Proposition 2: Related optimality conditions

(a) Let C be a convex set and $f: C \to \mathbb{R}$ a differentiable (not necessarily convex) function. If $x^* \in C$ is a local minimum of f over C, then
$$(\nabla f)(x^*)^T (x - x^*) \geq 0 \quad \forall x \in C.$$

(b) Let C be convex and $f: C \to \mathbb{R}$ convex and differentiable. If $(\nabla f)(x^*)^T (x - x^*) \geq 0$, then
$$x^* = \underset{x \in C}{\operatorname{argmin}}\, f(x).$$

Proof sketch:

(a) Suppose $\exists x \in C$ $\quad \overbrace{(\nabla f)(x^*)^T (x - x^*)}^{= D_v f \text{ -- directional derivative}} < 0. \quad (*)$
Then consider f along the straight line connecting x^* and x (in the direction of $v := x - x^*$). Because of $(*)$, there is a point along this path with smaller cost than $f(x^*)$.

(b) From convexity of f:
$$f(x) \geq f(x^*) + \underbrace{(\nabla f)(x^*)^T (x - x^*)}_{\geq 0} \quad \forall x \in C.$$

Remark: Proposition 2 and Proposition 1 are related: consider $f(x) = \frac{1}{2}\|x-y\|^2$

$$\rightsquigarrow (\nabla f)(x) = x - y$$

The inequality $(\nabla f)(x^*)^T(x-x^*) \geq 0$

thus becomes $(x^*-y)^T(x-x^*) \geq 0$

$$\Leftrightarrow \underline{\underline{(y-x^*)^T(x-x^*) \leq 0}}$$

Inequality from Prop. 1 !

Theorem 2: Separating Hyperplane between S and $\{y\}$.

Let S be a nonempty closed convex set in $\mathbb{R}^n$ and $y \notin S$. Then there exists $p \neq 0$ $(p \in \mathbb{R}^n)$ and $\alpha \in \mathbb{R}$ such that the hyperplane

$$H := \{x \in \mathbb{R}^n : p^T x = \alpha\}$$

strongly separates the set S from the set $\{y\}$, i.e.,

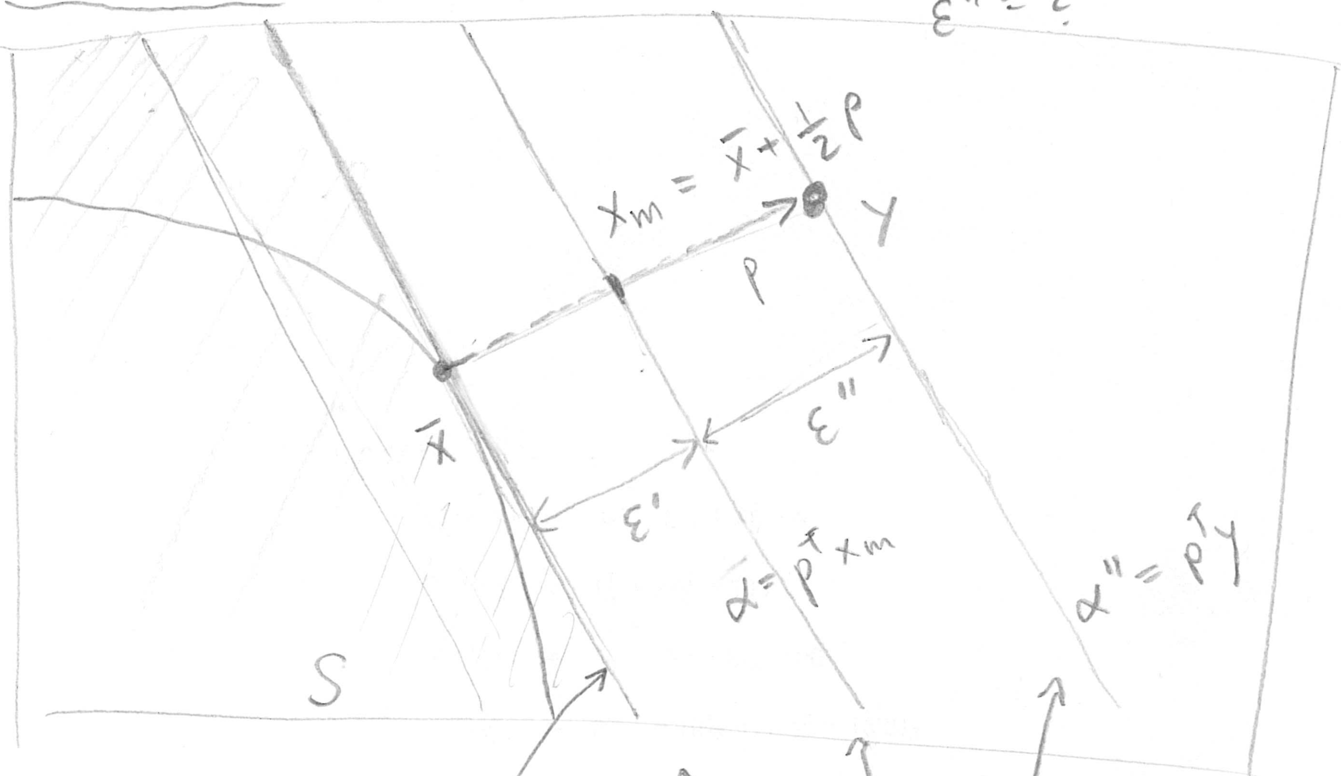
$$p^T x \leq \alpha - \varepsilon \quad \forall x \in S$$

$$p^T y \geq \alpha + \varepsilon$$

with some $\varepsilon > 0$.

Theorem 2: Illustration

$\varepsilon' = ?$
$\varepsilon'' = ?$



Level sets of $h: x \mapsto p^T x$

$$\{x \in \mathbb{R}^n : p^T x \leq \overbrace{p^T \bar{x}}^{=: \alpha'}\} = H_{p,\bar{x}}$$

increasing in the direction of $p^T = \nabla h$

$$h(x + \underbrace{\lambda}_{\geq 0} p) = p^T x + \underbrace{\lambda p^T p}_{\geq 0} \geq p^T x = h(x)$$

Claim: $S \subset H_{p,\bar{x}}$

From Proposition 1, we know that $\underbrace{(y - \bar{x})}_{p}{}^T (x - \bar{x}) \leq 0 \;\; \forall x \in S$

$$\Leftrightarrow \quad p^T (x - \bar{x}) \leq 0 \quad \forall x \in S$$
$$\Leftrightarrow \quad p^T x \leq p^T \bar{x} \quad \forall x \in S$$
$$\Leftrightarrow \quad x \in H_{p,\bar{x}} \quad \forall x \in S.$$

$H_{p,\bar{x}}$ is called a supporting hyperpla[ne] of S at $\bar{x}$

The <u>Separating Hyperplane Theorem</u> is crucial in proving two very important <u>theorems on alternatives for linear inequalities</u>, namely

- Farkas' Lemma
- Key Lemma.

These two results will turn the geometric <u>description</u> into algebraic ones!

$F_0 \cap G_0 \cap H_0 = \emptyset$

<u>Theorem 3</u> : Farkas' Lemma

Let $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$. Exactly one of the following two systems has a solution:

(i) $Ax \leq 0$, $c^T x > 0$

(ii) $A^T y = c$, $y \geq 0$.

(i)	(ii)
0	0
0	1
1	0
1	1

remains

<u>Remark:</u> Only two alternatives:

- (i) has a solution, then (ii) does not.
- (ii) has a solution, then (i) does not.
- (i) and (ii) can not both have solutions

we show these two

- (i) has no solution, then (ii) has solution
- (ii) has no solution, then (i) has solution.

Proof: If (i) and (ii) both have solutions, then $\exists x \in \mathbb{R}^n$ $c^T x > 0 \wedge Ax \leq 0$.

$c = A^T y$ by (ii) with $y \geqslant 0$

$\rightarrow (y^T A) x > 0$

But now $y^T (Ax) > 0$ with $\underset{(ii)}{y \geqslant 0}$ and $\underset{(i)}{Ax \leq 0}$, a contradiction.

Now suppose (ii) has **NO** solution. Define

$$S = \{ A^T y : y \geqslant 0 \},$$ which is convex & closed.

Then, with this definition of S, $c \notin S$.

By the Separating Hyperplane Theorem, there exists $p \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$c^T p > \alpha \quad \text{and} \quad p^T x \leq \alpha \ \forall x \in S.$$

By def. of S: $x \in S \Leftrightarrow \exists y \geqslant 0 \ A^T y = x$, so

$$\underbrace{p^T (A^T y)} \leq \alpha$$

$$(*) \ (Ap)^T y \leq \alpha \quad \text{for all } y \geqslant 0$$

If $(Ap)_i > 0$ for some index i, one could set $y_i \geqslant 0$ sufficiently large to get $(Ap)^T y > \alpha$ ↯ Thus $\underline{Ap \leq 0}$.

Setting $y = 0$ in $(*)$ shows $\alpha \geqslant 0$, so: $\underline{c^T p > 0}$

i.e. p is a solution of (i

Lemma 1 (Key Lemma)

Given matrices $\bar{A}$, B, and H of appropriate dimensions, exactly one of the following two systems has a solution:

(i) $\bar{A}x < 0$, $Bx \leq 0$, $Hx = 0$ (↑ cf. $F_0 \cap G_0 \cap H_0 = \emptyset$)

(ii) $\bar{A}^T u + B^T w + H^T v = 0$, $u \geq 0$, $w \geq 0$, $e^T u = 1$.
(v arbitrary) $(1, 1, \ldots, 1)$

Proof: (i) & (ii) can't both have sol'ns ↗ HW 7!

Suppose (i) has no solution. Then

$$\begin{aligned} \bar{A}x + e\theta &\leq 0, \quad \theta > 0 \\ Bx &\leq 0 \\ Hx &\leq 0 \\ -Hx &\leq 0 \end{aligned}$$

has no solution. This is equivalent to

$$\begin{pmatrix} \bar{A}, & e \\ B & 0 \\ H & 0 \\ -H & 0 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix} \leq 0, \quad (0, \ldots, 0, 1) \begin{pmatrix} x \\ \theta \end{pmatrix} > 0$$

having no solution. By Farkas' Lemma $\exists (u, w, v^1, v^2) \geq 0$:

$$\left(\begin{array}{cccc} \bar{A}^T & B^T & H^T & -H^T \\ \hdashline e^T & 0 & 0 & 0 \end{array}\right) \begin{pmatrix} u \\ w \\ v^1 \\ v^2 \end{pmatrix} = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hdashline 1 \end{array}\right)$$

$$\Rightarrow \bar{A}^T u + B^T w + H^T \underbrace{(v^1 - v^2)}_{=: v} = 0 \ \wedge \ e^T u = 1.$$

$\blacksquare$

Theorem 4: Fritz John Necessary Conditions

Let $\bar{x}$ be a feasible solution of (P) $\begin{cases} \min f(x) \\ \text{s.t. } g_i(x) \le 0 \quad i=1,\dots,m \\ \quad\quad h_i = 0 \quad i=1,\dots,l \end{cases}$

If $\bar{x}$ is a local minimum of (P), then there exists $(u_0, u, v) \neq 0$ such that

$$u_0 (\nabla f)(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0$$

$$u_0, u \geq 0$$

$$u_i g_i(\bar{x}) = 0 \quad , i=1,\dots,m$$

Proof: If $\nabla h_i(\bar{x})$, $i=1,\dots,l$ are linearly dependent, there exist $v \neq 0$ such that

$$\sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0.$$

Setting $u_0 = 0$ and $u = 0$ yields the result.

So now suppose $(\nabla h_i)(\bar{x})$ are linearly independent. Theorem 1 then applies, which asserts that

$$F_0 \cap G_0 \cap H_0 = \emptyset.$$

Assume for simplicity that the indices are chosen so that the index set of active constraints I is given as $I = \{1, 2, \dots, p\}$.

Let
$$\bar{A} = \begin{pmatrix} \nabla f(\bar{x})^T \\ \nabla g_1(\bar{x})^T \\ \vdots \\ \nabla g_p(\bar{x})^T \end{pmatrix}, \quad H = \begin{pmatrix} \nabla h_1(\bar{x})^T \\ \vdots \\ \nabla h_\ell(\bar{x})^T \end{pmatrix}$$

By Theorem 1, there is no vector d that satisfies
$$\bar{A} d < 0, \quad Hd = 0.$$

By the Key Lemma, there exist $(u_0, u_1, u_2, \dots, u_p)$ and $(v_1, \dots, v_\ell)$ such that
$$(*) \quad u_0 \nabla f(\bar{x}) + \sum_{i=1}^{p} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x}) = 0,$$
with $u_0 + u_1 + \dots + u_p = 1$ and $u_0, u_1, \dots, u_p \geq 0$.

Define $u_{p+1}, \dots, u_m = 0$. Then

$(*)$ is equal to the condition in the theorem.

Moreover $u_i \, g_i(\bar{x}) = 0$ is indeed true.
$i = 1, \dots, m$ ▨

Called COMPLEMENTARY SLACKNESS CONDITION

(KKT)

Theorem 5: Karush-Kuhn-Tucker Necessary Conditions)

Let $\bar{x}$ be a feasible solution of $(P): \begin{cases} \min_x & f(x) \\ \text{s.t.} & g_i(x) \leq 0 \quad i=1,\dots,m \\ & h_i(x) = 0 \quad i=1,\dots,l \end{cases}$

Let $I = \{i : g_i(\bar{x}) = 0\}$.

Further, suppose that

$$\nabla h_1(\bar{x}), \dots, \nabla h_l(\bar{x}), (\nabla g_i(x))_{i \in I}$$

are linearly independent.

LINEAR INDEPENDENCE CONSTRAINT QUALIFICATION (LICQ)

If $\bar{x}$ is a local minimum, there exists (u,v) such that

$$\nabla f(\bar{x}) + \nabla g(\bar{x})^T u + \nabla h(\bar{x})^T v = 0,$$

$$u \geq 0,$$

$$u_i g_i(\bar{x}) = 0 \quad , \quad i=1,\dots,m.$$

Proof: $\bar{x}$ must satisfy the Fritz John conditions.

Case $u_0 > 0$: Redefine $u^{KKT} = u^{FJ}/u_0$, $v^{KKT} = v^{FJ}/u_0$ ✓

Case $u_0 = 0$:

$$\underbrace{\sum_{i=1}^{m} u_i \nabla g_i(\bar{x})}_{= \sum_{i \in I} u_i \nabla g_i(\bar{x})} + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0$$

$= \sum_{i \in I} u_i \nabla g_i(\bar{x})$ since $u_i = 0 \ \forall i \notin I$, due to complementary slackness.

This would mean linear dependence violating LICQ. ∎

<u>Theorem 6</u>: KKT sufficient Conditions for CONVEX PROBLEMS

Suppose $(P): \begin{cases} \min_x f(x) \\ \text{s.t. } g(x) \leq 0 \\ h(x) = 0 \end{cases}$ is a convex problem (f, g convex and h affine fcns)

Let $\bar{x}$ be a feasible solution of (P) and suppose $\bar{x}$ together with multipliers (u, v) satisfies

$$\nabla f(\bar{x}) + \nabla g(\bar{x})^T u + \nabla h(\bar{x})^T v = 0$$

$$u \geq 0$$

$$u_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m,$$

then $\bar{x}$ is a global optimal solution of (P).

<u>Proof</u>: $\{x : g(x) \leq 0\}$ and $\{x : h(x) = 0\}$ are both convex sets when g convex & h affine.

Let $I := \{i : g_i(\bar{x}) = 0\}$ and let x be any other feasible point different from $\bar{x}$. Then <u>for all $i \in I$</u>:

$$g_i(x) \leq 0 = g_i(\bar{x}).$$

By convexity of g_i, $(i \in I)$, it holds that

$$\nabla g_i(\bar{x})^T (x - \bar{x}) \leq 0 \quad \forall i \in I. \quad \circledast \quad (A)$$

For the equality constraints:

$$\nabla h_i(\bar{x})^T (x - \bar{x}) = 0 \quad \forall i = 1, \dots, l \quad (B)$$

$\circledast$ $g_i(x) \geq g_i(\bar{x}) + (\nabla g)(\bar{x})^T (x - \bar{x}) \quad \forall x \in \mathcal{F} \Rightarrow \overbrace{g_i(x) - g_i(\bar{x})}^{0 \geq} \geq \nabla g(\bar{x})^T (x - \bar{x}) \quad \forall x \in \mathcal{F}$

Transposing the stationarity condition and right-multiplying $(x-\bar{x})$:

$$\nabla f(\bar{x})^T(x-\bar{x}) = \underbrace{\left(-\sum_{i\in I} u_i \nabla g_i(\bar{x}) - \sum_{i=1}^{\ell} v_i \nabla h_i(\bar{x})\right)^T (x-\bar{x})}_{\text{from (A) \& (B)}}$$

Thus $\nabla f(\bar{x})^T(x-\bar{x}) \geq 0$ for all x in the feasible set $(x \neq \bar{x})$

By Proposition 2, $\bar{x}$ is a global optimal solution.
(part (b))

$\blacksquare$