LECTURE 5: CONSTRAINED OPTIMIZATION - INTRODUCTION

1. Basic terminologies
2. KKT conditions – motivation
3. Background knowledge

Constrained optimization

- General form:

$$
\begin{aligned}
&\text{minimize} \quad f(x) \\
&\text{s.t.} \\
&\left.\begin{aligned}
h_1(x) &= 0 \\
&\vdots \\
h_m(x) &= 0 \\
g_1(x) &\le 0 \\
&\vdots \\
g_p(x) &\le 0
\end{aligned}\right\} \text{functional constraints (explicit)} \\
&x \in \Omega \subset E^n, \ (\text{set/implicit constraints})
\end{aligned}
$$

where $m \le n$ and $f,\ h_i,\ g_j \in C^k$ (most likely $k = 1$ or 2).

Short form:

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{s.t} \quad & h(x) = 0 \\
& g(x) \le 0 \\
& x \in \Omega,
\end{aligned}
$$

where $h(x) = (h_1(x), \cdots, h_m(x))^T$ and $g(x) = (g_1(x), \cdots, g_p(x))^T$.

Basic terminologies

- Definition:

  $x \in \Omega$ is a <u>feasible solution</u> (<u>feasible point</u>) if $h(x) = 0$ and $g(x) \leq 0$.

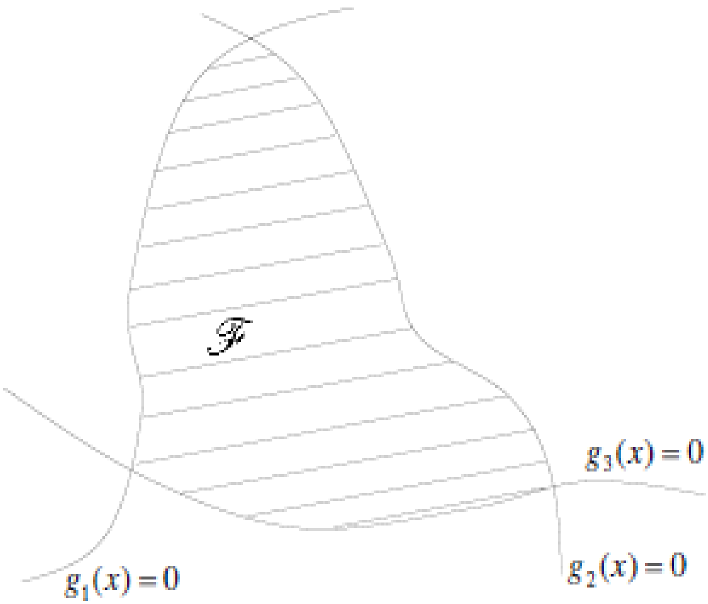
  $\mathscr{F} = \{x \in E^n \mid h(x) = 0,\ g(x) \leq 0, x \in \Omega\}$ is the <u>feasible domain</u> (<u>feasible solution set</u>).

- Definition:

  Let $\bar{x} \in \mathscr{F}$. If $g_j(\bar{x}) = 0$, then the constraint $g_j(x) \leq 0$ is <u>active</u> at $\bar{x}$. If $g_j(\bar{x}) < 0$, then the constraint is <u>inactive</u> at $\bar{x}$.

Property of active sets

- Observation



Let $\bar{x} \in \mathscr{F}$ and $J(\bar{x}) = \{j \mid g_j(\bar{x}) = 0,\ j = 1, \cdots, p\}$. Then there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that

$$J(x) \subseteq J(\bar{x}) , \quad \forall\, x \in \mathscr{F} \cap N(\bar{x}).$$

This means that at a feasible point near $\bar{x}$, we don't need to worry about those inactive constraints at $\bar{x}$. Therefore, we say that the inactive constraints at $\bar{x}$ play no role in optimization around $\bar{x}$.

Question

Given a constrained optimization problem, we all heard about "Lagrange multipliers," "Lagrangian method," "K-K-T conditions," "Lagrangian dual," etc.

In particular, we may know something like

(1) $l(x, \lambda, \mu) \triangleq f(x) + \lambda^T h(x) + \mu^T g(x),$

with $\lambda \in E^m, \quad \mu \in E_+^p$;

(2)

$$\nabla_x l(x^*) \triangleq \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0,$$

$$\mu^T g(x^*) = 0 \ ;$$

(3) $L(x^*) \triangleq F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$.

How do these concepts play together to form the main body of constrained optimization?

Necessary conditions for constrained optimization

- More precisely, we have

$$
(\text{NLP}) \quad \begin{aligned} \text{maximize} \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in E^n \end{aligned}
$$

<u>Theorem</u>: (K-K-T Conditions)

Let x^* be a relative minimum point for (NLP) that is a regular point. Then $\exists$ a vector $\lambda \in E^m$ and a vector $\mu \in E^p_+$ s.t.

$$
(*) \begin{cases} \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0 \\ \mu^T g(x^*) = 0. \end{cases}
$$

- But, Why? How?

Intuition and speculation

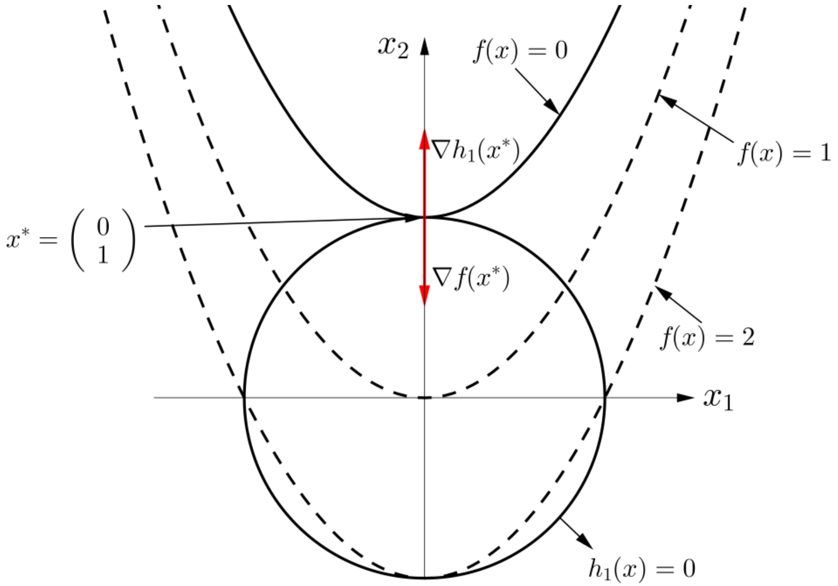
- (I-a) NLP with one equality constraint

$$\text{minimize} \quad f(x_1, x_2) \triangleq x_1^2 - x_2 + 1$$

s.t.

$$h_1(x_1, x_2) \triangleq x_1^2 + x_2^2 - 1 = 0$$

$$x \in \Omega \triangleq E^2$$



1. For $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$,

$$\nabla f(x^*) = (2x_1, -1) \mid_{x^*} = (0, -1),$$
$$\nabla h_1(x^*) = (2x_1, 2x_2) \mid_{x^*} = (0,\ 2),$$
$$\nabla f(x^*) + \tfrac{1}{2}\nabla h_1(x^*) = 0\ .$$

2. Since the equality constraint can be written as

$$\bar{h}_1(x) \triangleq -h_1(x) = -x_1^2 - x_2^2 + 1 = 0$$

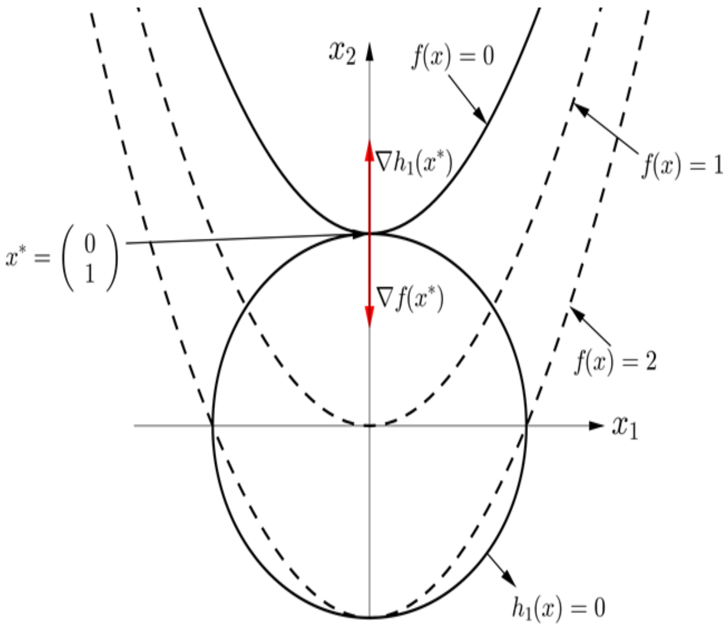
and

$$\nabla \bar{h}_1(x^*) = -\nabla h_1(x^*) = (0, -2), \quad \text{we have}$$

$$\nabla f(x^*) - \frac{1}{2}\nabla \bar{h}_1(x^*) = 0.$$

$$\Rightarrow \nabla f(x^*) + \lambda_1 \nabla h_1(x^*) = 0, \text{ for some } \lambda_1 \in R.$$

(I-a) - continue



3. For $\bar{x}^* = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \end{pmatrix}$, this is a local maximum point.

   $\nabla f(\bar{x}^*) = (2x_1, -1)\,|_{\bar{x}^*} = (\sqrt{3}, -1)$

   $\nabla h_1(\bar{x}^*) = (2x_1, 2x_2)\,|_{\bar{x}^*} = (\sqrt{3}, -1)$

   and

   $$\nabla f(x^*) - \nabla h(\bar{x}^*) = 0,$$

   $\Rightarrow$ necessary condition only !

4. Note that $\mathscr{F}$ is not a convex set while $f(x)$ is convex on E^2.

(I-b) NLP with two equality constraints

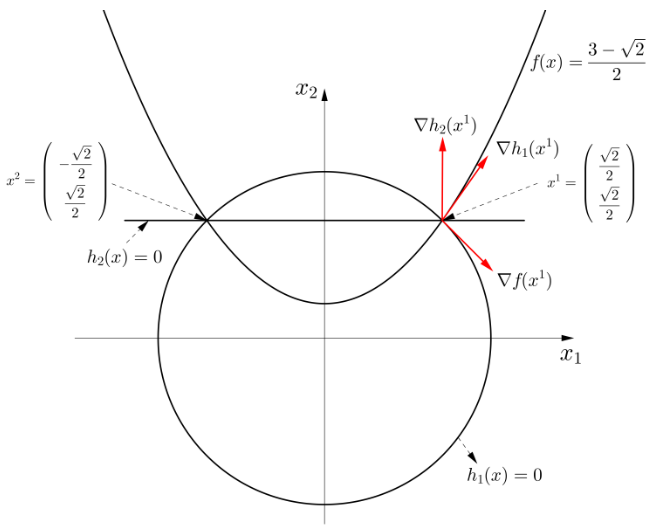
$$\text{minimize} \quad f(x) \triangleq x_1^2 - x_2 + 1$$

s.t.

$$h_1(x) \triangleq x_1^2 + x_2^2 - 1 = 0$$

$$h_2(x) \triangleq x_2 - \frac{\sqrt{2}}{2} = 0$$

$$x \in \Omega \triangleq E^2$$



1. For $x^1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$,

$$\nabla f(x^1) = (2x_1, -1)\,|_{x^1} = (\sqrt{2}, -1),$$

$$\left.\begin{array}{l} \nabla h_1(x^1) = (2x_1, 2x_2)\,|_{x^1} = (\sqrt{2}, \sqrt{2}), \\ \nabla h_2(x^*) = (0, 1)\,|_{x^1} = (0, 1). \end{array}\right\} \quad \begin{array}{c}\text{linearly} \\ \text{independent}\end{array}$$

We have

$$\nabla f(x^1) + (-1)\nabla h_1(x^1) + (1+\sqrt{2})\nabla h_2(x^1) = 0,$$

or

$$\nabla f(x^1) + (1)\nabla \bar{h}_1(x^1) + (-1-\sqrt{2})\nabla \bar{h}_2(x^1) = 0,$$

or

$$\nabla f(x^1) + (-1)\nabla h_1(x^1) + (-1-\sqrt{2})\nabla \bar{h}_2(x^1) = 0,$$

or

$$\nabla f(x^1) + (1)\nabla \bar{h}_1(x^1) + (1+\sqrt{2})\nabla h_2(x^1) = 0.$$

2. Similar results hold for $x^2 = \begin{pmatrix} \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$.

(I-c) NLP with multiple equality constraints

- General form (in guess)

$$\nabla f(x^*) + \underbrace{\lambda^T \nabla h(x^*)}_{\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)} = 0$$

where $\lambda \in E^m$.

- Observations:

1. Notice that the system

$$(*) \begin{cases} \nabla f(x) + \lambda^T \nabla h(x) = 0 \\ h(x) = 0 \end{cases}$$

has $n+m$ variables satisfying $n+m$ equations, that uniquely determines x^*.

For inequality constraints:

2. If we know which inequality constraints are inactive, then the problem becomes much simpler.

3. How to solve the system $(*)$?

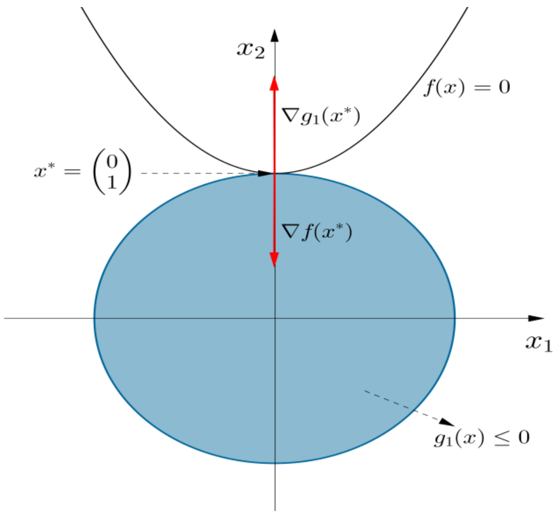
(II-a) NLP with one inequality constraints

$$\text{minimize} \quad f(x) \triangleq x_1^2 - x_2 + 1$$

s.t.

$$g_1(x) \triangleq x_1^2 + x_2^2 - 1 \le 0$$

$$x \in \Omega \triangleq E^2.$$



1. For $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$,

   $\nabla f(x^*) = (0, -1)$,

   $\nabla g_1(x^*) = (0, 2)$,

   $\nabla f(x^*) + \frac{1}{2}\nabla g_1(x^*) = 0.$

2. Note that

   (i) $g_1(x) \le 0 \not\Leftrightarrow \bar{g}_1(x) \triangleq -g_1(x) \le 0.$

   (ii) Both f on E^2 and $\mathscr{F}$ are convex.

3. For $\bar{x}^* = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$,

   $\nabla f(\bar{x}^*) = (0, -1)$,

   $\nabla g_1(\bar{x}^*) = (0, -2)$,

   $\Rightarrow \nabla f(\bar{x}^*) + \lambda_1 \nabla g_1(\bar{x}^*) \neq 0$ for any $\lambda_1 \ge 0.$

4. $\nabla g_1(x^*)$ has to be in the opposite direction of $\nabla f(x^*)$, i.e., $\exists\ \lambda_1 \ge 0$ s.t.

   $$\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) = 0.$$

   If not, then $\nabla g_1(x^*)$ and $\nabla f(x^*)$ are on the same side, then we can move along $-\nabla g_1(x^*)$ from x^* to keep feasibility while we reduce the objective value. ($\because -\nabla g_1(x^*)$ and $-\nabla f(x^*)$ points to the same direction !!)

(II-b) NLP with two inequality constraints

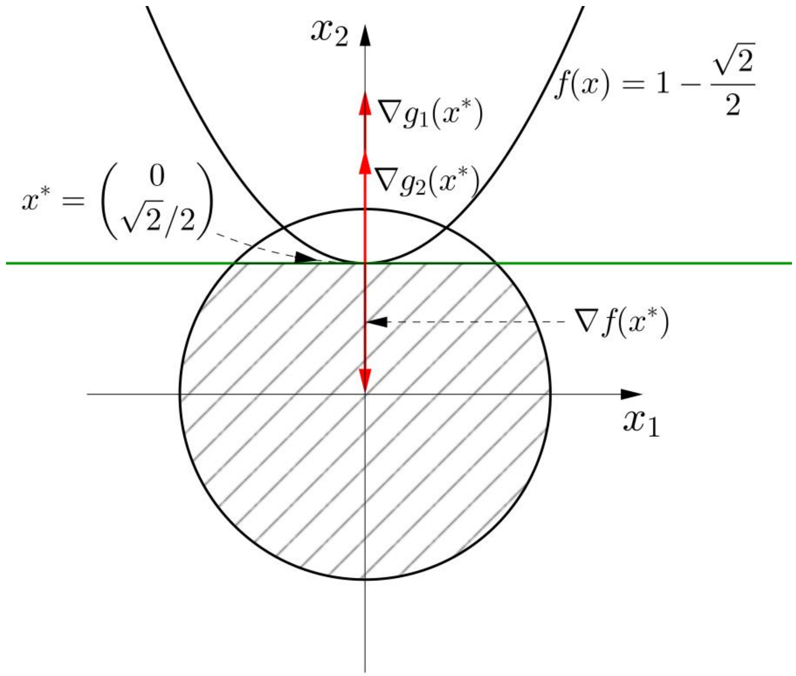
minimize $\quad f(x) \triangleq x_1^2 - x_2 + 1$

s.t.

$$g_1(x) \triangleq x_1^2 + x_2^2 - 1 \leq 0$$

$$g_2(x) \triangleq x_2 - \frac{\sqrt{2}}{2} \leq 0$$

$$x \in \Omega \triangleq E^2.$$



1. For $\quad x^* = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix}$,

$$\nabla f(x^*) = (2x_1, 1)\mid_{x^*} = (0, -1),$$

$$\nabla g_1(x^*) = (2x_1, 2x_2)\mid_{x^*} = (0, \sqrt{2}),$$

$$\nabla g_2(x^*) = (0, 1)\mid_{x^*} = (0, 1),$$

$$\nabla f(x^*) + (0)\nabla g_1(x^*) + (1)\nabla g_2(x^*) = 0.$$

2. Notice that

$$g_1(x) \leq 0 \text{ is inactive at } x^*.$$

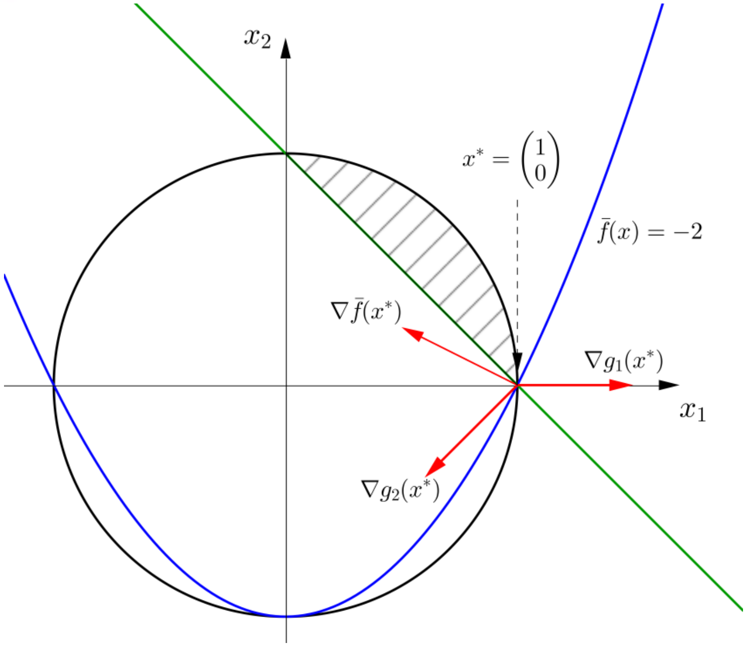
(II-b) - Continue

3. To get two active inequality constraints at an optimal solution, we consider

$$\bar{g}_2(x) \triangleq -x_1 - x_2 + 1 \leq 0$$

and

$$\bar{f}(x) \triangleq -f(x) = -x_1^2 + x_2 - 1.$$



4. For $x^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,

$$\left.\begin{array}{l} \nabla \bar{f}(x^*) = (-2x_1, 1)\,|_{x^*} = (-2, 1), \\ \nabla g_1(x^*) = (2x_1, 2x_2)\,|_{x^*} = (2, 0), \\ \nabla g_2(x^*) = (-1, -1)\,|_{x^*} = (-1, -1), \end{array}\right\} \text{linearly independent}$$

$$\nabla \bar{f}(x^*) + \tfrac{3}{2}\nabla g_1(x^*) + (1)\nabla g_2(x^*) = 0.$$

(II-c) NLP with multiple inequality constraints

- General form (in guess)

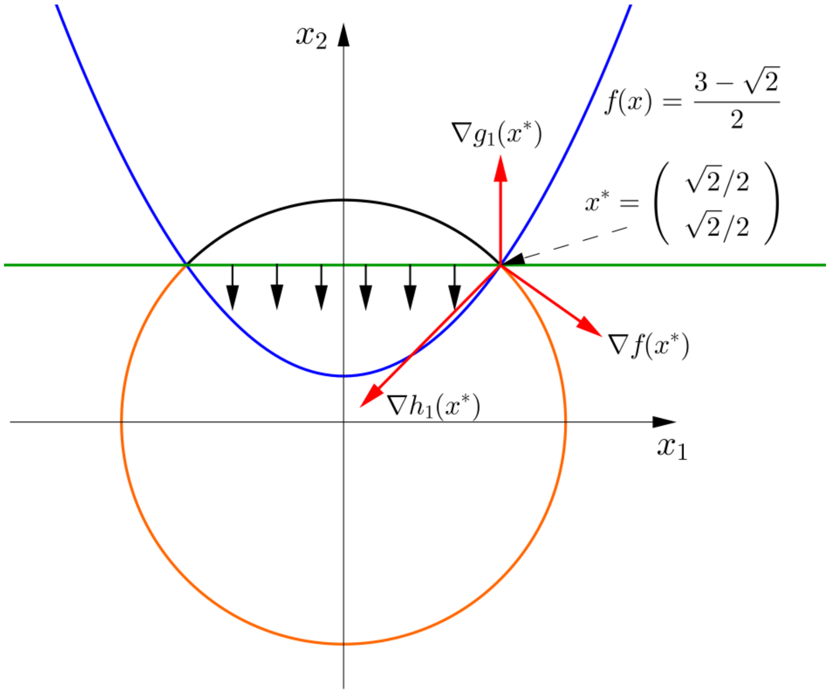
$$\nabla f(x^*) + \underbrace{\mu^T \nabla g(x^*)}_{\sum_{j=1}^{p} \mu_j \nabla g_j(x^*)} = 0$$

$$\text{where} \quad \mu \in E_+^m \quad \text{and}$$

$$\mu_j = 0 \quad \text{if} \quad g_j(x^*) < 0 \; .$$

(III-a) NLP with one equality and one inequality constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \triangleq x_1^2 - x_2 + 1 \\ \text{s.t.} \quad & \\ & \bar{h}_1(x) \triangleq -x_1^2 - x_2^2 + 1 = 0 \\ & g_1(x) \triangleq x_2 - \frac{\sqrt{2}}{2} \leq 0 \\ & x \in \Omega \triangleq E^2. \end{aligned}$$



For $x^* = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$,

$$\nabla f(x^*) = (2x_1, -1)\,|_{x^*} = (\sqrt{2}, -1),$$

$$\left.\begin{aligned} \nabla h_1(x^*) &= (-2x_1, -2x_2)\,|_{x^*} = (-\sqrt{2}, -\sqrt{2}), \\ \nabla g_1(x^*) &= (0, 1), \end{aligned}\right\} \text{ linearly independent}$$

$$\nabla f(x^*) + (1)\nabla h_1(x^*) + (1 + \sqrt{2})\nabla g_1(x^*) = 0.$$

(III-b) NLP with multiple equality/inequality constraints

- General form (in Guess)

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$

$$\text{where } \lambda \in E^m, \quad \mu \in E_+^p \quad \text{and}$$

$$\mu_j = 0 \quad \text{if} \quad g_j(x^*) < 0.$$

Question

- After learning these facts, can our speculation be realized in a mathematical theory?

- A story of G. B. Dantzig, J. von Neumann, A. W. Tucker, H. W. Kuhn, W. Karush and F. John.

Historical development

- G.B. Dantzig visited John von Neumann in Princeton in May 1948.
- John von Neumann circulated privately a short typewritten note "Discussion of Maximum Problem".
- H.W. Kuhn and A.W. Tucker (1951) published "Nonlinear Programming" in J. Neyman (ed.) "Proceeding of the 2nd Berkeley Symposium on Mathematical Statistics and Probability," UC Press, Berkeley, 481-492.
- W. Karush (1939), "Minima of Functions of Several Variables with Inequalities as Side Conditions," MS Thesis, Dept. of Mathematics, University of Chicago.
- F. John (1948), "Extremum Problems with Inequalities as Subsidiary Conditions," Studies and Essays presented to R. Courant on his 60th Birthday, Interscience, NY, 187-204.

Background knowledge

- Basic concepts from Taylor's theorem:

1. Moving along the direction of $-\nabla f(\bar{x})$ (or having a projected component of it) will reduce the objective function value.
2. To keep the feasibility of an equality constraint $h_i(x) = 0$ around $\bar{x}$, the moving direction $d \in E^n$ has to satisfy that

$$\nabla h_i(\bar{x})d = 0.$$

3. To keep the feasibility of an inequality constraint $g_j(x) \le 0$ around $\bar{x}$, the moving direction $d \in E^n$ better stays on the same side of $-\nabla g_j(\bar{x})$ (or having a projected component of it).

Implications

4. <u>Key idea of necessary conditions:</u>

" All feasible directions at $\bar{x}$ are not good direction for improvement." (i.e., $\nabla f(\bar{x})^T d_f \geq 0$.)

Equivalently, no feasible direction at $\bar{x}$ makes $\nabla f(\bar{x})^T d_f < 0$!!

5. For equality constraints,

$$-\nabla f(\bar{x}) = \sum_i \lambda_i \nabla h_i(\bar{x}), \quad \lambda_i \in R.$$

(Because $\nabla h_i(\bar{x})^T d_f = 0$.)

For inequality constraints,

$$-\nabla f(\bar{x}) = \sum_j \mu_j \nabla g_j(\bar{x}), \quad \mu_j \geq 0.$$

(Because d_f stays on the same side of $-\nabla g_j(\bar{x})$.)

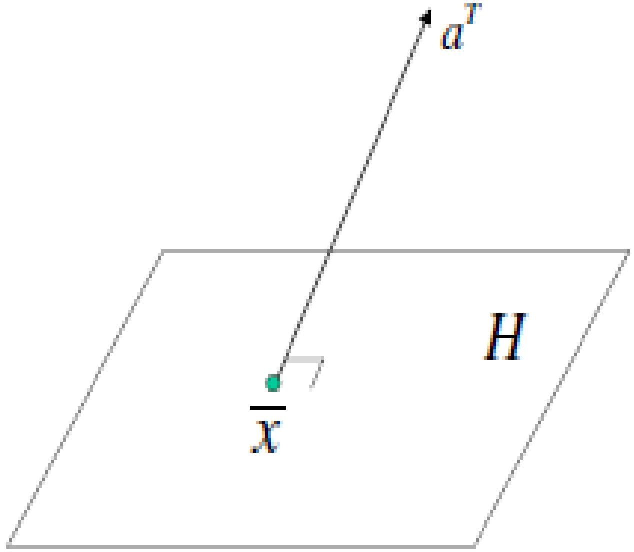
Background knowledge

- Basic concepts from Linear Programming

  - **Definition:**

    Let $a \in E^n$ and $\beta \in R$. Then

    $H \triangleq \{x \in E^n \mid a^T x - \beta = 0\}$ is a hyperplane.

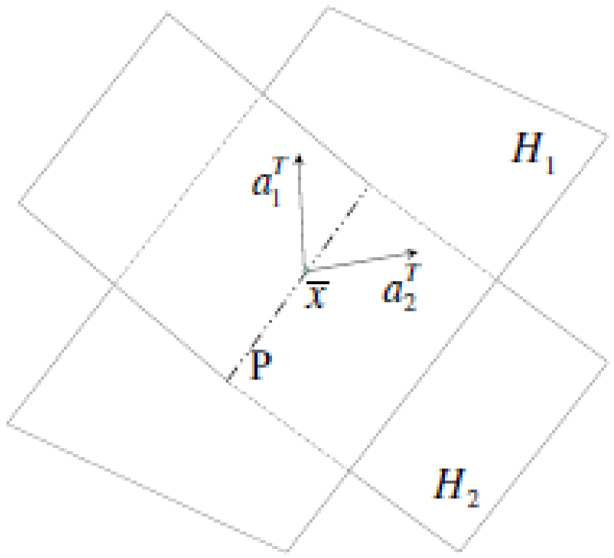


Observations

1. Let $a_i \in E^n$, $\beta_i \in R$, $i = 1, 2, \cdots, m$, and $A = [\ a_i^T\ ]_{m \times n}$, $b = (\beta_1, \cdots, \beta_m)^T$. Then $P \triangleq \{x \in E^n \mid Ax - b = 0\}$ is a <u>polyhedral set.</u>

2. Given $\bar{x} \in H$, a_1^T is orthogonal to H at $\bar{x}$. The feasible directions at $\bar{x}$ falls in $T \triangleq \{d \in E^n \mid a_1^T d = 0\}$.

3. Given $\bar{x} \in P$, the feasible direction at $\bar{x}$ falls in $T \triangleq \{d \in E^n \mid Ad = 0\}$.



4. In a linear programming problem with m equality constraints, a vertex $\bar{x}$ is "non-degenerated" if it is uniquely determined by m equations.

Also, we always assume that $\text{rank}(A) = m$ to start the study of LP. In this case, $\{a_1, a_2, \cdots, a_m\}$ are linearly independent (at $\bar{x}$ and so as to other points).