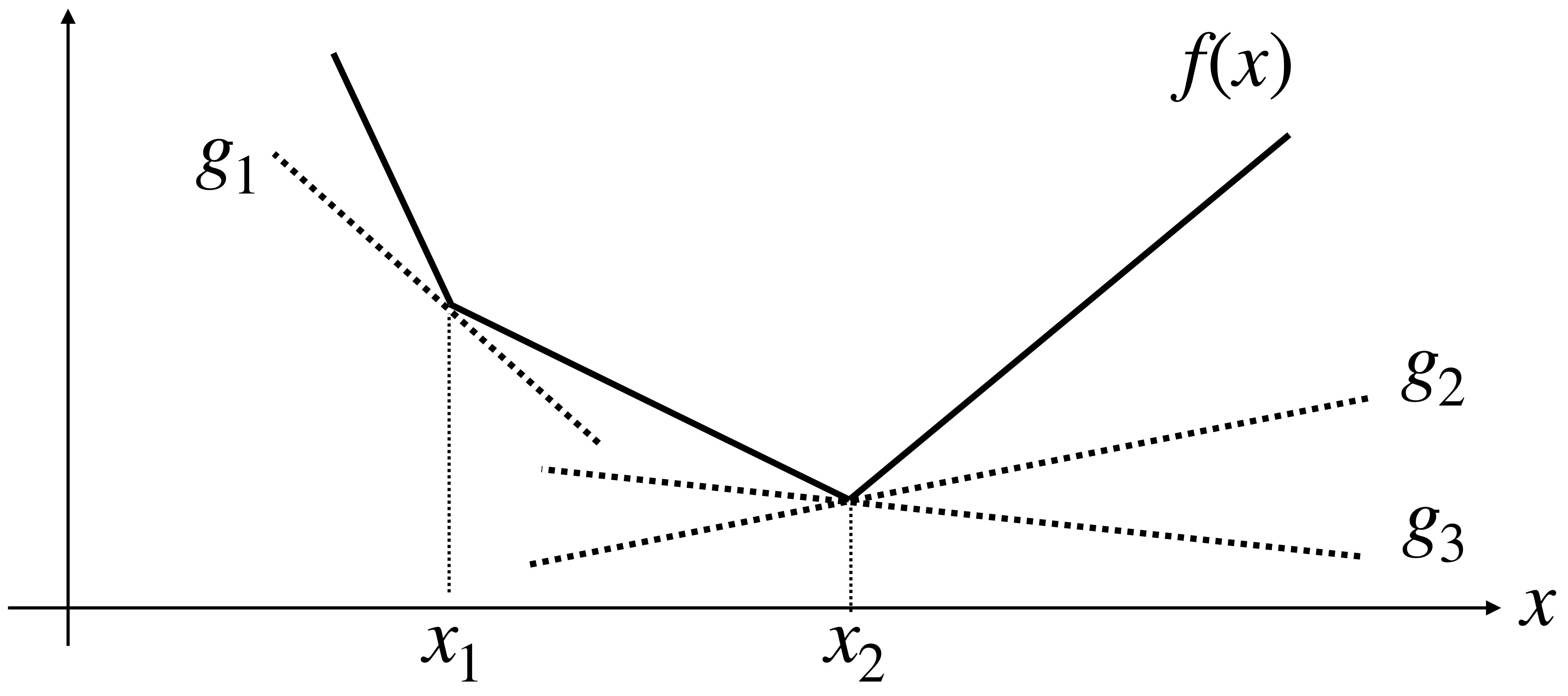
Could we extend the notion of gradients and optimality conditions?

(Figure Credit: https://python.plainenglish.io/descent-carefully-on-your-gradient-c0f030ddef81)

Subgradients and Subdifferential

(The slides are partially adapted from Stephen Boyd's EE364B)

Subgradients



(In plain English: $f(x) + g^\top(z - x)$ is a global underestimate)

Definition: $g \in \mathbb{R}^n$ is a *subgradient* of f (possibly non-convex) at x if

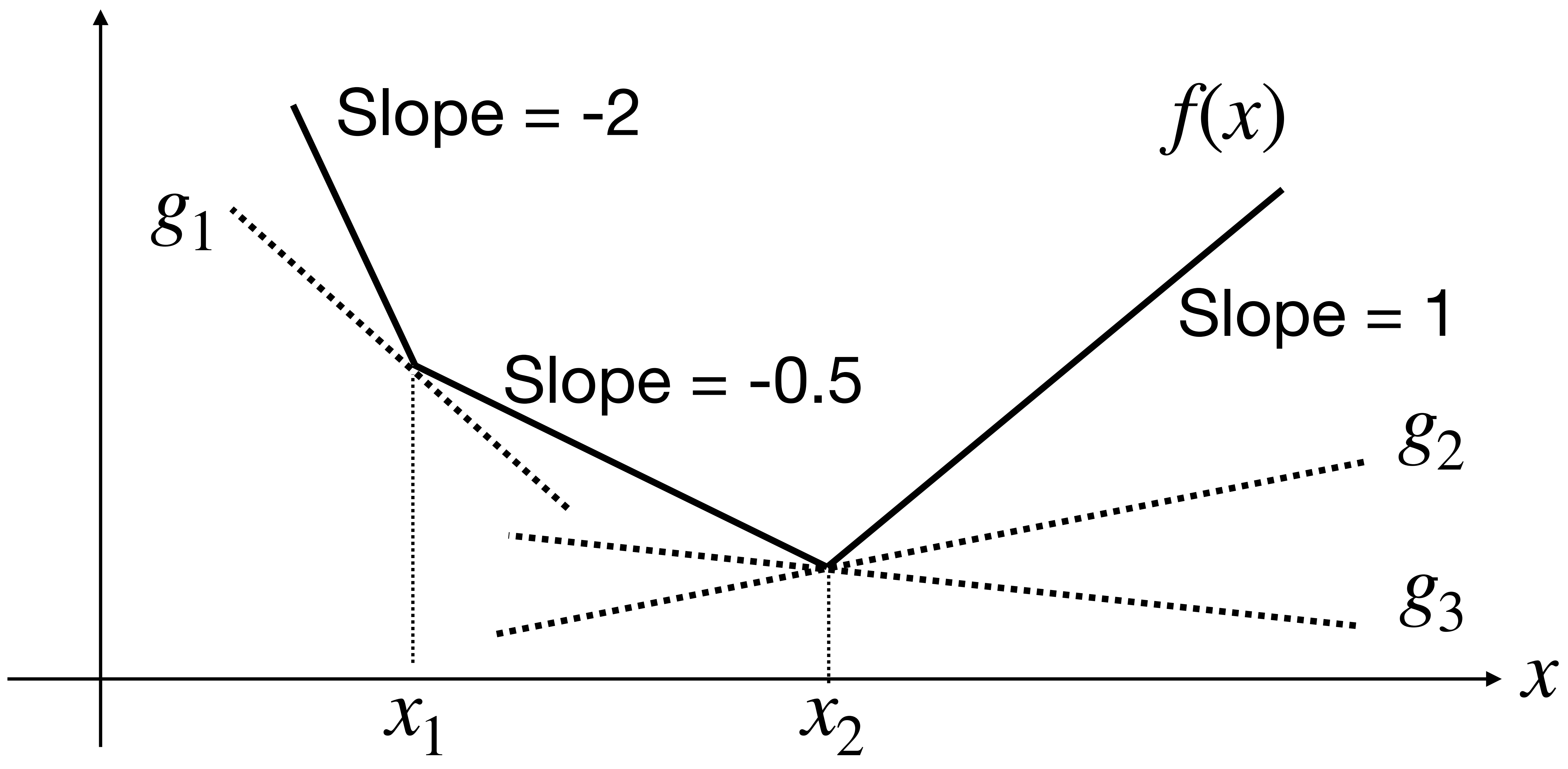
$$f(z) \geq f(x) + g^\top(z - x), \text{ for all } z \in X$$

Question: If f is differentiable, then could you find a natural subgradient?

Subdifferentials

Definition: The *subdifferential* of f at x, denoted by $\partial f(x)$, is defined as the set of all subgradients of f at x.

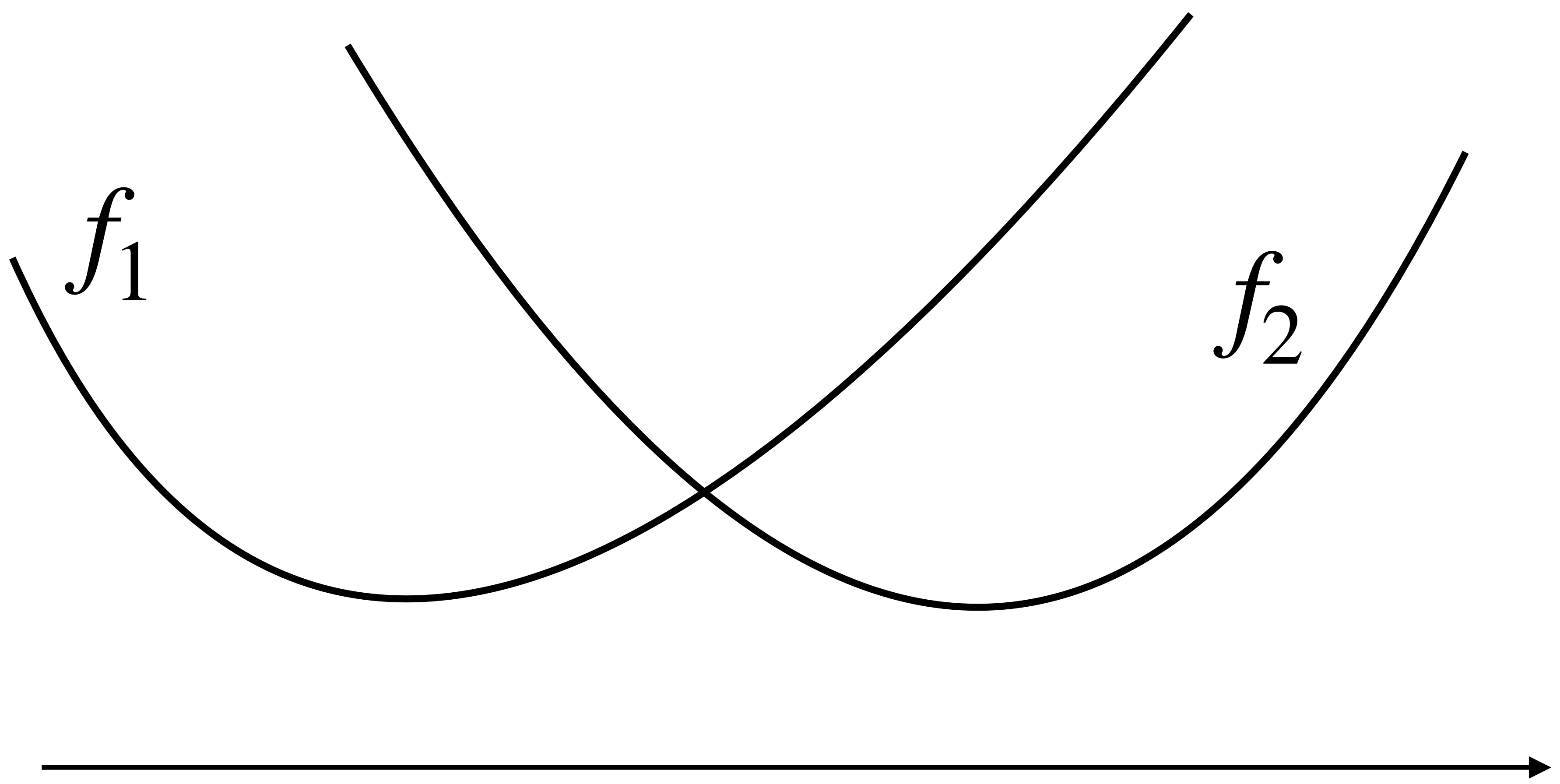
Example:



- Subdifferential at x_1?
- Subdifferential at x_2?

More Examples of Subdifferentials

Suppose $f = \max\{f_1, f_2\}$, where $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$ are both convex and differentiable



- For x with $f_1(x) > f_2(x)$:

- For x with $f_1(x) < f_2(x)$:

- For x with $f_1(x) = f_2(x)$:

Subdifferential of f ?

Subdifferentials of Convex Functions

If $f : X \to \mathbb{R}$ is convex, then $\partial f(x)$ has some nice properties

- If x is in the relative interior of X, then $\partial f(x) \neq \emptyset$

- If f is differentiable at x, then $\partial f(x) = \{ \nabla f(x) \}$

- If $\partial f(x) = \{g\}$, then f is differentiable and $g = \nabla f(x)$

- If f is differentiable at x, then $\partial f(x) = \{ \nabla f(x) \}$

Basic Calculus Rules of Subdifferentials

- Nonnegative scaling: For any $\alpha > 0$, $\partial(\alpha f)(x) = \{\alpha g : g \in \partial f(x)\}$

- Addition (General): $\partial f_1(x) + \partial f_2(x) \subset \partial(f_1 + f_2)(x)$ (Vice versa? See the next page)

  (Set addition / Minkowski addition)

- Addition (Convex cases): If f_1, f_2 are convex, then $\partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x)$

(For more properties of subdifferentials, please check Stephen Boyd's slides for Lecture 1 of EE364B)

We Do Not Have $\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$ in General

Let's construct two functions $f_1 : \mathbb{R} \to \mathbb{R}, f_2 : \mathbb{R} \to \mathbb{R}$

$$f_1(x) := \begin{cases} -\sqrt{x}, & \text{if } x \geq 0 \\ -x + \sqrt{-x}, & \text{if } x < 0 \end{cases}$$

$$f_2(x) := \begin{cases} x + \sqrt{x}, & \text{if } x \geq 0 \\ -\sqrt{-x}, & \text{if } x < 0 \end{cases}$$

Exercise: Try to verify the following

- $\partial f_1(0) = \emptyset$ and $\partial f_2(0) = \emptyset$
- $\partial(f_1 + f_2)(0) = [-1,1]$

C7. Optimality Conditions Revisited: Without Differentiability

Theorem (Fermat's Rule): Let $f : \mathbb{R}^n \to \mathbb{R}$ (not necessarily differentiable). Then, we have

$$\arg\min f = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}$$

Proof: (1) RHS $\subseteq$ LHS

(2) LHS $\subseteq$ RHS

Example: Indicator Function

Given any set $A \subset \mathbb{R}^n$, let $\mathbf{1}_A$ be the indicator function for A:

$$\mathbf{1}_A(x) := \begin{cases} 0, & \text{if } x \in A, \\ \infty, & \text{otherwise.} \end{cases}$$

Property: A is a convex set if and only if $\mathbf{1}_A(x)$ is convex

---

Question: Subdifferential of $\mathbf{1}_A(x)$ for any $x \in X$?

Connecting Indicator Functions and Constrained Problems

Given any set $A \in \mathbb{R}^n$, let $\mathbf{1}_A$ be the indicator function for A:

$$\mathbf{1}_A(x) := \begin{cases} 0, & \text{if } x \in A, \\ \infty, & \text{otherwise.} \end{cases}$$

Convert a *constrained* problem into an *unconstrained* one:

$$\min_x \ f(x) \quad \text{subject to } x \in A \subseteq X \qquad \longrightarrow \qquad \min_{x \in X} \ f(x) + \mathbf{1}_A(x)$$

An alternative derivation of "optimality condition for convex constrained problems"

- If f is convex and differentiable, then x^* is a global minimizer iff $0 \in \partial(f + \mathbf{1}_A)(x^*)$
- $\partial(f + \mathbf{1}_A)(x^*) = \nabla f(x^*) + \partial \mathbf{1}_A(x^*) = \nabla f(x^*) + \{g : g^\top (y - x^*) \leq 0, \forall y \in X\}$
- Hence, $\nabla f(x^*)^\top (y - x^*) \geq 0, \forall y \in X$!

Remark: Indicator Functions and Constrained Problems

- $0 \in \nabla f(x^*) + \partial \mathbf{1}_A(x^*)$ is an elegant and very general condition for optimality in convex optimization problems

- However, it is not always easy to play with

- We will discuss some simpler conditions (e.g., KKT conditions) later!

Appendix

Mean Value Theorem (For Multivariate Functions)

Mean Value Theorem: Let $f : X \to \mathbb{R}$ be a differentiable function. Let a, b be points in X such that the line segment of a, b lies in U. Then, there must exists some $z = \alpha a + (1 - \alpha)b$ with $\alpha \in [0,1]$ such that

$$f(b) - f(a) = \nabla f(z)^\top (b - a)$$

The proof can be found at: https://links.uwaterloo.ca/amath731docs/meanvalue.pdf