535520: Optimization Algorithms
Lecture 2 — Subgradients, Constrained Optimization, and Duality

Ping-Chun Hsieh (謝秉均)

September 9, 2024

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?

This Lecture

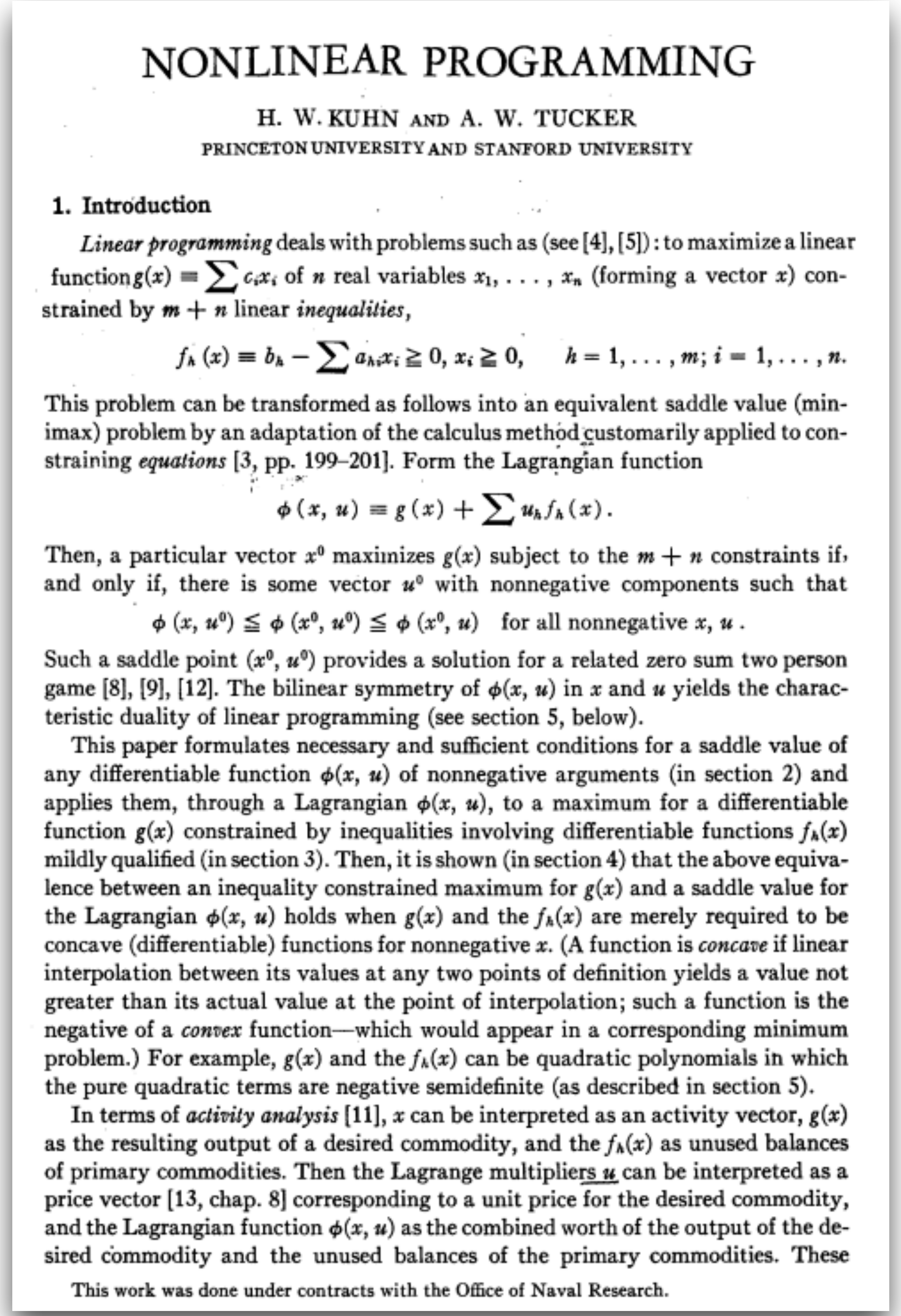
1. Subgradients and Subdifferentials

2. Constrained Optimization: Lagrangians and Duality

3. Karush-Kuhn-Tucker (KKT) Conditions

- Reading Material:
  - Chapters 2 & 5 of Dimitri Bertsekas's textbook "Nonlinear Programming"
  - Chapter 5.5-5.6 of Stephen Boyd's textbook "Convex Optimization"
  - Stephen Boyd's note: https://stanford.edu/class/ee364b/lectures/subgradients_slides.pdf

Historical Account on Karush-Kuhn-Tucker (KKT)



NONLINEAR PROGRAMMING

H. W. KUHN AND A. W. TUCKER
PRINCETON UNIVERSITY AND STANFORD UNIVERSITY

1. Introduction

Linear programming deals with problems such as (see [4], [5]): to maximize a linear function $g(x) \equiv \sum c_i x_i$ of n real variables $x_1, \ldots, x_n$ (forming a vector x) constrained by $m + n$ linear *inequalities*,

$$f_h(x) \equiv b_h - \sum a_{hi} x_i \geqq 0, \; x_i \geqq 0, \qquad h = 1, \ldots, m; \; i = 1, \ldots, n.$$

This problem can be transformed as follows into an equivalent saddle value (minimax) problem by an adaptation of the calculus method customarily applied to constraining *equations* [3, pp. 199–201]. Form the Lagrangian function

$$\phi(x, u) \equiv g(x) + \sum u_h f_h(x).$$

Then, a particular vector x^0 maximizes $g(x)$ subject to the $m + n$ constraints if, and only if, there is some vector u^0 with nonnegative components such that

$$\phi(x, u^0) \leqq \phi(x^0, u^0) \leqq \phi(x^0, u) \quad \text{for all nonnegative } x, u.$$

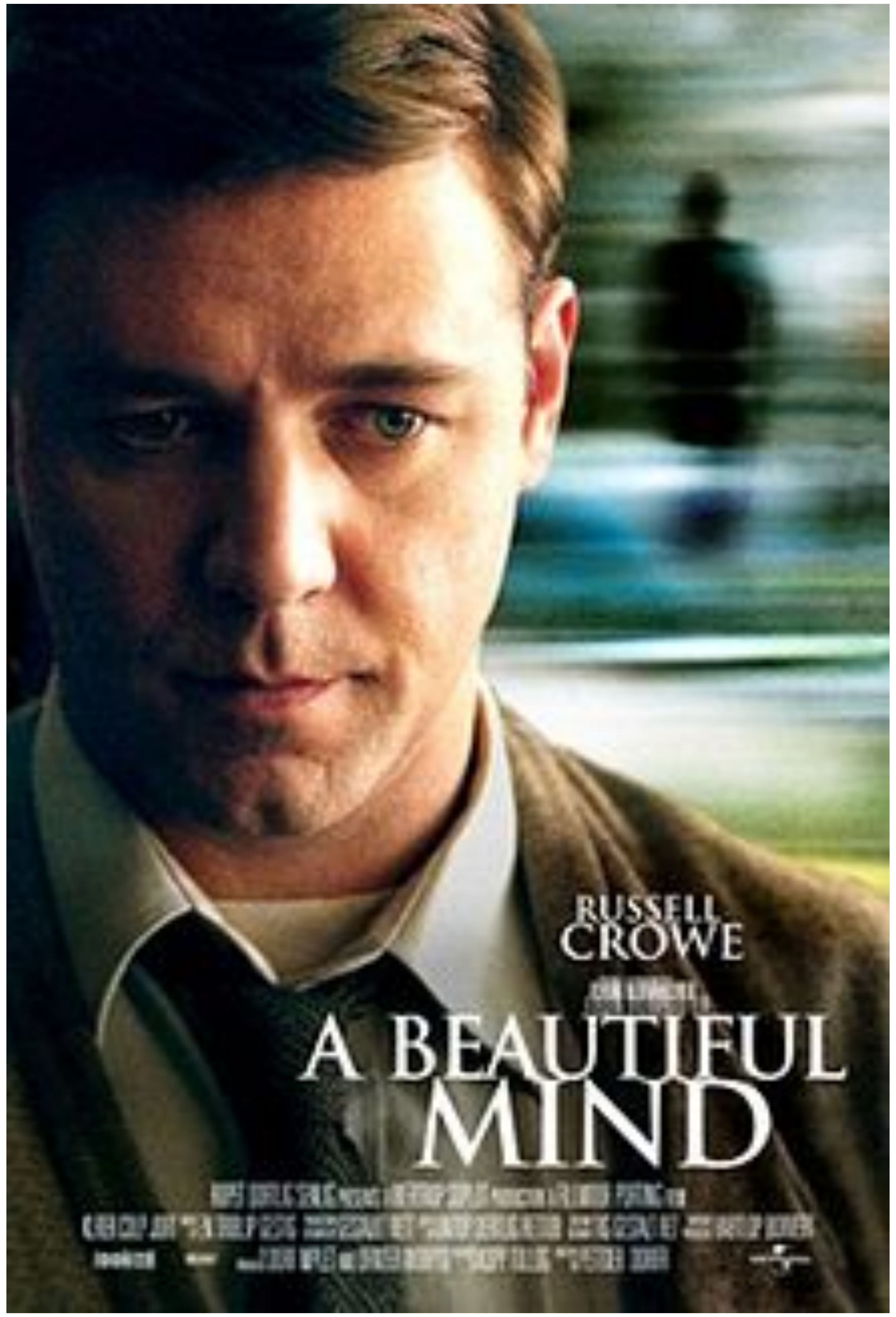
Such a saddle point (x^0, u^0) provides a solution for a related zero sum two person game [8], [9], [12]. The bilinear symmetry of $\phi(x, u)$ in x and u yields the characteristic duality of linear programming (see section 5, below).

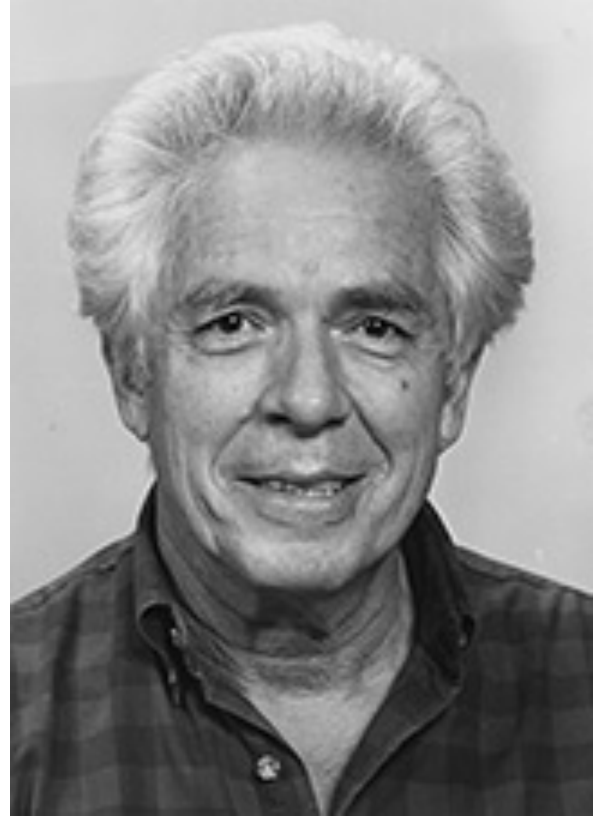
This paper formulates necessary and sufficient conditions for a saddle value of any differentiable function $\phi(x, u)$ of nonnegative arguments (in section 2) and applies them, through a Lagrangian $\phi(x, u)$, to a maximum for a differentiable function $g(x)$ constrained by inequalities involving differentiable functions $f_h(x)$ mildly qualified (in section 3). Then, it is shown (in section 4) that the above equivalence between an inequality constrained maximum for $g(x)$ and a saddle value for the Lagrangian $\phi(x, u)$ holds when $g(x)$ and the $f_h(x)$ are merely required to be concave (differentiable) functions for nonnegative x. (A function is *concave* if linear interpolation between its values at any two points of definition yields a value not greater than its actual value at the point of interpolation; such a function is the negative of a *convex* function—which would appear in a corresponding minimum problem.) For example, $g(x)$ and the $f_h(x)$ can be quadratic polynomials in which the pure quadratic terms are negative semidefinite (as described in section 5).

In terms of *activity analysis* [11], x can be interpreted as an activity vector, $g(x)$ as the resulting output of a desired commodity, and the $f_h(x)$ as unused balances of primary commodities. Then the Lagrange multipliers u can be interpreted as a price vector [13, chap. 8] corresponding to a unit price for the desired commodity, and the Lagrangian function $\phi(x, u)$ as the combined worth of the output of the desired commodity and the unused balances of the primary commodities. These

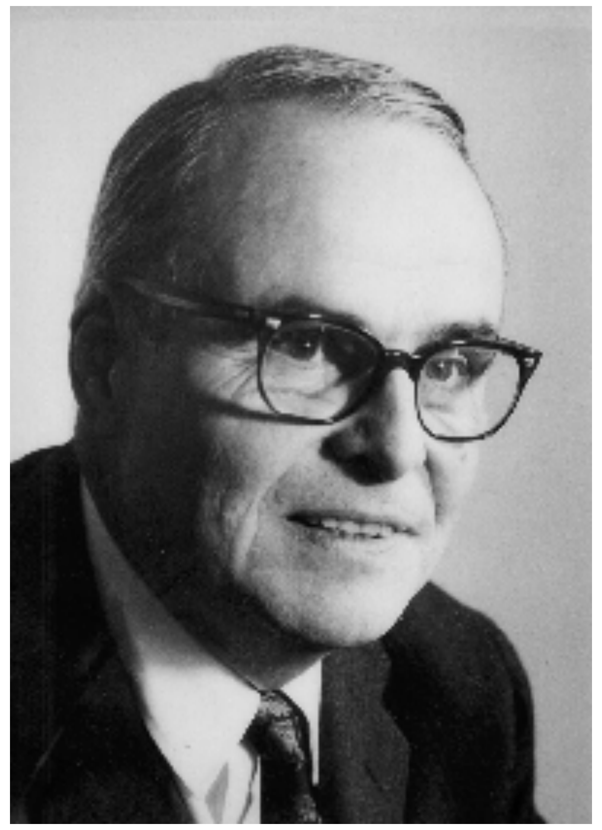
This work was done under contracts with the Office of Naval Research.

- H. W. Kuhn and A. W. Tucker, "Nonlinear programming," Proceedings of 2nd Berkeley Symposium, pp. 481–492, 1951.





Harold W. Kuhn



Albert W. Tucker

John Nash

THE UNIVERSITY OF CHICAGO

MINIMA OF FUNCTIONS OF SEVERAL VARIABLES WITH
INEQUALITIES AS SIDE CONDITIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

BY
WILLIAM KARUSH

CHICAGO, ILLINOIS
DECEMBER, 1939

- W. Karush, “Minima of Functions of Several Variables with Inequalities as Side Constraints,” M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois. 1939.

“…*Linear programming aroused interest in constraints in the form of inequalities and in the theory of linear inequalities and convex sets. The Kuhn-Tucker study appeared in the middle of this interest with a full recognition of such developments. However, the theory of nonlinear programming when the constraints are all in the form of equalities has been known for a long time - in fact, since Euler and Lagrange. The inequality constraints were treated in a fairly satisfactory manner already in 1939 by Karush. Karush’s work is apparently under the influence of a similar work in the calculus of variations by Valentine. Unfortunately, Karush’s work has been largely ignored*…” (by A. Takayama in 1974)

Review: Optimality Conditions for Unconstrained Problems

Unconstrained cases:

C1. FONC: f continuously differentiable on an ϵ-neighborhood of x^*, and x^* is a local minimizer $\Rightarrow$ ____________________

C2. SONC: f twice continuously differentiable on an ϵ-neighborhood of x^*, and x^* is a local minimizer $\Rightarrow$ ____________________

C3. FOSC: Suppose f differentiable and convex with a convex domain

____________________ $\Rightarrow$ $x^* \in X$ is a global minimizer

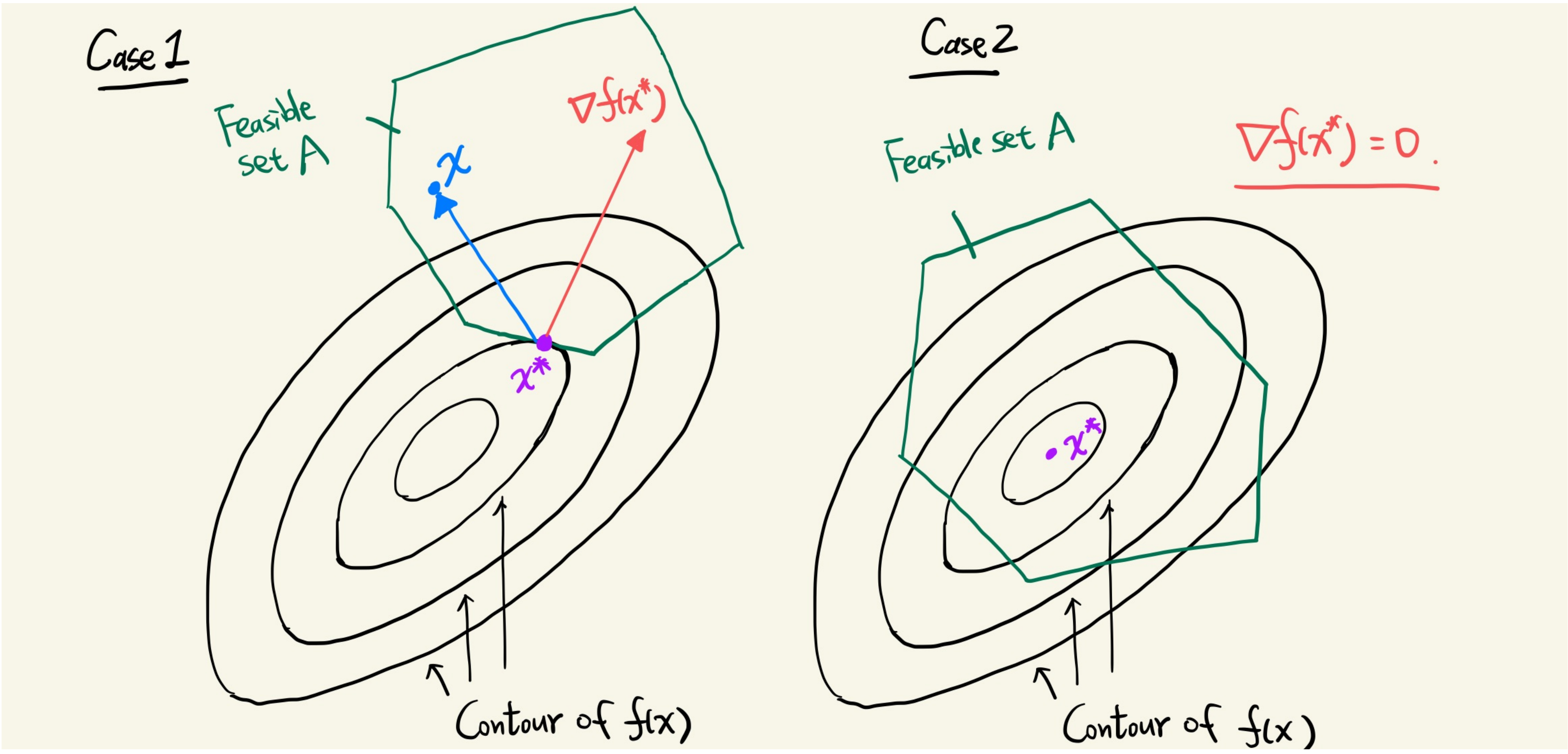
C4. SOSC: Suppose f twice continuously differentiable

____________________ $\Rightarrow$ $x^* \in X$ is a strict be a local minimizer

Corollary of C4: Suppose f is convex and twice continuously differentiable

____________________ $\Rightarrow$ $x^* \in X$ is a strict be a global minimizer

Review: Optimality Conditions for Constrained Problems



Constrained cases:

C5. FONC-C:

f continuously differentiable with feasible set $A \subseteq X$, and x^* is a local minimizer ➡ ____________________

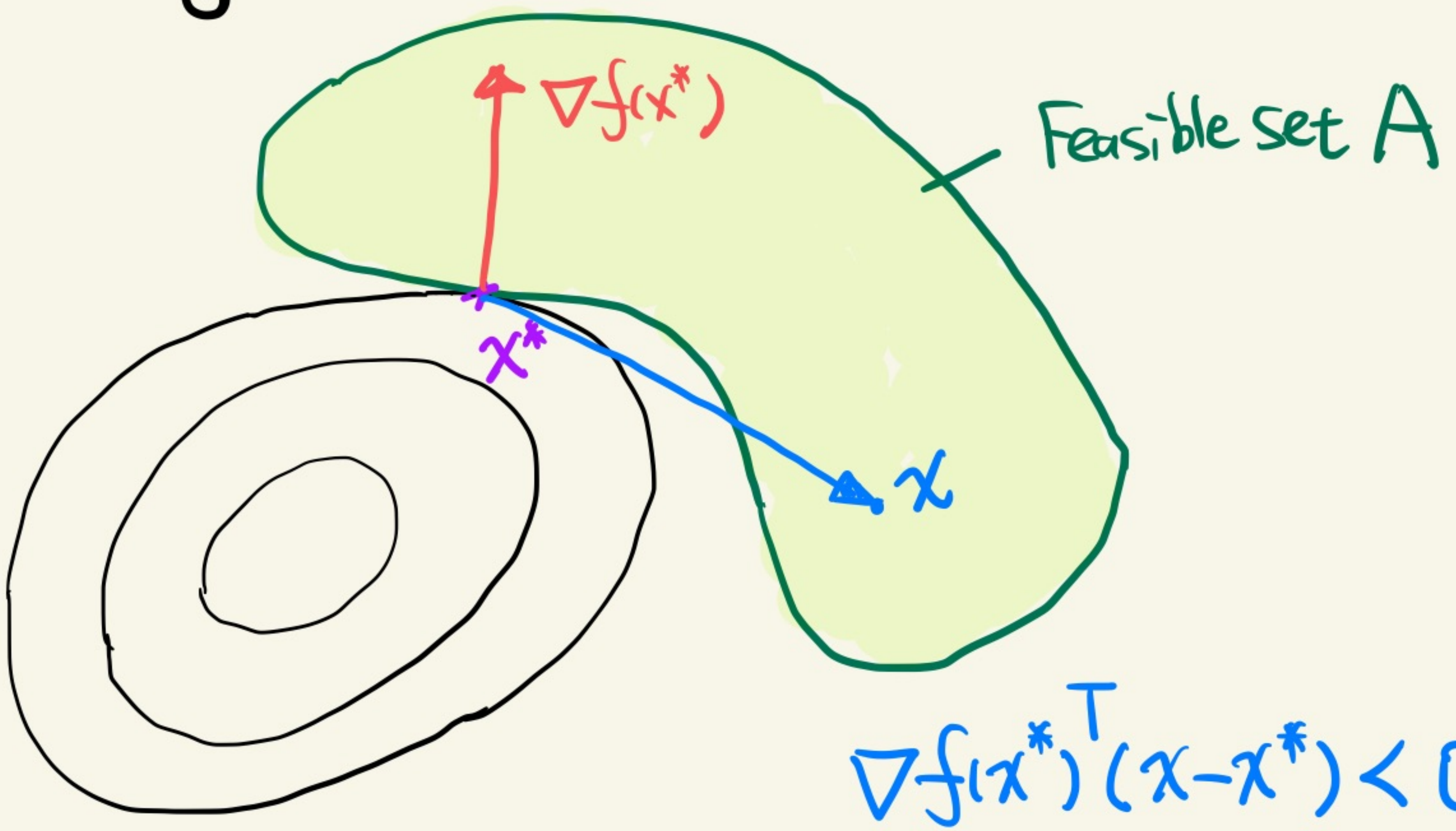
C6. FOSC-C:

f is convex and continuously differentiable with feasible set $A \subseteq X$, and x^* is a local minimizer

____________________ ➡ $x^* \in A$ is a global minimizer

Why "Convex" Feasible Sets?

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



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

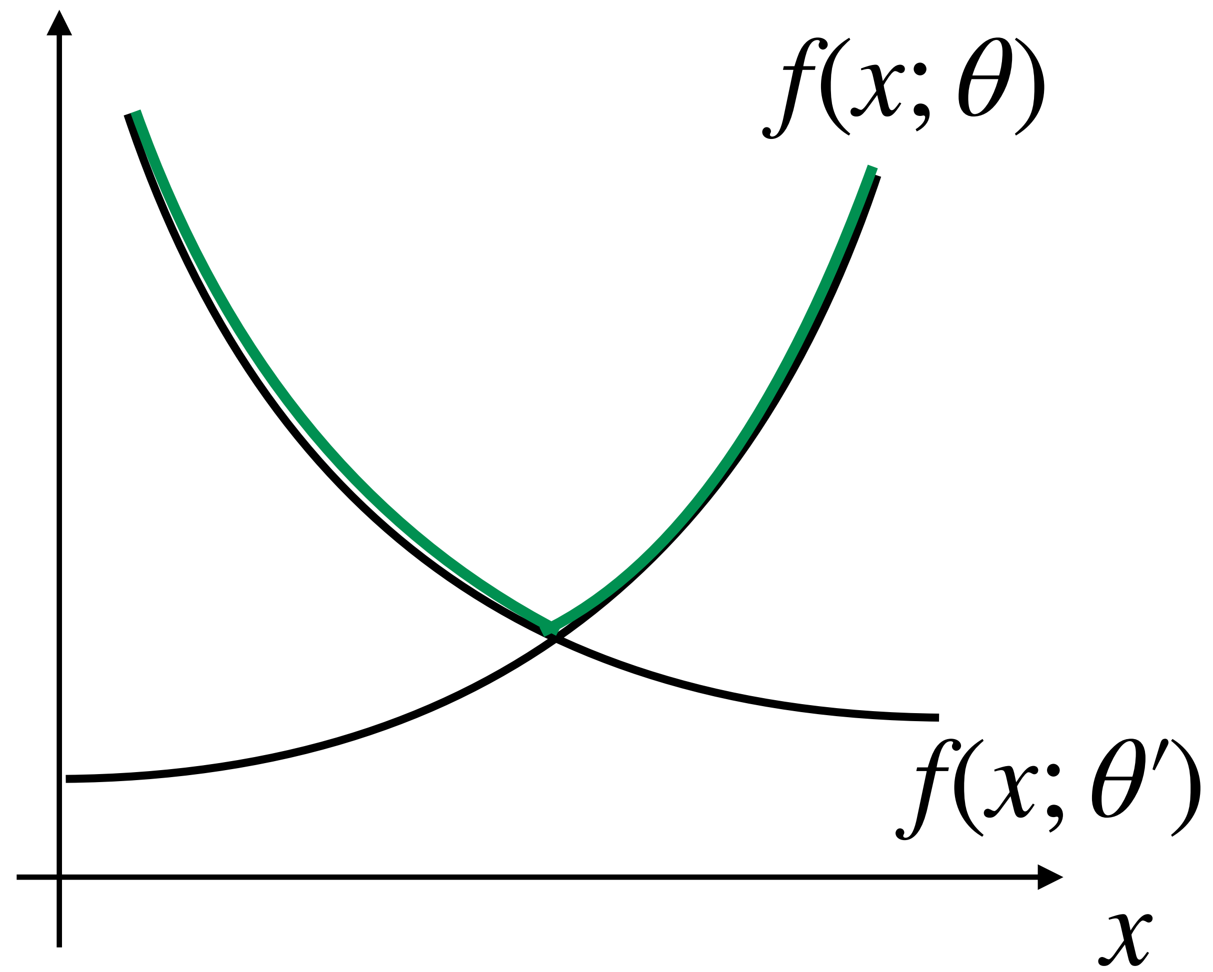
Example: Pointwise Maximum of Convex Functions

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.

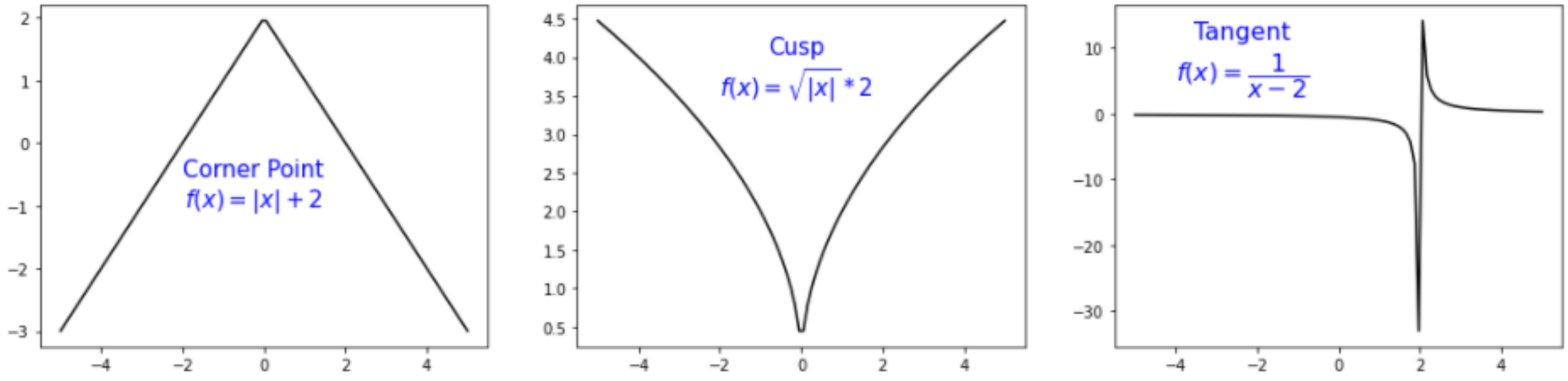
Define $$F(x) := \max_{\theta\in\Theta} f(x;\theta)$$

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



Optimality Conditions Beyond Differentiability?

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



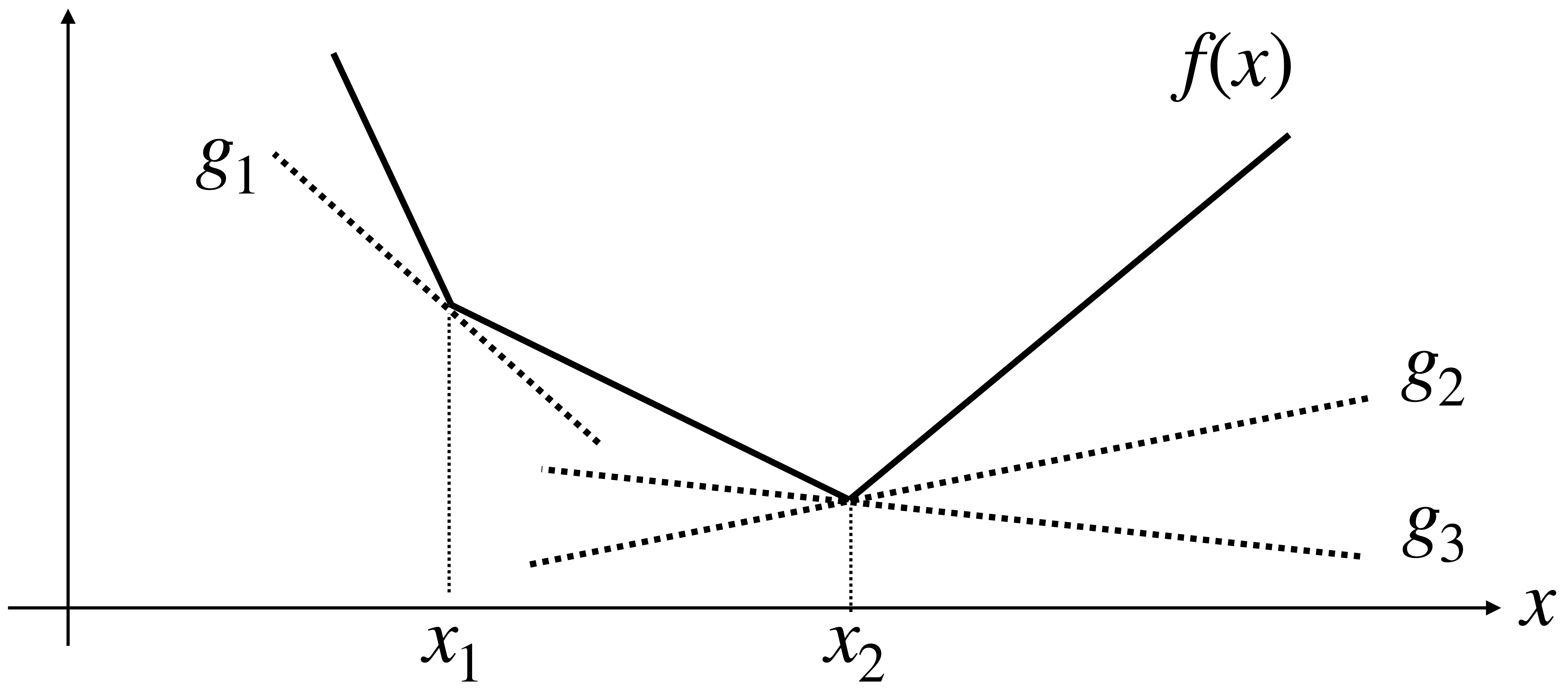
Could we extend the notion of gradients and optimality conditions?

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

1. 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

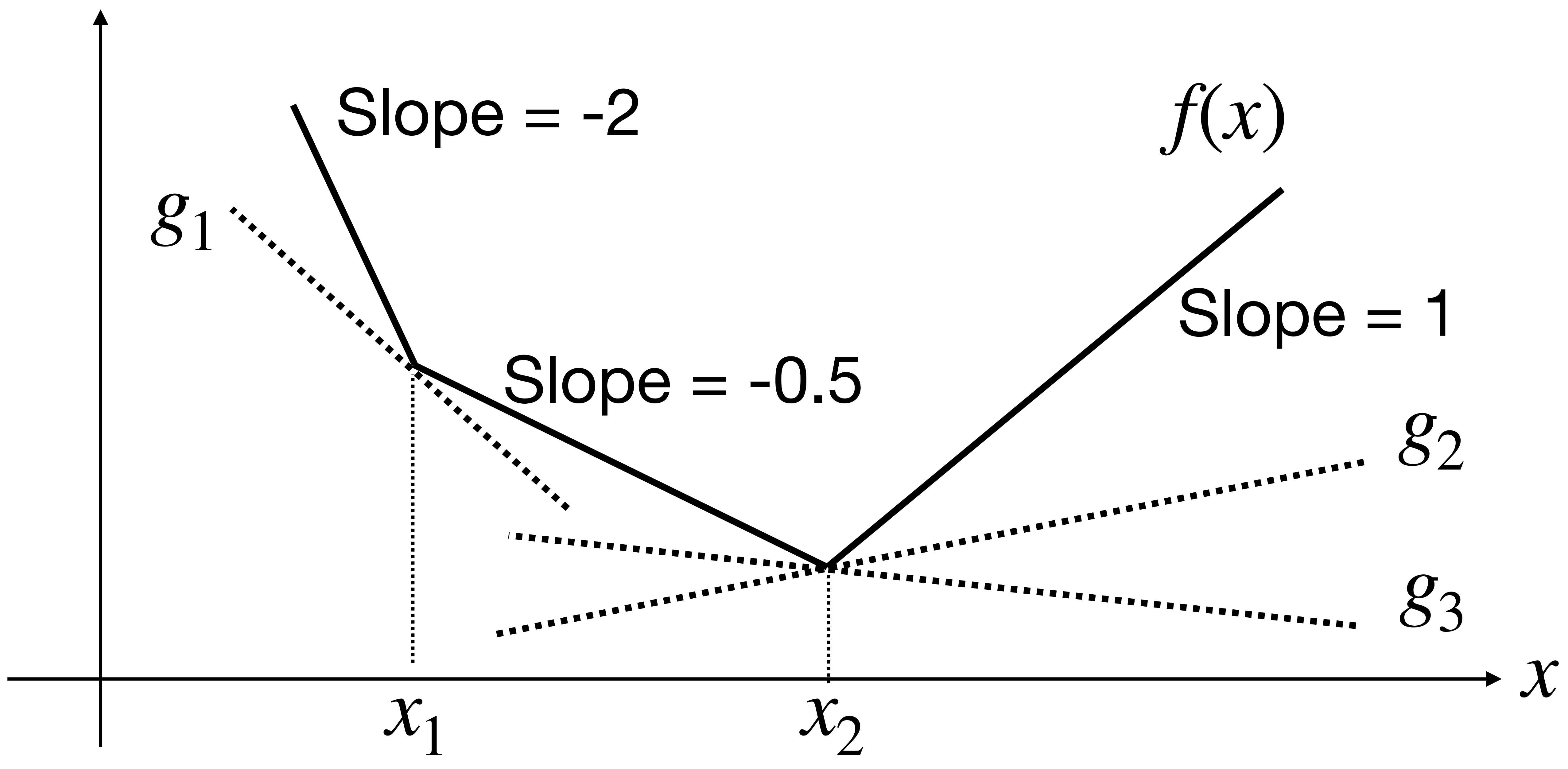
$$f(z) \geq f(x) + g^\top(z - x), \quad \text{for all } z \in X$$

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.

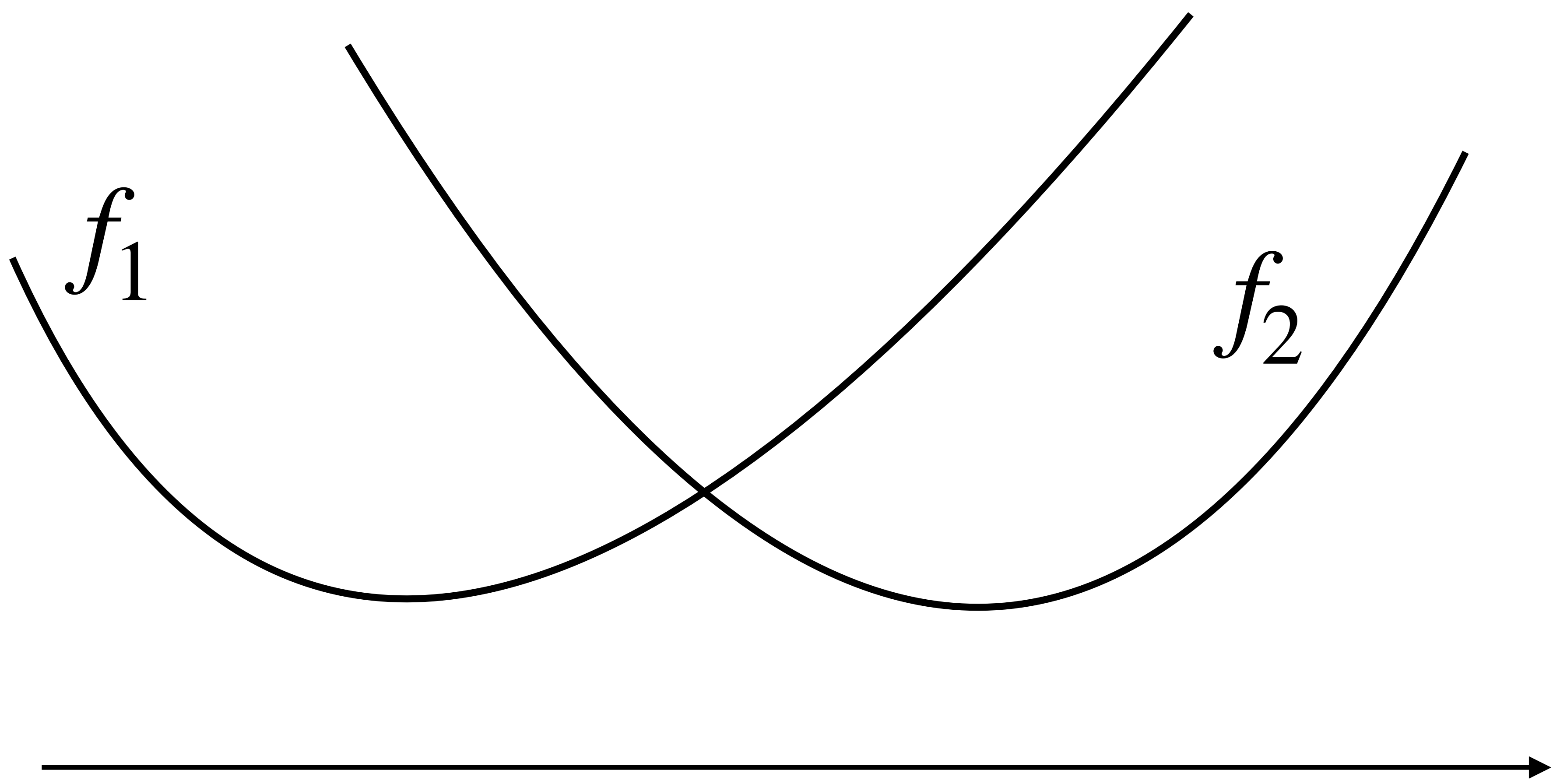
Example:



- 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 a convex function, 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 $g = \nabla f(x)$

You will prove this in HW0 :)

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 $x \in X$?

(1) $x \in A$

(2) $x \notin A$

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:

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

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 (why?)

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

2. *Constrained* Optimization & Lagrangians

Primal Problem

(P)

Primal variable

$$\min_{x} \quad f(x) \qquad f : \mathbb{R}^n \to \mathbb{R}$$

$$g_i : \mathbb{R}^n \to \mathbb{R}$$

subject to $\; g_i(x) \leq 0$, for all $i = 1, \cdots, n$

- Let "dom f"and "dom g_i" denote the domains of the corresponding functions

- Question: What is the domain of this constrained problem?

- Question: Is the expression above general enough? (i.e., can $g_i(x) \leq 0$ capture all types of constraints?)

- A vector x' is feasible if it is in the domain and satisfies all the constraints

Lagrangian (From Hard to Soft Constraints)

- Given the problem (P), construct the Lagrangian as

Lagrangian: $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^m_+ \to (-\infty, \infty]$

$$\mathscr{L}(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i \cdot g_i(x)$$

- $\lambda \equiv (\lambda_1, \cdots, \lambda_m)$ are called the Lagrange multipliers
- We presume $\lambda_i \geq 0$, for all i (why?)
- Question: Suppose x is feasible. Can we say anything about $f(x)$ and $\mathscr{L}(x, \lambda)$?

Lagrangian: Basic Properties (1/2)

- **Property 1**: $$\sup_{\lambda \geq 0} \mathscr{L}(x, \lambda) = \begin{cases} f(x), & \text{if } x \text{ is feasible,} \\ \infty, & \text{otherwise.} \end{cases}$$

- Intuition:

- Proof: (1) If x is feasible

(2) If x is not feasible:

Lagrangian: Basic Properties (2/2)

Let p^* denote the optimal value of the primal problem (P)

- **Property 2**: $$p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathscr{L}(x, \lambda)$$

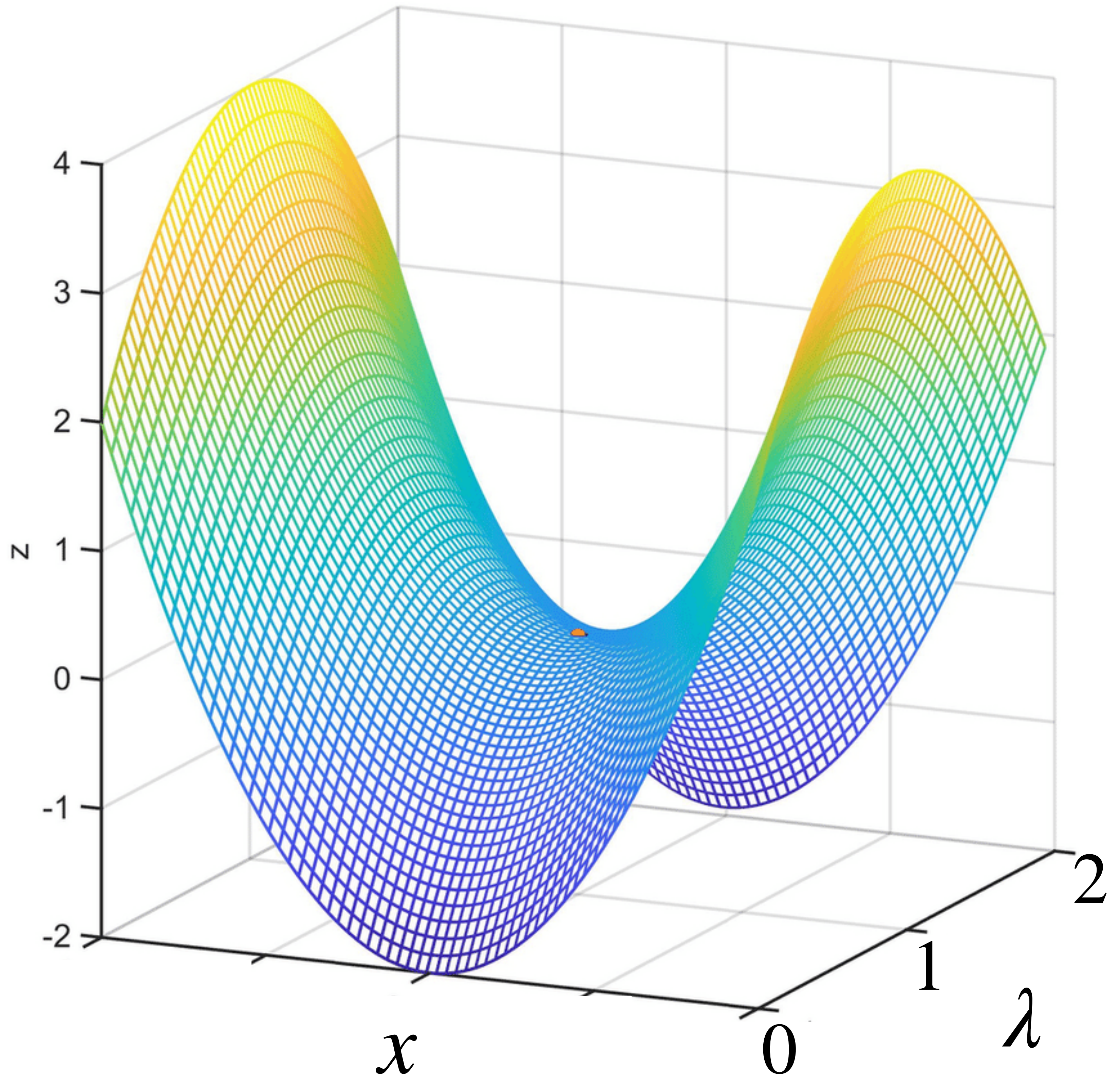
- **Question**: How to show this result?

A Natural Question to Ask

- Now we know that $p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathscr{L}(x, \lambda)$

A natural question is: Do we have

$$\inf_{x \in X} \sup_{\lambda \geq 0} \mathscr{L}(x, \lambda) = \sup_{\lambda \geq 0} \underbrace{\inf_{x \in X} \mathscr{L}(x, \lambda)}_{=: g(\lambda)} ?$$



If the above is true, then finding p^* is equivalent to solving $\sup_{\lambda \geq 0} g(\lambda)$

Dual Function and Dual Value

Recall that $p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathscr{L}(x, \lambda)$

Dual function: $g := \mathbb{R}^m \to \mathbb{R}$

$$g(\lambda) := \inf_{x \in X} \mathscr{L}(x, \lambda)$$

Dual problem and dual value:

$$d^* := \sup_{\lambda \geq 0} g(\lambda) = \sup_{\lambda \geq 0} \inf_{x \in X} \mathscr{L}(x, \lambda)$$

Example: Quadratic Program

Primal Problem $\quad \min_{x} \quad x^\top P x \qquad$ (Suppose $P \in \mathbb{R}^{n\times n}$ is pd)

subject to $\quad Ax \leq b \qquad$ ($A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^{m}$)

- Lagrangian? Dual function? Dual problem?

Example: Quadratic Program

Primal Problem $\min_x \quad x^\top P x$ (Suppose $P \in \mathbb{R}^{n \times n}$ is pd)

subject to $Ax \leq b$ ($A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$)

- Dual function

$$g(\lambda) = \inf_{x \in X} \left(x^\top P x + \lambda^\top (Ax - b) \right) = -\frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda - b^\top \lambda$$

- Dual problem

$$\sup_{\lambda \geq 0} \quad -\frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda - b^\top \lambda$$

Dual Function is Always Concave

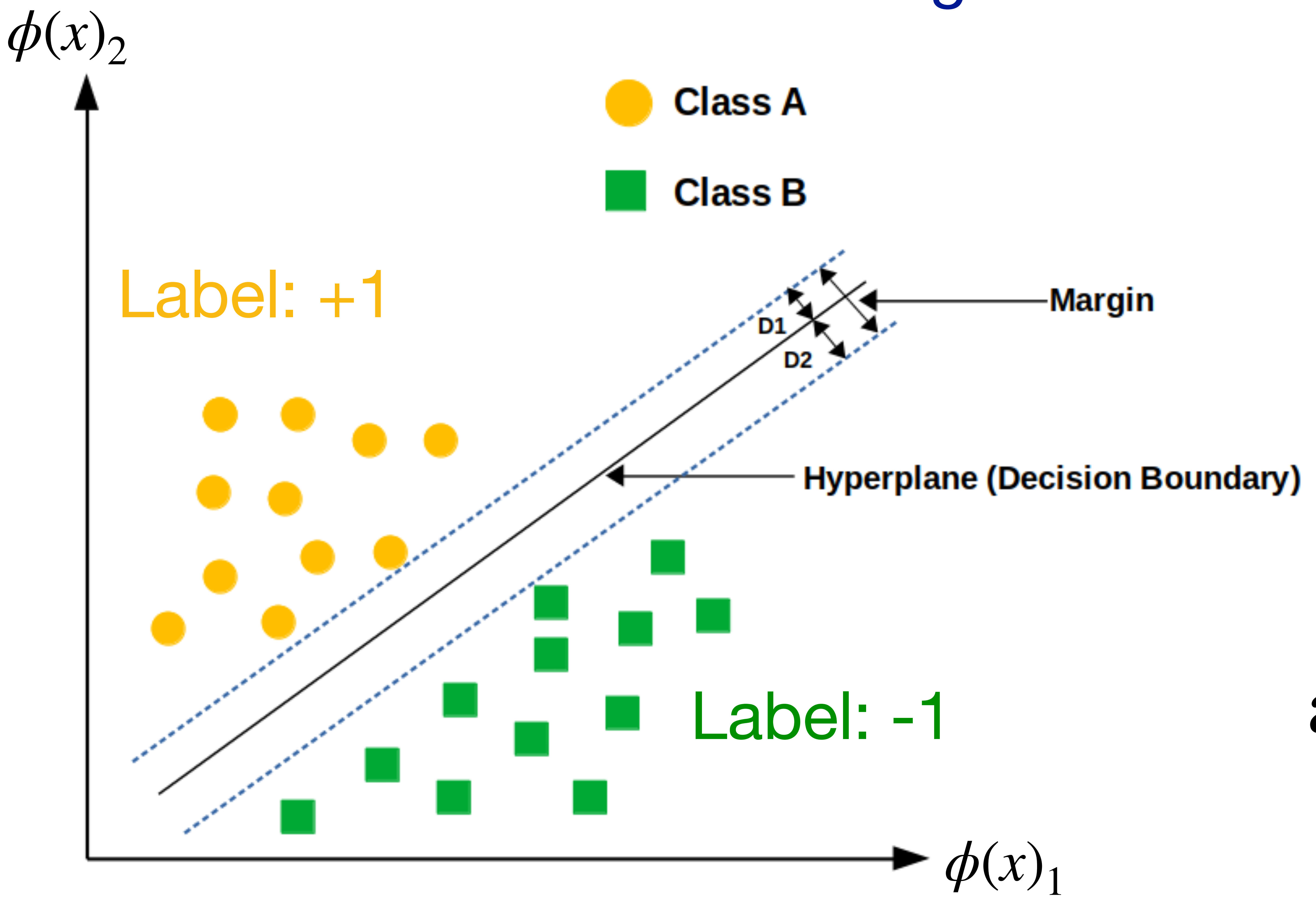
Theorem:

Let $g(\lambda)$ be the dual function of an arbitrary primal problem. Then, $g(\lambda)$ is always a *concave* function of λ (i.e., $-g(\lambda)$ is a convex function)

- Proof:

Example: Support Vector Machines (SVM)

SVM aka "maximum margin classifier"



- Linear model in SVM:

$$y_w(x) = w^\top \phi(x) + b$$

- Maximum margin solution

$$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_n [t_n \cdot \left(w^\top \phi(x^{(n)}) + b\right)] \right\}$$

Is this problem easy to solve?

Let $t_n \in \{-1, +1\}$ denote the true label of the n-th training sample

Example: Support Vector Machines (SVM)

- Original optimization problem of SVM

$$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_n [t_n \cdot \left(w^\top \phi(x^{(n)}) + b\right)] \right\}$$

Rescale w, b

$$\begin{cases} t_n\left(w^\top \phi(x^{(n)}) + b\right) = 1, & \text{if } x^{(n)} \text{ is the closest to the boundary,} \\ t_n\left(w^\top \phi(x^{(n)}) + b\right) > 1, & \text{otherwise.} \end{cases}$$

- An equivalent (and simpler) SVM formulation (aka “Dual SVM”)

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

subject to $t_n\left(w^\top \phi(x^{(n)}) + b\right) \geq 1$ (for all $n = 1, \cdots, N$)

Example: Support Vector Machines (SVM)

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{subject to} \quad t_n\left(w^\top \phi(x^{(n)}) + b\right) \geq 1 \qquad \text{(for all } n = 1,\cdots,N\text{)}$$

- Lagrangian? Dual function? Dual problem?

(Let $\phi(x^{(n)}) \equiv \phi_n$)

Example: Support Vector Machines (SVM)

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

subject to $t_n\left(w^\top \phi(x^{(n)}) + b\right) \geq 1$ (for all $n = 1, \cdots, N$)

- Lagrangian

(Let $\phi(x^{(n)}) \equiv \phi_n$)

$$\mathscr{L}(w, b, \lambda) = \frac{1}{2}\|w\|^2 + \sum_{n=1}^{N} \lambda_n \left(1 - t_n \left(w^\top \phi_n + b\right)\right)$$

- Dual function

$$g(\lambda) = \begin{cases} \sum_{n=1}^{N} \lambda_n - \frac{1}{2} \sum_{1 \leq n,m \leq N} \lambda_n \lambda_m t_n t_m \phi_n^\top \phi_m & \text{, } \textbf{if } \lambda \geq 0 \textbf{ and } \sum_{n=1}^{N} \lambda_n t_n = 0 \\ -\infty & \text{, } \textbf{otherwise} \end{cases}$$

3. Duality

Weak Duality

Recall that $p^* = \inf_{x \in X} \sup_{\lambda \geq 0} \mathscr{L}(x, \lambda)$

$$d^* := \sup_{\lambda \geq 0} g(\lambda) = \sup_{\lambda \geq 0} \inf_{x \in X} \mathscr{L}(x, \lambda)$$

Which one is larger?

Weak Duality (Formally)

Theorem (Weak Duality):

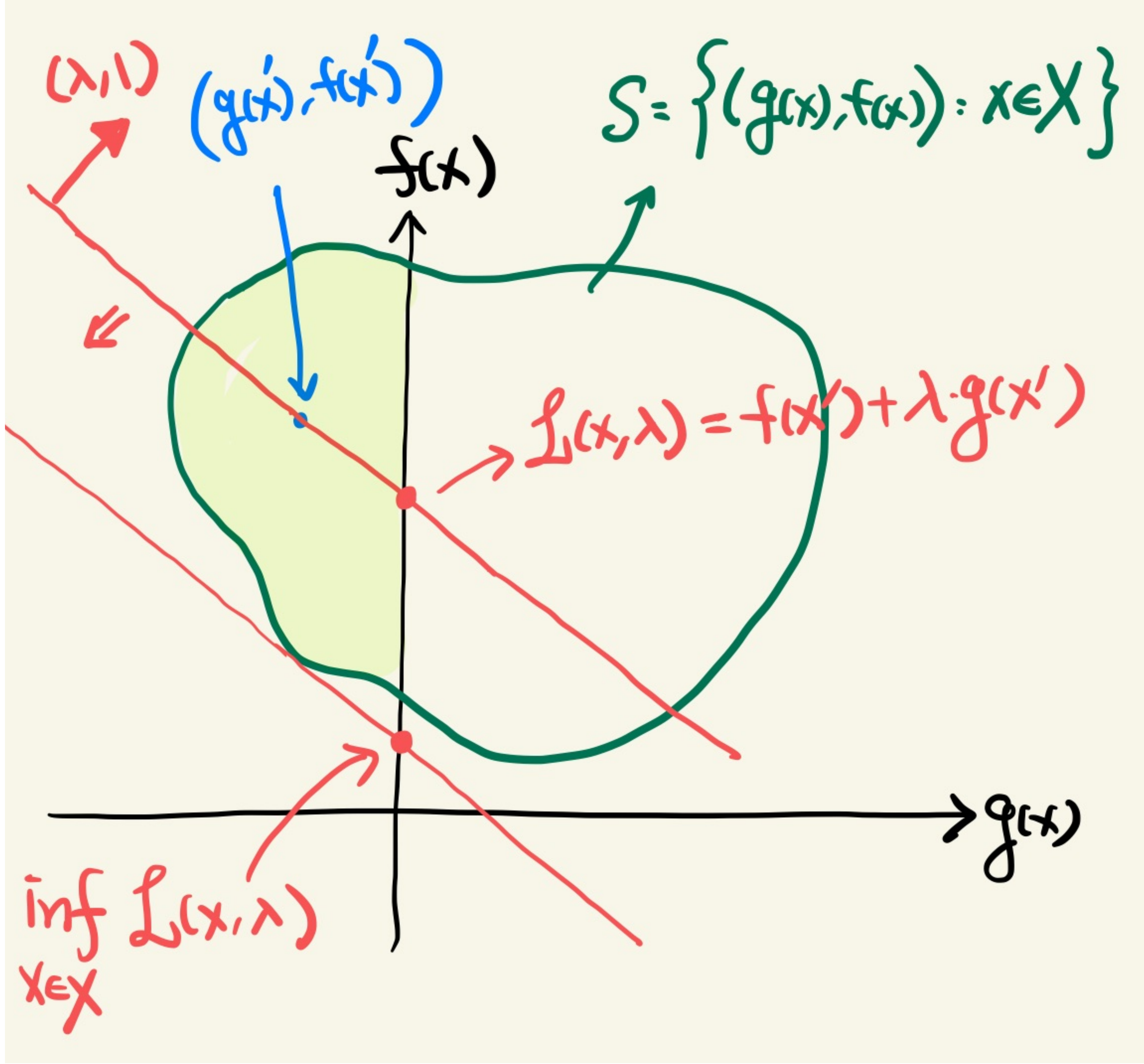
Given any primal and dual problem with optimal values p^*, d^*, we always have $p^* \geq d^*$

- **Remark**: One practical implication of weak duality is that if the primal problem is difficult, then we could solve the dual problem to get an "approximate solution" d^*

Geometric Interpretation

For ease of exposition, let's consider

$$\min_{x} \quad f(x)$$
$$\text{subject to} \quad g_1(x) \leq 0$$



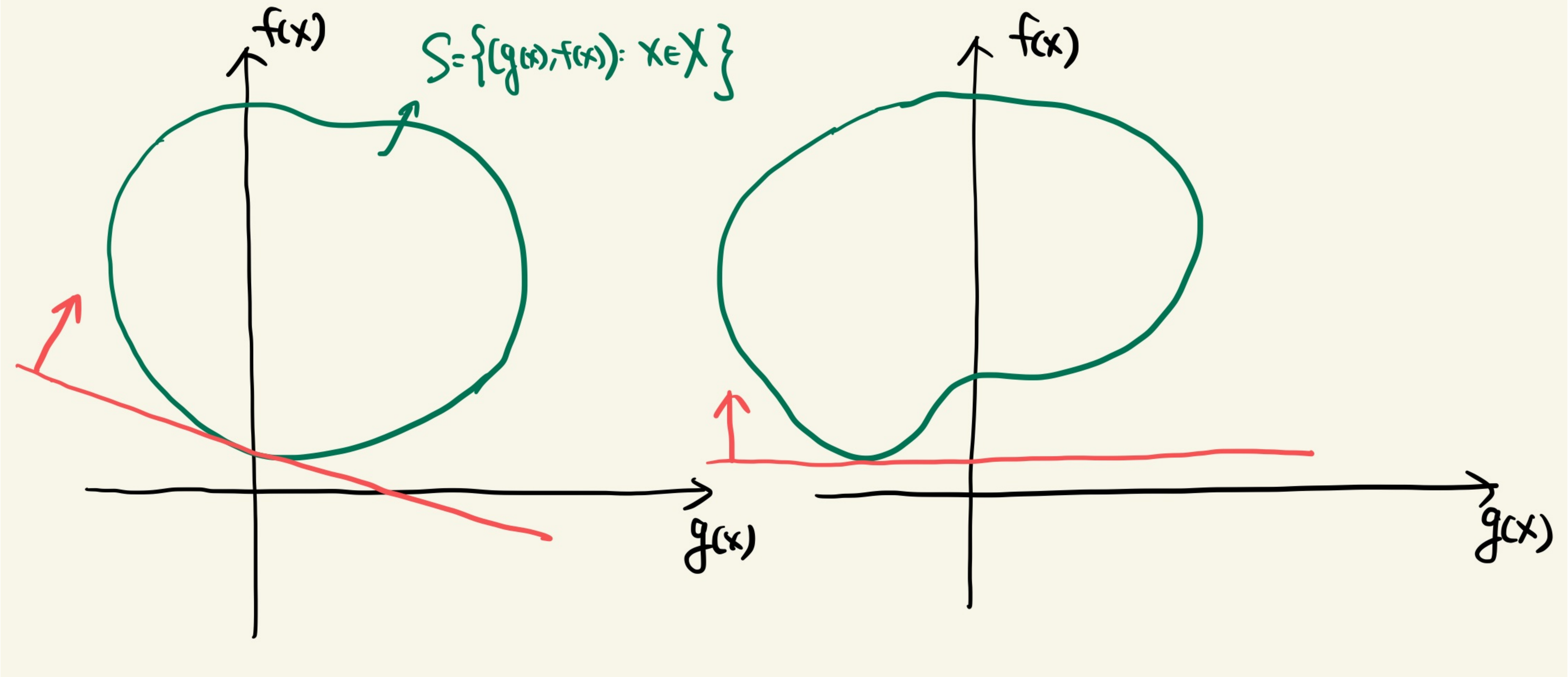
Visualize constraint-cost pair $(g_1(x), f(x))$

- Under each fixed λ, we have

$$\inf_{x \in X} \mathscr{L}(x, \lambda) = g(\lambda)$$

- To get d^*, we shall find a λ that gives us the maximum $g(\lambda)$

Geometric Interpretation: More Possible Scenarios



Proof of Weak Duality

Theorem (Weak Duality):

Given any primal and dual problem with optimal values p^*, d^*, we always have $p^* \geq d^*$

- Proof:

Step 1: We know $f(x') \geq \mathcal{L}(x', \lambda)$, for all $x' \in X$

Step 2: Therefore, for any $x \in X$, we have $f(x) \geq \inf_{x' \in X} \mathcal{L}(x', \lambda) = g(\lambda)$

Step 3: By taking "inf" over x, we have $\inf_{x \in X} f(x) \geq g(\lambda)$

Step 4: By taking "sup" over λ, we have $\sup_{\lambda \geq 0} \inf_{x \in X} f(x) \geq \sup_{\lambda \geq 0} g(\lambda)$

Could we have $p* = q*$?

(If so, under what conditions?)

Strong Duality and Duality Gap

- Definition (Duality gap): Consider a primal problem (P) with its dual (D). Let p^* and d^* denote the primal optimum and the dual optimum, respectively. Then. The duality gap Δ is defined as

$$\Delta := p^* - d^*$$

- Definition (Strong duality): We say that strong duality holds if $\Delta = 0$.

- There are various sufficient conditions for *strong duality*!
- We will discuss two popular conditions: Slater's and KKT

Slater's Constraint Qualification (CQ)

Consider the following "convex problem"

$$\min_{x} \quad f(x)$$
$(f(x)$ is a convex function)

$$\text{subject to} \quad g_i(x) \leq 0, \forall i$$
($g_i(x)$'s are convex functions)

Theorem (Slater's Condition):

Given a convex problem, if the problem is strictly feasible (i.e., $g_i(x) < 0$ for all i), then *strong duality* holds.

LAGRANGE MULTIPLIERS REVISITED

A Contribution to Non-Linear Programming

by Morton Slater

November 7, 1950

1. Introduction

The present paper was inspired by the work of Kuhn and Tucker [1] [1]. These authors transformed a certain class of constrained maximum problems into equivalent saddle value (minimax) problems.

Their work seems to hinge on the consideration of still a third type of problem. A very simple but illustrative form of this problem is the following: let $x \in$ positive orthant of some finite dimensional Euclidean space, and let f and g be real valued functions of x with the property that whenever $f \geq 0$, then also $g \geq 0$; under what conditions can one then conclude that $\exists$ a non-negative constant u such that $uf \leq g$ for all $x \geq 0$?

Kuhn and Tucker showed that if f is concave and differentiable, if g is convex and differentiable, and if the set $\{x: f(x) \geq 0\}$ satisfies certain regularity restrictions, then there does indeed exist such a u.

Two directions for generalization are presented:

First of all, the Kuhn-Tucker argument rests heavily on the

"… Cowles Commission Discussion Papers are preliminary materials circulated privately to stimulate private discussion and are not ready for critical comment for appraisal in publications. …"

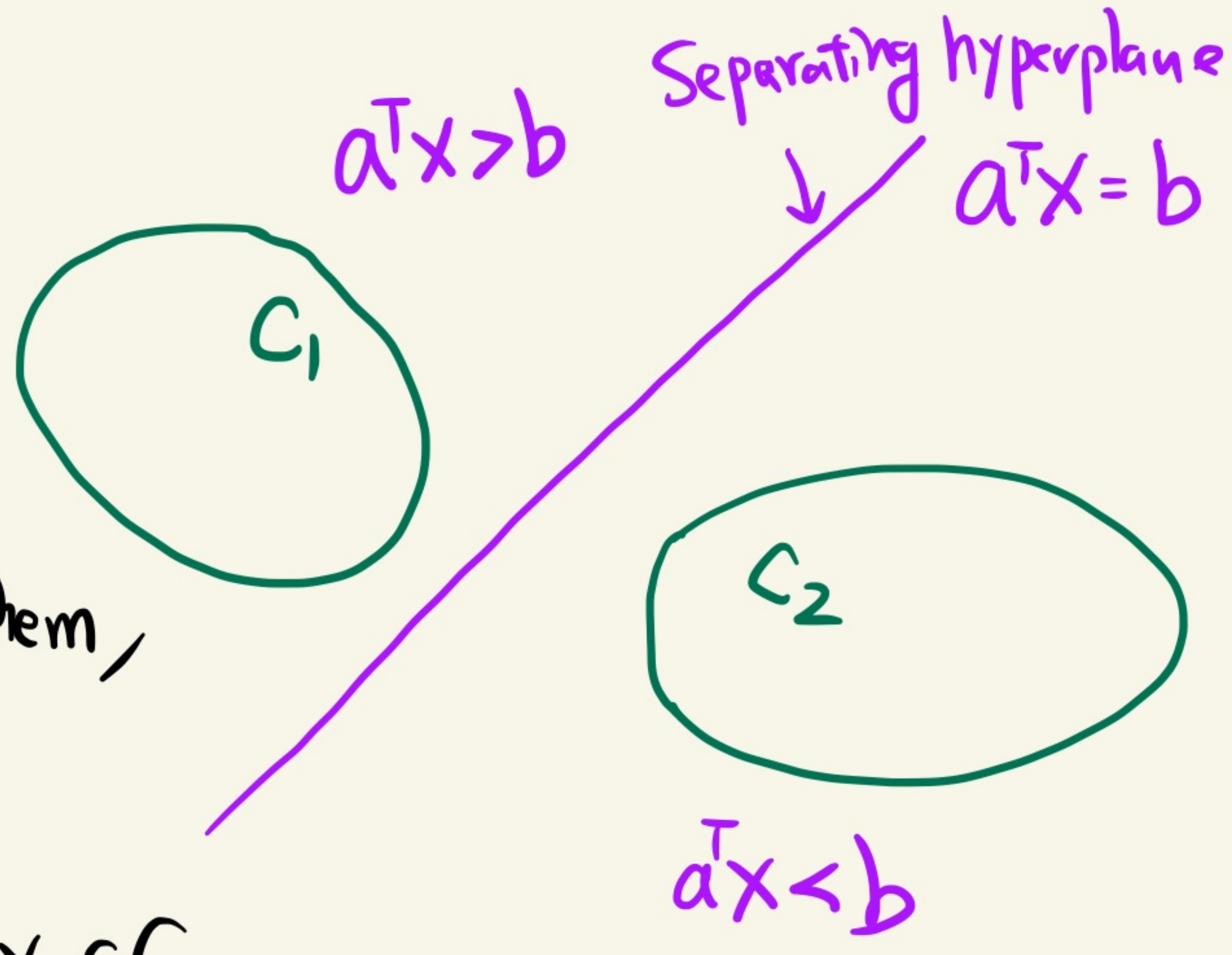
Morton Slater, "Lagrange Multipliers Revisited," *Cowles Commission Discussion Paper No. 403*, 1950

Separating Hyperplane Theorem

Theorem: If C_1, C_2 are two "nonempty" and "disjoint" convex subsets of $\mathbb{R}^n$.

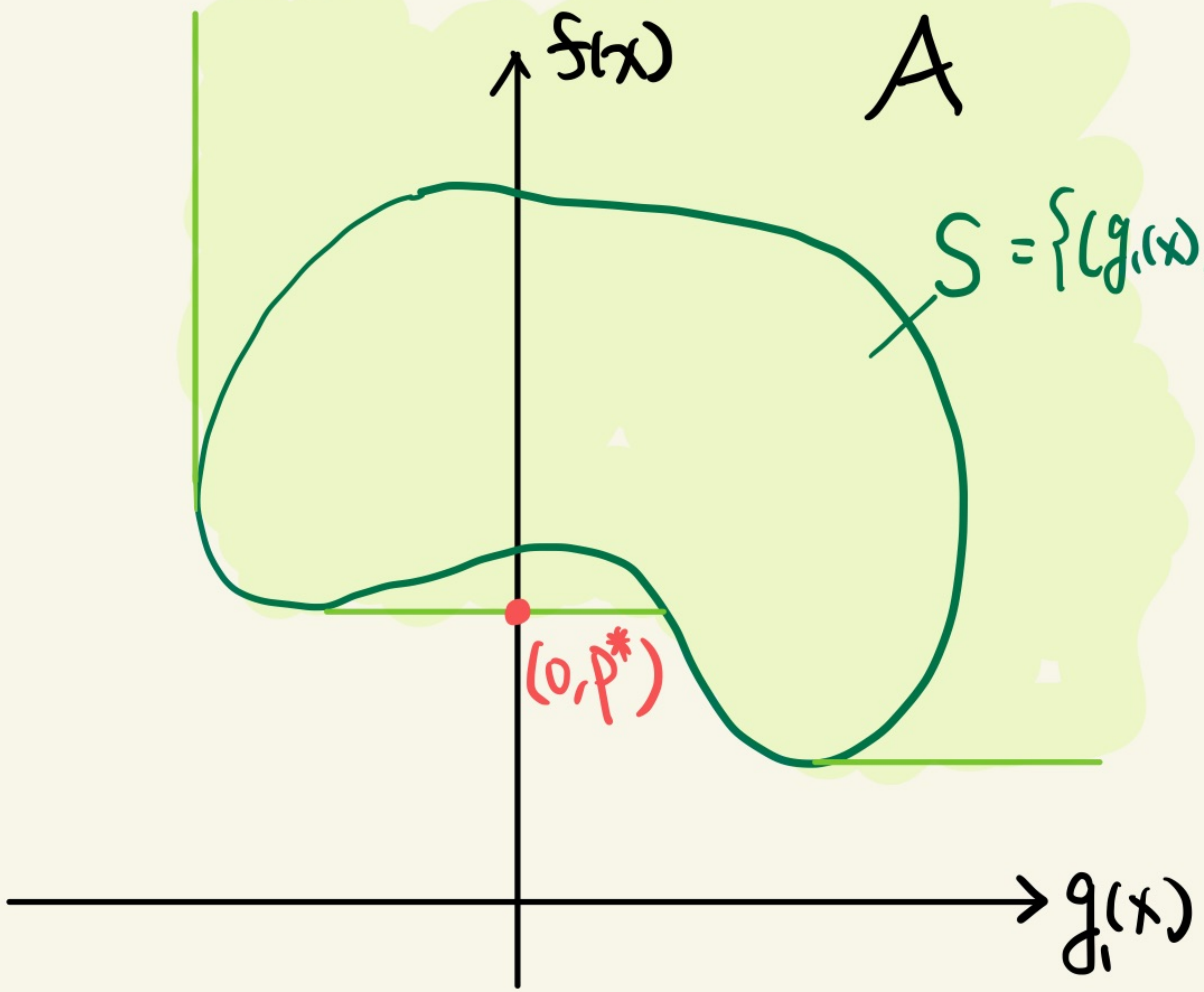
Then, there exists a hyperplane that separates them, i.e., a vector $a \neq 0$ such that

$$a^T x_1 \le a^T x_2 \ , \text{ for all } x_1 \in C_1, x_2 \in C_2$$



(Proof: See Chapter 2.5 of Stephen Boyd's textbook)

A Useful Set A



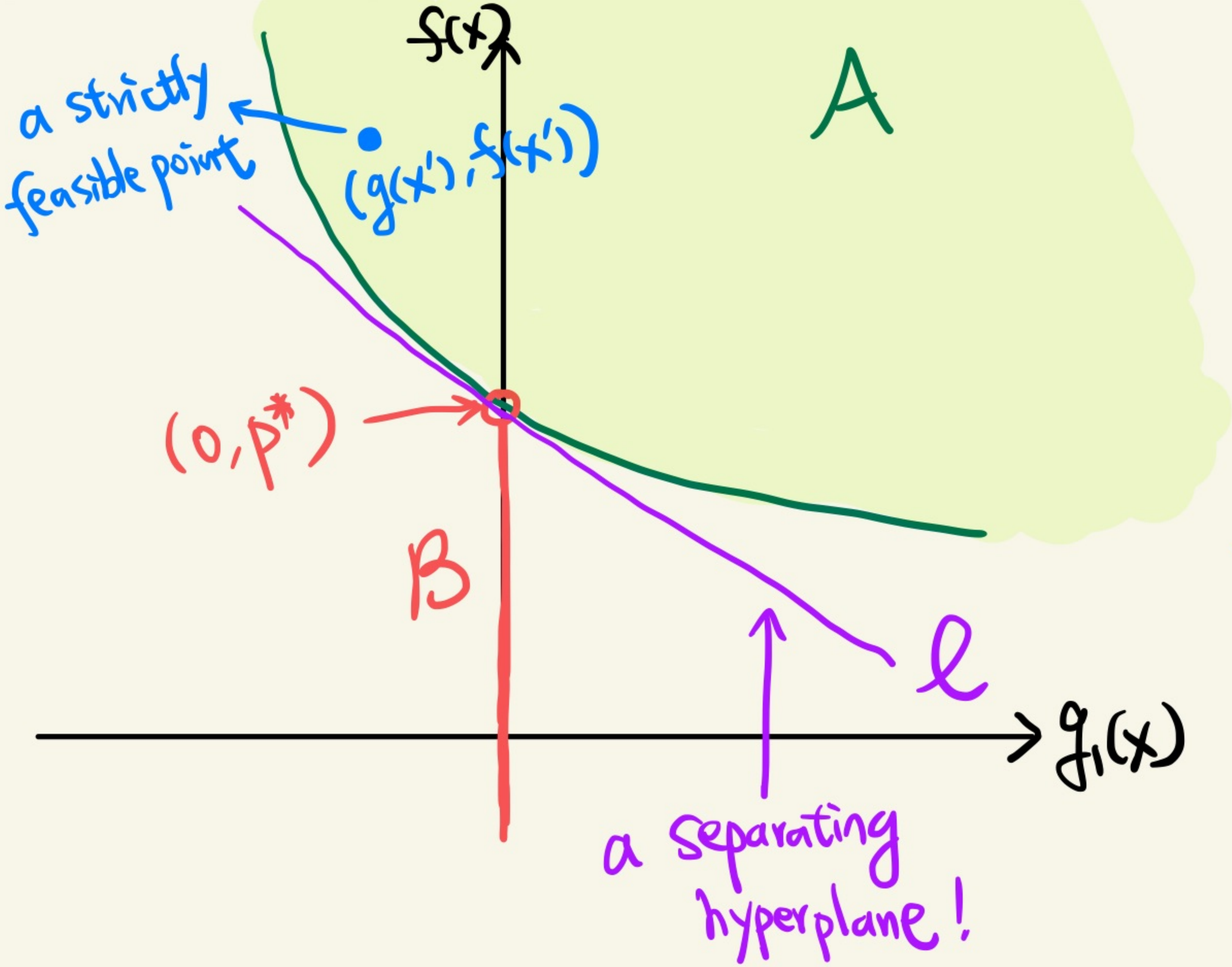
Recall the primal problem: $\min_x f(x)$

subject to $g_1(x) \leq 0$.

$A := \{(\alpha, \beta) : \exists\, x \in X$ such that $f(x) \leq \beta,\ g_1(x) \leq \alpha\}$

Property: A is a convex set if f and g_1 are convex function

Geometric Interpretation of Slater's Condition



$$A := \{(\alpha, \beta) : \exists x \in X \text{ such that } g_1(x) \le \alpha,\ f(x) \le \beta\}$$

$$B := \{(0, \beta) : \beta < p^*\}$$

Question: Are A and B convex sets?

Are A and B disjoint?

- In the 1-constraint case: Slater's condition implies that ℓ is not a vertical line.

Optimality Conditions by Lagrangian

(Primal Problem)

$$\min_x f(x)$$

subject to $g_i(x) \leq 0, \quad i=1,\cdots,m$

Recall: Optimality condition in Lec2

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \text{ for all feasible } x$$

Theorem (Lagrangian optimality condition, or LOC)

Suppose strong duality holds and that p^* and d^* can be attained.

Then, we have $x^* \in \underset{x \in X}{\operatorname{argmin}} \mathcal{L}(x, \lambda^*)$

Question: Why is this condition useful?

Proof of LOC

Want to show: $x^* \in \operatorname*{argmin}_{x \in X} \mathcal{L}(x, \lambda^*)$

As p^* and d^* can be attained, there exists x^* and λ^* such that

$$p^* = f(x^*) \quad \cdots\cdots \quad (\qquad\qquad)$$

$$= d^* \quad \cdots\cdots \quad (\qquad\qquad)$$

$$= g(\lambda^*) \quad \cdots\cdots \quad (\qquad\qquad)$$

$$= \min_x \mathcal{L}(x, \lambda^*) \quad \cdots\cdots \quad (\qquad\qquad)$$

$$\leq \mathcal{L}(x^*, \lambda^*) \quad \cdots\cdots \quad (\qquad\qquad)$$

$$\leq f(x^*) = p^* \quad \cdots\cdots \quad (\qquad\qquad)$$

From LOC to Complementary Slackness

Recall LOC: $x^* \in \underset{x \in X}{\operatorname{argmin}} \mathcal{L}(x, \lambda^*)$

Theorem (Complementary Slackness)

LOC implies that $\lambda_i^* g_i(x^*) = 0$, for all $i = 1, \cdots, m$.

Question: Why is it called "complementary slackness"?

From LOC to Complementary Slackness

Recall LOC: $x^* \in \underset{x \in X}{\operatorname{argmin}} \mathcal{L}(x, \lambda^*)$

Theorem (Complementary Slackness)

LOC implies that $\lambda_i^* g_i(x^*) = 0$, for all $i = 1, \cdots, m$.

Proof:

Since $\mathcal{L}(x^*, \lambda^*) = f(x^*)$, we have $\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0$

Moreover, given that $\lambda_i^* \geq 0$ and $g_i(x^*) \leq 0$, then we must have

$\lambda_i^* \cdot g_i(x^*) = 0$, for all i

4. KKT Conditions

Karush-Kuhn-Tucker (KKT) Optimality Conditions

KKT Conditions

① $g_i(x^*) \leq 0$, for all $i=1,\cdots,m$ (primal feasibility)

② $\lambda_i^* \geq 0$, for all $i=1,\cdots,m$ (dual feasibility)

③ $\lambda_i^* \cdot g_i(x^*) = 0$, for all $i=1,\cdots,m$ (complementary slackness)

④ $\nabla_x \mathcal{L}(x,\lambda^*)\big|_{x=x^*} = 0.$ (Lagrangian stationarity)

Theorem If strong duality holds and (x^*,λ^*) exists, the KKT conditions must hold.

Remark: This is true for "any" differentiable optimization problem

Next Question: When will KKT conditions be *sufficient*?

Convex problems + constraint qualification (e.g., Slater's)!

Sufficiency of KKT Conditions

Theorem Let f and g_i's be convex functions.

If the KKT conditions ①–④ hold under $(\bar{x}, \bar{\lambda})$,

then: (i) $\bar{x}$ and $\bar{\lambda}$ are primal and dual optimal solutions,

(ii) Strong duality holds

Recall: KKT Conditions

① $g_i(x^*) \le 0$, for all i

② $\lambda_i^* \ge 0$, for all i

③ $\lambda_i^* \cdot g_i(x^*) = 0$, for all i

④ $\nabla_x \mathcal{L}(x, \lambda^*) = 0$.

Proof: step 1: By ②, we know $\mathcal{L}(x, \bar{\lambda})$ is convex in x.

step 2: By ④, we know $\bar{x}$ is a minimizer of $\mathcal{L}(x, \bar{\lambda})$.

step 3: $$g(\bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \cdot g_i(\bar{x}) = f(\bar{x})$$

why?

This implies that the duality gap is zero.