TMA947 / MMG621 — Nonlinear optimization

Lecture 1 — Introduction to optimization

Emil Gustavsson, Zuzana Nedělková

October 20, 2017

What is optimization?

Optimization is a mathematical discipline which is concerned with finding the minima or maxima of functions, possibly subject to constraints.

Basic notation

- Vectors are written with bold face, i.e., $\boldsymbol{x} \in \mathbb{R}^n$.
- Elements in a vector are written as $x_j, \; j = 1, \ldots, n$.
- All vectors are column vectors.
- The inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is written as $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{a} = \sum_{j=1}^{n} a_j b_j$.
- The norm $|| \cdot ||$ denotes the Euclidean norm, i.e., $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \sqrt{\sum_{j=1}^{n} x_j^2}$.
- We utilize vector inequalities, $\boldsymbol{a} \leq \boldsymbol{b}$, meaning that $a_j \leq b_j$, $j = 1, \ldots, n$.

Optimization problem formulation

In order to introduce a general optimization problem, we need to define the following:

$\boldsymbol{x} \in \mathbb{R}^n$	: vector of decision variables $x_j, \; j = 1, \ldots, n$,
$f : \mathbb{R}^n \to \mathbb{R} \cup \pm\infty$	: objective function,
$X \subseteq \mathbb{R}^n$	: ground set,
$g_i : \mathbb{R}^n \to \mathbb{R}$	: constraint function defining restriction on $\boldsymbol{x}$,
$g_i \geq 0, \; i \in \mathcal{I}$	: inequality constraints,
$g_i = 0, \; i \in \mathcal{E}$	: equality constraints.

A general **optimization problem** is to

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}), && \text{(1a)} \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq 0, \quad i \in \mathcal{I}, && \text{(1b)} \\ & g_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E}, && \text{(1c)} \\ & \boldsymbol{x} \in X. && \text{(1d)} \end{aligned}$$

(If we consider a maximization problem, we change the sign of f to get a minimization problem.)

Classification of optimization problems

Linear Programming (LP):
- Linear objective function $f(\boldsymbol{x}) = \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} = \sum_{j=1}^{n} c_j x_j$,
- Affine constraint functions $g_i(\boldsymbol{x}) = \boldsymbol{a}_i{}^{\mathsf{T}}\boldsymbol{x} - b_i,\ i \in \mathcal{I} \cup \mathcal{E}$
- Ground set X defined by affine equalities/inequalities.

Nonlinear programming (NLP):
- Some functions $f, g_i,\ i \in \mathcal{I} \cup \mathcal{E}$ are nonlinear.

Unconstrained optimization:
- $\mathcal{I} \cup \mathcal{E} = \emptyset$,
- $X = \mathbb{R}^n$.

Constrained optimization:
- $\mathcal{I} \cup \mathcal{E} \neq \emptyset$, and/or
- $X \subset \mathbb{R}^n$.

Integer programming (IP):
- $X \subseteq \mathbb{Z}^n$ or $X \subseteq \{0,1\}^n$.

Convex programming (CP):
- $f, g_i,\ i \in \mathcal{I}$ are convex functions,
- $g_i,\ i \in \mathcal{E}$ are affine,
- X is closed and convex.

Conventions

Let $S = \{\boldsymbol{x} \in \mathbb{R}^n \ \mid g_i(\boldsymbol{x}) \leq 0,\ i \in \mathcal{I},\ g_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E},\ \boldsymbol{x} \in X\}$ denote a feasible set.

What do we mean by solving the problem to $\underset{\boldsymbol{x} \in S}{\text{minimize}}\ f(\boldsymbol{x})$?

Let

$$f^* := \underset{\boldsymbol{x} \in S}{\text{infimum}}\ f(\boldsymbol{x})$$

denote the infimum value of f over the set S. If the value f^* is attained at some point $\boldsymbol{x}^*$ in S, we can write

$$f^* := \underset{\boldsymbol{x} \in S}{\text{minimum}}\ f(\boldsymbol{x}),$$

and have $f(\boldsymbol{x}^*) = f^*$. Another well-defined operator defines the set of minimal solutions to the problem

$$S^* := \arg \underset{\boldsymbol{x} \in S}{\text{minimum}}\ f(\boldsymbol{x}),$$

where $S^* \subseteq S$ is nonempty if and only if the infimum value f^* is attained at some point $\boldsymbol{x}^*$ in S.

Now we can define what we mean by the problem to $\underset{\boldsymbol{x} \in S}{\text{minimize}}\ f(\boldsymbol{x})$.

"to $\underset{\boldsymbol{x} \in S}{\text{minimize}}\ f(\boldsymbol{x})$" means "find f^* and an $\boldsymbol{x}^* \in S^*$"

If we have an optimization problem

$$P: \quad \underset{\boldsymbol{x} \in S}{\text{minimize}}\ f(\boldsymbol{x})$$

- A point $\boldsymbol{x}$ is **feasible** in problem P if $\boldsymbol{x} \in S$. The point is **infeasible** in problem P if $\boldsymbol{x} \notin S$.
- The problem P is feasible if there exist a $\boldsymbol{x} \in S$ and the problem P is infeasible if $S = \emptyset$.
- A point $\boldsymbol{x}^*$ is an **optimal solution** to P if $\boldsymbol{x}^* \in \arg \underset{\boldsymbol{x} \in S}{\text{minimum}}\ f(\boldsymbol{x})$.
- f^* is an **optimal value** to P if $f^* = \underset{\boldsymbol{x} \in S}{\text{minimum}}\ f(\boldsymbol{x})$.

Examples

I. Consider the problem to

$$\begin{aligned} \text{minimize} \quad & (x+1)^2, \\ \text{subject to} \quad & x \in \mathbb{R}, \end{aligned}$$

Easy problem, $(x+1)^2$ is convex, no constraints. Just solve $f'(x) = 0$, and get the optimal solution $x^* = -1$ and the optimal value $f^* = 0$.

(Convex, quadratic, unconstrained optimization problem)

II. A more complicated problem is to

$$\begin{aligned} \text{minimize} \quad & (x+1)^2, \\ \text{subject to} \quad & x \geq 0. \end{aligned}$$

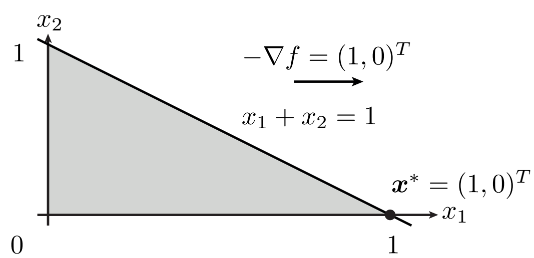
Now the "$f'(x) = 0$" trick does not work and we need to consider the boundary. We get the optimal solution $x^* = 0$ and the optimal value $f^* = 1$.

(Convex, quadratic, constrained optimization problem)

III. Consider the problem to

$$\begin{aligned} \text{minimize} \quad & -x_1, \\ \text{subject to} \quad & x_1 + x_2 \leq 1, \\ & x_1, x_2 \geq 0. \end{aligned}$$

We solve this graphically. So optimal solution is $\boldsymbol{x}^* = (1,0)^T$ and the optimal value if $f^* = -1$.



The diet problem

As a first example of an real optimization problem, we consider the **diet problem** (first formulated by George Stigler).

> For a moderately active person, how much of each of a number of foods should be eaten on a daily basis so that the person's intake of nutrients will be at least equal to the recommended dietary allowances (RDAs), with the cost of the diet being minimal?

Good example to show

- how to model a real optimization problem,
- why a realistic model sometimes can be difficult to achieve.

We consider the case when the only allowed foods can be found at McDonalds.

> For a moderately active person, how much of each of a number of McDonald foods (see Table 1) should be eaten on a daily basis so that the person's intake of nutrients will be at least equal to the recommended dietary allowances (RDAs), with the cost of the diet being minimal?

Food	Calories	Carb	Protein	Vit A	Vit C	Calc	Iron	Cost
Big Mac	550 kcal	46g	25g	6%	2%	25%	25%	30kr
Cheeseburger	300 kcal	33g	15g	6%	2%	20%	15%	10kr
McChicken	360 kcal	40g	14g	0%	2%	10%	15%	35kr
McNuggets	280 kcal	18g	13g	0%	2%	2%	4%	40kr
Caesar Sallad	350 kcal	24g	23g	160%	35%	20%	10%	50kr
French Fries	380 kcal	48g	4g	0%	15%	2%	6%	20kr
Apple Pie	250 kcal	32g	2g	4%	25%	2%	6%	10kr
Coca-Cola	210 kcal	58g	0g	0%	0%	0%	0%	15kr
Milk	100 kcal	12g	8g	10%	4%	30%	8%	15kr
Orange Juice	150 kcal	30g	2g	0%	140%	2%	0%	15kr
RDA	2000 kcal	350g	55g	100%	100%	100%	100%	

Table 1: Given data for the diet problem

We define the **sets**

$$\begin{aligned} \text{Foods} := \{&\text{Big Mac, Cheeseburger, McChicken, McNuggets, Caesar Sallad} \\ &\text{French Fried, Apple Pie, Coca-Cola, Milk, Orange Juice}\}, \\ \text{Nutrients} := \{&\text{Calories, Carb, Protein, Vit A, Vit C, Calc, Iron.}\} \end{aligned}$$

Then we define the **parameters**

$$\begin{aligned} a_{ij} &= \text{Amount of nutrient } i \text{ in food } j,\ i \in \text{Nutrients},\ j \in \text{Foods}, \\ b_i &= \text{Recommended daily amount (RDA) of nutrient } i,\ i \in \text{Nutrients}, \\ c_j &= \text{Cost for food } j,\ j \in \text{Foods}, \end{aligned}$$

and the **decision variables**

$$x_j = \text{Amount of food } j \text{ we should eat each day},\ j \in \text{Foods}.$$

The **model** of the diet optimization problem is then to

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in \text{Foods}} c_j x_j, && \text{(2a)} \\ \text{subject to} \quad & \sum_{j \in \text{Foods}} a_{ij} x_j \geq b_i, \quad i \in \text{Nutrients}, && \text{(2b)} \\ & x_j \geq 0, \quad j \in \text{Foods}. && \text{(2c)} \end{aligned}$$

(2a) We minimize the total cost, such that

(2b) we get enough of each nutrient, and such that

(2c) we don't sell anything to McDonalds.

The optimal solution is then

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 7.48 \\ 0 \\ 0 \\ 0.27 \\ 0 \\ 3.03 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Total cost = 118.47 kr.
Total intake of calories = 3093.51 kcal.

If we add the constraint that x_j should be integer, the solution is

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \\ 0 \\ 0 \\ 1 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Total cost = 150 kr.
Total intake of calories = 3200 kcal.

Now consider going on a diet, meaning that we would like to eat as few calories as possible. We reformulate our model to

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in \text{Foods}} a_{\text{Calories},j} x_j, && \text{(3a)} \\ \text{subject to} \quad & \sum_{j \in \text{Foods}} a_{ij} x_j \geq b_i, \quad i \in \text{Nutrients} \setminus \{\text{Calories}\}, && \text{(3b)} \\ & x_j \geq 0, \quad j \in \text{Foods}. && \text{(3c)} \end{aligned}$$

The optimal solution is then

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3.96 \\ 12.41 \\ 0.36 \end{pmatrix}.$$

Total cost	= 251.01 kr.
Total intake of calories	= 2127.47 kcal.

If we add the constraint that x_j should be integer, the solution is

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 11 \\ 6 \end{pmatrix}.$$

Total cost	= 270 kr.
Total intake of calories	= 2210 kcal.

The real diet problem

When first studied by the Stigler, the problem concerned the US military and had 77 different foods in the model. He didn't managed to solve the problem to optimality, but almost. The near optimal diet was

- Wheat flour
- Evaporated milk
- Cabbage
- Spinach
- Dried navy beans

at a cost of \$0.1 a day in 1939 US dollars.

Course material

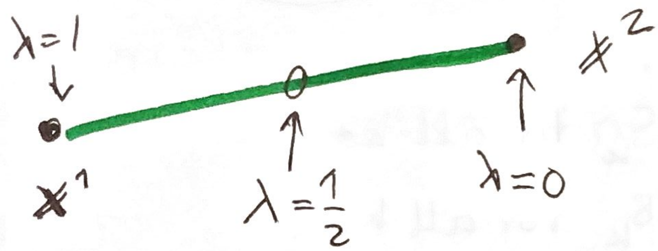
Lecture 1	Define and model optimization problems, classification
Lecture 2	Convexity of sets, functions, optimization problems
Lecture 3	Optimality conditions, introduction
Lecture 4	Unconstrained optimization, methods, classification.
Lecture 5	Optimality conditions, continued
Lecture 6	The Karush-Kuhn-Tucker conditions
Lecture 7	Lagrangian duality
Lecture 8	Linear programming, introduction
Lecture 9	Linear programming, continued
Lecture 10	Duality in linear programming
Lecture 11	Convex optimization
Lecture 12	Integer programming
Lecture 13	Nonlinear optimization methods, convex feasible sets
Lecture 14	Nonlinear optimization methods, general sets
Lecture 15	Overview of the course

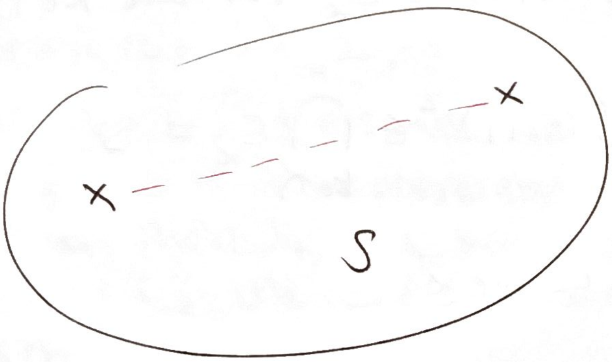
TISDAG
3 september
16.30

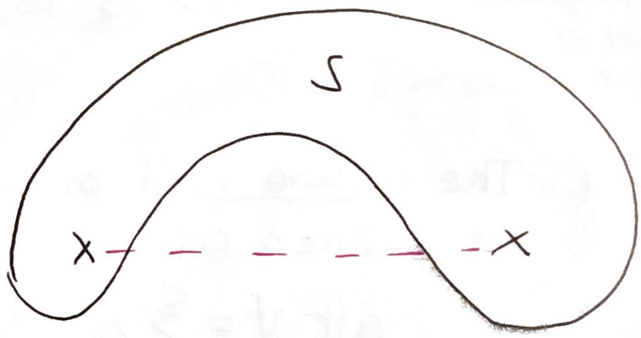
LECTURE 2

DEF. The set $S \subseteq \mathbb{R}^n$ is <u>convex</u> if

$$\left.\begin{array}{l} x^1, x^2 \in S \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow \lambda x^1 + (1-\lambda)x^2 \in S$$







<u>EXAMPLES</u>

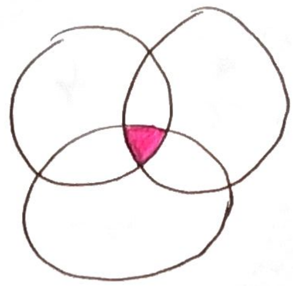
- $\emptyset$ (EMPTY SET) is convex
- The set $\{x \in \mathbb{R}^n : \|x\| \le \alpha\}$ is convex for any $\alpha \in \mathbb{R}$
- The set $\{x \in \mathbb{R}^n : \|x\| = \alpha\}$ is not convex for $\alpha > 0$.
- $\{1, 2, 3, 4\}$ is not convex.

PROP.

Let S_k, $k \in K$ be a collection of convex sets.
Then the intersection

$$S = \bigcap_{k \in K} S_k$$

is convex!



PROOF:

Let $x^1, x^2 \in S$. $\Rightarrow$ $x^1 \in S_k$ for all $k \in K$
$x^2 \in S_k$ for all $k \in K$

Let $\lambda \in (0,1)$ $\Rightarrow$ $\lambda x^1 + (1-\lambda) x^2 \in S_k$ for all $k \in K$
since S_k is convex.

$$\Rightarrow \lambda x^1 + (1-\lambda) x^2 \in \bigcap_{k \in K} S_k = S$$

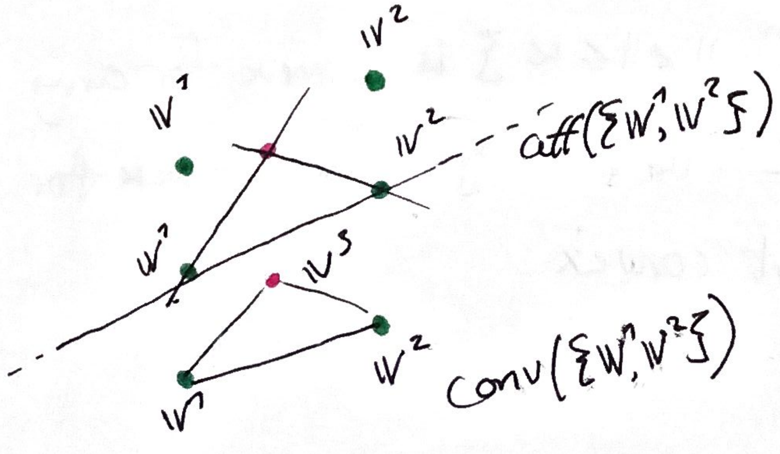
$\Rightarrow$ S is convex

DEF. The <u>affine hull</u> of a finite set $V = \{v^1, \dots, v^k\}$ is defined as

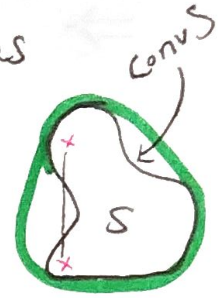
$$\text{aff } V = \{ \lambda_1 v^1 + \dots + \lambda_k v^k \mid \lambda_1, \dots, \lambda_k \in \mathbb{R} : \sum_{i=1}^{k} \lambda_i = 1 \}$$

The <u>convex hull</u> —"—

$$\text{conv } V = \{ \lambda_1 v^1 + \dots + \lambda_k v^k \mid \lambda_1, \dots, \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1 \}$$



In general, the convex hull of a set S is defined as



a) the unique minimal convex set containing S

b) the intersection of all convex sets containing S

c) the set of all convex combinations of points in S

Thm. (CARATHEODORY'S THM)

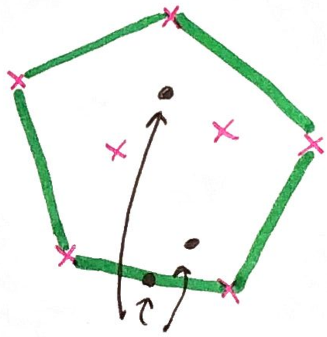
Let $x \in \operatorname{conv} S$, where $S \subseteq \mathbb{R}^n$. Then x can be expressed as a convex combination of $n+1$ or fewer points of S.

PROOF.

We know that

$$x = \lambda_1 a^1 + \ldots + \lambda_m a^m,$$

where $\lambda_i \geq 0$, $\sum_{i=1}^{m} \lambda_i = 1$ and $a^1, \ldots, a^m \in S$.
Assume that this representation is minimal. Then $\lambda_1, \ldots, \lambda_m > 0$ and $a^i \neq a^j$ for any i,j. We need to show that $m \leq n+1$.



can be expressed with less than or exactly 3 ×

Assume $m > n+1 \Rightarrow$

$\{a^1, \ldots, a^m\}$ is affinely dependent

$\exists\, \alpha_1, \ldots, \alpha_m$

$$\left.\begin{aligned} \alpha_1 a^1 + \ldots + \alpha_m a^m &= 0 \\ \alpha_1 + \ldots + \alpha_m &= 0 \end{aligned}\right\}$$

"stronger than linearly dependent"

Let $\varepsilon > 0$ such that

$$\lambda_1 + \varepsilon\alpha_1, \ldots, \lambda_m + \varepsilon\alpha_m$$

are non-negative and at least one is zero,
(which is always possible since at least 1 α_i is negative and all $\lambda_i > 0$.)

$$\Rightarrow x = \lambda_1 a^1 + \ldots + \lambda_m a^m + \varepsilon \underbrace{(\alpha_1 a^1 + \ldots + \alpha_m a^m)}_{0} =$$

$$= (\lambda_1 + \varepsilon \alpha_1) a^1 + \ldots + (\lambda_m + \varepsilon \alpha_m) a^m$$

A new representation of x but where one weight is 0.

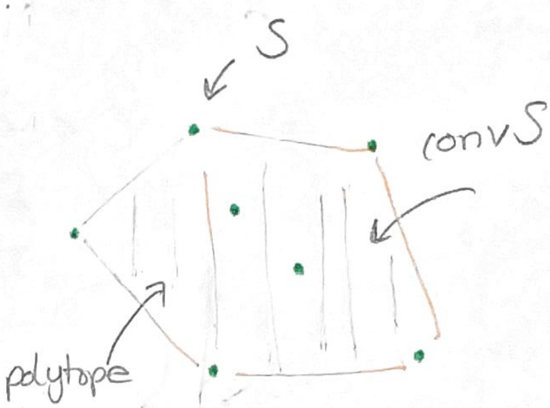
CONTRADICTION to minimal representation. $\Rightarrow$

$$m \leq n+1$$

MÅNDAG
9 september
8:00

continuing LECTURE 2

DEF A set $P \in \mathbb{R}^n$ is a polytope if it is the convex hull of finitely many points in $\mathbb{R}^n$.



Ex. A cube or tetrahedon in $\mathbb{R}^3$ are polytopes.

DEF A point $v \in \mathbb{R}^n$ in a convex set P is an extreme point if

$$\left.\begin{array}{l} v = \lambda x^1 + (1-\lambda)x^2 \\ x^1, x^2 \in P \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow v = x^1 = x^2$$

THM Let P be the polytope conv$V$, where $V = \{v^1, \dots, v^k\}$. Then P is equal to the convex hull of its extreme points.

DEF A set $P \in \mathbb{R}^n$ is a polyhedron if there exists a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ such that

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

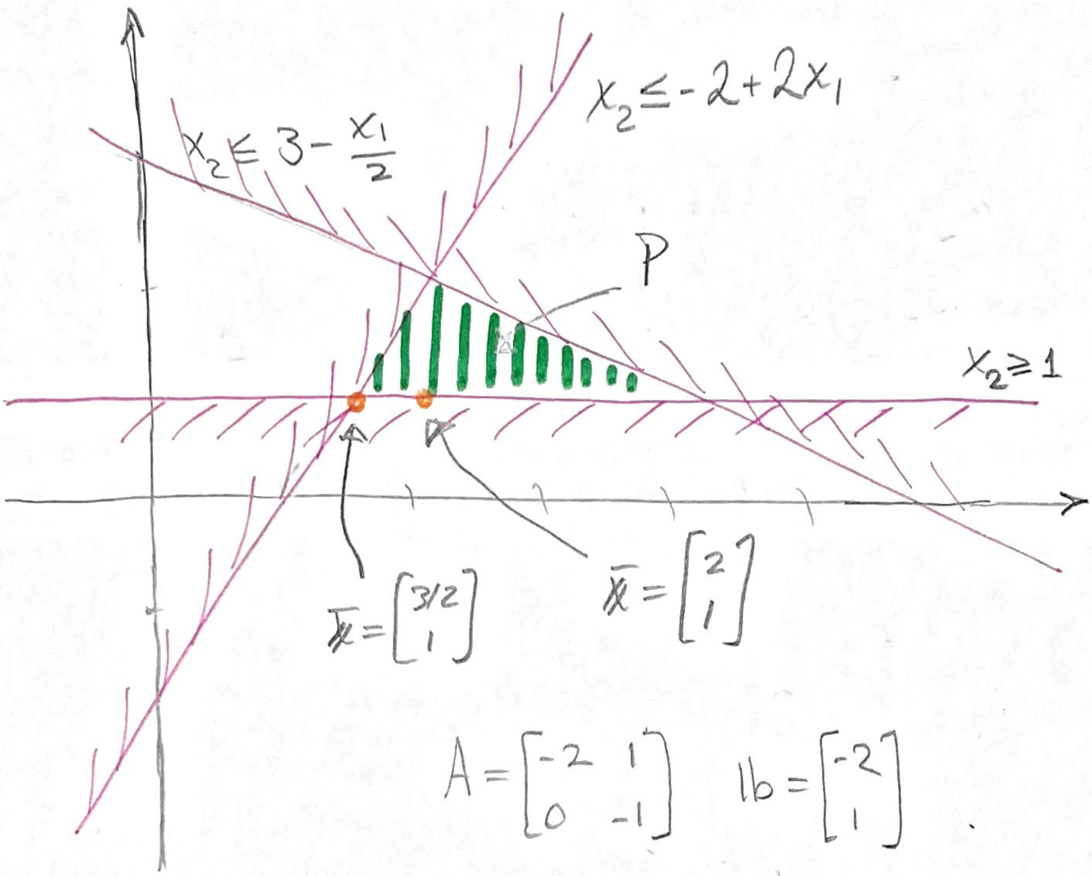
- $Ax \leq b \iff a_i x \leq b_i, \quad i = 1, \dots, m$
- $\{x \in \mathbb{R}^n \mid a_i x \leq b\}$ are half-spaces
- P is intersection of half-spaces

Ex. Let

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \\ 0 & -1 \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ -2 \\ -1 \end{bmatrix}$$

This means

$$x_1 + 2x_2 \leq 6$$
$$-2x_1 + x_2 \leq -2$$
$$-x_2 \leq -1$$



$$A = \begin{bmatrix} -2 & 1 \\ 0 & -1 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

- Polytope = convex hull of finitely many points
- Polyhedron = intersection of finitely many half-spaces

THM Let $\bar{x} \in P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ where $A \in \mathbb{R}^{m \times n}$, rank $A = n$ and $b \in \mathbb{R}^m$. Further, let

$$\bar{A}\bar{x} = \bar{b}$$

be the equality subsystem of $Ax \leq b$ then

$$\text{rank}\, \bar{A} = n \Leftrightarrow \bar{x} \text{ extreme point in } P.$$

Ex. Let

$$\bar{x} = \begin{bmatrix} 3/2 \\ 1 \end{bmatrix}$$

$$Ax \le b \iff \begin{aligned} x_1 + 2x_2 &\le 6 \\ -2x_1 + x_2 &\le -2 \\ -x_2 &\le -1 \end{aligned}$$

Plug in $\bar{x}$

$$\left.\begin{aligned} &\cancel{3/2 + 2\cdot 1 = 3{,}5 < 6} \\ &-2\cdot\tfrac{3}{2} + 1 = -2 = -2 \\ &-1 = -1 \end{aligned}\right\} \Rightarrow$$

$$\bar{A} = \begin{bmatrix} -2 & 1 \\ 0 & -1 \end{bmatrix}$$

$$\text{rank } \bar{A} = n \Rightarrow$$

$\bar{x}$ extreme point.

Ex. Let

$$\bar{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Plug in $\bar{x}$

$$\left.\begin{aligned} &\cancel{2 + 2\cdot 1 = 4 < 6} \\ &\cancel{-2\cdot 2 + 1 = -3 < -2} \\ &-1 = -1 \end{aligned}\right\} \Rightarrow$$

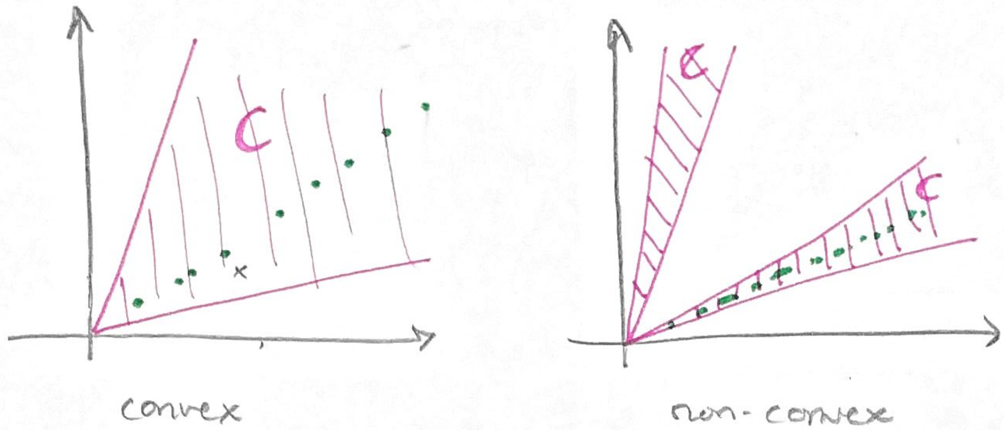
$$\bar{A} = [0 \ -1], \ b = [-1]$$

$$\text{rank } \bar{A} = 1 < n \implies$$

$\bar{x}$ not extreme point.

9⁰⁰

DEF A set $C \subseteq \mathbb{R}^n$ is a <u>cone</u> if $\lambda x \in C$ whenever $x \in C$ and $\lambda > 0$.



THM (REPRESENTATION THM)

Let the polyhedron

$$Q = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

and let $\{v^1, \dots, v^k\}$ extreme points.

Define:

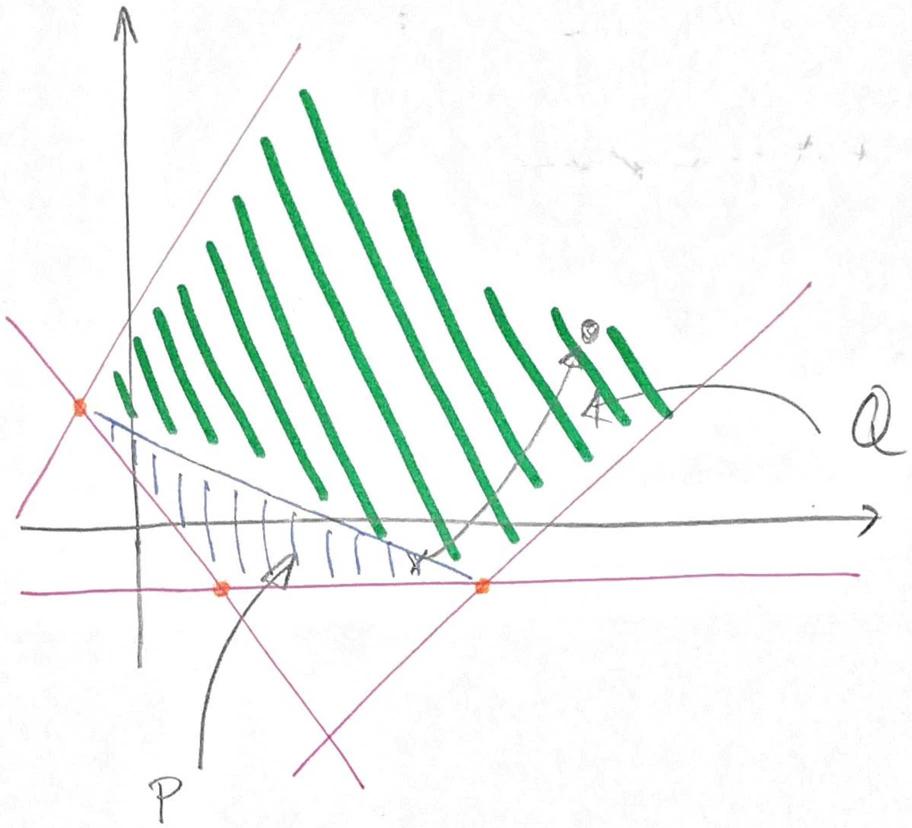
$$P := \operatorname{conv}\left(\{v^1, \dots, v^k\}\right)$$

and let

$$C = \{x \in \mathbb{R}^n \mid Ax \leq 0\}.$$

Then

$$Q = P + C = \{x \in \mathbb{R}^n \mid x = u + w,\ u \in P,\ w \in C\}$$



CONVEX FUNCTIONS

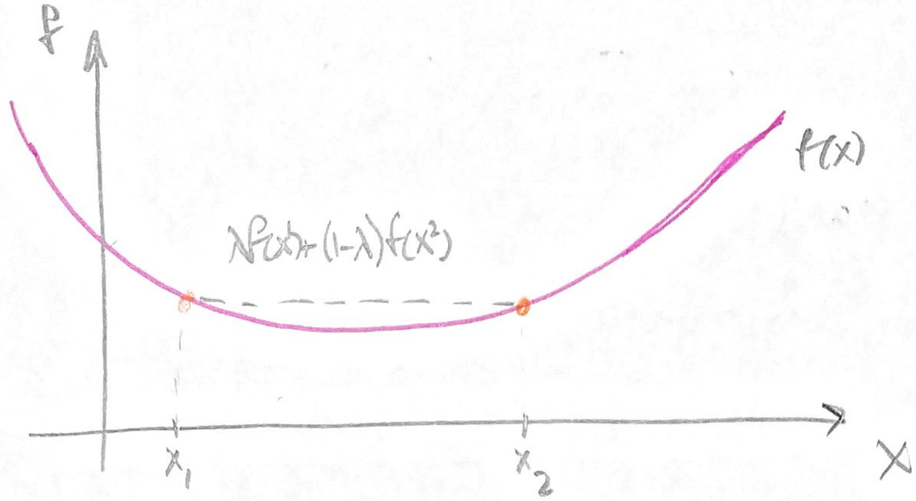
DEF suppose $S \subseteq \mathbb{R}^n$ is convex. A function

$$f: \mathbb{R}^n \longrightarrow \mathbb{R},$$

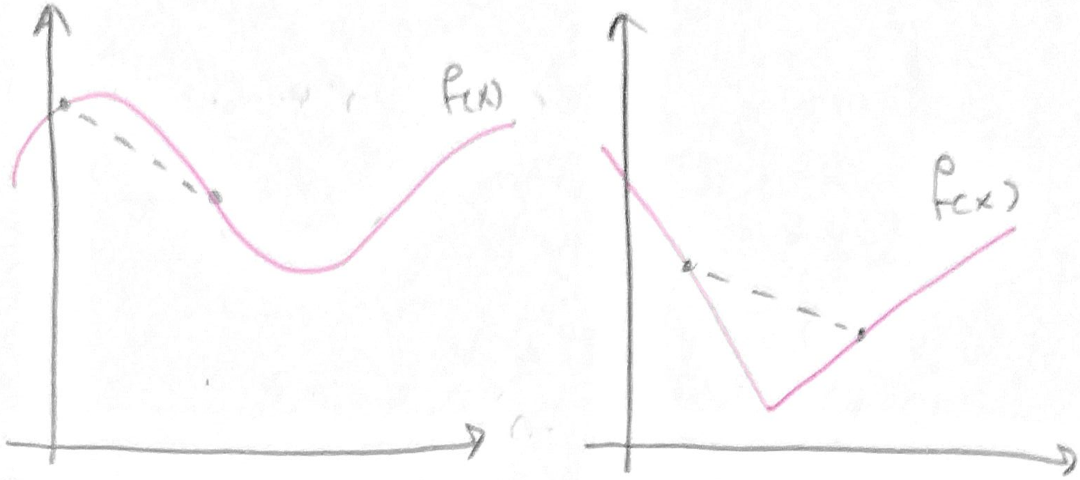
is convex if

$$\left.\begin{matrix} x^1, x^2 \\ \lambda \in (0,1) \end{matrix}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2)$$

Ex.



Ex.



Non-convex Convex

- $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \mathbf{b}$ is convex and concave
- $f(\mathbf{x}) = \|\mathbf{x}\|$ is convex
- $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is strictly convex

A function is strictly convex if the inequality is strict.

A function is concave if $-f$ is convex.

THM Let $S \subseteq \mathbb{R}^n$ be convex and let $f_k, k \in K$ be a set of convex functions. Let $\alpha_k \geq 0$, $k \in K$. Then

$$f(\mathbf{x}) = \sum_{k \in K} \alpha_k f_k(\mathbf{x})$$

is convex.

PROOF Let $\mathbf{x}, \mathbf{y} \in S$ and $\lambda \in (0,1)$. Then

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \sum \alpha_k f_k(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq$$

$$\leq \sum_i \alpha_k \left(\lambda f_k(x) + (1-\lambda) f_k(y)\right) \overset{f_k \text{ convex}}{=}$$

$$= \lambda \sum_i \alpha_k f_k(x) + (1-\lambda) \sum_i \alpha_k f_k(y) =$$

$$= \lambda f(x) + (1-\lambda) f(y)$$

$\square$

THM Let $g: \mathbb{R}^n \to \mathbb{R}$ be convex

$f: \mathbb{R} \to \mathbb{R}$ be convex and non-decreasing.

$\Rightarrow$

$f(g)$ is convex.

EX.

$f(x) = x^2 + e^x$

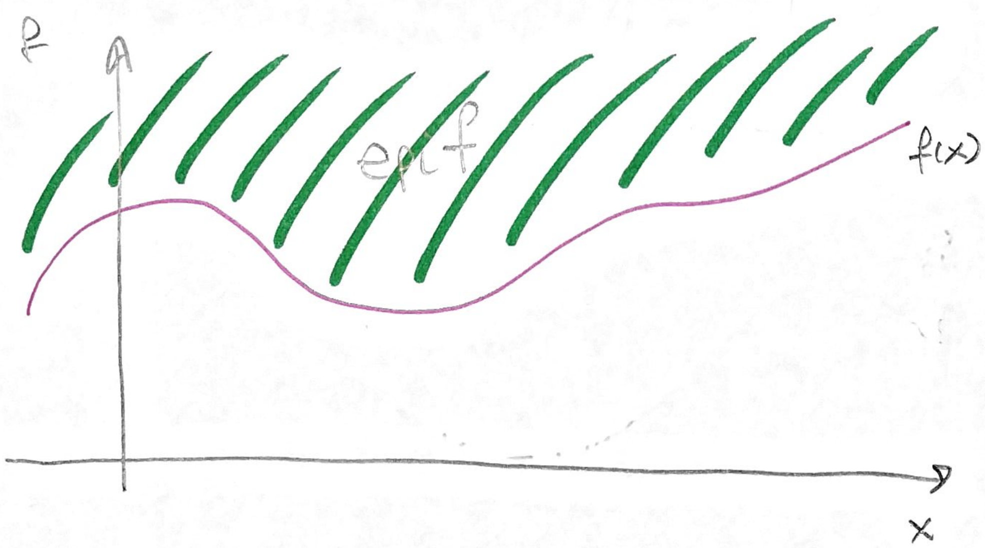
$f(x) = e^{x^2}$

DEF The <u>epigraph</u> of a function

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

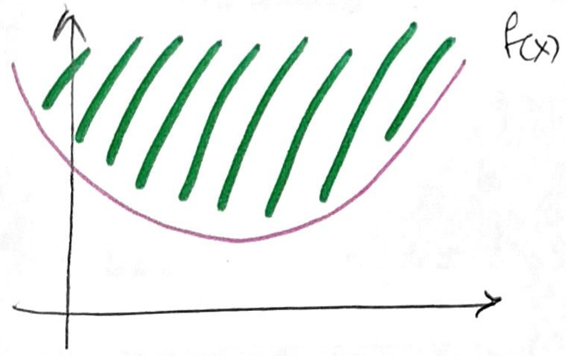
is

$$\text{epi} f = \{ (x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha \}$$



THM Let $f: \mathbb{R}^n \to \mathbb{R}$. Then

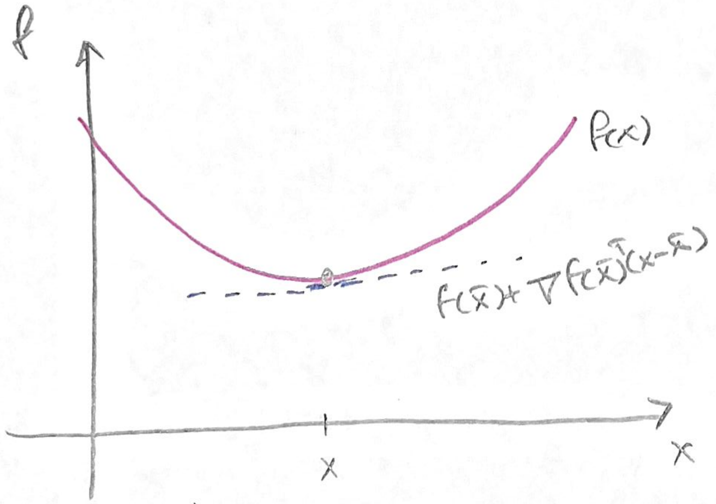
$$f \text{ is convex} \iff \text{epi} f \text{ is a convex set}$$



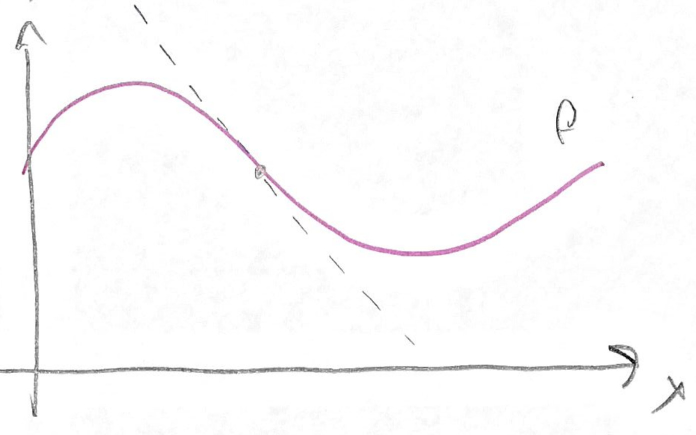
PROOF see book.

THM Let $f \in C^1$ on an open convex set S. Then

$$f \text{ is convex on } S \iff$$
$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \text{ for all } x, \bar{x} \in S.$$



Convex



non-convex

PROOF

$\Rightarrow$: Let $x^1, x^2 \in S$ and let $\lambda \in (0,1)$.

$$\lambda f(x^1) + (1-\lambda) f(x^2) \geq f(\lambda x^1 + (1-\lambda) x^2) \quad \Rightarrow$$

$$f(x^1) - f(x^2) \geq \frac{1}{\lambda}\left[f(\lambda x^1 + (1-\lambda)x^2) - f(x^2)\right] =$$

$$= \frac{f(x^2 + \lambda(x^1 - x^2)) - f(x^2)}{\lambda}$$

looks like derivative

Let $\lambda \to 0 \Rightarrow$

$$\nabla f(x^2)^T (x^1 - x^2)$$

So

$$f(x^1) = f(x^2) + \nabla f(x^2)^T (x^1 - x^2)$$

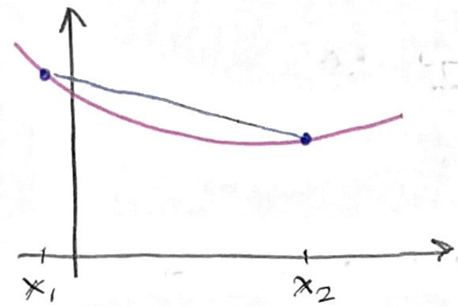
Thuesday
10 september
15-18

LECTURE 3

Recap:

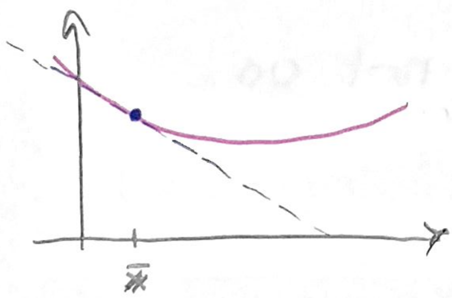
DEF A function f is <u>convex</u> on S (which is convex) if

$$\left.\begin{array}{l} x^1, x^2 \in S \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2)$$



THM Let $f \in C^1$ on an open set S.

$$f \text{ is convex} \iff f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \quad \forall x, \bar{x} \in S$$



THM Let $f \in C^2$

a) f is convex $\iff \nabla^2 f(x) \succeq 0 \ \forall x$

$(A \succeq 0 \iff p^T A p \geq 0 \ \forall p)$

positive semidefinite

b) $\nabla^2 f(x) \succ 0 \Rightarrow f$ is strictly convex $\forall x$

(ex. $f(x)=x^4$)

PROBLEM (P)

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{S.T. } g_i(x) \le 0, \quad i=1,\dots,m \\ &\qquad h_j(x)=0, \quad j=1,\dots,k \\ &\qquad x \in X \end{aligned}$$

P is a <u>convex optimization problem</u> if

- f is a convex function
- g_i, $i=1,\dots,m$ are convex functions
- h_j, $j=1,\dots,k$ are affine functions
- X is a convex set

LECTURE 3 - starts here

Today:

$$\boxed{\begin{aligned} (P) \quad &\text{minimize } f(x) \\ &\text{S.T.} \quad x \in S \end{aligned}}$$

DEF x^* is a <u>global minimum</u> to P if

$f(x^*) \le f(x)$

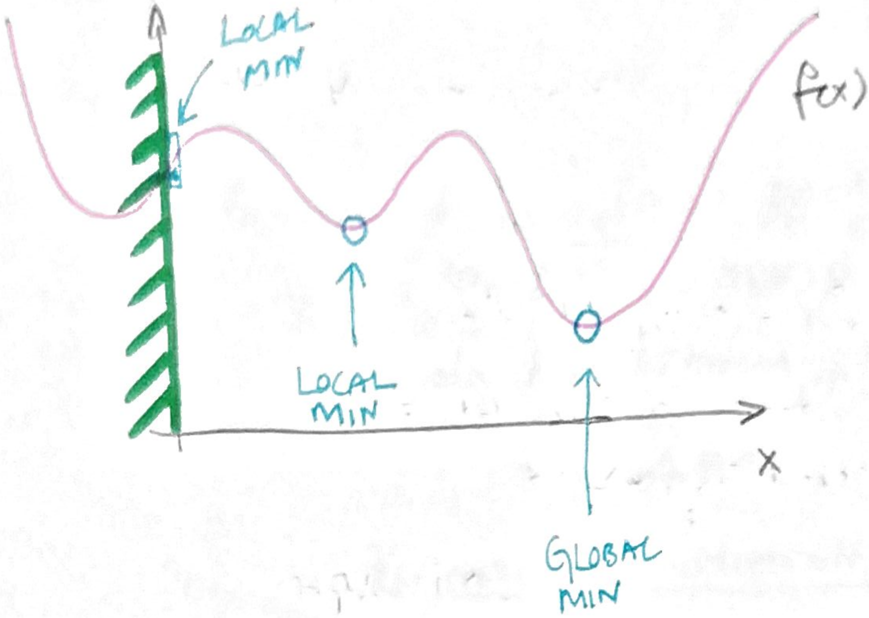
for all $x \in S$.

DEF x^* is a <u>local minimum</u> to P if $\exists\, \varepsilon > 0$ such that

$f(x^*) \le f(x)$

for all $x \in S \cap B_\varepsilon(x^*)$ where

$B_\varepsilon(x^*) = \{ x \in \mathbb{R}^n \mid \|x - x^*\| < \varepsilon \}$



important

THM (Fundamental thm of optimization)

If f is convex and S is convex, then

$$x^* \text{ Local minimum} \Rightarrow x^* \text{ Global minimum}$$

PROOF Assume x^* is a local minimum but not a global minimum.

Let $\bar{x} \in S$ be a point that satisfies

$$f(\bar{x}) < f(x^*).$$

Take $\lambda \in (0,1)$

$$\Rightarrow \quad \lambda \bar{x} + (1-\lambda) x^* \in S$$

Since f is convex:

$$f(\lambda \bar{x} + (1-\lambda) x^*) \leq \lambda f(\bar{x}) + (1-\lambda) f(x^*) <$$

$$< \lambda f(x^*) + (1-\lambda) f(x^*) = f(x^*)$$

Choose λ really small!

But this contradicts that x^* is a local min. $\square$

So whe have proven that a point arbitrarily close to x^* is "better" than x^*, but that can not be the case.

DEFINITIONS

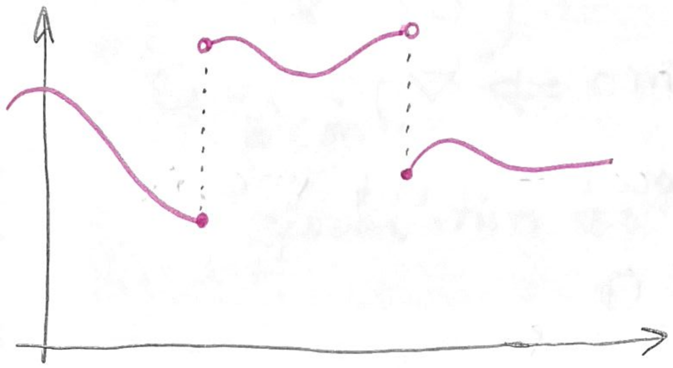
- A set $S \subseteq \mathbb{R}^n$ is <u>open</u> if for every $x \in S$ there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset S$.
- A set S is <u>closed</u> if $\mathbb{R}^n \setminus S$ is open.
- A set $S \subseteq \mathbb{R}^n$ is <u>bounded</u> if there exists a constant $C > 0$ such that $\|x\| < C$ for all $x \in S$
- If a set is closed and bounded we say it is <u>compact</u>.

DEF A function f is <u>weakly coercive</u> with respect to a set S if either S is bounded or

$$\lim_{\substack{\|x\| \to \infty \\ x \in S}} f(x) = \infty$$

DEF A function f is <u>lower semi-continuous</u> at x if

$$x_k \to x \Rightarrow f(x) \leq \liminf_{k \to \infty} f(x_k)$$



THM (Weierstrass thm)

Consider problem (P). If

- S is nonempty and closed
- f is lower semi-continuous on S
- f weakly coercive w.r.t. S

$\Rightarrow$ there exists a nonempty and compact set of optimal solutions to (P).

Ex1

$$\min \ 1/x \quad \text{s.t.} \quad x \geq 1$$

not weakly coercive

Ex2

$$\min \ x^2 \quad \text{s.t.} \quad x > 0$$

the set is open

16.15

OPTIMALITY CONDITIONS WHEN $S = \mathbb{R}^n$

Necessary optimization conditions

$$x^* \text{ local min} \Rightarrow (*)$$

Sufficient optimization conditions

$$(*) \Rightarrow x^* \text{ is local min}$$

THM If $f \in C^1$. Then

$$x^* \text{ local min} \Rightarrow \nabla f(x^*) = 0$$

PROOF Suppose x^* is a local min but

$$\nabla f(x^*) \neq 0$$

Let

$$p = -\nabla f(x^*)$$

Taylor expansion:

$$f(x^*+\alpha p) = f(x^*) + \alpha \nabla f(x^*)^T p + O(\alpha) =$$

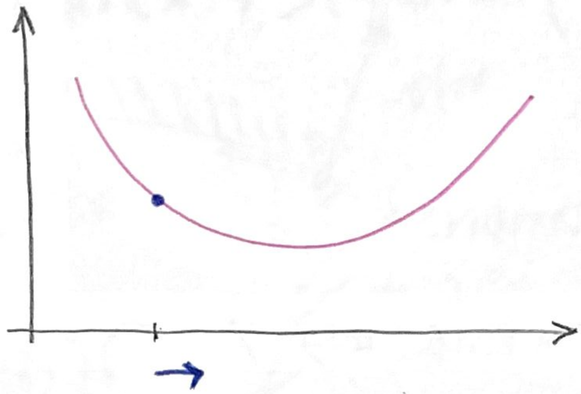
$$= f(x^*) - \alpha \|\nabla f(x)\|^2 + O(\alpha)$$

Choose α small:

$$< f(x^*)$$

For small $\alpha > 0$.

$\Rightarrow$ x^* can not be lokal min! $\square$



THM If $f \in C^2$, then

$$x^* \text{ local min} \Rightarrow \begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succeq 0 \end{cases}$$

THM If $f \in C^2$, then

$$\left.\begin{array}{l} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succ 0 \end{array}\right\} \Rightarrow x^* \text{ strict local min}$$

THM If $f \in C^1$ is convex, then

$$x^* \text{ global min} \Leftrightarrow \nabla f(x^*) = 0$$

OPTIMALITY CONDITIONS WHEN $S \subseteq \mathbb{R}^n$

DEF A

Let $x \in S$. A vector $p \in \mathbb{R}^n$ is a <u>feasible direction</u> at x if

$$\exists \delta > 0 : x + \alpha p \in S \quad \forall \alpha \in [0, \delta]$$

DEF B

A vector $p \in \mathbb{R}^n$ is a <u>descent direction</u> with respect to f if

$$\exists \delta > 0 : f(x + \alpha p) < f(x) \quad \forall \alpha \in (0, \delta]$$

"NECESSARY OPT. CONDITIONS"

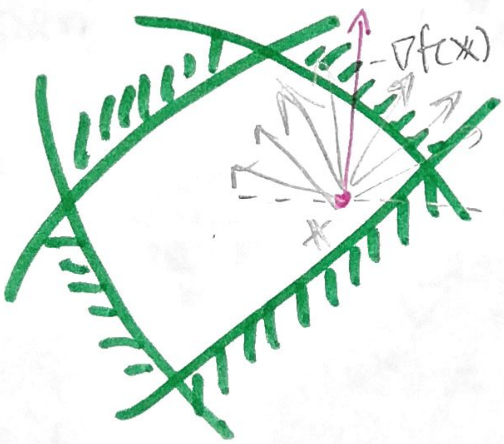
$$x^* \text{ is a local minima} \Rightarrow A(x^*) \cap B(x^*) = \emptyset$$

NOTE suppose $f \in C^1$.

If a vector p satisfies

$$\nabla f(x)^T p < 0$$

then p is a descent direction at x.



$$-\nabla f(x)^T p > 0$$

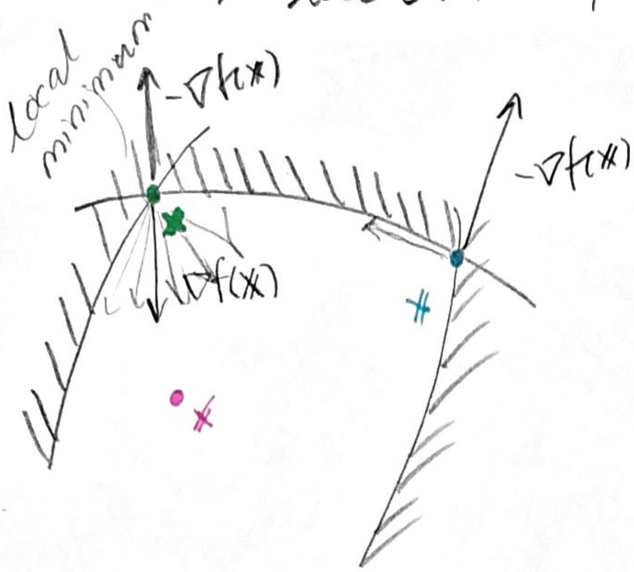
NOTE If $\nabla f(x) \neq 0$, then $-\nabla f(x)$ is a descent direction.

THM Suppose $S \subseteq \mathbb{R}^n$ and that $f \in C^1$.

a) x^* local min $\Rightarrow$ $\nabla f(x^*)^T p \geq 0$ for all feasible directions p.

b) Suppose S is a convex set.

x^* local min $\Rightarrow$ $\nabla f(x^*)^T (x - x^*) \geq 0$ for all $x \in S$



Monday
16 september
2019

LECTURE 3

RECAP Optimality conditions

NECESSARY CONDITIONS

x^* Local min $\Rightarrow$ $(\star)$

SUFFICIENT CONDITIONS

$(\star) \Rightarrow x^*$ Local min

$$(P) \quad \min f(x) \quad \text{s.t. } x \in S$$

WHEN $S = \mathbb{R}^n$

THM if $f \in C^1$, then

x^* Local min $\Rightarrow \nabla f(x^*) = 0$

THM if $f \in C^1$ is convex, then

x^* global min $\Leftrightarrow \nabla f(x^*) = 0$

WHEN $S \subseteq \mathbb{R}^n$

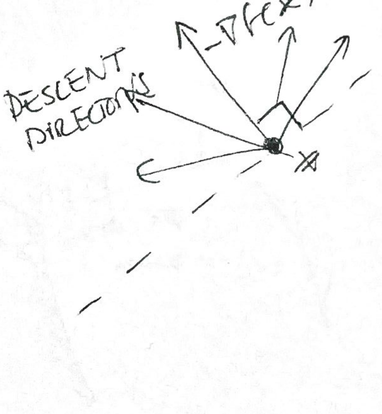
FEASIBLE DIRECTIONS AT x:

$p \in \mathbb{R}^n : x + \alpha p \in S \ \forall \alpha \in [0, \delta]$

DESCENT DIRECTIONS AT x:

$p \in \mathbb{R}^n : f(x + \alpha p) < f(x) \ \forall \alpha \in (0, \delta]$

NOTE If $\nabla f(x)^T p < 0$, then p is a descent direction at x



INTUITIVE THM

x^* Local min $\Rightarrow$ "Should not exist vector p which is both a feasible direction and a descent direction at x^*"

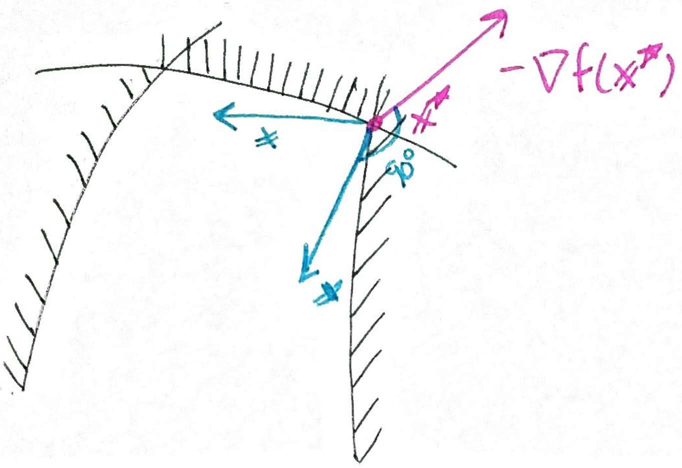
continuing:

THM Suppose $S \subseteq \mathbb{R}^n$ and that $f \in C^1$.

a) x^* local min $\Rightarrow \nabla f(x^*)^T p \geq 0$
for all feasible directions p at x^*

b) Suppose S is convex.

x^* local min $\Rightarrow \nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in S$



THM If $f \in C^1$ and S convex

x^* local min $\Rightarrow$ x^* stationary

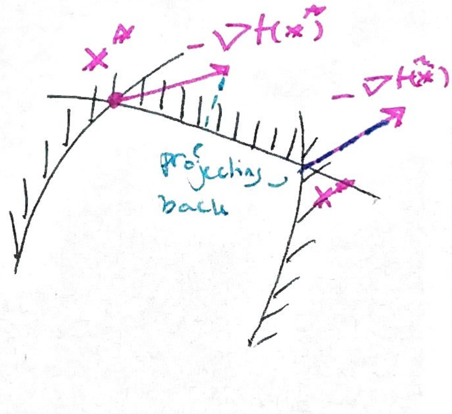
DEF A point $x^* \in S$ (S convex) is <u>stationary</u> if one of the equivalent statements hold.

a) $\nabla f(x^*)^T (x - x^*) \geq \forall x \in S$

b) $\min\limits_{x \in S} \nabla f(x^*)^T (x - x^*) = 0$

c) $x^* = \mathrm{Proj}_S (x^* - \nabla f(x^*))$

d) $-\nabla f(x^*) \in N_S(x^*)$

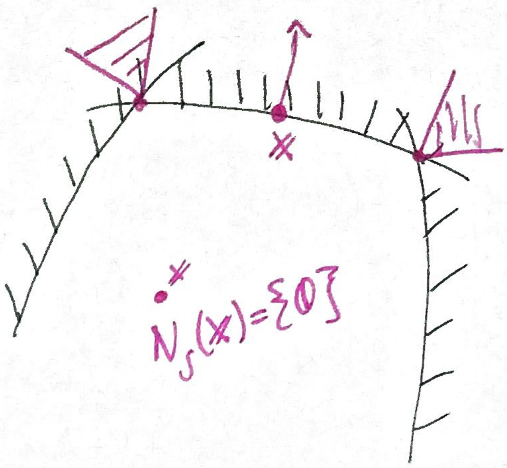


The <u>normal cone</u> to S at x is

$$N_S(x) = \{ p \in \mathbb{R}^n \mid p^T(y-x) \leq 0, \ y \in S \}$$

THM If $f \in C^1$, S is convex and f is convex

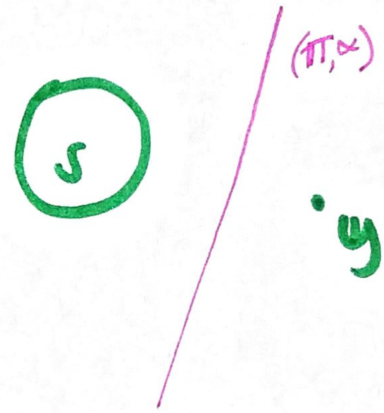
x^* global min $\Longleftrightarrow$ x^* stationary



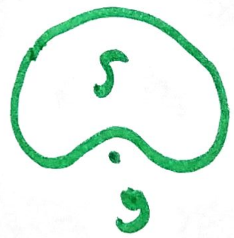
THM <u>SEPARATION THM</u>

Suppose $S \subseteq \mathbb{R}^n$ is closed and convex, and that $y \notin S$.
Then there exists a vector $\pi \neq 0$ and a scalar $\alpha \in \mathbb{R}$ such that

$$\pi^T y > \alpha$$
$$\pi^T x \leq \alpha \quad \forall x \in S$$



convex set S



non-convex set S

9

LECTURE 4

$$\boxed{\mathcal{P} \quad \begin{array}{l} \min f(x) \\ \text{s.t. } x \in \mathbb{R}^n \end{array}}$$

DEF **LINE SEARCH TYPE ALG**

step 0: Starting point $x_0 \in \mathbb{R}^n$.
Let $k := 0$

step 1: Find search direction $p_k \in \mathbb{R}^n$.

step 2: Perform line search, i.e. find $\alpha_k > 0$ such that

$$f(x_k + \alpha_k p_k) < f(x_k)$$

step 3: Let $x_{k+1} = x_k + \alpha_k p_k$

step 4: Check termination criteria. If not fullfilled, let $k := k+1 \longrightarrow (1)$.

Step 1: Let $f \in C^1$. Then we know that $p_k = -\nabla f(x_k)$ is a descent direction this direction is called the steepest descent direction because it solves

$$\min_{\substack{p \in \mathbb{R}^n \\ \|p\|=1}} \nabla f(x_k)^T p$$

- for Q symmetric and positive definite

$$p_k = -Q \nabla f(x_k)$$

is also a descent direction because

$$\nabla f(x_k)^T p_k = -\nabla f(x_k)^T Q \nabla f(x_k) < 0$$

steepest descent: $Q = I$

Newton: $Q = [\nabla^2 f(x_k)]$

NEWTONS METHOD

First assume that $\nabla^2 f(x_k) > 0$

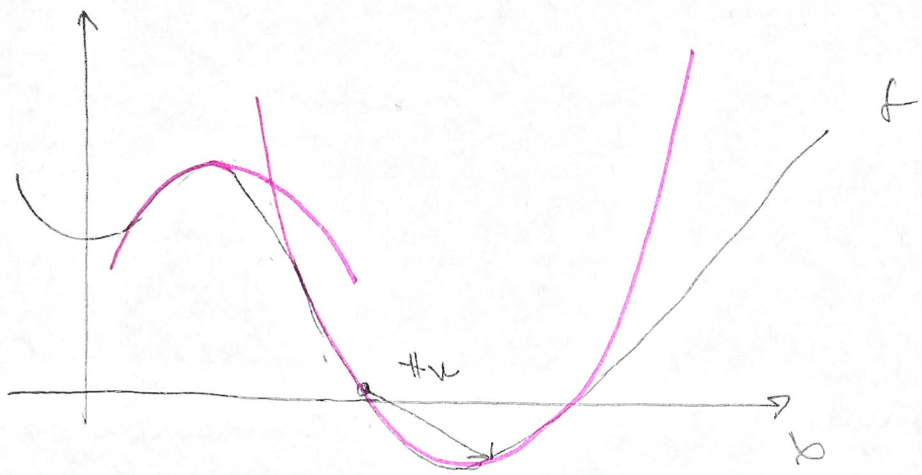
$$f(x_k + p) \approx f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla f(x_k) p = \varphi(p)$$

"Use $\nabla = 0$ trick"

$$\nabla_p \varphi(p) = \nabla f(x_k) + \nabla^2 f(x_k) p = 0$$

$$\nabla^2 f(x_k) p = -\nabla f(x_k)$$

$$p = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$



Levensberg–Marquart modification

$$p_k = -\left[\nabla^2 f(x_k) + \gamma I\right]^{-1} \nabla f(x_k)$$

$\gamma = 0$ Newton

$\gamma = \infty$ steepest descent

Steepest descent $\quad \mathbb{P}_k = -\nabla f(x_k)$

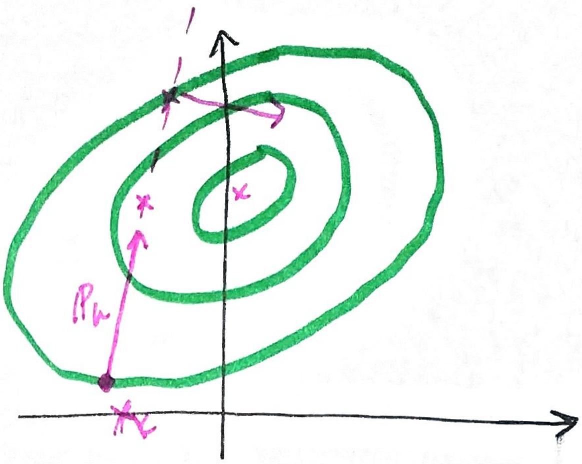
Newton $\quad \mathbb{P}_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$

Levensberg-Marquardt $\quad \mathbb{P}_k = -[\nabla^2 f(x_k) + \gamma I]^{-1} \nabla f(x_k)$

Quasi-Newton $\quad \mathbb{P}_k = -B_k \nabla f(x_k)$

Step 2: $\quad \varphi(\alpha) := f(x_k + \alpha \mathbb{P}_k)$

$$\min_{\alpha \geq 0} \varphi(\alpha)$$



How to choose α?

- Interpolation. Use $f(x_k)$, $\nabla f(x_k)$, $\nabla^2 f(x_k)$ to approximate $\varphi(\alpha)$ and then solve analytically.
- Newton's method

$$\alpha = \alpha - \varphi'(\alpha)/\varphi''(\alpha)$$

- Golden section: Derivative free method which shrinks an interval until you know $\varphi'(\alpha)=0$ is within that interval.
- $\alpha_k = \alpha \qquad \alpha = 0.1$
- Armijo's step length

Step 4:

a) $\|\nabla f(x_k)\| \leq \varepsilon_1 (1 + |f(x_k)|)$

b) $|f(x_{k+1}) - f(x_k)| \leq \varepsilon_2 (1 + |f(x_k)|)$

c) $\|x_{k+1} - x_k\| \leq \varepsilon_3 (1 + \|x_k\|)$

ASSUMPTIONS ON p_k

a) $$-\frac{\nabla f(x_k) p_k}{\|\nabla f(x_k)\| \|p_k\|} \geq s_1$$

b) $\|p_k\| \geq s_2 \|\nabla f(x_k)\|$

c) $\|p_k\| \leq M$

for some $s_1, s_2 > 0$.

THM Suppose $f \in C^1$ and for the starting point x_0 it holds that

$$\{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$$

is bounded.

Let α_k be choosen by armijo's rule. Then

a) $\{x_k\}$ is bounded

b) $\{f(x_k)\}$ descending

c) Every limit point of $\{x_k\}$ is stationary.

Thursday
17th september
15^{15}

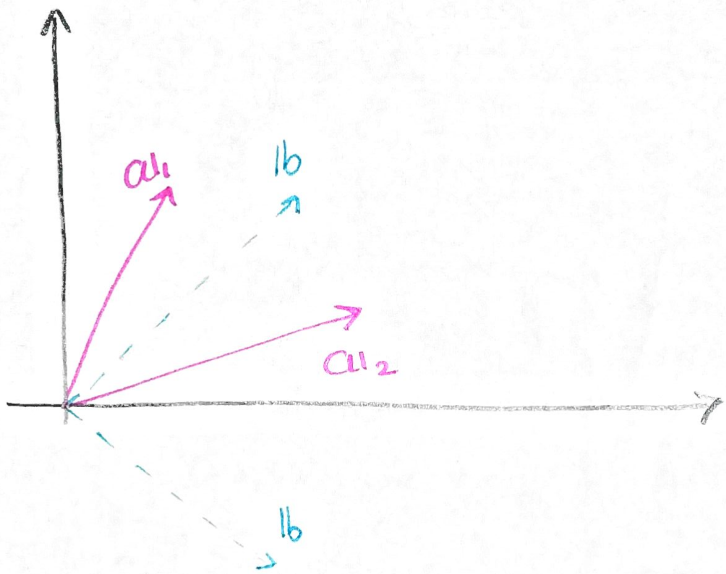
LEMMA Farkas' lemma

For any $A \in \mathbb{R}^{m \times n}$ and $\mathbb{b} \in \mathbb{R}^m$, exactly one of the systems

(I) $A\mathbb{x} = \mathbb{b}$
$\mathbb{x} \geq \mathbb{0}$

(II) $A^T\mathbb{y} \leq \mathbb{0}$
$\mathbb{b}^T\mathbb{y} > 0$

is feasible, and the other one is not.



LECTURE 5

OPTIMALITY CONDITIONS

$$\boxed{(P) \quad \begin{array}{ll} \min & f(\mathbb{x}) \\ \text{s.t.} & \mathbb{x} \in S \end{array}}$$

LAST WEEK

Assume S is convex

THM

x^* local min $\Longrightarrow$ x^* stationary

DEF

x^* is stationary if

a) $\nabla f(x^*)^T(x - x^*) \geq 0 \quad \forall x \in S$

b) $\min_{x \in S} \nabla f(x^*)^T (x - x^*) = 0$

c) $x^* = \text{Proj}_S(x^* - \nabla f(x^*))$

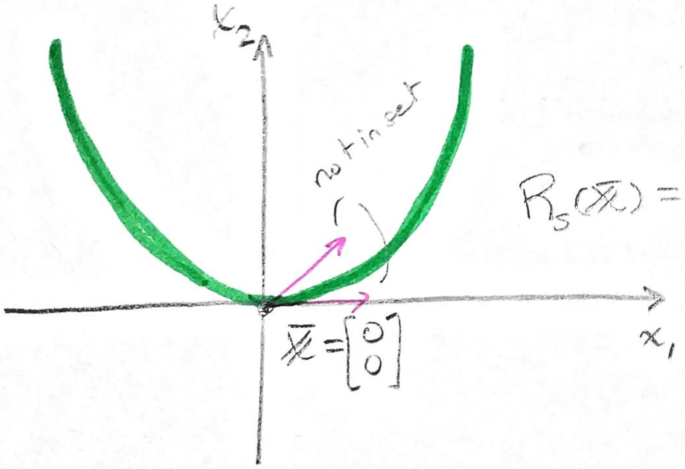
d) $\boxed{-\nabla f(x^*) \in N_S(x^*)}$

DEF The cone of feasible directions at x is

$$R_S(x) = \{ p \in \mathbb{R}^n \mid \exists \delta > 0 : x + \alpha p \in S \ \forall \alpha \in [0, \delta] \}$$

Ex.

$$S = \{ x \in \mathbb{R}^2 \mid x_2 = x_1^2 \}$$

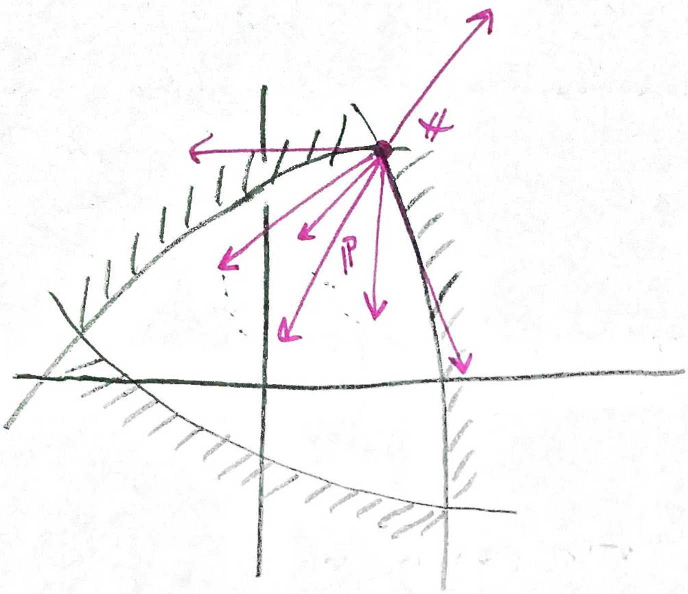


$R_S(\bar{x}) = \emptyset$

$T_S(\bar{x})$

DEF The <u>tangent cone</u> at x is defined as

$$T_S(x) = \left\{ p \in \mathbb{R}^n \;\middle|\; \exists \{x_k\} \subset S, \{\lambda_k\} \subset (0,\infty), \text{ such that } \lim_{k\to\infty} x_k = x, \; \lim_{k\to\infty} \lambda_k (x_k - x) = p \right\}$$



$$\boxed{\operatorname{cl} R_S(x) \subseteq T_S(x)}$$

DEF The <u>cone of descent directions</u> at x is defined as

$$\mathring{F}(x) = \left\{ p \in \mathbb{R}^n \;\middle|\; \nabla f(x)^T p < 0 \right\}$$

THM Let $f \in C^1$, then

$$x^* \text{ local min} \Rightarrow \mathring{F}(x^*) \cap T_S(x^*) = \emptyset$$

LEMMA

If the cone $C(x) \subseteq T_S(x)$ for all $x \in S$, then

$$x^* \text{ local min} \Rightarrow \mathring{F}(x^*) \cap C(x^*) = \emptyset$$

$$\boxed{\begin{array}{ll} (P) & \min f(x) \\ & \text{s.t. } g_i(x) \leq 0,\ i=1,\dots,m \end{array}}$$

The set of <u>active constraints</u> at x is

$$I(x) = \{ i \in \{1,\dots,m\} : g_i(x) = 0 \}$$

DEF The <u>inner gradient</u> cone is

$$\mathring{G}(x) = \{ p \in \mathbb{R}^n \mid \nabla g_i(x)^T p < 0,\ i \in I(x) \}$$

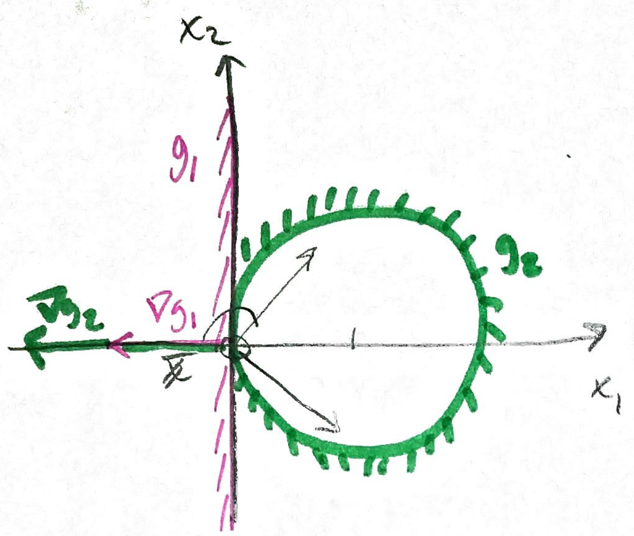
The <u>gradient cone</u> is

$$G(x) = \{ p \in \mathbb{R}^n \mid \nabla g_i(x)^T p \leq 0,\ i \in I(x) \}$$

$$\boxed{\mathrm{cl}\, \mathring{G}(x) \subseteq \mathrm{cl}\, R_S(x) \subseteq T_S(x) \subseteq G(x)}$$

16^{15}

Ex. $S = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} -x_1 \le 0 \\ (x_1-1)^2 + x_2^2 - 1 \le 0 \end{array} \right\}$ $\begin{array}{l} g_1 \\ g_2 \end{array}$



$\bar{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$I(\bar{x}) = \{1, 2\}$ $\nabla g_2(x) = \begin{bmatrix} 2(x_1-1) \\ 2x_2 \end{bmatrix}$

$\nabla g(\bar{x}) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

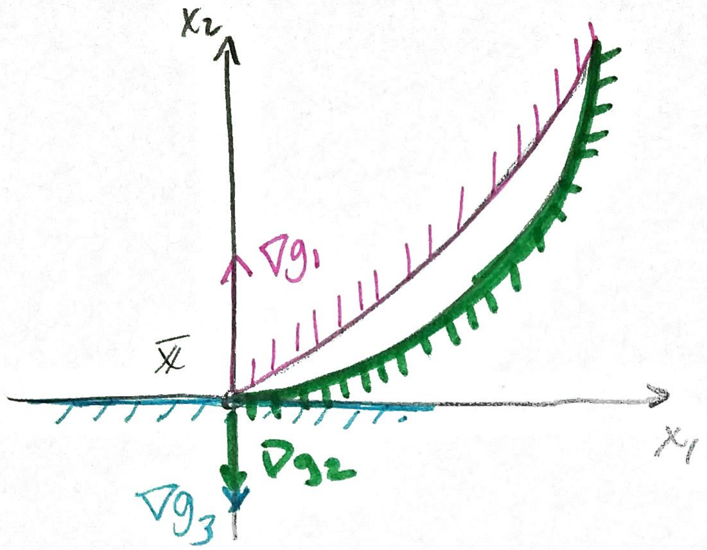
$R_S(\bar{x}) = \{ p \in \mathbb{R}^2 : p_1 > 0 \}$

$T_S(\bar{x}) = \{ p \in \mathbb{R}^2 : p_1 \ge 0 \}$

$\mathring{G}(\bar{x}) = \{ p \in \mathbb{R}^2 \mid p_1 > 0 \}$

$G(\bar{x}) = \{ p \in \mathbb{R}^2 \mid p_1 \ge 0 \}$

Ex. $S = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} -x_1^3 + x_2 \leq 0 \\ x_1^5 - x_2 \leq 0 \\ -x_2 \leq 0 \end{array} \right\} \begin{array}{l} g_1 \\ g_2 \\ g_3 \end{array}$



$I(\bar{x}) = \{1, 2, 3\}$

$\nabla g_1(x) = \begin{bmatrix} -3x_1^2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\nabla g_2(x) = \begin{bmatrix} 5x_1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$\nabla g_3(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$

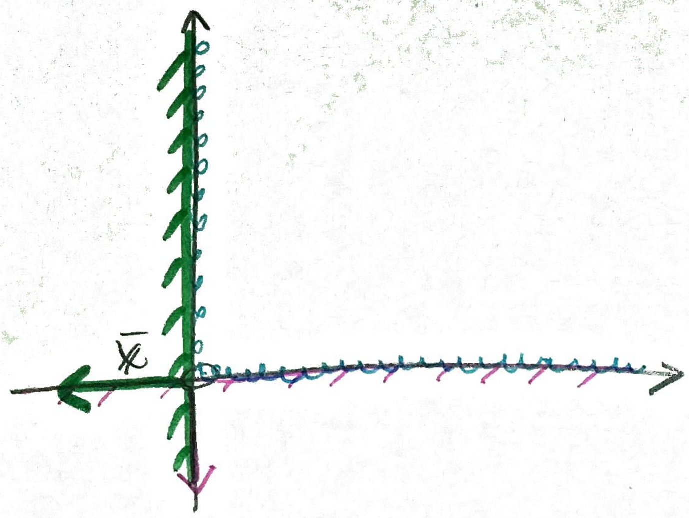
$R_S(\bar{x}) = \emptyset$

$T_S(\bar{x}) = \{ p \in \mathbb{R}^2 \mid p_1 > 0,\ p_2 = 0 \}$

$\mathring{G}(\bar{x}) = \emptyset$

$G(\bar{x}) = \{ p \in \mathbb{R}^2 \mid p_2 = 0 \}$

Ex. $S = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} -x_1 \le 0 \\ -x_2 \le 0 \\ x_1 x_2 \le 0 \end{array} \right\}$ $\begin{array}{l} g_1 \\ g_2 \\ g_3 \end{array}$



can only be on the axis

$\bar{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$I(x) = \{1, 2, 3\}$

$\nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$\nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$\nabla g_3(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$R_S(\bar{x}) = S$

$T_S(\bar{x}) = S$

$\mathring{G}(\bar{x}) = \emptyset$

$G(\bar{x}) = \{ p \in \mathbb{R}^2 \mid p_1 \ge 0, p_2 \ge 0 \}$

CONDITIONS FRITZ-JOHN CONDITIONS

x^* local min $\Rightarrow$ $\exists \mu_0 \in \mathbb{R}$ and $\mu \in \mathbb{R}^m$ such that

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0$$

$$\mu_i g_i(x^*) = 0, \quad i=1,\dots,m$$

$$\mu_0, \mu_i \geq 0, \quad i=1,\dots,m$$

not all zero

$\Longleftrightarrow$

$\mathring{F}(x^*) \cap \mathring{G}(x^*) = \emptyset \iff \nabla f(x^*)^T p < 0$

$\nabla g_i(x^*)^T p < 0 \quad \forall i \in I(x^*)$

CONDITIONS KKT-CONDITIONS

x^* local min $\Rightarrow$ $\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0$

$$\mu_i g_i(x^*) = 0 \quad \forall i$$

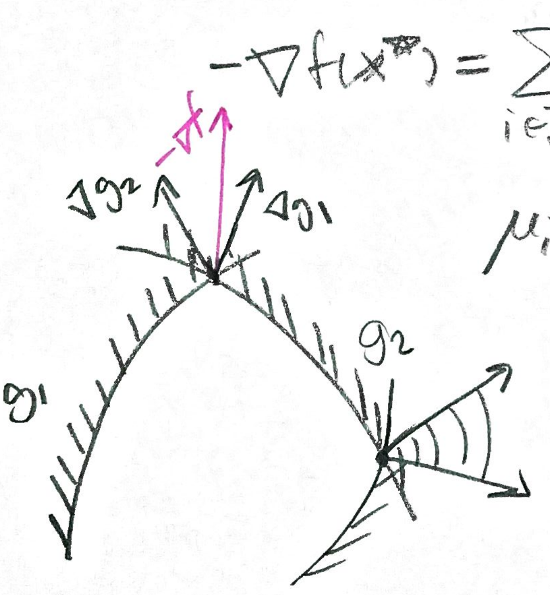
$$\mu_i \geq 0$$

is solvable for μ.

$\Longleftrightarrow$

$$-\nabla f(x^*) = \sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*)$$

$$\mu_i \geq 0 \quad i \in I(x^*)$$



Monday
23 september
8:00

LECTURE 6

The Karech-Kuhn-Tucker (KKT) conditions

Let $f: \mathbb{R}^n \to \mathbb{R}$, $g_i: \mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,m$, all in C^1. We consider

$$\min f(x)$$
$$\text{s.t. } g_i(x) \le 0 \quad i=1,\dots,m$$

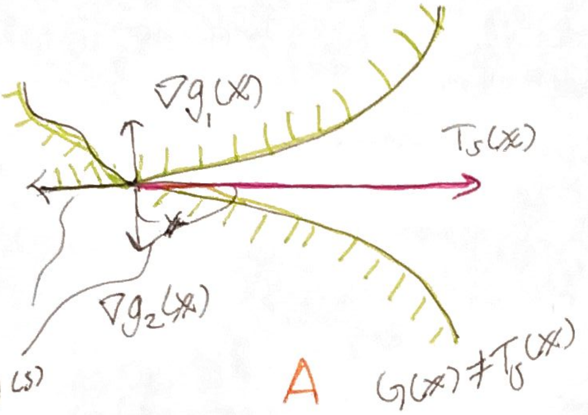
$$S := \{x \in \mathbb{R}^n \mid g_i(x) \le 0,\ i=1,\dots,m\}$$

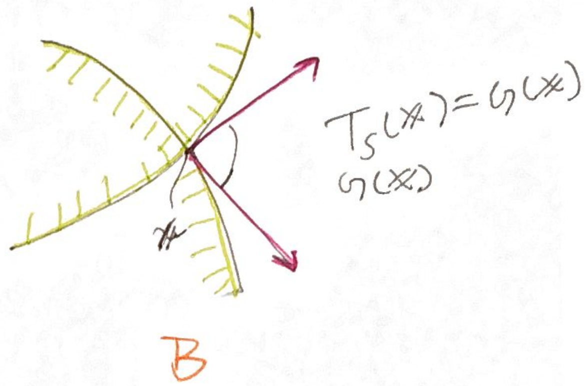
Recall

Def Tangent cone.,

$$T_S(x) = \{ p \in \mathbb{R}^2 \mid \exists \{x^k\} \subset S, \{\lambda_k\} \subset \{0,\infty\};$$
$$x_k \to x,\ \lambda_k(x_k - x) = p \}$$

Ex.





Def. Gradient cone

$$G(x) := \{ p \in \mathbb{R}^n \mid \nabla g_i(x)^T p \le 0,\ \forall i \in I(x) \}$$

where

$$I(x) := \{ i=1,\dots,m \mid g_i(x) = 0 \}$$

DEF Abadie's constraint qualification (Abadie's CQ)

If $T_S(x) = G(x)$, $x \in S$, we say that Abadie's CQ holds for x.

Note Abadie's CQ ensure that S is "well-behaving" in x.

THM KKT conditions

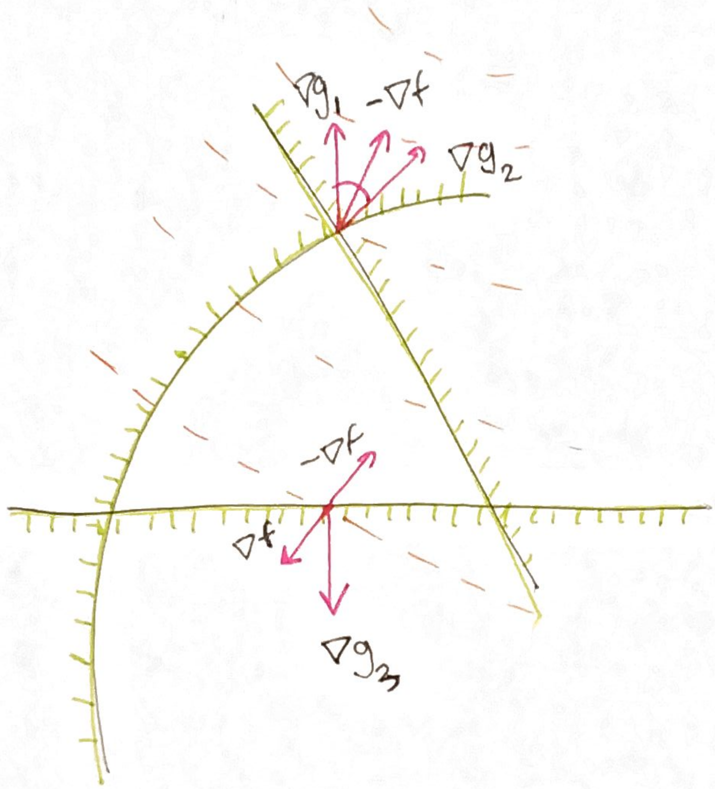
Assume that Abadie's CQ holds in $x^* \in S$.

$$x^* \text{ local min} \Rightarrow \nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0$$

$\mu_i \geq 0 \quad i=1,\dots,m$

$\mu_i g_i(x^*) = 0 \quad i=1,\dots,m \quad \Leftrightarrow \mu_i = 0 \ \forall i \notin \mathcal{I}(x)$

Proof See book, thm 5.29.



Same picture as front page of book.

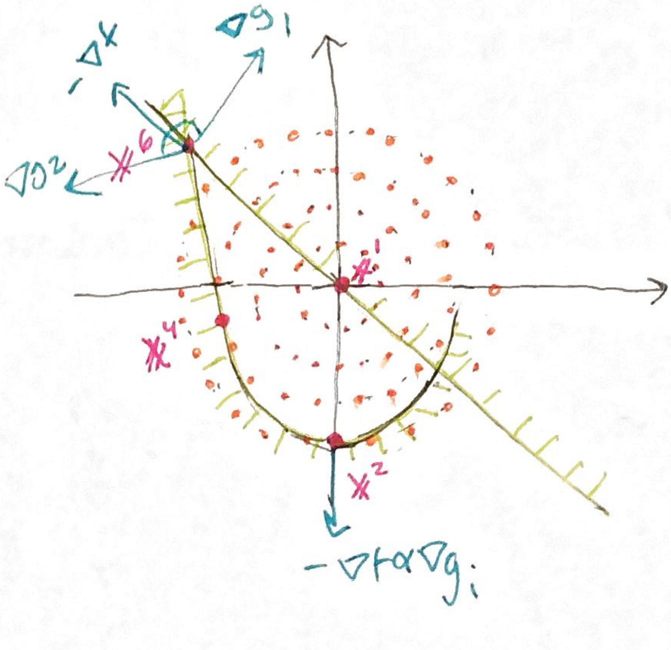
Note A KKT point x^* is such that $-\nabla f(x)$ is in the cone spanned by $\nabla g_i(x)$, $i \in \mathcal{I}(x)$.

Ex

$$\min -x_1^2 - x_2^2$$
$$x_1 + x_2 \leq 0$$
$$x_1^2 - x_2 \leq 2$$

$\nabla f = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$, $\nabla g_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\nabla g_2 = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}$



KKT conditions for the problem

$$\begin{cases} \nabla f(x) + \mu_1 \nabla g_1(x) + \mu_2 \nabla g_2(x) = 0 \Rightarrow \begin{cases} -2x_1 + \mu_1 + 2\mu_2 x_1 = 0 \\ -2x_2 + \mu_1 - \mu_2 = 0 \end{cases} \\ \mu_i = 0 \ \ i \notin I(x) \\ \mu_i \geq 0 \ \ \forall i \end{cases}$$

Solve for each possible $I(x)$:

$\underline{I(x) = \emptyset}$: $(\mu_1 = \mu_2 = 0)$, solve $\nabla f(x) = 0 \Longrightarrow$

$x^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $x^1 \in S$, x^1 is a KKT point (but $I(x) = \{1\}$).

$\underline{I(x) = \{1\}}$: $(\mu_2 = 0)$, solve $\nabla f(x) + \mu_1 \nabla g_1(x) = 0$, $g_1(x) = 0$

$$\left.\begin{cases} \left.\begin{array}{l} -2x_1 + \mu_1 = 0 \Rightarrow \mu_1 = 2x_1 \\ -2x_2 + \mu_1 = 0 \Rightarrow \mu_1 = 2x_2 \end{array}\right\} \Rightarrow x_1 = x_2 \\ x_1 + x_2 = 0 \Rightarrow x_1 = -x_2 \end{cases}\right\} \Rightarrow x_1 = x_2 = 0$$

$\underline{I(x) = \{2\}}$; $(\mu_1 = 0)$, Solve $\nabla f(x) + \mu_2 \nabla g_2(x) = 0$, $g_2(x) = 0$

$$\begin{cases} -2x_1 + 2\mu_2 x_1 = 0 \Rightarrow 2x_1(\mu_2 - 1) = 0 \\ -2x_2 - \mu_2 = 0 \quad ① \ x_1 = 0 \Rightarrow x_2 = -2 \rightarrow \mu_2 = 1/4 \geq 0 \\ x_1^2 - x_2 = 2 \end{cases}$$

$x^2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$ is a KKT point.

② $\mu_2 = 1$, $x_2 = -1/2 \Rightarrow x_1 = \pm\sqrt{3/2}$,

$x^3 = \begin{pmatrix} \sqrt{3/2} \\ -1/2 \end{pmatrix}$, $x^4 = \begin{pmatrix} -\sqrt{3/2} \\ -1/2 \end{pmatrix}$

$x^3 \notin S$ is <u>not</u> a KKT point $(g_1(x^3) > 0)$

$x^4 \in S$ is a KKT point.

$\underline{\mathcal{I}(x) = \{1,2\}}$

$$\begin{cases} \nabla f(x) + \mu_1 \nabla g_1(x) + \mu_2 \nabla g_2(x) = 0 \\ g_1(x) = 0 \\ g_2(x) = 0 \end{cases}$$

$$\begin{cases} g_1(x) = 0 \\ g_2(x) = 0 \end{cases} \Rightarrow x^5 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad x^6 = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$$

$$x^5: \begin{cases} -2 + \mu_1 + 2\mu_2 = 0 \\ 2 + \mu_1 - \mu_2 = 0 \end{cases} \Rightarrow \dots \Rightarrow \mu_1 = -2/3 \not\geq 0$$

not a KKT point

$$x^6: \begin{cases} 4 + \mu_1 - 4\mu_2 = 0 \\ -4 + \mu_1 - \mu_2 = 0 \end{cases} \Rightarrow \begin{array}{l} \mu_1 = 20/3 \\ \mu_2 = 8/3 \end{array} \Rightarrow$$

is a KKT point.

Constraint qualification

For KKT thm to hold we need to verify Abadie's CQ $\forall x \in S$.

[inner gradient cone

$$\mathring{G}(x) := \{ p \in \mathbb{R}^n \mid \nabla g_i(x)^T p < 0 \},$$

We give four criterias for Abadie's CQ.

1. The Mangasarin-Fromovitz CQ (MFCQ) holds at $x \in S$ if $\mathring{G}(x) \neq \emptyset$ (extra conditions of equality constraints are persistent)
2. The linear independence CQ (LICQ) holds at $x \in S$ if ∇g_i, $i \in \mathcal{I}(x)$, $\underbrace{\nabla h_j(x), j = 1, \dots, l}_{\text{from equality constraints.}}$ are linearly independent

3. Slater CQ hold if $g_i(x)$, $i=1,\dots,m$ are convex, h_j are affine and $\exists x \in S: g_i(x)<0, i=1,\dots,m$

4. Affine CQ holds if g_i, $i=1,\dots,m$, h_j, $j=1,\dots,l$ are affine.

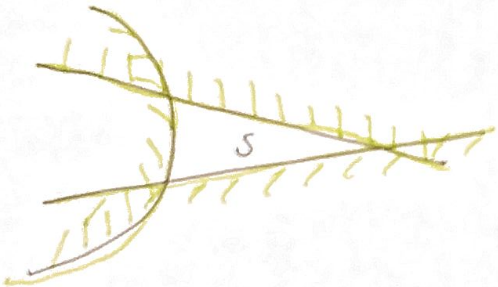
Note Abadie's CQ, MFCQ, LICQ holds at specific points. Slaters CQ, Affine CQ holds at every point.

Implications (see book for proof).

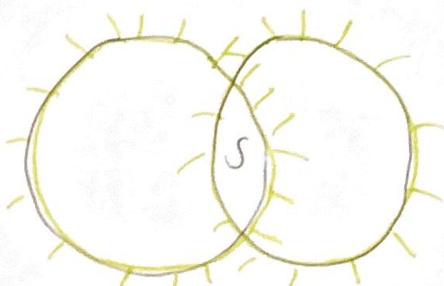
For $x \in S$: Abadie's CQ (at x) $\Leftarrow$ MFCQ (at x) $\Leftarrow$ LICQ (at x)

Abadie's CQ $\forall x \in S \Leftarrow$ Slater's CQ
$\Leftarrow$ Affine CQ

Which imply Abadie's CQ?

① 

LICQ

② 

Slaters CQ
LICQ

③ 

Affine CQ
slater CQ

Tuesday
24 september
15/15

LECTURE 7

Lagrangian Relaxation and Duality

Consider a generic problem

$$(1) \quad \min f(x) \;//\; f^*$$
$$\text{s.t } x \in S$$

If (1) is difficult, we can always replace it with something simpler.

Def A relaxation of (1) is a problem of the form

$$(1_R) \quad \min f_R(x) \;//\; f_R^*$$
$$\text{s.t. } x \in S_R$$

where we require

$$S \subseteq S_R$$

and

$$f_R(x) \leq f(x) \quad \forall x \in S$$

Ex

$$\min f(x)$$
$$\text{s.t } g_i(x) \leq 0 \quad i=1,\dots,m$$
$$x \in X$$

deleting constraints relaxation

$$\min f(x)$$
$$\text{s.t } x \in X$$

Ex

$$\min f(x)$$
$$\text{s.t. } g_i(x) \le 0 \quad i=1,\dots,m$$
$$x_j \in \{0,1\} \quad j=1,\dots,n$$

continuous relaxation

$$\min f(x)$$
$$\text{s.t. } g_i(x) \le 0 \quad i=1,\dots,m$$
$$0 \le x_j \le 1 \quad j=1,\dots,m$$

Thm For the problem (1) and (1_R)

a) $f_R^* \le f^*$

b) If (1_R) infeasible $\Longrightarrow$ (1) infeasible

c) If x_R^* is optimal in (1_R) and feasible in (1), then x_R^* is optimal in (1)

Proof

b) is left as an excercise.

a)

$$f_R^* = \min_{x \in S_R} f_R(x) \overset{\text{since } S \subseteq S_R}{\le} \min_{x \in S} f_R(x)$$

$$\le \left\{ \begin{array}{l} \text{since } f_R(x) \le f(x) \\ \text{for } \forall x \in S \end{array} \right\} \le \min_{x \in S} f(x) = f^*$$

c) Assume that

$$x_R^* \in \underset{x \in S_R}{\arg\min}\, f_R(x)$$

But then

since x_R^* minimizes f_R

since $f_R(x) \le f(x)$ for any $x \in S$

$$f(x_R^*) = f_R(x_R^*) \le f_R(x) \le f(x)$$

for any $x \in S$

since

$f(x_R^*) \le f(x)$

for any $x \in S$ and $x_R^* \in S$ we are done $\square$.

Lagrangian relaxation

Consider the

$$f^* = \min f(x) \qquad (P)$$
$$\text{s.t } g_i(x) \le 0 \quad i=1,\dots,m$$
$$x \in X$$

Idea Instead of just deleting the constraints g_i, include them in the objective function.

The Lagrangian relaxation of the constraints $g_i(x) \le 0$ is the problem

$$q(\mu) \quad \min f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$
$$\text{s.t.} \quad x \in X$$

We call $q(\mu)$ the dual function.

Thm (Weak duality)

For any $\mu \ge 0$ and any

$$x \in \{x \in X \mid g_i(x) \le 0, i=1,\dots,m\}$$

we have

$$q(\mu) \le f(x)$$

Note This is a restatement that for $\mu \ge 0$ Lagrangian relaxation is a relaxation.

Think of μ_i as a price for violating $g_i(x) \le 0$.
We require $\mu_i \ge 0$ to define a relaxation, since
we need to make sure that

$$\mu_i g(x) \begin{cases} \le 0, & \text{when } g_i(x) \le 0 \\ \ge 0, & \text{when } g_i(x) > 0 \end{cases}$$

Proof Can be done by showing that it is indeed a relaxation.
Excercise.

Let $\mu \ge 0$. Then

$$q(\mu) = \min_{z \in X} f(z) + \sum_{i=1}^{m} \mu_i g_i(z)$$

$$\underset{\substack{\text{since} \\ x \in X}}{\le} f(x) + \underbrace{\sum_{i=1}^{m} \underbrace{\mu_i}_{\ge 0} \underbrace{g_i(x)}_{\le 0}}_{\le 0} \le f(x)$$

Ex

$$f^* = \min x^2$$
$$\text{s.t. } x \ge 1$$

Let us relax the constraint

$$q(\mu) = \min_{x \in \mathbb{R}} f(x) + \mu \underbrace{(1-x)}_{\substack{\text{rewrite } x \ge 1 \\ \text{to } 1-x \le 0}}$$

The relaxed problem is an unconstrained convex problem
To find $q(\mu)$, solve

$$2x - \mu = 0$$

So

$$x = \frac{\mu}{2}$$

And

$$q(\mu) = \frac{\mu^2}{4} + \mu - \frac{\mu^2}{2} = -\frac{\mu^2}{4} + \mu$$

Plug in some values for μ.

$$q(0) = 0.$$

Hence

$$f^* \geq 0$$

$$q(1) = -\frac{1}{4} + 1 = \frac{3}{4} \leq f_* = 1$$

16^{15}

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g_i(x) \leq 0 \quad i = 1, \dots, m \\ & x \in X \end{aligned}$$

$$q(\mu) = \min_{x \in X} f(x) + \sum_i \mu_i g_i(x)$$

And we always have weak duality.
For $\mu \geq 0$ and feasible x in (P)

$$q(\mu) \leq f(x)$$

since

$$q(\mu) < f(x)$$

it makes sense to look for the largest value of q

Def The dual problem to (P) is the problem

$$q^* = \max_{\mu \geq 0} q(\mu)$$

Note We call (P) the <u>primal</u> problem.

Ex (continued)

We calculated

$$q(\mu) = -\frac{\mu^2}{4} + \mu$$

Then

$$q'(\mu) = -\frac{\mu}{2} + 1$$

So

$$q'(\mu) = 0 \Rightarrow \mu = 2$$

So

$$q^* = \max_{\mu \geq 0} q(\mu) = q(2) = -\frac{4}{4} + 2 = 1$$

So

$$q^* = 1 \leq f^* \leq f(1) = 1$$

But what if we try $x = 1$ in the primal problem
Note that $x=1$ is feasible in (P), with objective value $f(1) = 1$.

So we can bound

$$1 \leq f^* \leq 1 \Rightarrow f^* = 1 \Rightarrow x = 1 \text{ is optimal}$$

Dual function gives bounds from below.

Any feasible x gives bound from above.

In the example we had

$$q^* = f^* \longleftarrow$$ we call this no duality gap

This is in general not true.

In general

$$f^* - q^* \geq 0$$

is called the duality gap.

Ex.

$$f^* = \min x$$
$$\text{s.t. } x \geq \frac{1}{2}$$
$$x \in \{0,1\}$$

Note that $f_* = 1$ at $x=1$.

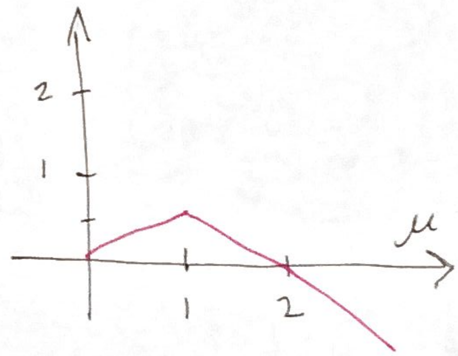
Relax constraints:

pick x based on the sign of $1-\mu$.

$$q(\mu) = \min_{x \in \{0,1\}} x + \mu\left(\frac{1}{2} - x\right) = \min_{x \in \{0,1\}} x(1-\mu) + \frac{\mu}{2}$$

So

$$q(\mu) = \begin{cases} \frac{\mu}{2} \quad (x=0) & \mu \leq 1 \\ 1 - \frac{\mu}{2} \quad (x=1) & \mu \geq 1 \end{cases}$$



$$q^* = \max_{\mu \geq 0} q(\mu) = q(1) = \frac{1}{2} < 1 = f^*$$

Practicality

It only makes sense to construct the dual problem

$$q^* = \max_{\mu \geq 0} q(\mu)$$

if it is somehow <u>easier</u> than the primal.

Thm $q(\mu)$ is a concave function.

Note The dual problem is <u>always</u> a convex problem.

Proof Take μ_1, μ_2 and let $\lambda \in (0,1)$.

We want to show that

$$q(\lambda\mu_1 + (1-\lambda)\mu_2) \geq \lambda q(\mu_1) + (1-\lambda) q(\mu_2).$$

$$q(\lambda\mu_1 + (1-\lambda)\mu_2) = \min_{x \in X} f(x) + (\lambda\mu_1 + (1-\lambda)\mu_2)^T g(x) =$$

$$= \min_{x \in X} \left[\lambda f(x) + \lambda \mu_1^T g(x)\right] + \left[(1-\lambda) f(x) + (1-\lambda)\mu_2^T g(x)\right] =$$

$$\geq \left\{ \begin{array}{l} \text{since} \\ \min a(x) + b(x) \geq \min a(x) + \min b(x) \end{array} \right\}$$

$$\geq \min \lambda \left[f(x) + \mu_1 g(x)\right] + \min (1-\lambda)\left[f(x) + \mu_2 g(x)\right]$$

since $\lambda, 1-\lambda > 0$

$$\geq \lambda q(\mu_1) + (1-\lambda) q(\mu_2)$$

Two things left

1) Can we recover x^* given
$$q^* = q(x^*) \,?$$
(optimality conditions similar to KKT)

2) Can we characterize when
$$f^* = q^* \,?$$

Strong duality

Recall weak duality $q(\mu) \leq f(x) \Rightarrow q^* \leq f^*$

Strong duality is when
$$q^* = f^*$$
Hopefully no surprise that this require convexity.

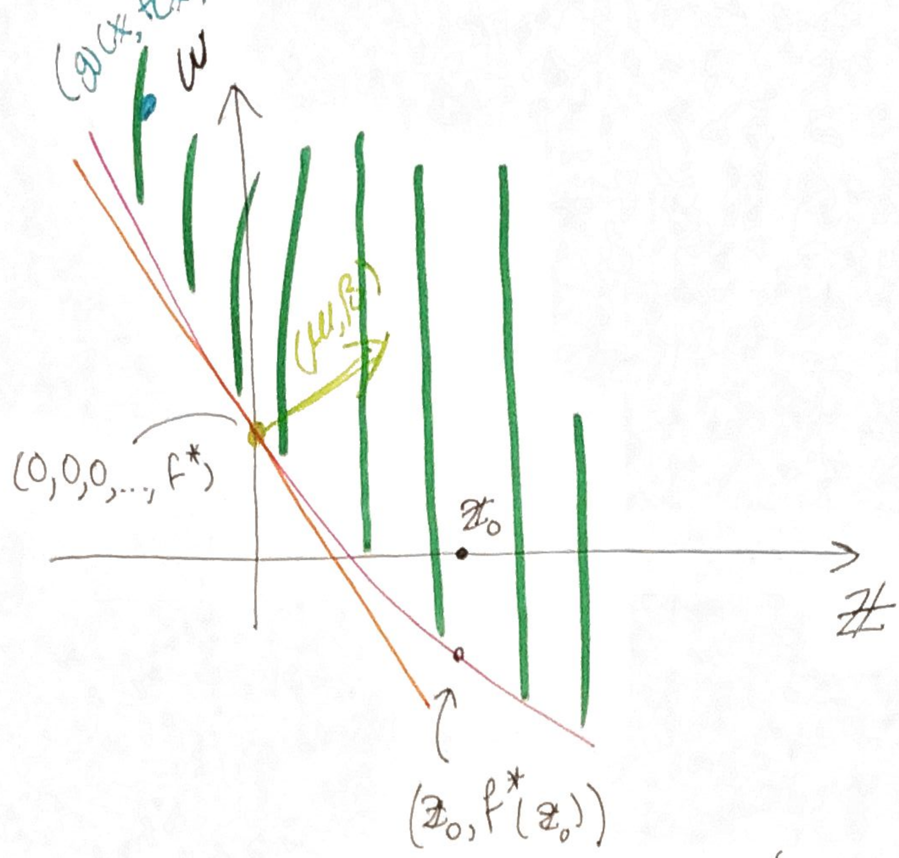
Thm (strong duality theorem)

Assume $f, g_i, i=1,\dots,m$ are convex and that X is convex, and that $\exists \, x \in X$ s.t.h. $g_i(x) < 0 \quad i=1,\dots,m$.

(Slater condition)

Then $$f^* = q^*$$

Proof Define the set
$$S = \left\{ \begin{array}{ll} (z, w) \in \mathbb{R}^m \times \mathbb{R} & \\ \exists \, x \in X \text{ with} & \begin{array}{l} g(x) \leq z \\ f(x) \leq w \end{array} \end{array} \right\}$$



where $f^*(z_0) = \min f(x)$
$\quad$ s.t. $g(x) \le z$
$\quad x \in X$

Claims

1) S is convex

2) $(0,0,0,\dots,f^*)$ is on the boundary of S since $(0,0,0,\dots,f^*-\varepsilon) \notin S$ for any $\varepsilon > 0$.

Can find supporting hyperplane $(\mu, \beta) \neq 0$ at $(0,0,0,\dots,f^*)$

so

$$\mu^T z + \beta w \geq \beta f^* \quad \forall (z, w) \in S$$

3) $\mu \geq 0, \ \beta \geq 0$

Proof

If $(z, w) \in S$, then $(z, w+\gamma) \in S \ \forall \gamma > 0$.

$$\mu^T z + \beta(w + \gamma) \geq \beta f^*$$

So if $\beta < 0$, LHS $\xrightarrow[\gamma \to +\infty]{} -\infty$

Contradiction.

Same argument using $(z+\gamma, w)$

4) Actually $\beta > 0$

Proof

Let $\bar{x} \in X$ s.t.h $g(\bar{x}) < 0$. Then

$$(g(\bar{x}), f(\bar{x})) \in S$$

Assume $\beta = 0$.

Then

$$\overset{\geq 0}{\mu^T} \underbrace{g(\bar{x})}_{<0} \geq 0$$

must hold.

So

$$\mu^T g(\bar{x}) \geq 0 \Longrightarrow \mu = 0$$

Contradiction since $(\mu, \beta) \neq 0$

Now rescale the inequality

$$\mu^T z + \beta w \geq \beta f^* \quad \text{by } \frac{1}{\beta} > 0.$$

yields

$$\underbrace{\left(\frac{\mu}{\beta}\right)}_{\mu^*}^T z + w \geq f^* \quad \forall (z, w) \in S.$$

Let $x \in X$ $(g(x), f(x)) \in S$

Then

$$f^* \leq f(x) + (\mu^*)^T g(x)$$

taking min over $x \in X$ yields

$$f^* \leq \min_{x \in X} f(x) + (\mu^*)^T g(x) = q(\mu^*) \leq q^* \leq f^* \Rightarrow$$

$$q^* = f^*$$

weak duality. $\square$

Tuesday

1 October

15^{15}

Primal recovery and optimality conditions

How to get primally opt x^* from dually optimal μ^k.
In general not possible!
But if $q^* = f^*$ (strong duality) we can.

Def We call μ^* a Lagrange multiplier if

$$f^* = \min_{x \in X} f(x) + \sum_{i=1}^{m} \mu_i^* g_i(x)$$

Thm Consider the pair (x^*, μ^*).
Then x^* is optimal in the primal an μ^* is a Lagrange multiplier if and only if

a) $$x^* \in \operatorname{argmin} f(x) + \sum_{i=1}^{m} \mu_i^* g_i(x)$$

b) $$\mu^* \geq 0$$

c) $$x^* \in X, \; g_i(x^*) \leq 0$$

d) $$\mu_i^* g_i(x^*) = 0$$

We call

a) Lagrangian optimality

b) Dual feasability

c) Primal feasability

d) Complementary slackness

(Note if $f, g_j \in C^1$ and convex and $X = \mathbb{R}^n$, a) becomes

$$\nabla f(x^*) + \sum \mu_i^* \nabla g_j(x^*) = 0 \;)$$

Proof Assume that x^* optimal μ^* a Lagrange multiplier.

$\Rightarrow$ b), c) are trivial.

a) is the definition of Lagrange multiplier

To prove d):

$$f^* \underset{\substack{\uparrow \\ x^* \text{ opt.}}}{=} f(x^*) \overset{\mu_i^* g_i(x^*) \leq 0}{\geq} f(x^*) + \sum_{i=1}^{m} \mu_i^* g_i(x^*) \geq$$

$$\underset{\substack{\uparrow \\ x^* \in X}}{\geq} \min_{x \in X} f(x) + \sum \mu_i^* g_j(x) \underset{\substack{\uparrow \\ \text{by def of L.M.}}}{=} f^*$$

$$\therefore f(x^*) = f(x^*) + \sum_{i=1}^{m} \mu_i^* g_i(x^*)$$

$$\Rightarrow \sum \mu_i^* g_i(x^*) = 0$$

$$\Rightarrow \mu_i^* g_i(x^*) = 0 \quad \forall i = 1, \dots, m$$

Ex (cont. of ex. from last time)

$$\min x$$
$$\text{s.t. } x \geq \frac{1}{2}$$

$$0 \leq x \leq 1$$

Relax $x \geq \frac{1}{2}$

$$q(\mu) = \min_{x \in [0,1]} x(1-\mu) + \frac{\mu}{2} = \begin{cases} \mu/2 & \mu \leq 1 \\ 1-\mu/2 & \mu \geq 1 \end{cases}$$

$\therefore \quad q^* = q(1) = \frac{1}{2}$

Recover x^*.

Look for $x^* \in \underset{x \in [0,1]}{\operatorname{argmin}}\ x\overbrace{(1-\mu)}^{=0} + \frac{\mu}{2} =$

$$= \underset{x \in [0,1]}{\operatorname{argmin}} \frac{\mu}{2} = [0,1]$$

Here, a) is useless.

But using complementary slackness d)

$$\mu^* \left(\frac{1}{2} - x^*\right) = 0$$

since $\mu^* = 1 \Rightarrow x^* = \frac{1}{2}$

Proof $\Leftarrow$ Now assume a) - d) holds

But then

$$q(\mu^*) = \min_{x \in X} f(x) + \sum_{i=1}^{m} \mu_i^* g_i(x) =$$

$$\overset{a)}{=} f(x^*) + \sum_{i=1}^{m} \mu_i^* g_i(x^*) =$$

$$\overset{d)}{=} f(x^*)$$

Hence,

$$f(x^*) = q(\mu^*) \overset{b)}{\leq} f^*$$

by weak duality.

Since x^* feasible (by c)), x^* optimal. □

LINEAR PROGRAMMING

Geometry

very, very very useful!

Def. A linear program is a problem of the form

$$\min \ c^T x$$
$$\text{s.t.} \quad x \in P$$

where P is a polyhedron, and $c \in \mathbb{R}^n$

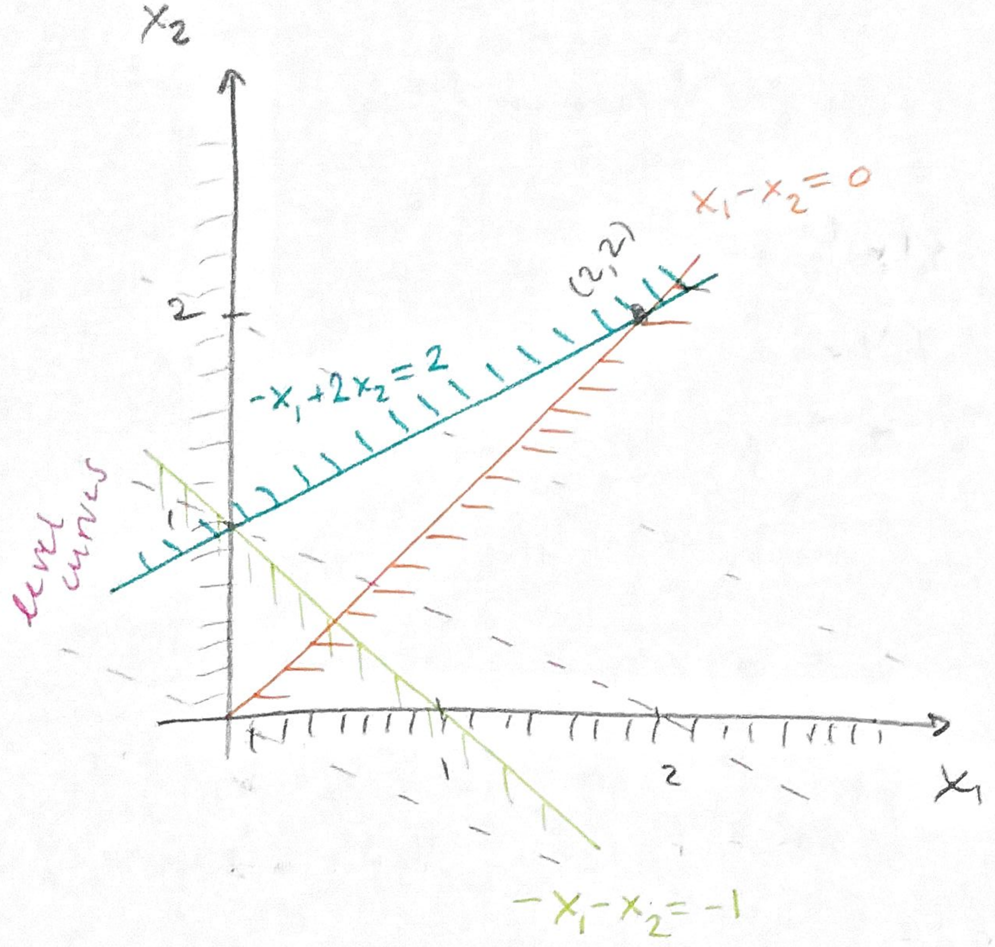
Recall. A polyhedron is the intersection of a set of half spaces, i.e.,

$$P = \{ x \in \mathbb{R} \mid Ax \leq b \}$$

where $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$.

Ex

$$\min \ -x_1 - 2x_2$$
$$\text{s.t.} \ -x_1 - x_2 \leq -1$$
$$x_1 - x_2 \leq 0$$
$$-x_1 + 2x_2 \leq 2$$
$$x_1, x_2 \geq 0$$



If we want to minimize $c^T x$ we should "push" the level curve as far as possible.

What we want to do is look for the optimum among the extreme points of P.

Aim Understand extreme points better (and algebraically).

16^{15} Reformulations of LPs

We will always work with standard form LPs.

Def An LP is in <u>standard form</u> if it is represented as

$$\begin{aligned} \min\ & c^T x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

Any LP can be written in standard form.

Ex.

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax \leq b \\ & x \geq 0 \end{aligned}$$

Trick add slack variables $s \in \mathbb{R}^m$

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax + s = b \\ & x \geq 0 \\ & s \geq 0 \end{aligned} \quad = \quad \begin{aligned} \min\ & c^T x \\ \text{s.t}\ & [A, I] \begin{bmatrix} x \\ s \end{bmatrix} = b \\ & \begin{bmatrix} x \\ s \end{bmatrix} \geq 0 \end{aligned}$$

Ex 2

$$\begin{aligned} \min\ & c^T x \\ \text{s.t}\ & Ax = b \end{aligned}$$

no sign restriction on x

Trick Replace: $x = x^+ - x^-$,
and the conditions

$$x^+ \geq 0$$
$$x^- \geq 0$$

The problem becomes

$$\begin{aligned} \min\ & c^T x^+ - c^T x^- \\ \text{s.t}\ & Ax^+ - Ax^- = b \\ & x^+, x^- \geq 0 \end{aligned}$$

So from now on, will only look at problems

$$\text{(LP)} \quad \begin{aligned} &\min \mathbb{c}^T \mathbb{x} \\ &\text{s.t. } A\mathbb{x} = \mathbb{b}, \quad A \in \mathbb{R}^{m \times n}, \quad \mathbb{b} \in \mathbb{R}^m \\ &\mathbb{x} \geq 0 \end{aligned}$$

and where $\text{rank}(A) = m$

the m rows of A are linearly independent (no redundant constraints)

Def A basic solution $\mathbb{x}$ of LP is a solution of $A\bar{\mathbb{x}} = \mathbb{b}$, where the columns A corresponding to non-zero elements at $\bar{\mathbb{x}}$ are linearly independent

Ex 1 (continued)

First, write the problem in standard form.

$$\begin{aligned} \min \quad & -x_1 - 2x_2 \\ \text{s.t.} \quad & -x_1 - x_2 + s_1 = 1 \\ & x_1 - x_2 + s_2 = 0 \\ & -x_1 + 2x_2 + s_3 = 2 \\ & x_1, x_2, s_1, s_2, s_3 \geq 0 \end{aligned}$$

To produce a basic solution:

Pick 3 linearly independent columns of

$$\begin{array}{ccccc} x_1 & x_2 & s_1 & s_2 & s_3 \end{array}$$
$$\begin{bmatrix} -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ -1 & 2 & 0 & 0 & 1 \end{bmatrix}$$

First pick columns corresponding to

$$(x_1, x_2, s_1).$$

We call this a basis.

Partition

$$A = [B \ N]$$

correspond to basic var.

correspond to nonbasic var.

Here

$$B = \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Solve

$$B x_B = b$$

$\begin{bmatrix} x_1 \\ x_2 \\ s_1 \end{bmatrix}$

What I'm really interested in is

$B x_B + N x_N = b \Rightarrow$

$B x_B = b$

because we solve with $x_N = 0$

$$x_B = B^{-1} b = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

So $x = (2, 2, 3, 0, 0)^T$ is a basic solution. (And a BFS).

Pick (s_1, s_2, s_3) as a basis:

$$B = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}, \quad N = \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 2 \end{bmatrix} \end{array}$$

Then

$$B^{-1} b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

$\Rightarrow x = (0, 0, -1, 0, 2)^T$ is a basic solution.

Note. This is not feasible (since $s_1 < 0$).

Not a BFS.

Def.

A basic feasible solution (BFS) is a basic solution that is feasible.

Thm.

Assume $A \in \mathbb{R}^{m \times n}$, rank$(A) = m$, $\mathbf{b} \in \mathbb{R}^m$, and let

$$P = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \begin{array}{l} A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq 0 \end{array} \right\}.$$

Then $\bar{\mathbf{x}}$ is an extreme point of P if and only if it is a BFS.

Proof Lets assume $\bar{\mathbf{x}}$ is a BFS, i.e, we have a partition

$$A = [B, N]$$

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{\mathbf{x}}_B \\ \bar{\mathbf{x}}_N \end{bmatrix}, \quad \bar{\mathbf{x}} \geq 0$$

where $\bar{\mathbf{x}}_N = 0$.

Assume that we have $\lambda \in (0,1)$, $\mathbf{x}^1, \mathbf{x}^2 \in P$. s.t.

$$\bar{\mathbf{x}} = \lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2,$$

Separate this into basic and nonbasic parts.

$$0 = \bar{\mathbf{x}}_N = \lambda \mathbf{x}^1_N + (1-\lambda)\mathbf{x}^2_N$$

But $\mathbf{x}^1_N, \mathbf{x}^2_N \geq 0 \Rightarrow \mathbf{x}^1_N = \mathbf{x}^2_N = 0$

Further, we have

$$A\mathbf{x}^1 = \mathbf{b}$$

since $\mathbf{x}^1_N = 0$ we have

$B x_B^1 = b$

{since B is invertible}

$$x_B^1 = B^{-1} b = \bar{x}$$

Same for x_B^2.

$\therefore\ \bar{x} = x^1 = x^2 \Rightarrow \bar{x}$ is an extreme point.

We skip the $\Rightarrow$ part.

Summary

- We "understood" that solutions to LPs are extreme points of polyhedra.
- We showed that extreme point $\Leftrightarrow$ BFS

Tomorrow

Develop the simplex algorithm.

Just searches all BFSs in a clever way.

Wednesday
2 October
1315

LECTURE

Recall the representation thm
(here in standard form):

Let

$$P = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{array} \right\}.$$

Then every $\mathbf{x} \in P$ can be written as

$$\mathbf{x} = \sum_{i=1}^{K} \alpha_i \mathbf{v}^i + \sum_{j=1}^{R} \beta_j \mathbf{d}^j$$

where

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

$$\alpha_i, \beta_j \geq 0$$

where

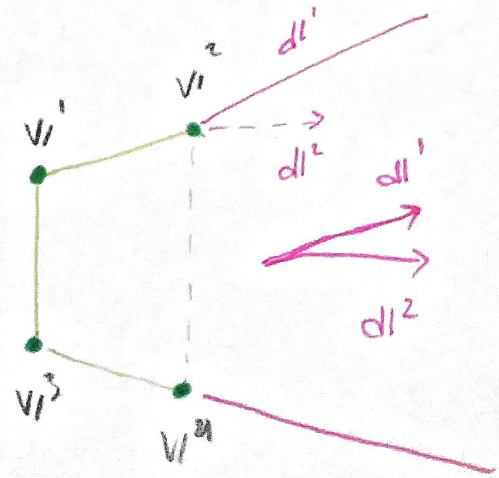
$$V = \{ \mathbf{v}^i, i = 1, \dots, k \}$$

are the extreme points of P and

$$D = \{ \mathbf{d}^j, j = 1, \dots, R \}$$

are the extreme directions of

$$C = \{ A\mathbf{p} = 0, \mathbf{p} \geq 0 \}$$



Thm Consider (LP)

a) If P is nonempty and $c^T d \geq 0$ for all $d \in D$, then (LP) has an optimal solution.

b) If (LP) has an optimal solution, then there exists an optimal solution among the extreme points.

Proof.

a) Let $x \in P$.
Representation thm yields

$$x = \sum_{i=1}^{k} \alpha_i v^i + \sum_{j=1}^{R} \beta_j d^j$$

We can then view (LP) as an opt. problem over α_i, β_j. Note

$$c^T x = \sum_{i=1}^{k} \alpha_i c^T v^i + \sum_{j=1}^{R} \beta_j c^T d^j$$

If any $c^T d^j < 0$, we can let $\beta_j \to \infty$ and then

$$c^T x \xrightarrow[\beta_j \to +\infty]{} -\infty .$$

On the other hand if $c^T d^j \geq 0 \ \forall j$, then taking $\beta_j = 0$ is optimal.

so in this case

$$\min_{s.t.\ P} c^T x = \min \sum_{i=1}^{k} \alpha_i c^T v^i$$

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

$$\alpha_i \geq 0$$

optimization over a compact set $\Rightarrow$ solution exists.

b) Now assume $\exists x^*$ optimal. As in a) we know that

$$x^* = \sum_{i=1}^{k} \alpha_i^* v^i$$

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

$$\alpha_i^* \geq 0 .$$

Now let $a \in \underset{i \in \{1,\dots,k\}}{\operatorname{argmin}} c^T v^i$

But then

$$c^T v^a = c^T v^a \sum_{i=1}^{k} \alpha_i^* = \sum_{i=1}^{k} \alpha_i^* c^T v^a \leq$$

$$\leq \sum_{i=1}^{k} \alpha_i^* c^T v^i = c^T x^*$$

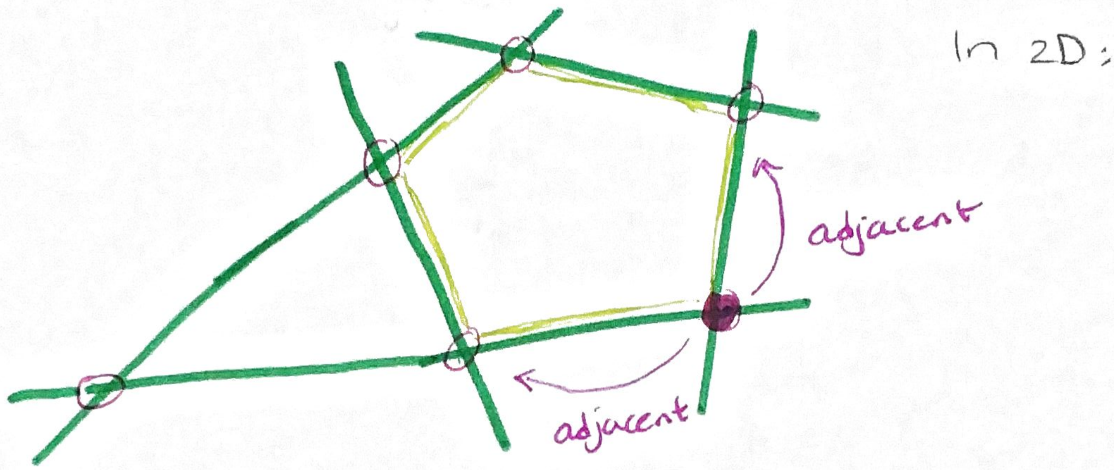
$\therefore$ v^a is an optimum.

Adjacent extreme points

Recall An extreme point of a polyhedron

$$P = \{ x \in \mathbb{R}^n \mid Ax \leq b \}$$

is where n lin. indep. hyperplanes meet.



Extreme points are adjacent if they differ by 1 hyperplane.

Look at BFS $\bar{x}$:

$$B\bar{x}_B = b \quad \text{(m equations)} \qquad \bar{x}_B \geq 0$$

$$x_N = 0 \quad \text{(n-m equations)}$$

Replacing a basic variable with a non-basic corresponds to adjecent extreme points.

SIMPLEX ALGORITM

Outline

1) Assume an initial BFS: $\bar{x}$.

2) Is $\bar{x}$ optimal?

3) Find a search direction.
   Towards an adjecent extreme point. (Geometric)
   (select a non-basic variable to include in basis)

4) Move along search direction
   until we find an extreme point.
   (until a basic variable becomes 0)

5) Update $\bar{x}$, go to 1).

Let us rewrite our (LP) relative to a BFS $\bar{x}$, corresponding to $A = [B, N]$

$$\begin{array}{ll} \min \ c^T x & \\ \text{s.t.} \ Ax = b & \\ \quad\ x \geq 0 & \end{array} = \begin{array}{ll} \min & c^T x_B + c^T x_N \\ \text{s.t.} & Bx_B + Nx_N = b \\ & x_N \geq 0 \\ & x_B \geq 0 \end{array}$$

{Rewrite $x_B = B^{-1}(b - Nx_N)$ }

$$= \begin{array}{ll} \min & c^T B^{-1} b + \left(c_N^T - c_B^T B^{-1} N\right) x_N \\ \text{s.t.} & x_N \geq 0 \\ & B^{-1} b - B^{-1} N x_N \geq 0 \end{array}$$

Observation If
$$c_N^T - c_B^T B^{-1} N \geq 0.$$
Then $x_N = 0$ is optimal.

14^{15} **Def.** We call
$$\tilde{c}_N^T = (c_N^T - c_B^T B^{-1} N)$$
the <u>reduced cost</u>.

Optimality condition

If reduced cost $\tilde{c}_N \geq 0$, current BFS is optimal.

3) <u>Search directions:</u>

Use directions given by including one non-basic variable j

i.e.
$$d^j = \begin{pmatrix} -B^{-1}N_j \\ e_j \end{pmatrix} \begin{array}{l} \leftarrow \text{basic part} \\ \leftarrow \text{non-basic part} \end{array}$$

Note
$$c^T d^j = (\tilde{c}_N)_j.$$

For search direction (or <u>incoming variable</u>) choose a non-basic variable with negative reduced cost.

Ex (continued)

Yesterday, we looked at basis (x_1, x_2, s_1).

$$B = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

BFS.

$$\bar{x} = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 0 \\ 0 \end{bmatrix}$$

Calculate reduced cost

$$\tilde{c}_N^T = \underset{s_2 \ \ s_3}{(0,0)} - \underset{x_1 \ \ x_2 \ \ s_1}{(2,-3,0)} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 0 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = (\underset{<0}{-1}, \underset{>0}{1})$$

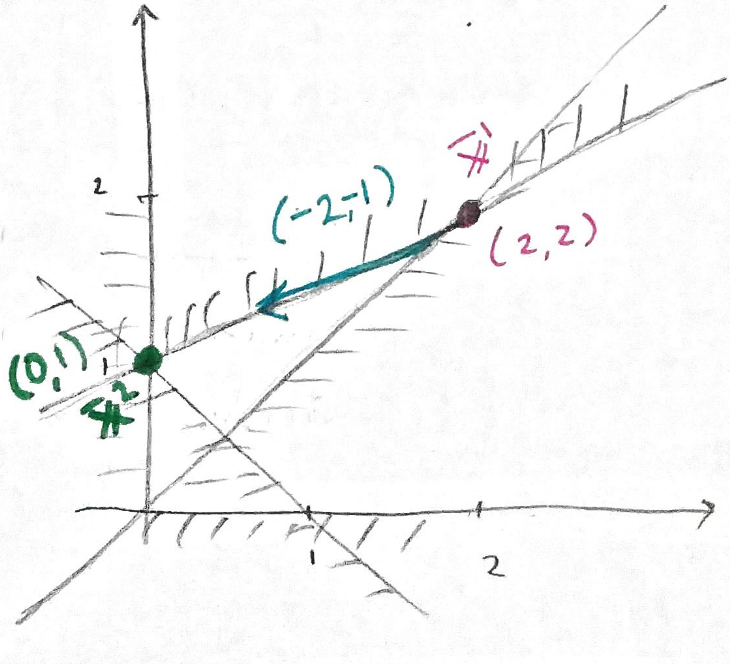
Incoming variable: s_2

Further look at

$$B^{-1}N = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 0 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 3 & 2 \end{bmatrix}$$

So the search direction

$$d^{s_2} = \begin{bmatrix} -2 \\ -1 \\ -3 \\ 1 \\ 0 \end{bmatrix}$$



4) <u>Step length</u>

We want to move as far as possible (while staying feasible).

Lets consider

$$\bar{x} + \theta d^{j}$$

want

$$\theta^{*} = \max \theta \quad \text{s.t} \quad \bar{x} + \theta d^{j} \in P$$

minimum ratio test

$$\theta^{*} = \min_{k : (B^{-1}N^{j})_k > 0} \frac{(B^{-1}b)_k}{(B^{-1}N^{j})_k}$$

Ex (continued)

$$\left.\begin{array}{l} B^{-1}b = \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} \\ (B^{-1}N_{s_2}) = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \end{array}\right\} \Rightarrow \begin{array}{l} 2/2 = 1 \\ 2/1 = 2 \\ 3/3 = 1 \end{array}$$

Here outgoing variable x_1 (or s_1).

Select outgoing variable by the <u>minimum ratio test</u>.

5) <u>Update $\bar{x}$</u>

Included a non-basic j into basis: dropped a basic variable by minimum ratio.

$\Rightarrow$ new basis, with new partition $A = [B, N]$.

Ex (continued)

New basis became (s_2, x_2, s_1)

$$
\begin{array}{cc}
 & \begin{array}{ccc} s_2 & x_2 & s_1 \end{array} \\
B = & \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 2 & 0 \end{bmatrix}
\end{array}
$$

$$B^{-1}b = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$ check this!

Warning

In the example we got a BFS where a basic variable was 0.

Call this a degenerate BFS.

Problem: If $(B^{-1}b)_k = 0$

So minimum ratio test can give $\theta^* = 0 \Longrightarrow$

Can lead to problems, but doesn't in practice (and there are ways to deal with it).

Simplex algorithm (explicit)

1) Assume known BFS $\bar{x}$, partition $A = [B, N]$.

2) Compute

$$\tilde{c}_N^T = c_N^T - c_B^T B^{-1} N$$

If $\tilde{c} \geq 0$: current $\bar{x}$ optimal. Stop

If $\tilde{c} \ngeq 0$: select inc. var. j^*

s.th. $(\tilde{c}_N)_{j^*} < 0$

3) Compute $B^{-1}N_{j^*}$

If $(B^{-1}N_{j^*}) \leq 0$ for all k, problem is unbounded.
stop.

If $(B^{-1}N_{j^*}) \nleq 0$: choose k^* by minimum ratio test.

4) Construct new basis by replacing j^* by k^*, (go to 1).

One thing left: How to find initial BFS?

Use the simplex algorithm.

consider

$$\begin{aligned} \min \quad & \mathbb{c}^T \mathbb{x} \\ \text{s.t.} \quad & A\mathbb{x} = \mathbb{b} \\ & \mathbb{x} \geq \mathbb{0}, \end{aligned} \qquad \text{(Phase II)}$$

assume $\mathbb{b} \geq 0$

Introduce artificial variables $\mathbb{a}$

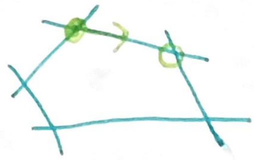
$$\begin{aligned} w^* = \min \quad & \sum_{i=1}^{m} a_i \\ \text{s.t.} \quad & A\mathbb{x} + I\mathbb{a} = \mathbb{b} \\ & \mathbb{x}, \mathbb{a} \geq \mathbb{0} \end{aligned} \qquad \text{(Phase I)}$$

Claim: Since $\mathbb{b} \geq 0$, $\mathbb{a} = \mathbb{b}$ is a BFS.

So we can solve (Phase I) by using simplex.

If $w^* = 0$, then opt. solution to Phase I is a BFS for (Phase II).

2. Find outgoing variable (min ratio test)



conceptually it's a line search type of algorithm.

LECTURE 10

LP duality

(P)
$$\begin{aligned} \min\ & c^T x \\ \text{s.t}\ & Ax = b \\ & x \geq 0 \end{aligned}$$

(D)
$$\begin{aligned} \max\ & b^T y \\ \text{s.t.}\ & A^T y \leq c \\ & y \in \mathbb{R}^m \end{aligned}$$

$$\begin{aligned} z^* = \min\ & c^T x \\ \text{s.t.}\ & Ax = b \\ & x \geq 0 \end{aligned}$$

$A \in \mathbb{R}^{m \times n}$

$$q(y) = \min_{\text{s.t } x \geq 0} \left[c^T x + y^T (b - Ax) \right] =$$

$$= b^T y + \min_{\text{s.t. } x \geq 0} \left[(c - A^T y)^T x \right] =$$

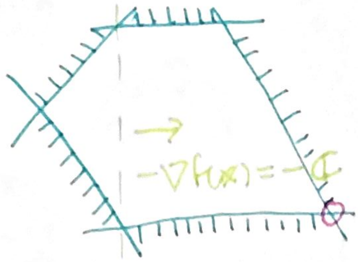
$$= \begin{cases} b^T y & \text{if } A^T y \leq c \\ -\infty & \text{otherwise} \end{cases}$$

Monday 7 October 800

Recap from last week

LINEAR PROGRAMMING

- "optimizing linear objective function over a polyhedron"



- standard form

$$\begin{cases} \min c^T x \\ \text{s.t. } Ax = b \quad (b \geq 0) \\ \quad x \geq 0 \end{cases}$$

- optimal solutions in extreme points

- extreme point $\iff$ BFS

- <u>BFS</u>: $A = [B, N]$, $B \in \mathbb{R}^{m \times m}$, $N \in \mathbb{R}^{m \times (n-m)}$, $B^{-1} b \geq 0$

$x = [x_B, x_N]$

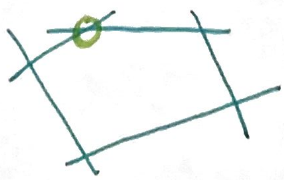
$$\begin{cases} x_B = B^{-1} b \\ x_N = 0 \end{cases}$$

SIMPLEX METHOD

0. Initial BFS,

$A = [B, N]$

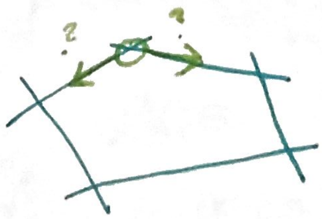
$x = [x_B, x_N]$



what simplex method does is jumps to adjecent extreme points.

1. Find incoming variable.

$\tilde{c}_N^T = c_N^T - c_B^T B^{-1} N$



so

$$q^* = \max q(y) =$$

$$= \boxed{\begin{array}{l} \max \mathbb{b}^T y \\ \text{s.t. } A^T y \leq \mathbb{c} \end{array}}$$

Weak duality

$$\left.\begin{array}{l} x \text{ feasible (P)} \\ y \text{ feasible (D)} \end{array}\right\} \mathbb{c}^T x \geq \mathbb{b}^T y$$

Proof

$$\mathbb{c}^T x \geq (A^T y)^T x \qquad \left[\begin{array}{c} \mathbb{c} \geq A^T y \\ x \geq \mathbb{0} \end{array}\right]$$

$$= y^T A x \qquad [A x = \mathbb{b}]$$

$$= y^T \mathbb{b}$$

$$= \mathbb{b}^T y$$

q^{∞}

	PRIMAL		DUAL
OBJ	min	OBJ	max
CONSTRAINTS		VARIABLES	
	$\geq$ canonical		≥ 0
	$\leq$ non-canonical		≤ 0
	$=$		free
VARIABLES		CONSTRAINTS	
	≥ 0		$\leq$
	≤ 0		$\geq$
	free		$=$

ex.

PRIMAL

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \le 1 \quad (y_1) \\ & 3x_1 - 4x_2 = 2 \quad (y_2) \\ & x_1 \ge 0 \end{aligned}$$

DUAL

$$\begin{aligned} \max \quad & y_1 + 2y_2 \\ \text{s.t.} \quad & y_1 + 3y_2 \le 1 \quad (x_1) \\ & 2y_1 - 4y_2 = 1 \quad (x_2) \\ & y_1 \le 0 \end{aligned}$$

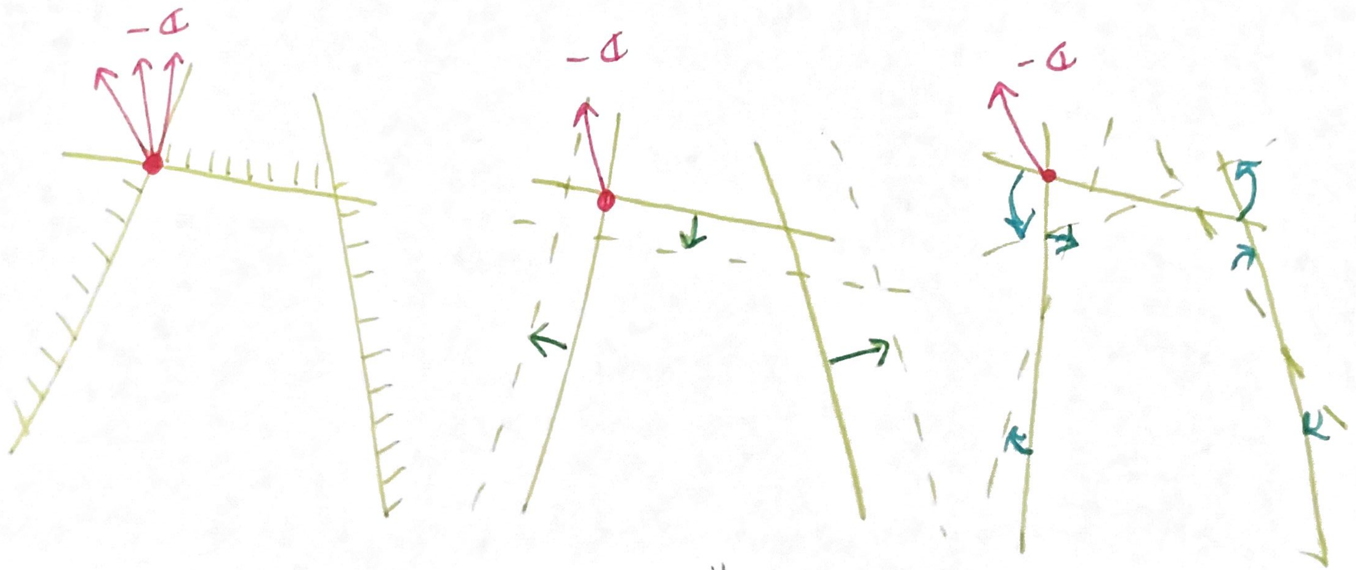
$$\mathbb{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbb{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$$

Complementary slackness - Very important!

SENSITIVITY ANALYSIS

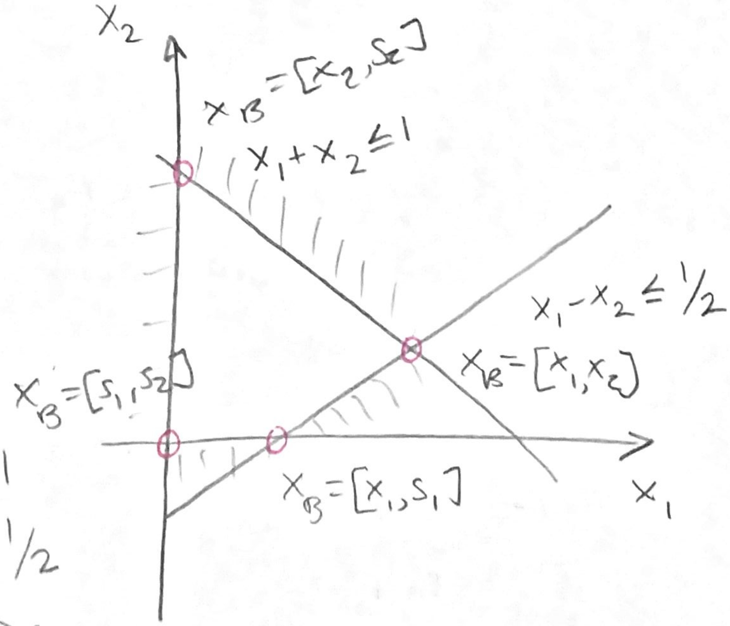


change $\mathbb{c}$. change $\mathbb{b}$. change A.

ex.

$$\min \quad x_1 + x_2$$
$$\text{s.t.} \quad x_1 + x_2 \le 1$$
$$x_1 - x_2 \le 1/2$$
$$x_1, x_2 \ge 0$$

$$\min \quad x_1 + x_2$$
$$x_1 + x_2 + s_1 = 1$$
$$x_1 - x_2 + s_2 = 1/2$$
$$x_1, x_2, s_1, s_2 \ge 0$$



Tuesday
8 October
15'15

LECTURE 11

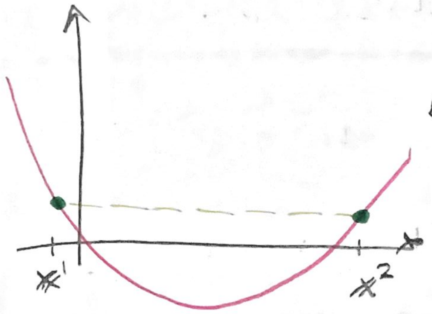
Convex optimization

A set S is <u>convex</u> if

$$\left.\begin{array}{l} x^1, x^2 \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow \lambda x^1 + (1-\lambda)x^2 \in S$$

A function f is convex on a convex set S if

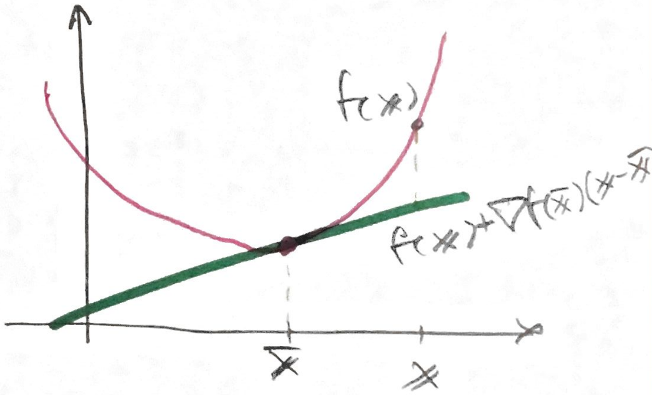
$$\left.\begin{array}{l} x^1, x^2 \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda)x^2) \le \lambda f(x^1) + (1-\lambda) f(x^2)$$



- If $f \in C^1$ then f is convex on S if

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$

for all $x, \bar{x} \in S$.



- An optimization problem is convex if

$$\text{(P)} \quad \min f(x) \quad \text{s.t. } x \in S$$

f is a convex function on S and S is a convex set.

- (P) $\min f(x)$
  s.t. $g_i(x) \leq 0, \quad i=1,\dots,m$
  $h_j(x) = 0, \quad j=1,\dots,k$
  $x \in X$

- (P) is convex if f is convex, $g_i, i=1,\dots,m$ is convex, $h_j, j=1,\dots,k$ is affine, and X is convex.

Thm Consider a convex optimization problem (P).

$$\boxed{\begin{array}{l} x^* \text{ local min to (P)} \Rightarrow \\ x^* \text{ global min to (P)} \end{array}}$$

- Most algorithms assume smoothness of f ($f \in C^1$)

- In convex optimization, we can drop this assumption.

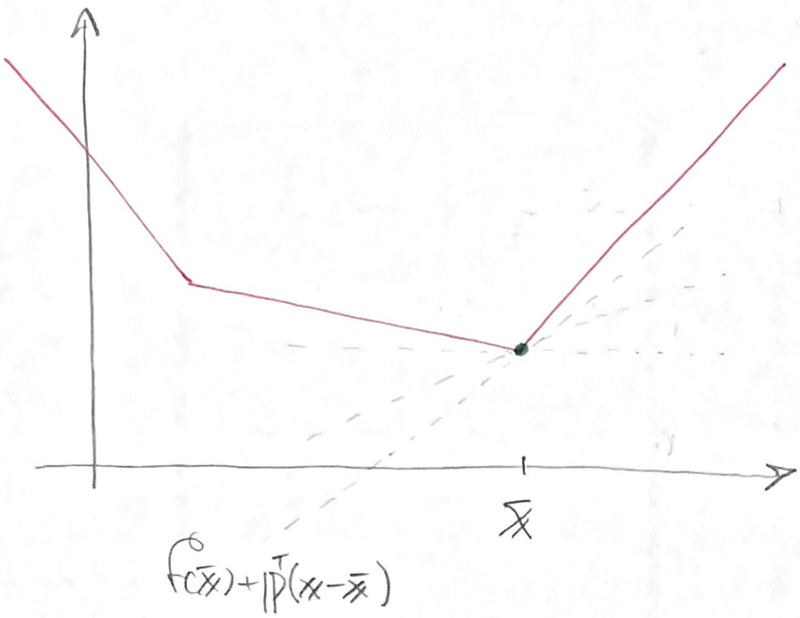
- In convex optimization, we can instead consider <u>subgradients</u>.

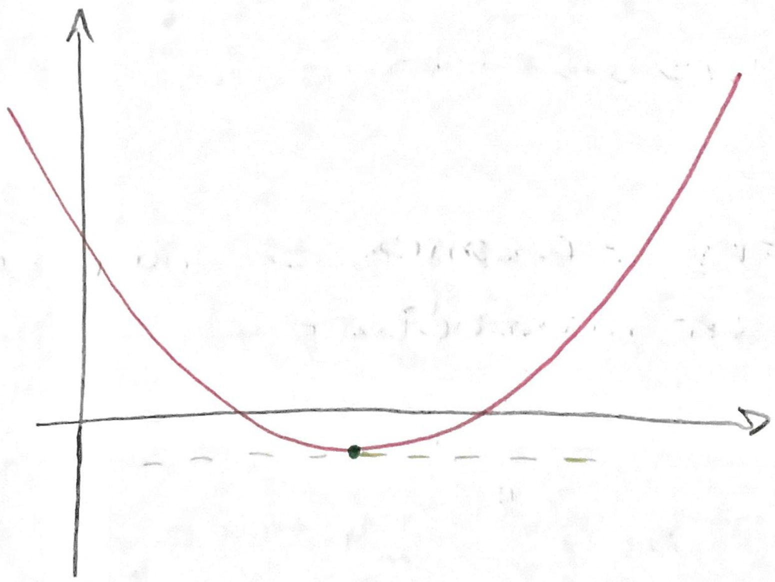
- $x^{k+1} = x^k + \alpha_k p^k$
  where p^k are subgradients.

Def. Let $S \subseteq \mathbb{R}^n$ be a convex set and f a convex function. Then $p \in \mathbb{R}^n$ is a <u>subgradient</u> to f in the point $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + p^T(x - \bar{x})$$

for all $x \in S$.



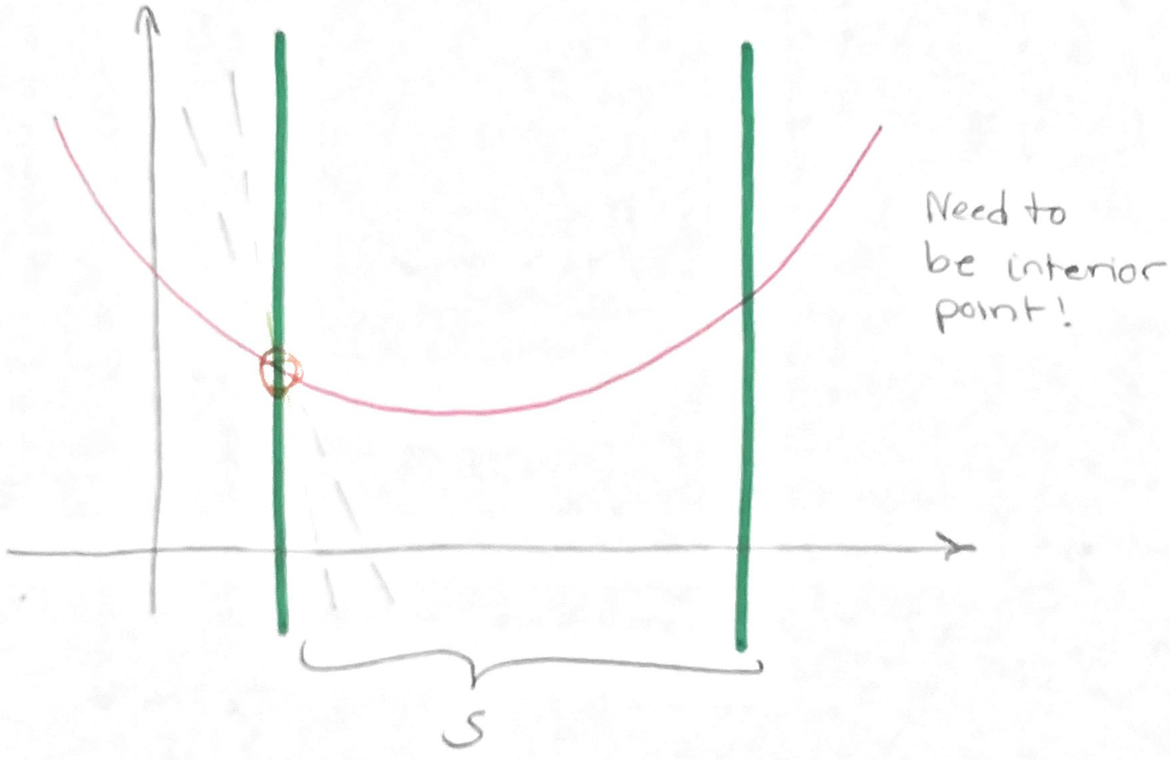


- The set of all subgradients to f at $\bar{x}$ is called the subdifferential and is denoted

$$\partial f(\bar{x}) = \{ \mathbb{P} \mid f(x) \geq f(\bar{x}) + \mathbb{P}^T(x-\bar{x}) \ \forall x \in S \}$$

Lemma If $\bar{x} \in \text{int}(S)$ and $f \in C^1$, then

$$\partial f(\bar{x}) = \{ \nabla f(\bar{x}) \}$$



Thm (*) Let $S \subseteq \mathbb{R}^n$ be a convex set and f a convex function on S. Then, for each $\bar{x} \in \text{int}(S)$, there exists a subgradient to f.

Def. The epigraph of f with respect to the set S is

$$\text{epi}_S f = \{ (x, \alpha) \in S \times \mathbb{R} \mid f(x) \leq \alpha \}$$

Thm

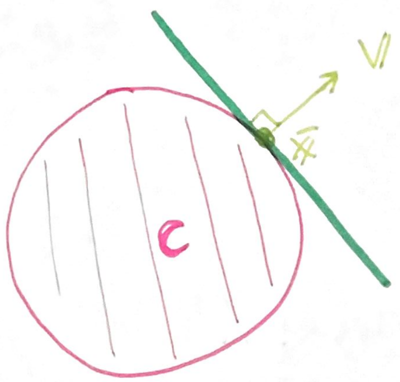
f is convex on $S \iff \text{epi}_S f$ is a convex set

Thm

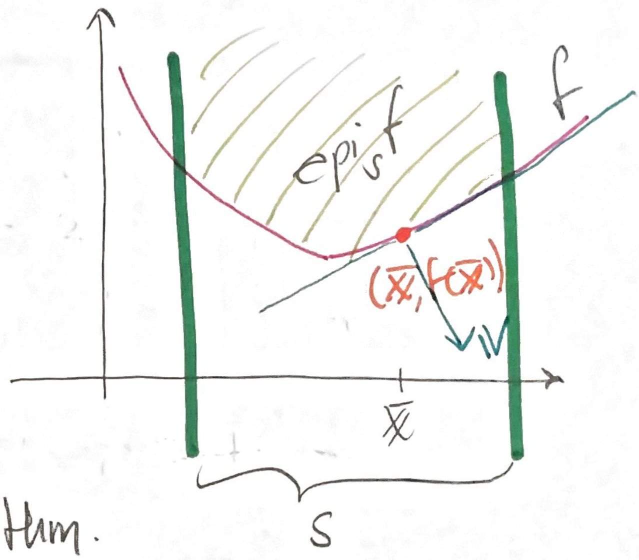
Let $C \subseteq \mathbb{R}^n$ be nonempty and convex. Let $\bar{x}$ be a point on the boundary of C.

Then there exists a supporting hyperplane to C at $\bar{x}$, meaning that $\exists v \neq 0$

$$v^T(x - \bar{x}) \leq 0 \quad \forall x \in C$$



Proof of (*)



- We know $\text{epi}_S f$ is convex
- For $\bar{x} \in \text{int}(S)$, $(\bar{x}, f(\bar{x}))$ is a boundary point to $\text{epi}_S f$.
- Use supporting hyperplane thm.
- There exists a vector $v \in \mathbb{R}^{n+1}$ such that

$$v^T\left(\begin{bmatrix} x \\ z \end{bmatrix} - \begin{bmatrix} \bar{x} \\ f(\bar{x}) \end{bmatrix}\right) \leq 0 \quad \forall (x, z) \in \text{epi}_S f$$

Write $v = (u, t)$, where $u \in \mathbb{R}^n$, $t \in \mathbb{R}$

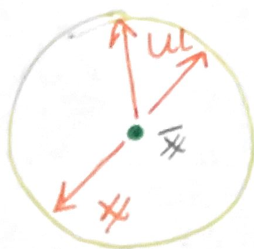
$\Rightarrow \quad u^T(x - \bar{x}) + t(z - f(\bar{x})) \leq 0$

CLAIM: $t \leq 0$

Choose $(x, z) = (\bar{x}, f(\bar{x}) + 1)$,
and note that $t \leq 0$.

Assume: $t = 0$

$u^T(x - \bar{x}) \leq 0 \quad \forall x \in S$.
Impossible, since
$\bar{x} \in \mathrm{int}(S)$.



$\Rightarrow t < 0$.
Then take $z = f(x)$
(OK since $(x, f(x)) \in \mathrm{epi}_S f$)
and rearrange.

$$\boxed{f(x) \geq f(\bar{x}) - \left(\frac{u}{t}\right)(x - \bar{x})}$$

for all $x \in S$.

$\Rightarrow \quad -\frac{u}{t} \in \partial f(\bar{x})$

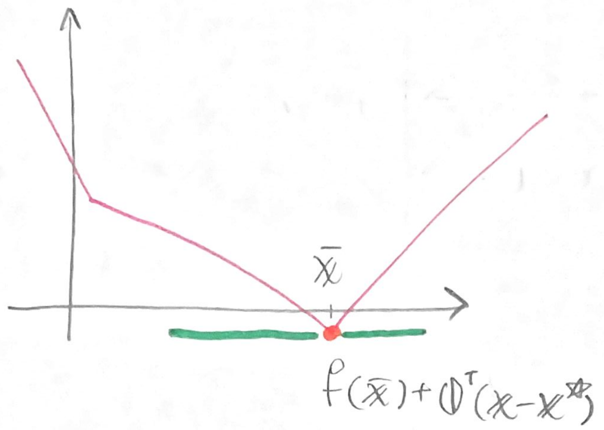
DONE!

16^{15}

Prop.

Let f be a convex function. Then

$$x^* \text{ global min to } f \text{ over } \mathbb{R}^n \iff 0 \in \partial f(x^*)$$



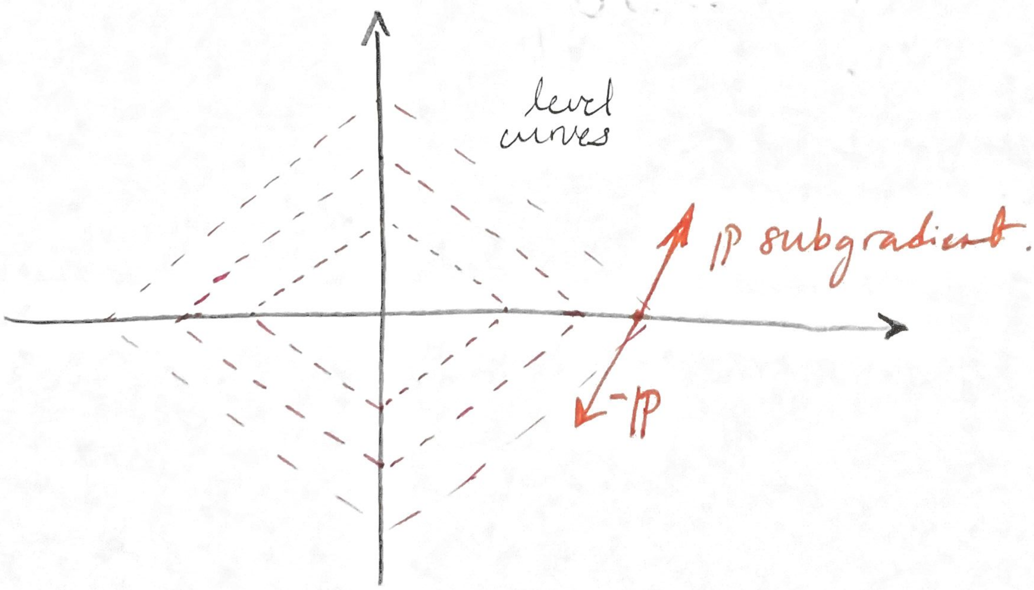
- <u>Subgradient method</u>

$$x^{k+1} = x^k - \alpha_k p^k$$

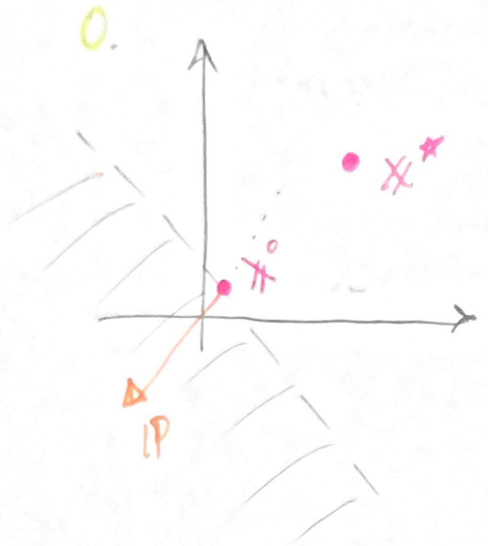
where p^k is a subgradient to f at x^k.

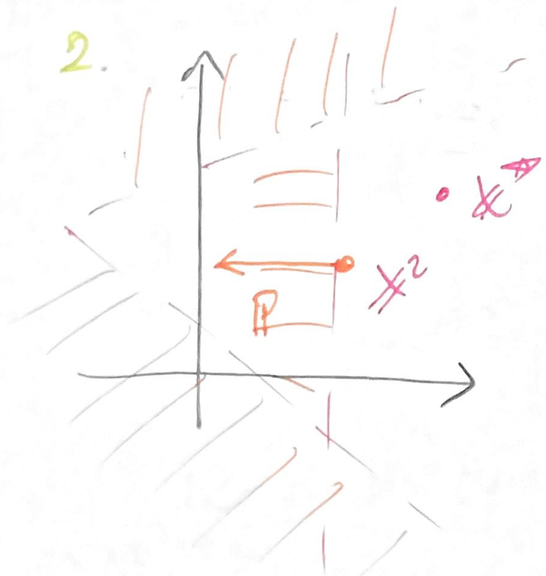
NOTE <u>steepet descent</u> always decreased the objective value.

- Subgradients are not necessarily descent directions.



- However, subgradients define cutting planes for where the optimum exists.





Subgradient method

0. Initiate $x^0 \in S$.
   Let $f_{BEST} = f(x^0)$. Let $k=0$

1. Find subgradient to f at x_k
   p^k

2. Update

$$x^{k+1} = \mathrm{Proj}_{S}(x^k - \alpha_k p^k)$$

3. $$f_{BEST} = \min(f_{BEST}, f(x^{k+1}))$$

4. Check termination criteria, go to 1.

NOTE

$\alpha_k = \alpha$ (constant)

$$\alpha_k : \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=0}^{\infty} \alpha_k = \infty$$

ex.

$$\alpha_k = \frac{1}{k+1}$$

Recap: Lagrange duality

$$f^* = \inf f(x) \quad \text{s.t. } g(x) \leq 0, \quad x \in X$$

$$\Rightarrow L(x, \mu) = f(x) + \mu^T g(x)$$

$$q(\mu) = \inf_{x \in X} [f(x) + \mu^T g(x)]$$

$$\boxed{\begin{aligned} q^* = \sup\ & q(\mu) \\ \text{s.t.}\ & \mu \geq 0 \end{aligned}}$$

- q is concave
- Maximizing a concave function over $\mu \geq 0$ means that this problem is convex.
- To evaluate $q(\mu)$ we need to solve

$$q(\mu) = \inf_{x \in X} \left[f(x) + \mu^T g(x) \right]$$

- Let $x(\mu)$ be a solution to this problem.

$\Rightarrow$ $\boxed{g(x(\mu)) \text{ is a subgradient to } q \text{ at } \mu}$

Proof Take $\bar{\mu} \geq 0$. Need to show

$$q(\bar{\mu}) \leq q(\mu) + g(x(\mu))^T (\bar{\mu} - \mu)$$

RHS

$$q(\mu) + g(x(\mu))^T (\bar{\mu} - \mu) =$$

$$= f(x(\mu)) + \cancel{\mu^T g(x(\mu))} +$$

$$+ g(x(\mu))^T (\bar{\mu} - \cancel{\mu}) =$$

$$= f(x(\mu)) + \bar{\mu}^T g(x(\mu)) \geq \underbrace{q(\bar{\mu})}_{\text{LHS}}$$

$\square$

Dual subgradient method

0. Initiate $\mu^0 \geq 0$
1. Solve the problem

$$q(\mu^k) = \inf_{x \in X} \left[f(x) + \mu^{k^T} g(x) \right]$$

2. Let solution be x^k.
   Update

$$\mu^{k+1} = \left[\mu^k + \alpha_k g(x^k) \right]$$

3. Check termination, go to 1.

Monday
14 october
8^{00}

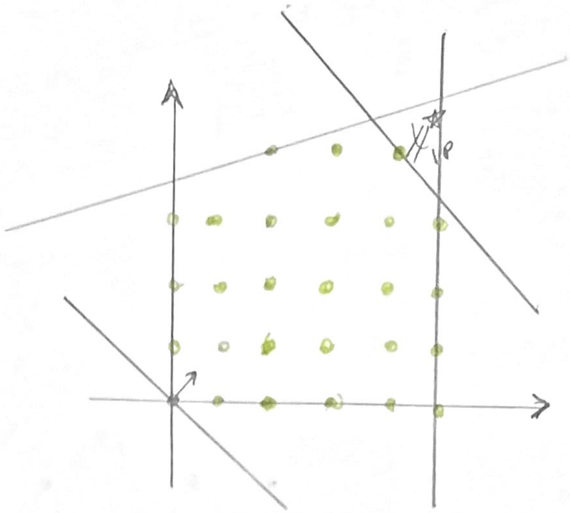
LECTURE

- Integer linear optimization

MILP - mixed integer linear programming.

Often binary, not whole $\mathbb{Z}$

Could use nonlinear constraints in many cases, but linear is easier.



- Sudoku problem

9^{00}

Branch and bound method

- Divide feasible set F into $F_1, \dots, F_k$

  Instead of solving

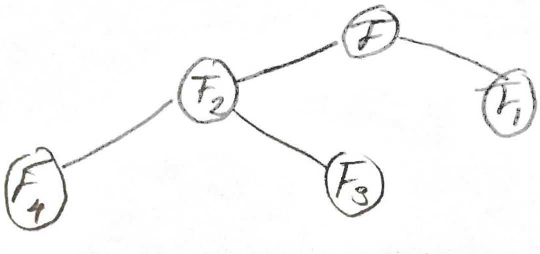
$$\min_{x} \; c^T x \quad \text{s.t.} \quad x \in F$$

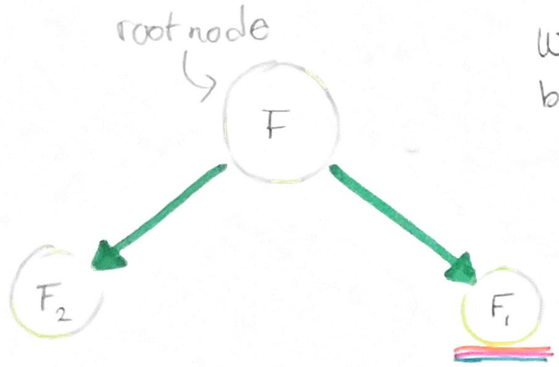
  solve

$$\min_{x} \; c^T x \quad \text{s.t.} \quad x \in F_i$$

  $\forall i$.

- May need to recursively divide F_i $i = 1, \dots, k$. This is branching.





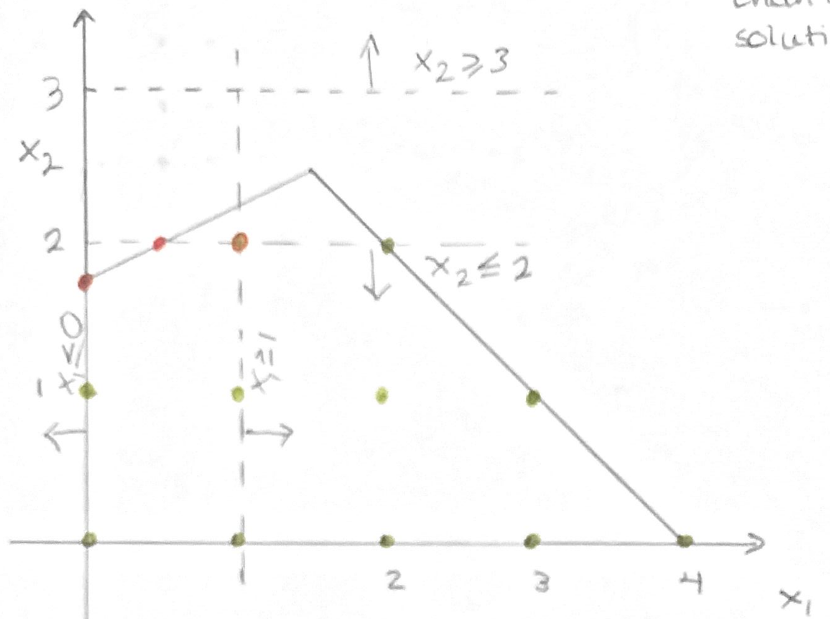
When can we stop the branching in a node?

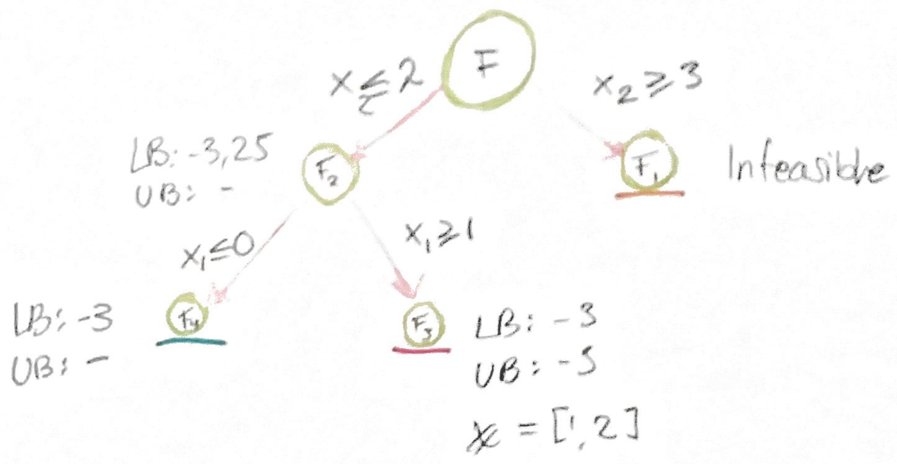
A) Infeasible.

B) Managed to solve the subproblem.

C) If we find a lower bound that is larger than some found solution.

ex.



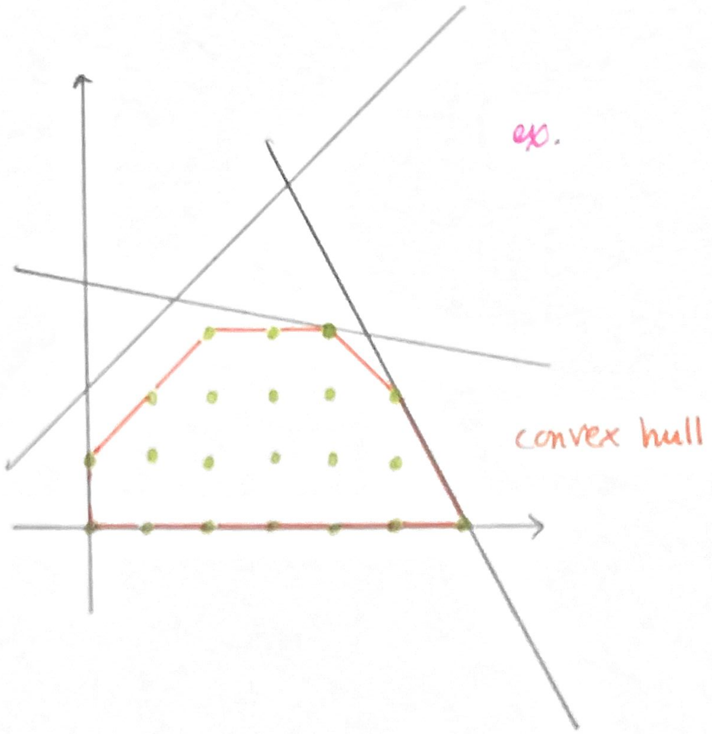


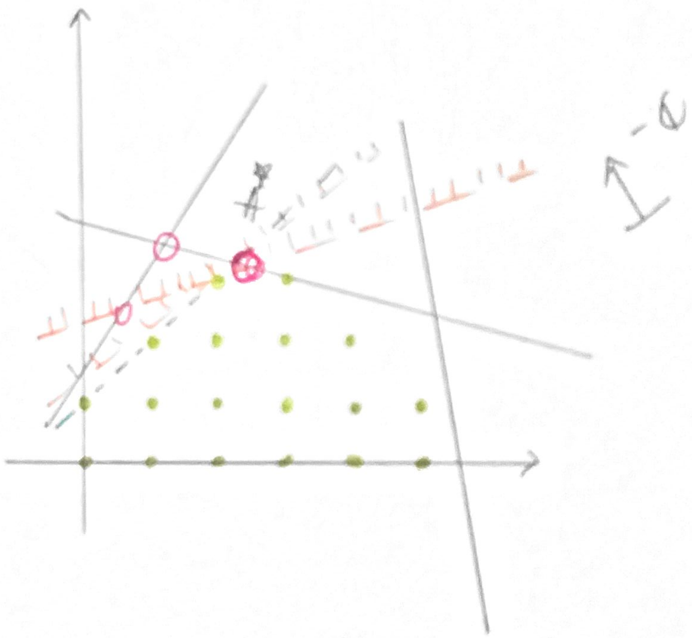
$UB_{BEST} = -3$

Cutting plane algorithm

- Add cuts

ex.





Tuesday
October 22th
15^{15}

SUMMARY

L2: (CONVEXITY)

convex sets

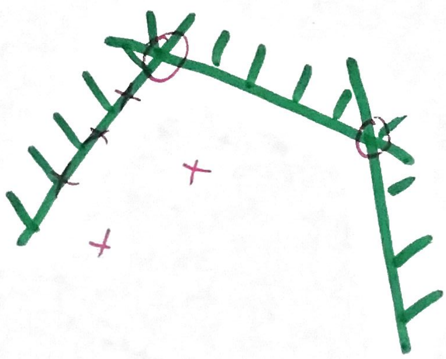
- $S \subseteq \mathbb{R}^n$ is convex if

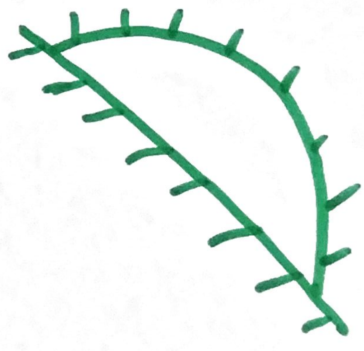
$$\left.\begin{array}{r} x^1, x^2 \in S \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow \lambda x^1 + (1-\lambda) x^2 \in S$$

- Intersection of convex sets is a convex set.
- If $S = \{ x \in \mathbb{R}^n \mid g_i(x) \leq 0 \}$ then if all g_i are convex functions we know that S is convex.

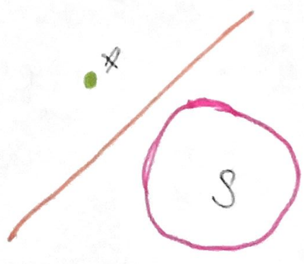
not the other way around.

- Polytope = convex hull of finitely many points.
- Polyhedron = intersection of finitely many half-spaces.
- An extreme point to a convex set is a point which can not be expressed as a convex combination of two other points.





- Separation theorem: Either a point lies in a convex set or it can be separated from the set by a half-space (hyperplane).



- Farkas' lemma: $A \in \mathbb{R}^{m \times n}$, $\mathbb{b} \in \mathbb{R}^m$

$$\begin{cases} A\mathbb{x} = \mathbb{b} \\ \mathbb{x} \geq \mathbb{0} \end{cases} \qquad \begin{cases} A^T \mathbb{y} \geq \mathbb{0} \\ \mathbb{b}^T \mathbb{y} < 0 \end{cases}$$

$\Rightarrow$ exactly one system is solvable.

convex functions

- f is convex if

$$f(\lambda \mathbb{x}^1 + (1-\lambda)\mathbb{x}^2) \leq \lambda f(\mathbb{x}^1) + (1-\lambda) f(\mathbb{x}^2)$$

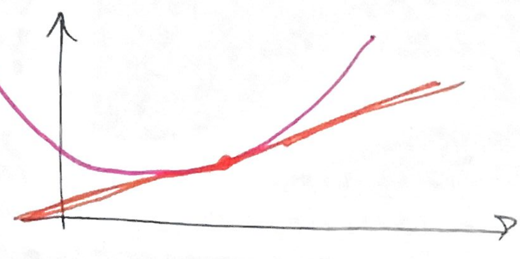
- The sum of convex functions is a convex function.

not the other way around. Instead: use counter example.

- If f is convex, $-f$ is concave.

- If $f \in C^1$

f is convex $\Leftrightarrow f(\mathbb{x}) \geq f(\bar{\mathbb{x}}) + \nabla f(\bar{\mathbb{x}})^T (\mathbb{x} - \bar{\mathbb{x}})$



- If $f \in C^2$

f is convex $\Leftrightarrow \nabla^2 f(\mathbb{x}) \succcurlyeq 0 \quad \forall \mathbb{x}$

convex problem

- (P) $\min f(x)$
  s.t. $x \in S$

(P) is convex if f is convex on S, and S is convex.

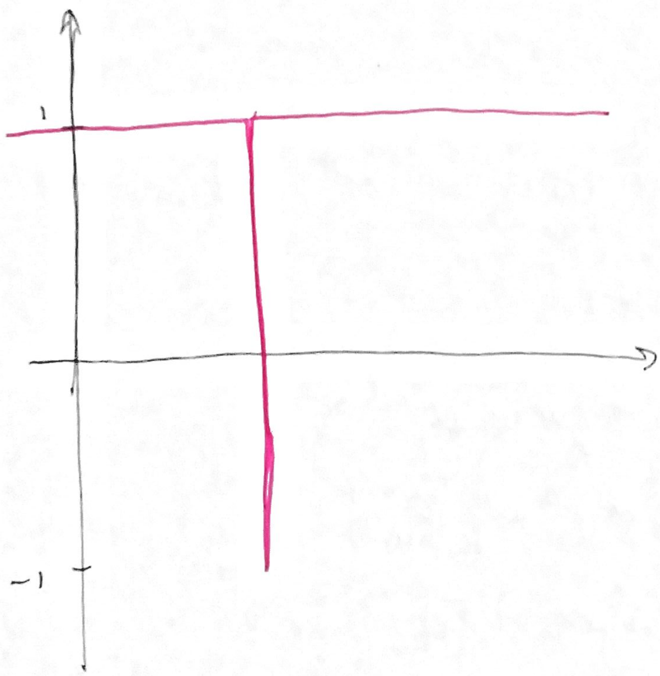
L3 (OPTIMALITY CONDITIONS)

$$\boxed{\begin{array}{l}\min f(x) \\ \text{s.t. } x \in S\end{array}}$$

- Global minima
- Local minima
- <u>Fundamental thm. of opt.</u>: If (P) is convex, then

$$x^* \text{ Local min} \Rightarrow x^* \text{ global min.}$$

ex.



- Weierstrass' thm.

unconstrained opt. $S = \mathbb{R}^n$

- x^* local min $\Longrightarrow$ $\nabla f(x^*) = 0$ $\quad (f \in C^1)$
- $f \in C^2$, x^* local min $\Longrightarrow$ $\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \succeq 0 \end{cases}$

- If f is convex, x^* global min $\Longleftrightarrow$ $\nabla f(x^*) = 0$

constrained opt. $S \subseteq \mathbb{R}^n$

- x^* local min $\Longrightarrow$ "it should not be possible to find a feasible descent direction at x^*"

- x^* local min $\Longrightarrow$ $\nabla f(x^*)^T p \geq 0$ $\forall$ feasible dir. p

- Suppose S convex:

  x^* local min $\Rightarrow$ x^* stationary

24 (UNCONST. OPT.)

- Line search type alg.

  $$x_{k+1} = x_k + \alpha_k p_k$$

- steepest descent

  $$p_k = -\nabla f(x_k)$$

- Newton

$$\nabla^2 f(x_k)\,\mathbb{P}_k = -\nabla f(x_k)$$

- Levenberg-Marquardt

$$(\nabla^2 f(x_k) + \gamma I)\,\mathbb{P}_k = -\nabla f(x_k)$$

- Step length using armijo's rule

25/26 (KKT)

- Tangent cone $T_S(x)$: "all directions from which a feasible sequence can converge to x".

- Gradient cone $\mathring{F}(x)$

$$\mathring{F}(x) = \{\mathbb{P} \mid \nabla f(x)^T \mathbb{P} < 0\}$$

- Geometric opt. cond.

$$x^* \text{ local min} \Rightarrow T_S(x^*) \cap \mathring{F}(x^*) = \emptyset$$

KKT

$$S = \{x \mid g_i(x) \leq 0\}$$

- Constraint gradient cone

$$G(x) = \{\mathbb{P} \mid \nabla g_i(x)^T \mathbb{P} \leq 0 \;\; \forall i \in \mathcal{I}(x)\}$$

- $G(x) \supseteq T_S(x)$

- Abadie's CQ : $T_S(x) = G(x)$

1615

$\downarrow$

- $\Rightarrow$ x^* local min $\Rightarrow$ $\underbrace{G(x) \cap \mathring{F}(x) = \emptyset}$

linear inequality system that should be unsolvable

$\Downarrow$ FARKAS

Another system is solvable.

$$\left[\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) &= 0 \\ \mu_i \, g_i(x^*) &= 0 \\ \mu_i &\geq 0 \end{aligned} \right]$$

- <u>Nessecary opt. cond.:</u>

If Abadie's holds at x^* $\Big\}$ $\Rightarrow$ x^* KKT
x^* local min

- <u>Sufficient opt. cond.:</u>

x^* KKT point $\Big\}$ $\Rightarrow$ x^* global opt.
(P) convex

L7 (LAGRANGE DUALITY)

- Relaxation problems

$$\min f_R(x), \qquad f_R \leq f$$
$$\text{s.t. } x \in S_R. \qquad S_R \supseteq S$$

- "IDEA": Some constraints are complicating. Lift them to the objective function with a "penalty"

- $$q(\mu) = \min_{\text{s.t. } x \in X} f(x) + \mu^T g(x)$$

- Weak duality: For any feasible x and $\mu \geq 0$

  $q(\mu) \leq f(x)$

- Strong duality: If f is convex, X convex, g convex, and inner point exists

  $\Rightarrow q^* = f^*$ (wrong in the slides here)

  $$q^* = \max_{\mu \geq 0} q(\mu)$$

- $q(\mu)$ is concave.

28/29 (LINEAR PROGRAMMING)

- Minimize linear function over polyhedron.
- $-\nabla f(x) = -c$ constant
- Solution exists in extreme points
- Standard form

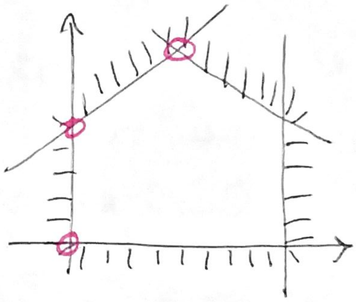
$$\begin{aligned} \min \ & c^T x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

- Extreme points $\Longleftrightarrow$ BFS

$$A = [\underbrace{B}_{\mathbb{R}^{m \times m}}, \underbrace{N}_{\mathbb{R}^{m \times (n-m)}}] \ : \ B^{-1}b \geq 0$$

remember to restructure the elements in x in the same way as in A.

- simplex. PLEASE DRAW! (on exam, in 2D) First, then compute.



- Phase I

210 (LP DUALITY)

(P) $\min c^T x$ s.t. $Ax = b$, $x \geq 0$

(D) $\max b^T y$ s.t. $A^T y \leq c$, $y \in \mathbb{R}^m$

- Dual program to every LP
- Translation table
- Every constraint in primal corresponds to variable in dual, and vice versa.
- <u>Weak duality</u>

$$\left.\begin{array}{l} x \text{ feasible in (P)} \\ y \text{ feasible in (D)} \end{array}\right\} \Longrightarrow c^T x \geq b^T y$$

- Strong duality

  If one problem has opt. sol $\Rightarrow c^T x^* = b^T y^*$

- Complementary slackness: Either the slack or the dual variable to a constraint is zero.

L11 (CONVEX OPT.)

(f convex)

- A subgradient p:

  $$f(x) \geq f(\bar{x}) + p^T(x - \bar{x}) \quad \forall x, \bar{x}$$

- Negative subgradients not always descent directions.
- But $x_{k+1} = x_k - \alpha_k p_k$ still works.

L12

- the methods from this lecture are not to be included in the exam. (But the modelling are).

2.13 (FEASIBLE DIRECTION METHODS)

(S polyhedron)

- Frank-Wolfe
- Simplicial decomposition
- Gradient projection. (even simpler sets than polyhedrons)

2.14 (PENALTY METHODS)

- Replace constrained problem with sequence of unconstrained problems.
- Exterior: Add penalty for being infeasible. Increase penalty.
- Interior: Add penalty for being close to constraints. Decrease penalty.
- SQP: Solve problem by a sequence of quadratic programs (QP).

Tip: check if point is feasible in computing alg!

Assignment 1:

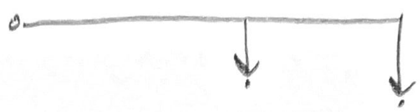


Table 4 $\Rightarrow$

Hint on assignments first task:

Modelling network flows:

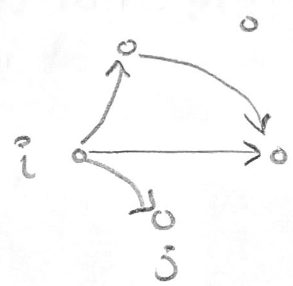
Let N be set of nodes.

Let $F(i)$ be destinations of node $i \in N$

$\beta(i)$ be the origins of $i \in N$

Flow balance

$$\sum_{j \in F(i)} f_{ij} - \sum_{j \in \beta(i)} f_{ij} = b_i$$



If

$b_i > 0 \rightarrow$ source

$b_i < 0 \rightarrow$ sink

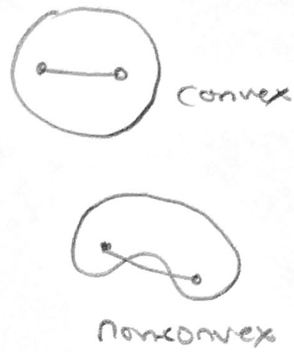
Måndag
9 september
10^{00}

PROBLEM SOLVING SESSION 2

CONVEXITY

DEF A set S convex if

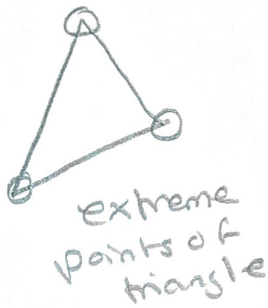
$$\left.\begin{array}{l} x^1, x^2 \in S \\ \lambda \in (0,1) \end{array}\right\} \Rightarrow \lambda x^1 + (1-\lambda) x^2 \in S$$



DEF A point $p \in S$ is called extreme if

$$\forall x^1, x^2 \in S, \lambda \in (0,1) : P = \lambda x^1 + (1-\lambda) x^2 \Rightarrow$$
$$x^1 = x^2 = P$$

THM For a polyhedron $P_1 (\{Ax \leq b\})$ we have $\hat{x} \in P$ is an extreme point iff $\tilde{A}$, in the equality subsystem of $A\tilde{x} \leq b$, has full rank.



3.4 Consider

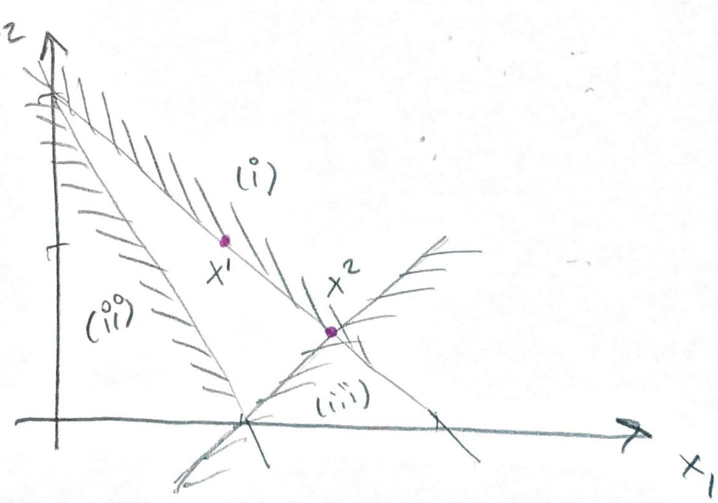
$$P = \begin{cases} x_1 + x_2 \leq 2 & (i) \\ 2x_1 + x_2 \geq 2 & (ii) \\ x_1 - x_2 \leq 1 & (iii) \end{cases}$$

Are

$$x^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix}$$

extreme points?

<u>Draw</u>!



x^{i}: only (i) yields equality

$\tilde{A} = [1 \ 1]$, rank $1 \neq 2$ $\Rightarrow$ x^1 is not ep.

x^{ii}: i & iii yields equality:

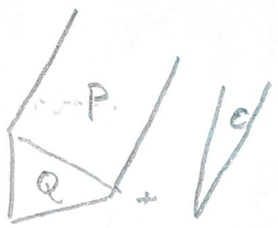
$$\tilde{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Gaussian elimination reviels $\tilde{A}$ has full rank (2).

THM Let

$C := \{Ax \leq 0\}$,

$Q = \text{conv}(\text{ext } P)$ $\Longrightarrow$ $\underset{\text{polyhedron}}{P} = Q + C$



8.6 Let

$$P := \{x \in \mathbb{R}^2 \mid -2x_1 + x_2 \leq 1;\ x_1 - x_2 \leq 1;\ -x_1 - x_2 \leq -1\}$$

$$C := \{x \in \mathbb{R}^2 \mid -2x_1 + x_2 \leq 0;\ x_1 - x_2 \leq 0;\ -x_1 - x_2 \leq 0\}$$

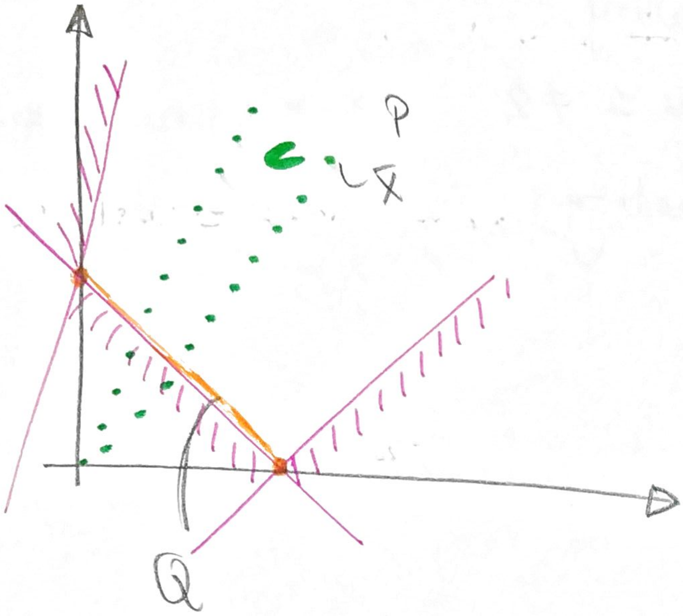
$$Q = \text{conv}(\text{ext } P)$$

show that

$$\bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

can be written as

$$\bar{x} = q + c, \quad q \in Q, \ c \in C.$$



$$\bar{x} = \underbrace{\sum_i \lambda_i q_i}_{q \in Q} + c, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0$$

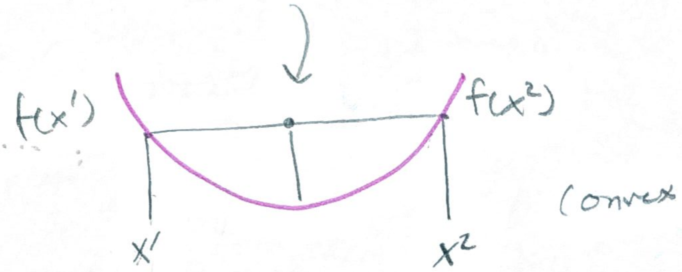
$$\iff \begin{pmatrix}1\\1\end{pmatrix} = \lambda \begin{pmatrix}1\\0\end{pmatrix} + (1-\lambda)\begin{pmatrix}0\\1\end{pmatrix} + c,$$

pick $c = \mu \begin{pmatrix}1\\1\end{pmatrix}$, $\mu \geq 0$

$$\Rightarrow \lambda = 1/2, \quad \mu = 1/2$$

DEF A function f is convex on S iff

$$\left.\begin{matrix} x^1, x^2 \in S \\ \lambda \in (0,1) \end{matrix}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2)$$



PROP. f_i convex $\Rightarrow \sum_i f_i$ convex

PROP. $f \in C^1$ on an open set $S \subseteq \mathbb{R}^n$. Then

f is convex $\Leftrightarrow \nabla^2 f(x)$ is positive semidefinite $\forall x \in S$

~~PRO~~

3,9 Determine if

$$f(x) = 2x_1^2 - 3x_1x_2 + 5x_2^2 - 2x_1 + 6x_2$$

is convex.

11^{00}

Method 1:

$$f(x) = x_1^2 + (x_1 - \tfrac{3}{2}x_2)^2 + \tfrac{11}{4}x_2^2 - 2x_1 + 6x_1 \Rightarrow$$

Sum of convex is convex.

Method 2:

$$\nabla f(x) = \begin{bmatrix} 4x_1 - 3x_2 - 2 \\ -3x_1 + 10x_2 \end{bmatrix},$$

$$\nabla^2 f(x) = \begin{bmatrix} 4 & -3 \\ -3 & 10 \end{bmatrix}$$

Our hessian is constant, nice!

our goal is to find the roots of

$$0 = \begin{vmatrix} 4-\lambda & -3 \\ -3 & 10-\lambda \end{vmatrix} = (4-\lambda)(10-\lambda) - 9 \Rightarrow$$

$$\lambda = 7 \pm \sqrt{18} > 0$$

$\nabla^2 f$ is positive definite $\Rightarrow f$ is strictly convex.

3.11 For $a>0$, which of the following functions are convex?

a) $f(x) = \ln x, \quad x>0$

$f'(x) = \frac{1}{x}$

$f''(x) = -\frac{1}{x^2} < 0 \longrightarrow$ strictly concave

d) $f(x) = \ln(1+e^{ax})$

$f'(x) = \frac{ae^{ax}}{1+e^{ax}},$

$f''(x) = a^2 e^{ax} \frac{(1-e^{ax}) - a^2 e^{2ax}}{(1+e^{ax})^2} = \frac{a^2 e^{ax}}{(1+e^{ax})^2} > 0 \Longrightarrow$

f is strictly convex.

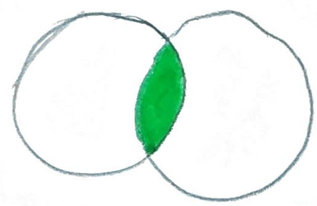
e) $f(x) = e^{ax},$

$f'(x) = ae^{ax}$

$f''(x) = a^2 e^{ax} > 0 \longrightarrow$ strictly convex

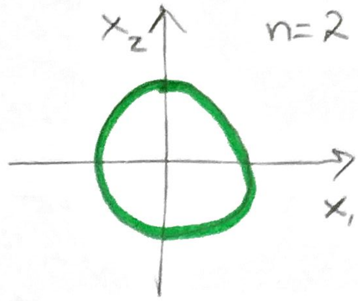
prop S_i convex set $\Longrightarrow \bigcap_i S_i$ convex

prop $\left.\begin{matrix} g(x) \text{ convex} \\ b \in \mathbb{R} \end{matrix}\right\} \Longrightarrow \{x \in \mathbb{R}^n \mid g(x) \leq b\}$ convex.



3.15 Are the sets convex or not?

c) $\{ x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i^2 = 1 \}$



Draw figure to know which path where going; proving for or against.

not a convex set. How can we show this? Look for counter example.

Not convex since

$$\frac{1}{2}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} + \frac{1}{2}\begin{pmatrix}-1\\0\\\vdots\\0\end{pmatrix} = \mathbb{0} \notin S$$

$\in S \qquad \in S$

e) $S: \{ x \in \mathbb{R}^2 \mid x_1 - x_2^2 \geq 1;\ x_1^3 + x_2^2 \leq 10;\ 2x_1 + x_2 \leq 8;\ x_1 \geq 1;\ x_2 \geq 0 \}$

$S_1 := \{ x \in \mathbb{R}^2 \mid -x_1 + x_2^2 \leq -1 \}$ convex (see prop)

$S_2 := \{ x \in \mathbb{R}^2 \mid x_1^3 + x_2^2 \leq 10,\ x_1 \geq 1 \}$ $\frac{\partial}{\partial x}(x^3) = 3x^2$, $\frac{\partial^2}{\partial x^2}(x^3) = 6x > 0$ if $x > 0$

$S_3 := \{ x \in \mathbb{R}^2 \mid 2x_1 + x_2 \leq 8;\ x_2 \geq 0 \}$ polyhedron $\Rightarrow$ convex

$S := S_1 \cap S_2 \cap S_3$ is convex.

Monday
16 september
10^{00}

PROBLEM SOLVING SESSION 4

Unconstrained optimization alg.

Update: $x^{k+1} = x^k + \alpha p^k$

search direction

steplength

p^k at x^k is

- ascent dir. $\nabla f(x^k) p^k > 0$
- descent dir. $\nabla f(x^k) p^k < 0$

Steepest descent

$$p^k = -\nabla f(x^k)$$

exact line search: $\alpha_k = \underset{\alpha \geq 0}{\operatorname{argmin}} \; f(x^k + \alpha p^k)$

11.5

$$\min_{x \in \mathbb{R}^2} f(x) := (2x_1^2 - x_2)^2 + 3x_1^2 - x_2$$

a) Do one iterate of steepest descent from

$$x^0 = \begin{pmatrix} 1/2 \\ 5/4 \end{pmatrix},$$

with exact line search.

$$p^0: \nabla f(x) = \dots = \begin{pmatrix} 16x_1^3 - 8x_1x_2 + 6x_1 \\ -4x_1^2 + 2x_2 - 1 \end{pmatrix},$$

$$p^0 = -\nabla f(x^0) = \dots = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

$$\alpha^0: \min_{\alpha \geq 0} \underbrace{f(x^0 - \alpha \nabla f(x^0))}_{:= \varphi(\alpha)} : \varphi'(\alpha) = 0$$

$$\varphi'(\alpha) = \nabla f(x^0 - \alpha \nabla f(x^0))^T (-\nabla f(x^0))$$

$$= 0 - \tfrac{1}{2} \nabla f_2\left(\begin{pmatrix}1/2\\5/4\end{pmatrix} - \alpha\begin{pmatrix}0\\1/2\end{pmatrix}\right) = \ldots = \frac{\alpha}{2} - \frac{1}{4} = 0 \Rightarrow$$

$$\alpha^* = \frac{1}{2}$$

(looking at $\varphi'(\alpha)$ we see that α^* is the minimizer)

$$x' = \begin{pmatrix}1/2\\5/4\end{pmatrix} - \frac{1}{2}\begin{pmatrix}0\\2\end{pmatrix} = \begin{pmatrix}1/2\\1\end{pmatrix}$$

b) is the function convex at x'?

$\nabla^2 f(x')$ pos semi. def.?

$$\nabla^2 f \begin{pmatrix} 48x_1^2 - 8x_2 + 6, & -8x_1 \\ -8x_1 & , 2 \end{pmatrix}$$

$$\nabla^2 f(x') = \begin{pmatrix} 10 & -4 \\ -4 & 2 \end{pmatrix}, \Rightarrow \ldots \Rightarrow$$

$\lambda = 6 \pm \sqrt{32} > 0 \rightarrow f$ is convex at x'

c) Will st coherce to glob. opt?

solve $\nabla f(x) = 0 \Rightarrow \ldots \Rightarrow \begin{pmatrix}0\\1/2\end{pmatrix}$ is only stat. point.

check that $f(x) \xrightarrow[\|x\| \to \infty]{} \infty \Rightarrow$ Wierstrass holds $\rightarrow$

glob opt exists $\rightarrow$ S.T will converge.

check $f(x^1)-f(x^0) \leq \mu \alpha \nabla f(x^0)^T p^0$;

$$-\frac{1040}{81} \leq \frac{1}{10}\,\frac{64}{3}, \text{ holds} \rightarrow x^1 = \begin{pmatrix} 4/3 \\ 2/3 \end{pmatrix}$$

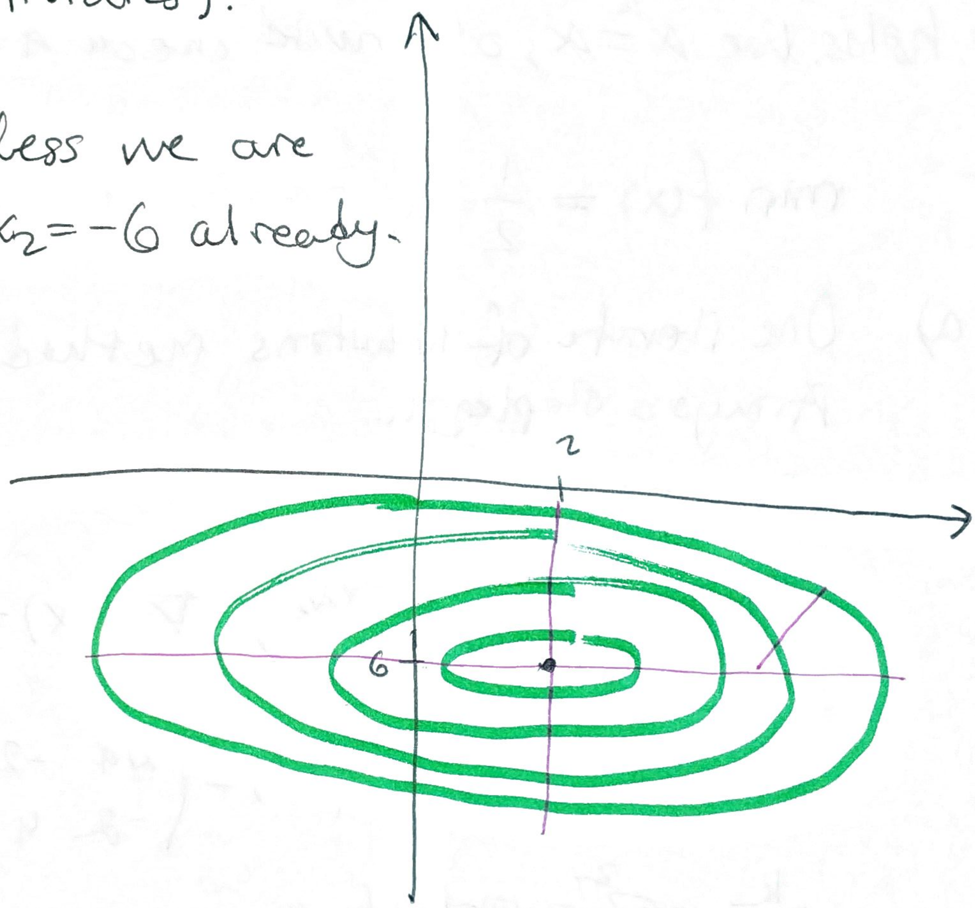
b) What $\mu \in (0,1)$ makes $\alpha = 1$ acceptable?

Solve

$$-\frac{1040}{81} \leq \mu \frac{64}{3} \Leftrightarrow \mu \leq \frac{64}{108}$$

11.4 Will S.D. reach opt. sol.? (in finite no. iterates).

no, not unless we are at $x_1 = 2$, $x_2 = -6$ already.



Newtons method:

$$p^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

Armijo steplength

$\mu \in (0,1)$, $\alpha = 1$.

If $\underbrace{f(x^k+\alpha p^k) - f(x^k)}_{\text{observed impr.}} \leq \mu \alpha \underbrace{\nabla f(x^k)^T p^k}_{\text{expected impr.}}$

holds. Use $\alpha^k = \alpha$, otherwise check $\alpha = \frac{\alpha}{2}$.

11.7 $\min f(x) = \frac{1}{2}(x_1 - 2x_2)^2 + x_1^4$

a) One iterate of Newtons method with Armijo's steplength.

$x^0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $\mu = 1/10$

$\nabla f(x) = \begin{pmatrix} 4x_1^3 + x_1 - 2x_2 \\ -2x_1 + 4x_2 \end{pmatrix}$, $\nabla^2 f(x) = \begin{pmatrix} 2x_1^2+1, & -2 \\ -2, & 4 \end{pmatrix}$

$\nabla f(x^0) = \begin{pmatrix} 32 \\ 0 \end{pmatrix}$, $\nabla^2 f(x^0) = \begin{pmatrix} 49 & -2 \\ -2 & 4 \end{pmatrix}$,

$p^k = -\nabla^2 f(x^0)^{-1} \nabla f(x^0) \underset{\text{gaussian elim.}}{= \dots =} \begin{pmatrix} -2/3 \\ -1/3 \end{pmatrix}$

$\alpha = 1$: $x' = x^0 + \alpha p^0 = \begin{pmatrix} 4/3 \\ 2/3 \end{pmatrix}$, $f(x^0) = 16$,

$f(x') = \frac{256}{81}$, $\nabla f(x_0)^T p_0 = -64/3$

PROBLEM SOLVING SESSION 5

KKT conditions

$$(P)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i=1,\dots,m \\ & h_j(x) = 0, \quad j=1,\dots,l \end{array} \Big\} S$$

Def Tangent cone

$$T_S(x) := \{ p \in \mathbb{R}^n \mid \exists \{x^k\} \subset S, \{\lambda_k\} \subset (0,\infty); x_k \to x, \lambda_k(x_k - x) \to p \}$$

Also; active constraints and gradient cone.

Constraint CQ:

Def Abadie's.
Def LICQ.
Affine
Slater

Implication

KKT conditions: $x^* \in S$

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*) = 0 \quad \Big\}$$

$$\mu_i \geq 0 \quad \forall i \quad \Big\} \text{ Dual feasibility}$$

$$\mu_i g_i(x^*) = 0 \quad \forall i \text{ - complem. slackness}$$

$$\left(\begin{array}{l} g_i(x^*) \leq 0 \quad \forall i \\ h_j(x^*) = 0 \quad \forall j \end{array} \right) \Big\} \text{ primal feasibility}$$

Thm If Abadie's CQ holds at $x^* \in S$, then

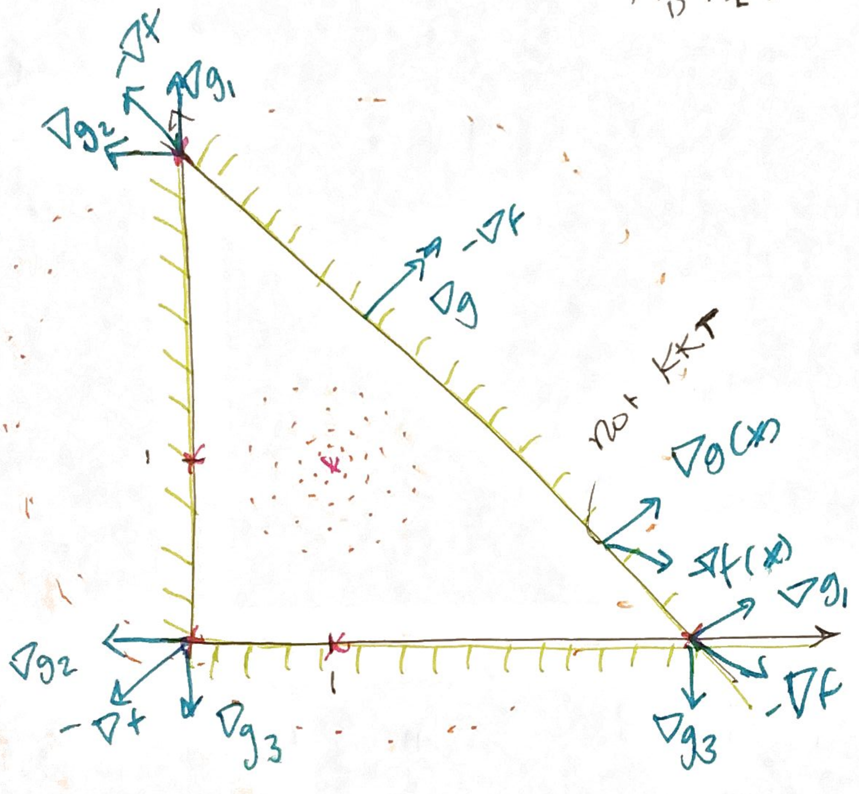
x^* loc. min. $\to$ x^* KKT point.

Thm If (P) convex, then

x^* KKT point $\Rightarrow$ x^* glob. min.

5.2

$$\min \; -(x_1-1)^2-(x_2-1)^2$$
$$\text{s.t.} \quad x_1+x_2 \le 4$$
$$x_1, x_2 \ge 0$$



e) draw all KKT points

E3.3

a) nessesary? Affine CQ $\Rightarrow$ Abadies CQ $\rightarrow$ KKT nessecary
convex problem $\Rightarrow$ KKT sufficient.

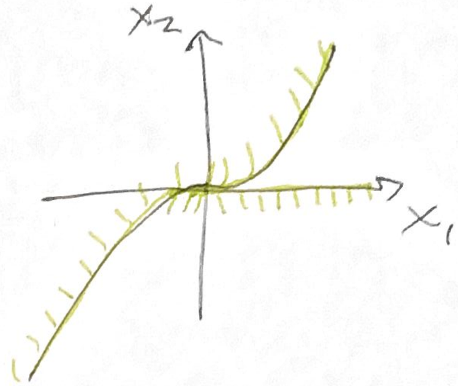
b) Same as a). Show it!

c) g_i convex.
$\bar{x}=\binom{1}{1} \Rightarrow g_i(\bar{x})<0 \;\forall i$ $\Big\} \Rightarrow$ Slater CQ $\Rightarrow$ KKT nessecary
f nonconvex $\Rightarrow$ no conclusion about sufficiency.

d) you can show

$T_S(x^*) = G(x^*)$

$\Rightarrow$ KKT not nessacary.



5.11 $\min f(x) = \sum_{j=1}^{n} x_j^2 / c_j$ $c_j > 0 \ \forall_j, \ D > 0$

s.t. $\sum_{j=1}^{n} x_j - D$

$x_j \geq 0, \ j = 1, \dots, n$

Find unique optimal solution

Affine CQ holds $\Rightarrow$ "x loc min $\Rightarrow$ KKT point" (necessity)

Sum of convex is convex $\Rightarrow$ f is convex
S polyhedron $\Rightarrow$ S convex. $\Bigg\} \Rightarrow$ problem is convex.

$\Rightarrow$ KKT suffient for global optimality.

$h(x) = \sum_{j=1}^{n} x_j - D, \ g_j(x) = -x_j, \ \nabla h = \mathbb{1}, \ \nabla g_i = -e_j$

$\nabla f_j(x) = \dfrac{2x_j}{c_j}$

$\dfrac{2x_j}{c_j} + \mu_i(-1) + \lambda = 0 \quad \Rightarrow \mu_j = \dfrac{2x_j}{c_j} + \lambda$

$\mu_i \geq 0 \quad \forall_j$

$\mu_j(-x_j) = 0 \quad \forall_j \Rightarrow \left(\dfrac{2x_j}{c_j} + \lambda\right)(-x_j) = 0$ $\nearrow x_j = 0$, $\searrow x_j = \dfrac{-\lambda c_j}{2}$

$\sum_{j=1}^{n} x_j = D$

$x_j \geq 0 \ \forall_j$

$$x = 0 \notin S \Rightarrow \exists j : \lambda = \frac{-2x_j}{c_j}, \quad x_j > 0$$

$$\mu_i \geq 0 : \left(\frac{2x_i}{c_i} - \frac{2x_j}{c_j}\right) \geq 0 \quad \frac{x_i}{c_i} \geq \frac{x_j}{c_j} > 0 \ \forall i$$

$$\Rightarrow x_i = \frac{-\lambda c_i}{2}$$

$$\sum_{j=1}^{n} x_j = D \Rightarrow \lambda = \frac{2D}{\sum_j c_j} \Rightarrow x_i = \frac{D c_i}{\sum_{j=1}^{n} c_j} \quad \forall i$$

$\square$

Friday
27 september
SO

PROBLEM SOLVING SESSION

5.2

$$\min -(x_1-1)^2-(x_2-1)^2$$

s.t. $g_1 = x_1+x_2 \leq 4$

$x_1, x_2 \geq 0$

e) Find all KKT points visually

$\nabla f = 0$

$\nabla f = \begin{bmatrix} -2(x_1-1) \\ -2(x_2-1) \end{bmatrix} \Rightarrow \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow$ KKT

f) Which points are global optimal

g_1 linear $\Rightarrow S$ is convex
(1,1) interior point
$\Big\} \Rightarrow$ slater CQ hold $\Rightarrow$ KKT nessesary

f is not convex $\Rightarrow$ KKT is not sufficient.

Check the objective function value of all KKT points.

at $(0,4)^T, (4,0)^T$ $f^* = -10$ are global opt.

5.11

$$\min f(x) = \sum_{j=1}^{n} \frac{x_j^2}{c_j}$$

st $\sum_{j=1}^{n} x_j = D$

$x_j \geq 0 \quad j=1,\dots,n$

where $c_j > 0 \quad j=1,\dots,n, \; D>0$

Find the global optimal solution.

Check KKT: all constraints affine $\Rightarrow$ Affine CQ holds $\longrightarrow$ KKT nessesary

Slater: does not hold.

KKT conditions:

$$\nabla f(x^*) + \sum_i \mu_i \cdot \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) = 0$$

$\mu_i \geq 0$ $\quad h_j(x^*) = 0$

$g_i(x^*) \leq 0$ $\quad \mu_i g_i(x^*) = 0$

$$\nabla f = \begin{pmatrix} \frac{2x_1}{c_1} \\ \vdots \\ \frac{2x_n}{c_n} \end{pmatrix} \quad \nabla g_i = \begin{pmatrix} 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{pmatrix} \text{ i row} \quad \nabla h = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

if all $x_i = 0$ contradict

at least $\exists x_j^* > 0$ $\mu_j = 0$

for ith row: we can get: $\frac{2x_j^*}{c_j} - \mu_i + \lambda = 0$

$$\lambda = -\frac{2x_j^*}{c_j} < 0$$

Suppose $\exists x_i^* = 0$, then for ith row:

$\frac{2x_i^*}{c_i} - \mu_i + \lambda = 0$ $\quad \mu_i = \lambda = -\frac{2x_j^*}{c_j} < 0$

0 "

Contradict!

all $x_i^* > 0 \Rightarrow \mu_i = 0$ for all i

$$\frac{2x_i^*}{c_i} + \lambda = 0 \qquad \frac{2x_i^*}{c_i} = \frac{2x_j^*}{c_j} = -\lambda$$

$$x_i = \frac{-\lambda c_i}{2} \qquad \frac{-\lambda(c_1 + \dots + c_n)}{2} = D$$

$$\lambda = \frac{-2D}{\sum_{j=1}^{n} c_j} \Rightarrow x_i^* = \frac{c_i D}{\sum_{j=1}^{n} c_j}$$

$$\left.\begin{array}{l} \text{KKT nessesary} \\ \text{only one KKT point} \end{array}\right\} \Rightarrow \text{The KKT point is global opt.}$$

Thm Global optimality conditions

(x^*, μ^*) is a pair of primal optimal solution Lagrange multiplier vector iff

$$\left\{\begin{array}{ll} \mu^* \geq 0^n & \text{(Dual feasability)} \\ x^* \in \arg\min_{x \in X} L(x, \mu^*) & \text{(Lagrange optimality)} \\ x^* \in X \quad g(x^*) \leq 0^m & \text{(Primal feasability)} \\ \mu_i^* g_i(x^*) = 0 & \text{(complementary slackness)} \end{array}\right.$$

6.10

min $-2x_1 + x_2$

s.t. $x_1 + x_2 \leq 3$ (1)

$$x \in \left\{ \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}4\\0\end{pmatrix}, \begin{pmatrix}0\\4\end{pmatrix}, \begin{pmatrix}4\\4\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix} \right\}$$

Lagrange relax (1)

Which optimality condition can't be fullfilled?

Calculate the optimality gap $\Gamma = f^* - q^*$

$$L(x,\mu) = -2x_1 + x_2 + \mu(x_1 + x_2 - 3)$$

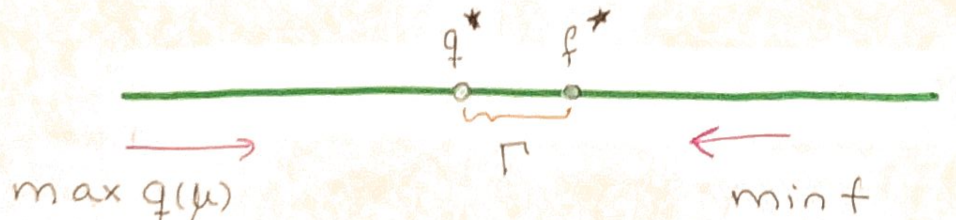
$$= -3\mu + x_1(\mu-2) + x_2(\mu+1)$$

$$q(\mu) = \min_{x \in X} -3\mu + x_1\underbrace{(\mu-2)} + x_2\underbrace{(\mu+1)} \qquad \mu \in [0,+\infty]$$

look at the sign.

$\mu \in [0,2]$ max x_1, min $x_2 \Rightarrow \binom{4}{0} \Rightarrow q(\mu) = \mu - 8$

$\mu \in (2,+\infty)$ min x_1, min $x_2 \Rightarrow \binom{0}{0} \Rightarrow q(\mu) = -3\mu$



$\mu \in [0,2]$ $\quad q^* = -6$

$\mu^* = 2$

$\mu \in (2, +\infty)$ $\quad q^* = -6$

max $q(\mu) = -6$ $\quad \mu = \mu^*$ on $\binom{0}{0}\binom{4}{0}$

(0,0) complementary slackness is not fullfilled.

(4,0) doesn't fullfill feas.

complementary slackness is not fullfilled.

plus in all feasible points and try.

$x^* = (2,1)^T$ $f^* = -3$, $\Gamma = f^* - q^* = 3$

5.11

correction:

$$\left.\begin{array}{l}\text{only one KKT point}\\ \text{KKT necessary}\\ \exists \text{ global opt}\end{array}\right\} \Rightarrow x^* \text{ is opt.}$$

6.4

$$\min f(x) = x_1^2 + 2x_2^2$$

s.t. $x_1 + x_2 \geq 2$ $\Rightarrow$ (1) $g_1 = -x_1 - x_2 + 2 \leq 0$

$x_1^2 + x_2^2 \leq 5$

Find optimal solution through lagrange duality.

We relax (1).

$$L(x, \mu) = x_1^2 + 2x_2^2 + \mu(-x_1 - x_2 + 2)$$

$$q(\mu) = 2\mu + \min(x_1^2 - \mu x_1) + \min(2x_2^2 - \mu x_2)$$

- $\min x_1^2 - \mu x_1$ $\nabla = 2x_1 - \mu$ $x_1^* = \frac{\mu}{2}$ $\nabla^2 = 2 > 0$
- $x_2^* = \frac{\mu}{4}$

$$q(\mu) = 2\mu - \frac{\mu^2}{4} - \frac{\mu^2}{8}$$

$$\max q(\mu) = -\frac{3}{8}\mu^2 + 2\mu$$

$\nabla q = -\frac{3}{4}\mu + 2 = 0$ $\mu^* = \frac{8}{3}$ $\underbrace{x_1^* = \frac{4}{3},\ x_2^* = \frac{2}{3}}$

$\nabla^2 q = -\frac{3}{4} < 0$

$\left(\frac{4}{3}\right)^2 + \left(\frac{2}{3}\right) = \frac{20}{9} \leq 5$

$\mu^* \geq 0$ ✓

$\frac{4}{3} + \frac{2}{3} - 2 = 0$

$x^* \in \underset{x \in X}{\operatorname{argmin}} L(x, \mu^*)$ ✓

$x^* \in X,\ g(x^*) \le 0$ ✓

$\mu_i^* g_i(x^*) = 0$ ✓

x^* is optimal.

Check conditions for all the constraints. Except in the last case, here use only the relaxed constraints

E3.7 A problem with objective function f, Lagrange relax some constraints. $q(\mu)$.

a) A primal feasible x' $f(x') = 6$.
μ' is a positive Lagrange mul. $q(\mu') = -2$
what can you say about f^*

$$q(\mu) \le f^* \le f \qquad -2 \le f^* \le 6$$

b) x^2, μ^2 $f(x^2) = 3$ $q(\mu^2) = -4$

$$q(\mu) \le f \le f \qquad -4 \le f^* \le 3 \Rightarrow -2 \le f^* \le 3$$

c) $q^* = 3$ $q(\mu) \le f^*$ $f^* \ge 3 \Rightarrow f^* = 3$

Monday
30 september
10^{00}

PROBLEM SOLVING SESSION 7

- Linear programming (LP) $\Leftrightarrow$ optimization with affine objective and constraints.

Def. Standard form LP

$$\begin{aligned} \min\ & z = c^T x \\ \text{s.t}\ & Ax = b \\ & x \geq 0 \end{aligned}$$

thm All LPs can be transformed into the standard form.

ex.
$$\begin{aligned} x_1 + x_2 &\leq 5 \\ x_1, x_2 &\geq 0 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} x_1 + x_2 + s_1 &= 5 \\ x_1, x_2, s_1 &\geq 0 \end{aligned}$$

E4.2 Write on standard form

$$\begin{aligned} \max\ & 3x_1 - 6x_2 \\ \text{s.t.}\ & 10x_1 - 3x_2 = 5 \\ & -x_1 - 3x_2 \geq 7 \\ & x_2 \geq 5 \end{aligned}$$

(1) minimize $-f \Rightarrow \min\ -3x_1 + 6x_2$

(2) Only non-negative variables.

$$x_1 = \underbrace{x_1^+}_{\geq 0} - \underbrace{x_1^-}_{\geq 0}, \quad \tilde{x}_2 = x_2 - 5 \geq 0$$

(3) Make equality $-x_1 - 3x_2 - s = 7$
$s \geq 0$

$$\begin{aligned} \min\ & -3(x_1^+ - x_1^-) + 6(\tilde{x}_2 + 5) \\ \text{s.t.}\ & 10(x_1^+ - x_2^+) - 3(\tilde{x}_2 + 5) = 5 \\ & -(x_1^+ - x_1^-) - 3(\tilde{x}_2 + 5) - s = 7 \\ & s \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & -3x_1^+ + 3x_1^- + 6\tilde{x}_2 + 30 \\ \text{s.t.} \quad & 10x_1^+ - 10x_1^- - 3\tilde{x}_2 = 20 \\ & -x_1^+ + x_1^- - 3\tilde{x}_2 - s = 22 \\ & x_1^+, \dots, \quad , s \geq 0 \end{aligned}$$

min without constant

Remember to transform back into the original problem.

8.1 Formulate as linear problems:

a) $$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m |(Ax-b)_i|$$

s.t. $$\max_{i=1,\dots,n} |x_i| \leq 1$$

(1) $|Ax-b| \leq y \in \mathbb{R}^m \iff -y \leq Ax-b \leq y$

$$\min \sum_{i=1}^m |Ax-b|_i \iff \min \sum_{i=1}^m y_i$$
$$\text{s.t. } -y \leq Ax-b \leq y$$

Works because we want y to be small

(2) $$\{x \in \mathbb{R}^n \mid \max_{i=1,\dots,n} |x_i| \leq 1\} \stackrel{?}{=}$$

$$\{x \in \mathbb{R}^n \mid -1 \leq x_i \leq 1,\ i=1,\dots,n\}$$

If you are to be correct, show that these are subsets of one another.
Lets just state

$$-1 \leq x_i \leq 1 \iff |x_i| \leq 1 \iff \max_i |\lambda_i| \leq 1$$

$$\min \sum_{i=1}^m y_i$$
$$\text{s.t. } -y \leq Ax+b \leq y$$
$$-1 \leq x \leq 1$$

Generalization of E3.10

$$\min \; cx$$
$$\text{s.t. } Ax < b$$
$$x \geq 0$$

$$q(\mu) = \min_{x \geq 0} c^T x + \mu^T (Ax - b) = -\mu^T b + \min_{x \geq 0} (c^T + \mu^T A) x$$

$$\max_{\mu \geq 0} q(\mu) = \max \; -\mu^T b$$
$$\text{s.t. } c^T + \mu^T A \geq 0$$
$$\mu \geq 0$$

Friday
4 october
8^{00}

PROBLEM SOLVING SESSION 8

– Simplex method

Def. for $Ax=b$
$x \geq 0$,

$\tilde{x}$ is a basic feasible solution (BFS), if

(1) $A\tilde{x}=b$
$\tilde{x} \geq 0$

(2) $\{A_j \mid \tilde{x}_j \neq 0\}$

are linearly independent.

Note BFS $\Longleftrightarrow$ extreme point.

8.4 Suppose a linear program includes a free variable x. Show that if

$$\begin{cases} x = x^+ - x^- \\ x^+, x^- \geq 0 \end{cases}$$

then no BFS can include x^+ & x^- as nonzero.

Proof If a is the column of x, then a and $-a$ are the columns of x^+ and x^-.

So since a and $-a$ are linearly dependent, then both $x^+ > 0$ and $x^- > 0$ cannot hold for a BFS.

Idea of simplex method

- Move to adjecent BFS in a descent direction.

Let $A = (B, N)$, $x = (x_B, x_N)$, $B \in \mathbb{R}^{m\times m}$

x_N: $= 0$ now

$$Ax = Bx_B + Nx_N = b \Rightarrow$$

$$x_B = B^{-1}b - B^{-1}Nx_N$$

$$z = c^T x = c_B^T x_B + c_N^T x_N =$$

$$= c_B^T B^{-1} b + \underbrace{(c_N^T - c_B^T B^{-1} N)}_{:= \bar{c}_N^T} x_N$$

Reduced cost

0. $x = (x_B, x_N)$ is BFS, $A \equiv (B, N)$ ($BFS \Rightarrow B^{-1}b \geq 0$)

1. If $\bar{c}_N \geq 0 \Rightarrow x$ is optimal

   Otherwise: Pick $j: (\bar{c}_N)_j < 0$ (most negative)

   column N_j enter the basis. $(x_N)_j$ increase.

2. step size (remain-feasible):

$$x_B = \underset{\geq 0}{B^{-1}b} - B^{-1}N_j \underset{= \theta}{x_j} \geq 0$$

- If $B^{-1}N_j \leq 0 \Rightarrow$ unbounded. Stop.
- $i \in \underset{k | (B^{-1}N_j)_k > 0}{\operatorname{argmin}} \dfrac{(B^{-1}b)_k}{(B^{-1}N_j)_k}$, B_i leave the basis.

Go to 1.

Phase I problem

Find initial BFS

$$\min w = \mathbb{1}^T a$$
$$\text{s.t.} \quad Ax + Ia = b \geq 0$$
$$x, a \geq 0$$

$\begin{matrix} Ax = b \\ x \geq 0 \end{matrix}$ feasible solution iff $w^* = 0$.

Hence we start with $x_B = a$.

$$A = [\tilde{A}, \pm I]$$

Whatever you do: You need to start in a BFS

Make sure b is positive, if not, change sign.

9.1 Show that

$$(P) = \begin{cases} 3x_1 + 2x_2 - x_3 \leq -3 \\ -x_1 - x_2 + 2x_3 \leq -1 \\ x_1, x_2, x_3 \geq 0 \end{cases}$$

is inconsistent.

Phase I problem:

We should always add artificial variables

$$\min \ a_1 + a_2$$
$$\text{s.t.} \quad (1) \ -3x_1 - 2x_2 + x_3 - s_1 + a_1 = 3$$
$$(2) \quad x_1 + x_2 - x_3 - s_2 + a_2 = 1$$
$$x_1, \dots, a_2 \geq 0$$

$w^* = 0 \iff P$ feasible.

$(1) + (2) = 4 = \underbrace{-2x_1 - x_2 - x_3 - s_1 - s_2}_{\leq 0} + \underbrace{a_1 + a_2}_{= w} \leq w$

$\Rightarrow w^* \geq 4$

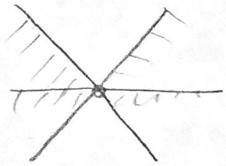
so (P) is inconsistent.

9^00

$\bar{c}_N > 0 \Rightarrow$ unique optimal solution

$\bar{c}_N \geq 0$ & $(\bar{B}^{-1} b) > 0 \Rightarrow$ unique opt.

$\exists_j (\bar{c}_N)_j = 0$



degenerate case

9.4 Solve

$$\min z = -x_1 + x_2$$

$$\text{s.t.} \quad -x_1 + 2x_2 \geq \tfrac{1}{2}$$

$$-2x_1 - 2x_2 \geq 1$$

$$x_2 \geq 0$$

with the simplex method.

(1) to std. form,

$$x_1 = x_1^+ - x_1^-, \; s_1 \geq 0, \; s_2 \geq 0$$

(2) Formulate Phase I.

$$\min \ \omega = a_1 + a_2$$

$$\text{s.t.} \quad -x_1^+ + x_1^- + 2x_2 - s_1 + a_1 = 1/2$$

$$-2x_1^+ + 2x_1^- - 2x_2 - s_2 + a_2 = 1$$

$$B = \begin{array}{cc} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ a_1 \ \ a_2 \end{array}, \quad N = \begin{array}{c} \begin{bmatrix} -1 & 1 & 2 & -1 & 0 \\ -2 & 2 & -2 & 0 & -1 \end{bmatrix} \\ x_1^+ \ \ x_1^- \ \ x_2 \ \ s_1 \ \ s_2 \end{array}$$

$$x_B = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \quad c_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad c_N^T = [0, 0, 0, 0, 0]$$

1. Recall $\bar{c}_N = c_N - c_B^T B^{-1} N$

$$\bar{c}_N^T = (0,0,0,0,0) - (1,1) B^{-1} N =$$

$$= 0 - (1,1) I \begin{bmatrix} -1 & 1 & 2 & -1 & 0 \\ -2 & 2 & -2 & 0 & -1 \end{bmatrix} =$$

$$= (3, \underline{\underline{-3}}, 0, 1, 1)$$

$$\min(\bar{c}_N) = -3 \Rightarrow x_1^- \text{ enter}$$

2. $B^{-1}N_2 = \binom{1}{2}$, Find $\max \theta : B^{-1}b - B^{-1}N_2 \theta \geq 0$

$$\theta = \min_{i \mid (B^{-1}N_2)_i > 0} \frac{(B^{-1}b)_i}{(B^{-1}N_2)_i} = \min\left\{ \frac{1/2}{1}, \frac{1}{2} \right\} = \frac{1}{2}$$

tied, but let a_1 leave.

$$B = \begin{array}{cc} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \\ x_1^- \; a_2 \end{array}, \quad N = \begin{array}{c} \begin{bmatrix} -1 & 2 & -1 & 0 & 1 \\ -2 & -2 & 0 & -1 & 0 \end{bmatrix} \\ x_1^+ \; x_2 \; s_1 \; s_2 \; a_1 \end{array},$$

$$x_B = B^{-1}b = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$$

$$c_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad c_N^T = (0,0,0,0,1)$$

1. $$\bar{c}_N^T = (0,0,0,0,1) - \underbrace{(0,1)\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}}_{(-2,1)} \begin{bmatrix} -1 & 2 & -1 & 0 & 1 \\ -2 & -2 & 0 & -1 & 0 \end{bmatrix} =$$

$$= (0, 6, \underline{\underline{-2}}, 1, 3) \Rightarrow s_1 \text{ enters the basis}$$

2. $$B^{-1}N_3 = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \begin{matrix} x_1^- \\ a_2 \end{matrix},$$

$$\theta = \min\left\{\frac{0}{2}\right\} = 0 \quad \Rightarrow \quad a_2 \text{ leaves the basis.}$$

we do not include the ones that are less than zero in our min ratio test.

$\Rightarrow x_B = (x_1^-, s_1)$ is a BFS.

(3) Phase II:

$$\begin{aligned} \min \quad & -x_1^+ + x_1^- + x_2 \\ \text{s.t.} \quad & -x_1^+ + x_1^- + 2x_2 - s_1 = 1/2 \\ & -2x_1^+ + 2x_1^- - 2x_2 - s_2 = 1 \\ & x_1^+, x_1^-, x_2, s_1, s_2 \geq 0 \end{aligned}$$

$$B = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} -1 & 2 & 0 \\ -2 & -2 & -1 \end{bmatrix},$$

$$\begin{array}{cc} x_1^- & s_1 \end{array} \qquad \begin{array}{ccc} x_1^+ & x_2 & s_2 \end{array}$$

$$x_B = B^{-1} b = \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$$

$$c_B^T = (1, 0), \quad c_N^T = (-1, 1, 0)$$

$$\bar{c}_N = c_N - c_B^T B^{-1} N = (-1, 1, 0) - \underbrace{(1, 0) \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix}}_{= (0, 1/2)} \begin{bmatrix} -1 & 2 & 0 \\ -2 & -2 & -1 \end{bmatrix} =$$

$$= (0, 2, 1/2) \geq 0 \Rightarrow$$

x_B is optimal.

$$\left.\begin{array}{l} x_1^- = 1/2 \\ s_1 = 0 \\ x_N = 0 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = -1/2 \\ x_2 = 0 \end{array}, \quad z^* = -\left(-\frac{1}{2}\right) + 0 = \frac{1}{2}$$

Multiple solutions to artificial problem, but unique to original.

Monday
7 October
10^{00}

PROBLEM SOLVING SESSION

PRIMAL (P)

$$\min \mathbb{c}^T \mathbb{x} \quad \text{s.t. } A\mathbb{x} \geq \mathbb{b}, \quad \mathbb{x} \geq \mathbb{0}$$

DUAL (D)

$$\max \mathbb{b}^T \mathbb{y} \quad \text{s.t. } A^T \mathbb{y} \leq \mathbb{c}, \quad \mathbb{y} \geq \mathbb{0}$$

Pos. variables $\Longleftrightarrow$ canonical constraints ($\leq$ in max, $\geq$ in min)

neg. variables $\Longleftrightarrow$ non-canonical constraints ($\geq$ in max, $\leq$ in min)

Free variables $\Longleftrightarrow$ equality constraints

E5.1 Formulate the dual to

(P)
$$\begin{aligned} \min\ & 3x_1 + 2x_2 \\ \text{s.t.}\ & x_1 + 2x_2 \leq 3 && (y_1) \\ & x_1 + x_2 \leq 10 && (y_2) \\ & 5x_1 - x_2 \geq 8 && (y_3) \\ & x_1 \geq 0 \\ & x_2 \leq 0 \end{aligned}$$

(D)
$$\begin{aligned} \max\ & 3y_1 + 10y_2 + 8y_3 \\ & y_1 + y_2 + 5y_3 \leq 3 \\ & 2y_1 + y_2 - y_3 \geq 2 \\ & y_1, y_2 \leq 0 \\ & y_3 \geq 0 \end{aligned}$$

A linear program is either:

- Feasible with optimal solution
- Infeasible
- Unbounded

Weak duality

$\left.\begin{array}{l} x \text{ feasible in (P)} \\ y \text{ feasible in (D)} \end{array}\right\} c^T x \geq b^T y$

Strong duality

x^* optimal in (P) $\iff$ x^* feasible in (P)
y^* optimal in (D) $\quad$ y^* feasible in (D)
$(x^*, y^*$ fullfill complementary slackness)

$\Rightarrow x_j(c_j - y^T A_j) = 0$
$y_i(A_i x - b_i) = 0$

E5.2

a) (P) is infeasible $\Rightarrow$ (D) infeasible or unbounded

b) (P) is feasible with optimal solution $\Rightarrow$
(D) is feasible with optimal solution

c) (P) is unbounded $\Rightarrow$ (D) infeasible

d) According to thm; an optimal primal-dual pair must satisfy
- Primal feasibility
- Dual feasibility
- Complementary slackness.

WHICH OF THESE CONDITIONS ARE SATISFIED DURING SIMPLEX?

In each iteration, we have a BFS

$$x = [x_B, x_N] \qquad y^T = c_B^T B^{-1}$$

$$A = [B, N]$$

- x is always feasible in simplex method
- x, y satisfies complementary slackness

$$x_j (c_j - y^T A_j) = 0 \quad \forall j$$

For x_N-variables, always ok.

For x_B-variables

$$c_B^T - y^T B = c_B^T - c_B^T B^{-1} B = 0$$

- y feasible?

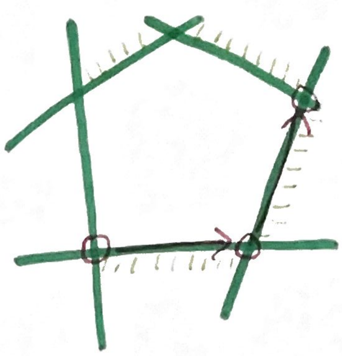
$$A^T y \leq c \; ? \qquad c^T - y^T A \geq 0$$

$$\Longleftrightarrow$$

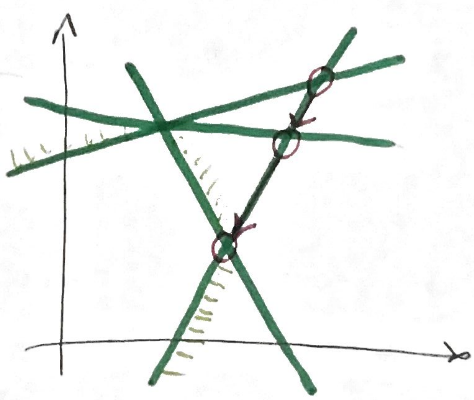
$$\begin{cases} c_B^T - c_B^T B^{-1} B = 0 \geq 0 \\ \tilde{c}_N = c_N^T - c_B^T B^{-1} N \geq 0 \end{cases}$$

Only feasible if x^* is optimal.

PRIMAL



DUAL (only conceptually)



10.13 Consider

$$\min \; x_1 - 18x_2 - c_3x_3 - c_4x_4$$

$$\text{s.t.} \quad x_1 + 2x_2 + 3x_3 + 4x_4 + s_1 = 3$$

$$-3x_1 + 4x_2 - 5x_3 - 6x_4 + s_2 = 1$$

$$x_1, x_2, x_3, x_4, s_1, s_2 \geq 0$$

For which c_3, c_4 is $x_B = [x_1, x_2]$ optimal?

Let

$x_B = [x_1, x_2]$ $\qquad B = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix}$

$x_N = [x_3, x_4, s_1, s_2]$ $\qquad N = \ldots$

$c_B^T = [1, -18]$ $\qquad c_N^T = [-c_3, -c_4, 0, 0]$

$$\Rightarrow \tilde{c}_N^T = c_N^T - c_B^T B^{-1} N = [5 - c_3, 8 - c_4, 5, 2]$$

$$\geq 0 \text{ if } \begin{cases} c_3 \leq 5 \\ c_4 \leq 8 \end{cases}$$

Friday
11 october
8:00

PROBLEM SOLVING SESSION

10.13

(LP) $\max z = -x_1 + 18x_2 + c_3x_3 + c_4x_4$

s.t $x_1 + 2x_2 + 3x_3 + 4x_4 \leq 3$

$-3x_1 + 4x_2 - 5x_3 - 6x_4 \leq 1$

$x_1, x_2, x_3, x_4 \geq 0$

Find values of c_3, c_4 such that the basic solution with $x_B = (x_1, x_2)$ is an opt BFS to the problem. SF:

$\min z = x_1 - 18x_2 - c_3x_3 - c_4x_4$

s.t. $x_1 + 2x_2 + 3x_3 + 4x_4 + s_1 = 3$

$-3x_1 + 4x_2 - 5x_3 - 6x_4 + s_2 = 4$

$x_B = (x_1, x_2)^T \quad x_N = (x_3, x_4, s_1, s_2)^T$

$B = \begin{pmatrix} 1 & 2 \\ -3 & 4 \end{pmatrix} \quad N = \begin{pmatrix} 3 & 4 & 1 & 0 \\ -5 & -6 & 0 & 1 \end{pmatrix}$

feasible?

$x_B = B^{-1}b = \frac{1}{10}\begin{pmatrix} 4 & -2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \geq 0$ ✓

opt?

$c_B^T = (1, -18) \quad c_N^T = (-c_3, -c_4, 0, 0)$

$\hat{c}_N^T = c_N^T - c_B^T B^{-1} N = \ldots = (5 - c_3, 8 - c_4, 5, 2)$

$$\tilde{C}_N^T \geq 0 \Rightarrow \left.\begin{matrix} 5 - C_3 \geq 0 \\ 8 - C_4 \geq 0 \end{matrix}\right\} \Rightarrow \begin{matrix} C_3 \leq 5 \\ C_4 \leq 8 \end{matrix}$$

Subgradient

Let f be a convex function. We say g is a subgradient at $x \in \mathbb{R}^n$ if

$$f(y) \geq f(x) + g^T(y-x) \quad y \in \mathbb{R}^n$$

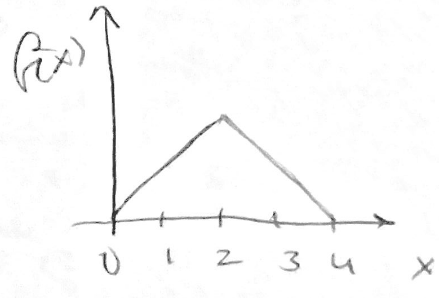
The set of g defines the sub differential of f at x and is denoted as $\partial f(x)$

$$\partial f(\bar{x}) = \{ p \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + p^T(x - \bar{x}),\ \forall x \in S \}$$

If f is differentiable: $\partial f(\bar{x}) = \nabla f(\bar{x})$

$$0^n \in \partial f(x^*) \Leftrightarrow f \text{ is global min at } x^* \in \mathbb{R}^n$$

E5.9 $\quad f(x) = \begin{cases} x & 0 \leq x \leq 2 \\ 4 - x & 2 \leq x \leq 4 \end{cases}$



a) Find the subdifferential of $f(x)$ at $x=1$. $f(x)$ is differentiable at $x=1 \Rightarrow \partial f(\bar{x}) = \nabla f(\bar{x}) = 1$.

b) Find the subdifferential of $f(x)$ at $x=2$

$$\lim_{x \to 2^-} \frac{f(x) - f(2)}{x - 2} = 1$$

$$\lim_{x \to 2^+} \frac{f(x) - f(2)}{x-2} = -1$$

so

$$\partial h(2) = [-1, 1]$$

Convex min problem feasible set closed and convex

Subgradient projection algorithm

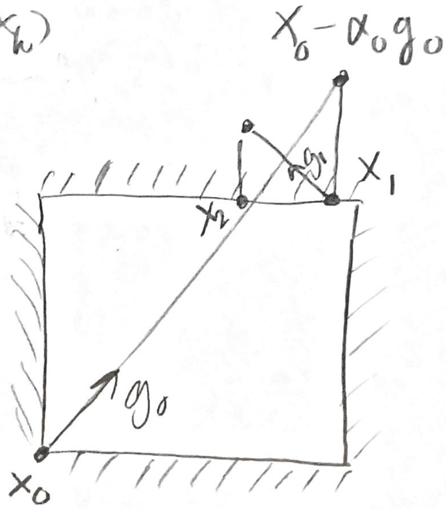
Step 0: x_0, $f^0_{Best} = f(x_0)$, $k=0$

Step 1: Find a subgradient $g_k \in \partial f(x_k)$

Step 2: $x_{k+1} = \text{Proj}_{\mathcal{X}}(x_k - \alpha_k g_k)$

Step 3: $f^{k+1}_{Best} = \min(f^k_{Best}, f(x_{k+1}))$

Step 4: Termination criteria fullfilled $\Rightarrow$ Stop.
Otherwise go to step 1, $k=k+1$.



E5.1

$$\min \|x\|_1$$
$$\text{s.t. } Ax = b$$
$$x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$$

A has full rank $m < n$ rank $A = m$

- Write down the projected subgradient update

Lemma 12.26: feasible descent direction from projected gradients

for problem:

$$\min f(x)$$
$$Ex = 0$$
$$Ax \le b \qquad M = \begin{pmatrix} A \\ E \end{pmatrix}. \text{ full rank}$$

x is a feasible point, $P = I^n - M^T(MM^T)^{-1}M$

then

$$p = -P\nabla f(x)$$

is a feasible descent direction.

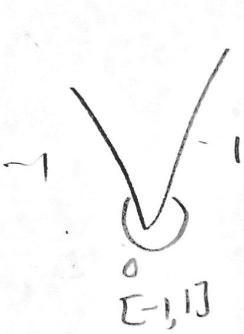
$$P = I^n - A^T(AA^T)^{-1}A$$

for first problem.

$$p = -P \cdot g \qquad g \in \partial f(x)$$

We look at $\tilde{f} = |x_i|$

$$\frac{\partial \tilde{f}}{\partial x_i} = \begin{cases} 1 & x_i > 0 \\ [-1,1] & x_i = 0 \\ -1 & x_i < 0 \end{cases}$$



$$g(x_i) = \begin{cases} 1 & x_i > 0 \\ 0 & x_i = 0 \\ -1 & x_i < 0 \end{cases} \qquad g(x_i) \in \partial f(x_i)$$

$$g_i(x_i) = \text{sign}(x_i)$$

$f = \|x\|_1$ $f = |x_1| + |x_2| + \dots + |x_n|$

$$\begin{pmatrix} \text{sign}(x_1) \\ \text{sign}(x_2) \\ \vdots \\ \text{sign}(x_n) \end{pmatrix} \in \partial f(x)$$

$g(x) = \text{sign}(x)$ is a subgradient of f at $\bar{x}_j$.

$$x^{k+1} = x^k - \alpha_k \left(I - A^T(AA^T)^{-1}A\right) \text{sign}(x^k)$$

goo

5.2

d) Which conditions of theorem 10.15 are satisfied during the iterations of the simplex algorithm by defining

$$y^T = c_B^T B^{-1} \quad \text{(shadow price)}$$

Theorem 10.15: primal feasibility, dual feasibility, complementary slackness:

$$\begin{cases} x_j (c_j - y^T A_{\cdot j}) = 0 \\ y_i (A_{i\cdot} x - b_i) = 0 \end{cases}$$

Primal feasibility: $\bar{x}$ feasible

Dual feasibility: (P) $\min z = c^T x$ s.t. $Ax = b$, $x \geq 0$ (D) $\max w = b^T y$ s.t. $A^T y \leq c$, y free

We want to know $(A^T y)^T \leq c^T$

$$y^T A \leq c^T$$

$$c_B^T B^{-1} A \overset{?}{\leq} c^T$$

Split $A = (B|N)$ $\quad c = (c_B, c_N)$

B: $\quad c_B^T B^{-1} B - c_B^T = c_B^T - c_B^T = 0 \leq 0$

N: $\quad c_B^T B^{-1} N - c_N^T = -\tilde{c}_N^T \overset{?}{\leq} 0$

So this condition holds only at the last iteration.
Dual feasibility is fullfilled only in the opt. point.

Complementary Slackness

$$x_j (c_j - y^T A_{\cdot j}) = 0, \quad j = 1, \dots, n$$

for non-basic variables $x_j = 0 \implies$ ✓

for basic variables: $c_j - y^T A_{\cdot j}$

$$c_B^T - y^T B = c_B^T - c_B^T B^{-1} B = c_B^T - c_B^T = 0$$

$\implies$ ✓

$y_{i\cdot}(A_{i\cdot}x - b_{i\cdot}) = 0$

$Ax = b$

comp. cda. is fullfilled

PROBLEM SOLVING SESSION

Monday
14 october
10^{00}

- Feasible-direction methods

$\min f(x)$, $\quad f: \mathbb{R}^n \to \mathbb{R}$, C^1 on X,

s.t $x \in X$, $\quad X$ Polyhedron

Frank-Wolfe algorithm (FW)

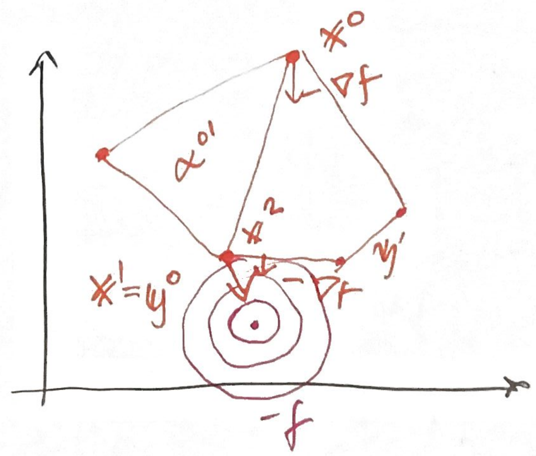
$\Rightarrow$ linearize f to find search direction

⓪ $x^0 \in X$, $k=0$

① y^k: solve $\min\limits_{y \in X} \nabla f(x^k)^T y$, $p^k = y^k - x^k$

② α_k: solve $\min\limits_{\alpha \in [0,1]} f(x^k + \alpha p^k)$ $\to$ exactly or approximately

③ $x^{k+1} = x^k + \alpha_k p^k$, $k := k+1$, go to ① unless $\alpha_k \approx 0$ or
$\nabla f(x^k)^T p^k \approx 0$ $\to$ stationary point

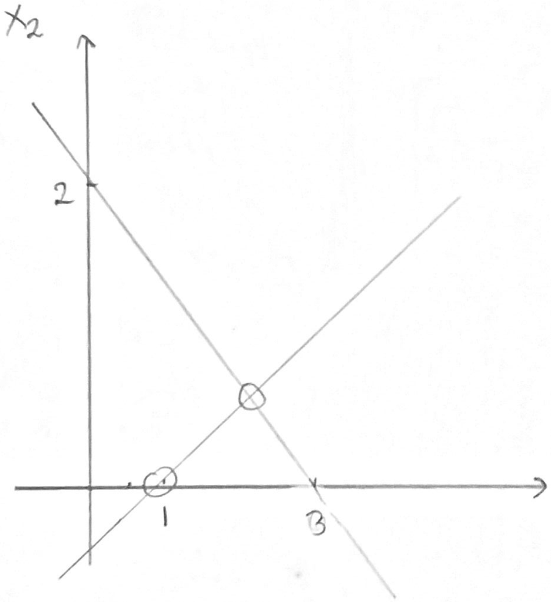


12.4 Use FW.

$$\min \ x_1^2 + 4x_2^2 - 16x_1 - 24x_2$$

$$\text{s.t} \quad \left.\begin{array}{l} x_1 + x_2 \leq 6 \\ x_1 - x_2 \leq 3 \\ x_1 , x_2 \geq 0 \end{array}\right\} = X$$

$$\nabla f(x) = \begin{pmatrix} 2(x_1-8) \\ 8(x_2-3) \end{pmatrix}$$



$\text{ext}(X) = \{\begin{pmatrix}0\\0\end{pmatrix}$

Let $x^0 = \begin{pmatrix}0\\0\end{pmatrix}$.

1 : (1) $\min\limits_{y \in X} \nabla f(x^0)^T y = \min\limits_{y \in X} -16y_1 - 24y_2 =$

$= \min\{0, -144, -48, -108\} \Rightarrow y^0 = \begin{pmatrix}0\\6\end{pmatrix}$

(2) $\min_{\alpha \in [0,1]} f(x^0 + \alpha p^0) = \min_{\alpha \in [0,1]} f\begin{pmatrix} 0 \\ 6\alpha \end{pmatrix} = \ldots =$

$= 144(\alpha - 1)\alpha \Rightarrow \alpha_0 = 1/2$

$x^1 = x^0 + \alpha_0 p^0 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$

2: (1) $\min_{y \in X} \nabla f(x^1)^T y = \min_{y \in X} -16 y_1 + 0 y_2, \Rightarrow \ldots$

$\Rightarrow y^1 = \frac{1}{2}\begin{pmatrix} 9 \\ 3 \end{pmatrix}, \; p^1 = \frac{1}{2}\begin{pmatrix} 9 \\ -3 \end{pmatrix}$

(2) $\min_{\alpha \in [0,1]} \underbrace{f(x^1 + \alpha p^1)}_{:= \varphi(\alpha)} : \varphi'(\alpha) = 0 \Rightarrow \ldots \Rightarrow \alpha > 1$

$\alpha = 1, \; x^2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 9 \\ -3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 9 \\ 3 \end{pmatrix}$

3: (1) $\min_{y \in X} \nabla f(x^2)^T y = \min_{y \in X} -7y_1 - 12y_2 \Rightarrow \ldots$

$\Rightarrow y^2 = \begin{pmatrix} 0 \\ 6 \end{pmatrix}, \; p^2 = \frac{1}{2}\begin{pmatrix} -9 \\ 9 \end{pmatrix}$

(2) $\min_{\alpha \in [0,1]} f(x^2 + \alpha p^2) : \varphi'(\alpha) = 0 \Rightarrow \ldots$

$\Rightarrow \alpha = \frac{1}{9} \in [0,1] \Rightarrow$

$x^3 = \frac{1}{2}\begin{pmatrix} 9 \\ 3 \end{pmatrix} + \frac{1}{9}\frac{1}{2}\begin{pmatrix} -9 \\ 9 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

4: $\min_{y \in X} \nabla f(x^3)^T y = \min_{y \in X} -8y_1 - 8y_2 \Rightarrow$

$y^3 = \begin{pmatrix} 0 \\ 6 \end{pmatrix} \Rightarrow p^3 = \begin{pmatrix} -4 \\ 4 \end{pmatrix} \Rightarrow$

$\nabla f(x^3)^T p^3 = 0 \rightarrow$

x^3 is a stationary point.

PROBLEM SOLVING SESSION

Friday
18 october
800

Recap Frank Wolfe algorithm

$$\min f(x), \quad f \in C^1$$
$$\text{s.t. } x \in X, \quad X \text{ polyhedra}$$

0. $x^0 \in X$, $k=0$

1. y^k: solve $\min_{y \in X} \nabla f(x^k)^T y$, $p^k = y^k - x^k$

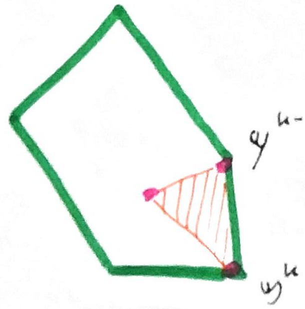
2. α_k: linesearch, $\alpha \in [0,1]$

3. termination criteria, p^k not descent dir.

Simplicial decomposition

generalise 3 by including old p^k.

$\Rightarrow$ multidim. line search

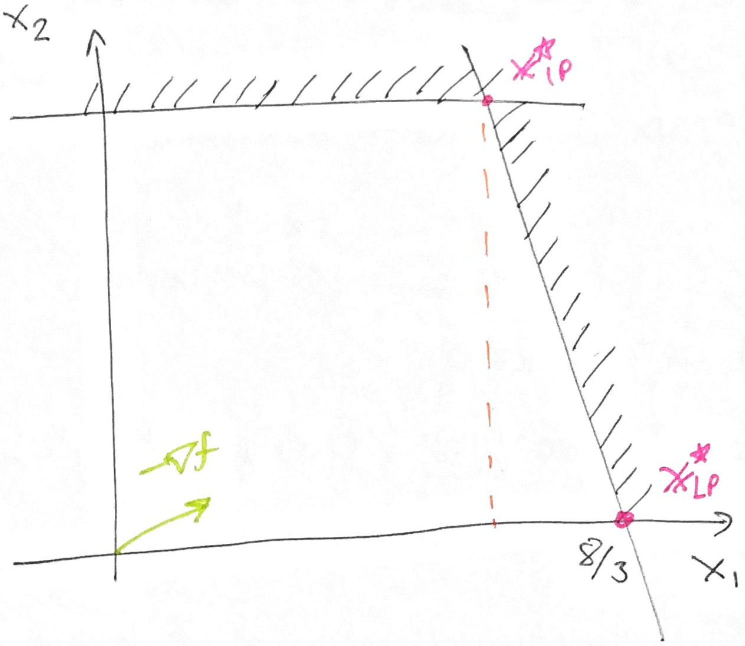


E6.4

Solve

$$\begin{aligned} \min \quad & -4x_1 - x_2 \\ \text{s.t.} \quad & 3x_1 + x_2 \leq 8 \\ & x_2 \leq 2 \\ & x_1, x_2 \geq 0 \\ & x_1, x_2 \in \mathbb{Z} \end{aligned}$$

a) Graphically, with or without integer constraint.



$x_{LP} = \begin{pmatrix} 8/3 \\ 0 \end{pmatrix}, \quad x_{IP} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

$z^*_{LP} = -\frac{32}{3} < -10 = z^*_{IP}$

b) Add $x_1 \leq 2$, what happens?

note: no integer point is removed,
& $x^*_{LP} = x^*_{IP}$

All extreme points are integer!

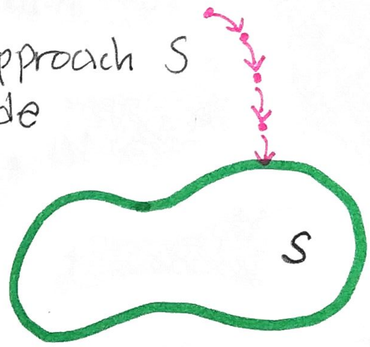
Penalty methods

$$\min f(x) \quad \text{s.t.} \quad x \in S \quad (1) \xrightarrow{\text{Idea}} \min_{x \in \mathbb{R}^n} f(x) + \chi_S(x),$$

$$\chi_S(x) = \begin{cases} 0, & x \in S \\ \infty, & x \notin S \end{cases}$$

Replace $\chi_S(x)$ with nice function

- Exterior penalty method
  - Penalize infeasible points
  - Increase penalty $\Rightarrow$ opt. sol. approach S from outside



Let $S := \{ x \in \mathbb{R}^n \mid g_i(x) \le 0,\ i=1,\dots,m$
$h_j(x) = 0,\ j = 1, \dots, l \}$

$\psi : \mathbb{R} \to \mathbb{R}_+$, $\psi(t) = 0$ iff $t = 0$,
e.g. $\psi(t) = t^2$.

$$\chi_S(x) \approx \nu \tilde{\chi}_S(x) = \nu \left(\sum_{i=1}^{m} \psi(\max(0, g_i(x))) + \sum_{j=1}^{l} \psi(h_j(x)) \right)$$

Penalty parameter

$$\min_{x \in \mathbb{R}^n} f(x) + \nu \tilde{\chi}_S(x) \quad (2)$$

Thm

$$\left. \begin{array}{l} x_\nu^* \text{ opt. in } (2) \\ x_\nu^* \xrightarrow[\nu \to \infty]{} x^* \end{array} \right\} \Rightarrow x^* \text{ opt. in } (1)$$

13.3 $\min f(x) := \frac{1}{2}(x_1^2 + x_2^2)$

s.t $x_1 = 1$

Solve using exterior penalty method.

$$\varphi(x,\nu) = \frac{1}{2}(x_1^2+x_2^2) + \overset{>0}{\nu}(x_1-1)^2 =$$

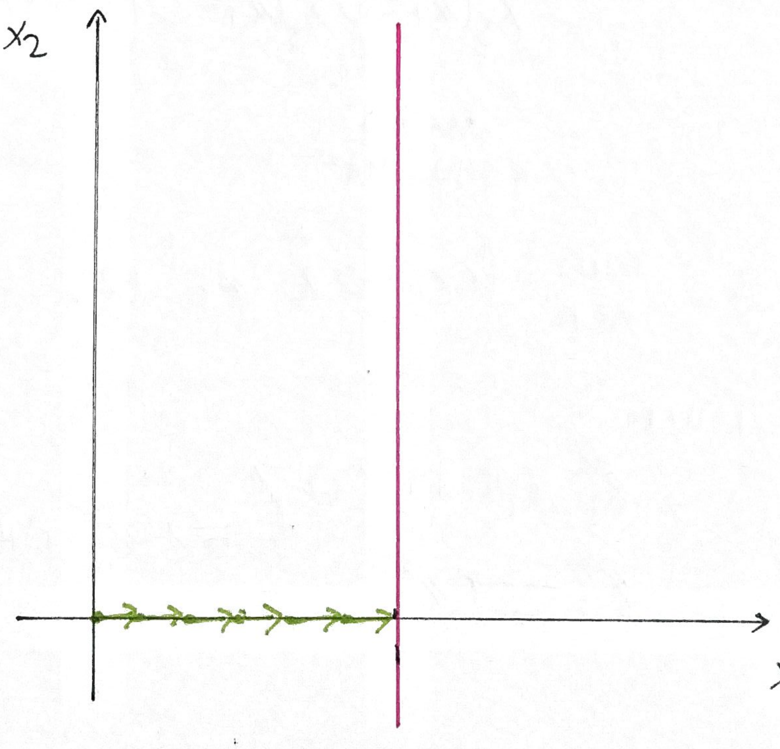
$$= \left(\frac{1}{2}+\nu\right)x_1^2 + \frac{x_2^2}{2} - 2\nu x_1 + \nu$$

$$\min_{x \in \mathbb{R}^2} \varphi(x,\nu) = ?$$

φ convex, enough to find $\nabla_x \varphi(x,\nu) = 0$.

$$\nabla_x \varphi(x,\nu) = 0 \Rightarrow \begin{cases} (1+2\nu)x_1 - 2\nu = 0 \\ x_2 = 0 \end{cases}$$

$$\Rightarrow x_1 = \frac{2\nu}{1+2\nu} = \frac{2}{\frac{1}{\nu}+2} \xrightarrow[\nu\to\infty]{} 1$$



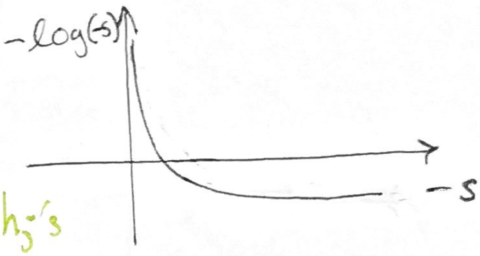
900

- Interior penalty method
  - add a penalty in interior of S: $\to \infty$ at the boundary, let it decrease in strict interior.
  - tend to opt. sol. from within S.

  Let $\phi: \mathbb{R}^- \to \mathbb{R}^+$ s.t. $\phi(s_k) \xrightarrow[s_k \to 0^-]{} \infty$,

  e.g. $\phi(s) = -\log(-s)$

  ↑ convex



We do not want to be on the boundary $\Rightarrow$ no h_j's

$$\chi_S(x) \approx \nu \tilde{\chi}_S(x) = \nu \sum_{i=1}^{m} \phi(g_i(x))$$

  - note: S has to have a strict interior point: $g_i(\bar{x}) < 0 \ \forall i$.

$$\min_{x \in \mathbb{R}^n} f(x) + \nu \tilde{\chi}_S(x) \qquad (3)$$

If S is nice (see book):

$$\left.\begin{array}{l} x_\nu^* \text{ opt. in } (3) \\ x_\nu^* \xrightarrow[\nu \to 0^+]{} x^* \end{array}\right\} \Rightarrow x^* \text{ is opt. in } (1)$$

13.5 Solve

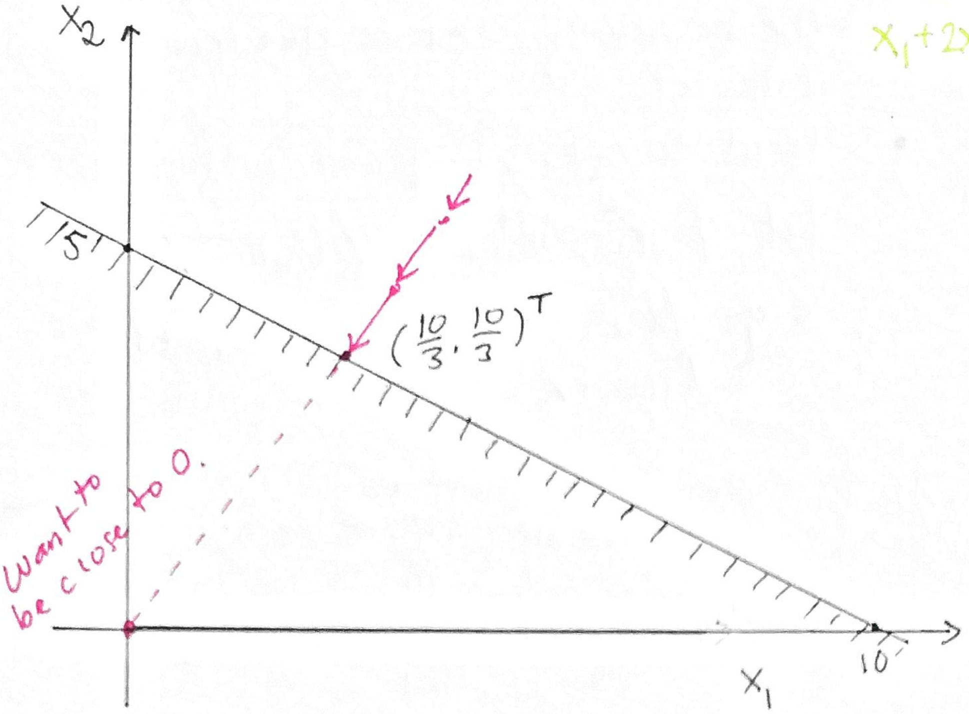
$$\min \frac{1}{2} x_1^2 + x_2^2$$

$$\text{s.t.} \quad x_1 + 2x_2 \geq 10.$$

use interior logaritmic barrier penalty method.

$$\varphi(x,\nu) = \underbrace{\tfrac{1}{2}x_1^2 + x_2^2}_{\text{convex}} \underbrace{- \nu \log(x_1 + 2x_2 - 10)}_{\text{convex}}$$

assume $x_1 + 2x_2 - 10 > 0$



$$\min_{x \in \mathbb{R}^2} \varphi(x,\nu) = ?$$

$$\nabla_x \varphi(x,\nu) = \begin{cases} x_1 - \dfrac{\nu}{x_1+2x_2-10} = 0 \\ 2x_2 - \dfrac{2\nu}{x_1+2x_2-10} = 0 \end{cases} \Rightarrow$$

$$\nu = x_1(x_1 + 2x_2 - 10) \Rightarrow$$

$$2x_2 - 2x_1 = 0 \Rightarrow x_1 = x_2$$

$$\nu = 3x_1^2 - 10x_1 \Rightarrow \dots \Rightarrow x_1 = \frac{5}{3} \pm \sqrt{\frac{25+3\nu}{9}}$$

$-$ gives infeasible sol.

$$x_1^*(\nu) = x_2^*(\nu) = \frac{5 + \sqrt{25+3\nu}}{3} \xrightarrow[\nu \to 0^+]{} \frac{5}{3} + \frac{5}{3} = \frac{10}{3}$$

(Last exercise that Edvin presents: extends it to include the most important things in the course):

Double check opt. by KKT !

- Problem is convex & strict interior points
  e.g. $x_1 = x_2 = 10 \Rightarrow 10 + 2 \cdot 10 = 30 > 10$
- KKT is sufficient for optimality
  "KKT" $\Rightarrow$ "opt."

$$\nabla f(x^*) = \begin{pmatrix} x_1^* \\ 2x_2^* \end{pmatrix} = \frac{10}{3} \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

$$\nabla g(x^*) = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

$$\nabla f(x^*) + \mu \nabla g(x^*) = 0 \Rightarrow \mu = \frac{10}{3} \geq 0$$

$$g(x^*) = \ldots = 0 \Rightarrow g(x^*)\mu = 0$$

$$\Rightarrow x^* \in S$$

$\Rightarrow$ KKT holds.

$\Rightarrow$ x^* is the optimum.