Lecture 16. Karush-Kuhn-Tucker Conditions

16.1 Active constraints in inequality constrained problems

We now consider general optimization problems with inequality constraints

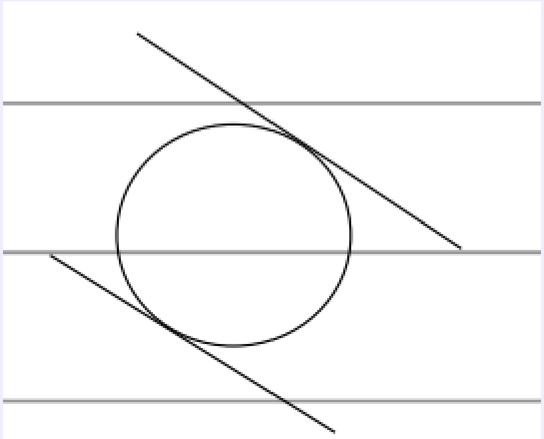
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0 \quad 1 \le i \le m\,, \\ & h_j(x) \le 0 \quad 1 \le j \le \ell\,. \end{aligned}$$

First, we study the optimality condition.

Example

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 \le 2 \end{aligned}$$

The feasible set of the above problem and the level sets of the objective function can be sketched as follows.



- Is $\begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}$ optimal? *No.*

  It satisfies $x_1^2 + x_2^2 = 2$ and is a regular point, but it does not satisfy the Lagrange multiplier condition. So it is even not optimal in the set $\{(x_1, x_2)^\top \mid x_1^2 + x_2^2 = 2\}$, which is a subset of the feasible set.

- Is $\begin{pmatrix}1\\1\end{pmatrix}$ optimal? *Possible.*

  At least it is optimal in the set $\{(x_1, x_2)^\mathsf{T} \mid x_1^2 + x_2^2 = 2\}$ because it is regular and has Lagrange multipliers.
- Is $\begin{pmatrix}-1\\-1\end{pmatrix}$ optimal? *Possible* for the same reason as $\begin{pmatrix}1\\1\end{pmatrix}$.
- Is $\begin{pmatrix}0\\0\end{pmatrix}$ optimal? *No.*

  It satisfies $x_1^2 + x_2^2 < 2$. Then, there exists $\varepsilon > 0$, such that for any $\begin{pmatrix}x_1\\x_2\end{pmatrix} \in \mathcal{B}(\mathbf{0}, \varepsilon)$, $x_1^2 + x_2^2 \leq 2$. If it is optimal, then it must be a local minimum in $\mathcal{B}(\mathbf{0}, \varepsilon)$. However, $\nabla f(0, 0) \neq \mathbf{0}$, which shows that it is not a local minimum.

From this example, we can find that different constraints provide different requirements. We have the following definition to distinguish them.

Definition (*Active and inactive constraints*)

Given $x_0 \in \Omega$, if a constraint $h_j(x) \leq 0$ is tight at x_0, namely, $h_j(x_0) = 0$, then it is called an *active constraint*, otherwise it is called an *inactive constraint*. Denote by $J(x_0) \triangleq \{j \mid h_j(x_0) = 0\}$ the set of indices of active constraints at x_0 .

16.2 Karush-Kuhn-Tucker conditions

If x^* is an optimal solution to

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0, 1 \leq i \leq m \\ & h_i(x) \leq 0, 1 \leq i \leq \ell, \end{aligned}$$

then x^* is also optimal to

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0, 1 \leq i \leq m \\ & h_i(x) = 0, 1 \leq i \leq \ell. \end{aligned}$$

If x^* is a regular point, then there exists λ^*, μ^*, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J(x^*)} \mu_j^* \nabla h_j(x^*) = 0.$$

If $j \notin J(x^*)$ (inactive), we set $\mu_j^* = 0$. Then we can rewrite above statement as follows. There exists $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^k$, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{\ell} \mu_j^* \nabla h_j(x^*) = 0$$

and for any j, $\mu_j^* h_j(x^*) = 0$.

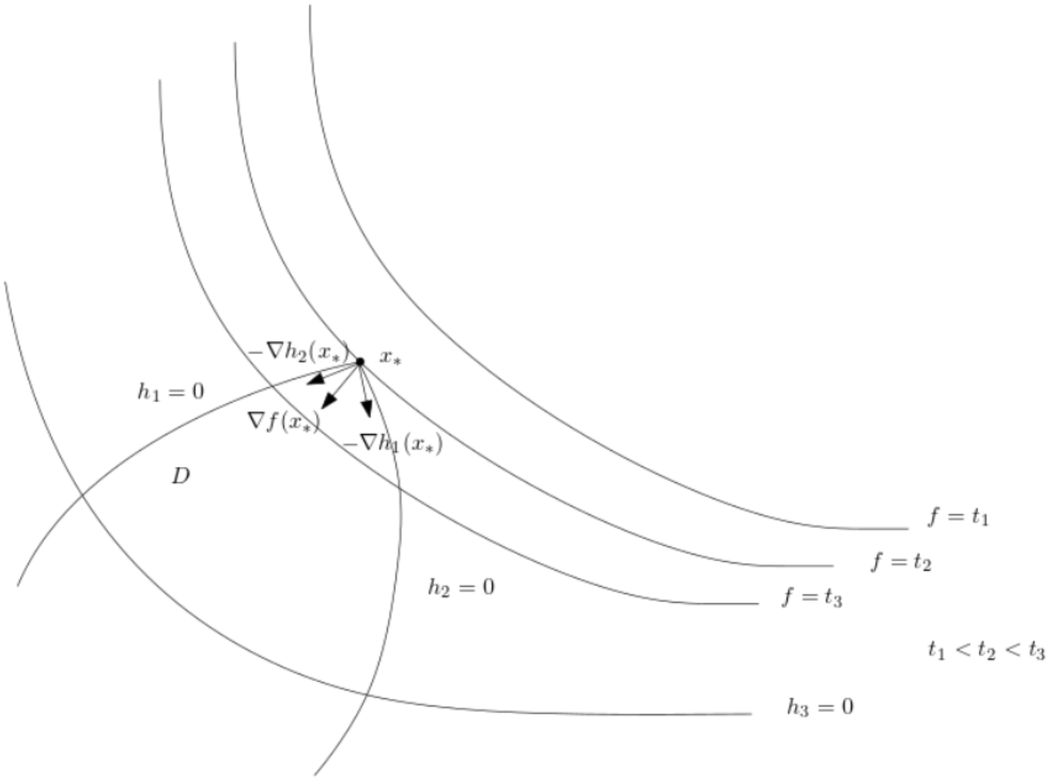
Consider the above example, there are two solutions $\begin{pmatrix}-1\\-1\end{pmatrix}$ and $\begin{pmatrix}1\\1\end{pmatrix}$ having such multipliers. However, only $\begin{pmatrix}-1\\-1\end{pmatrix}$ is optimal. We would like to rule out $\begin{pmatrix}1\\1\end{pmatrix}$.

Note that $f(x_1, x_2) = x_1 + x_2$ and $h(x_1, x_2) = x_1^2 + x^2 - 2$. So $\nabla f = \begin{pmatrix}1\\1\end{pmatrix}$ and $\nabla h = \begin{pmatrix}2x_1\\2x_2\end{pmatrix}$. Then

- for $\begin{pmatrix}1\\1\end{pmatrix}$, $\nabla f - \frac{1}{2}\nabla h = 0$.
- for $\begin{pmatrix}-1\\-1\end{pmatrix}$, $\nabla f + \frac{1}{2}\nabla h = 0$.

  We may force $\mu \geq 0$ to rule out $\begin{pmatrix}-1\\-1\end{pmatrix}$.

Intuitively, the requirement $\mu \geq 0$ is reasonable, since we hope $f(x) \geq f(x^*)$ and $h(x) \leq 0$ in the feasible set, namely, we hope $\nabla h(x^*)$ point outside the feasible set and $\nabla f(x^*)$ point inside it.



Now we can introduce the *Karush-Kuhn-Tucker conditions*.

Theorem (*Karush-Kuhn-Tucker conditions*)

Suppose x^* is a local minimum point of an inequality constrained problem

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0, 1 \le i \le m \\ & h_i(x) = 0, 1 \le i \le \ell. \end{aligned}$$

If x^* is regular for all equality constraints and active inequality constraints, then there exists Lagrange / KKT multipliers $\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_\ell^*$ such that

1. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{\ell} \mu_j^* \nabla h_j(x^*) = \mathbf{0}$.
2. $\mu_j^* h_j(x^*) = 0$, for all $j = 1, \ldots, \ell$.
3. $\mu_j^* \ge 0$ for all $j = 1, \ldots, \ell$.
4. $g_i(x^*) = 0$ for all $i = 1, \ldots, m$, and $h_j(x^*) \le 0$ for all $j = 1, \ldots, \ell$.

We can use KKT conditions to solve optimization problems.

Example 1

$$
\begin{aligned}
\min \quad & x_1^2 + x_2^2 \\
\text{subject to} \quad & x_1 + x_2 = 1 \\
& x_2 \le \alpha
\end{aligned}
$$

If $\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix}$ is optimal, then there are KKT multipliers such that

$$
\begin{cases}
2x_1^* + \lambda = 0 \\
2x_2^* + \lambda + \mu = 0 \\
\mu \ge 0 \\
\mu(x_2^* - \alpha) = 0 \\
x_1^* + x_2^* = 1 \\
x_2^* \le \alpha
\end{cases}
$$

which implies that

$$
2x_1^* + 2x_2^* + 2\lambda + \mu = 0
$$

and further gives that $2\lambda + \mu = -2$. So we have

$$
\begin{cases}
x_1^* = \frac{1}{2} + \frac{\mu}{4} \\
x_2^* = \frac{1}{2} - \frac{\mu}{4}
\end{cases}.
$$

- Case 1. $\alpha > \frac{1}{2}$. From the constraint of x_2 we have $x_2^* = \frac{1}{2} - \frac{\mu}{4} \le \alpha$, which is always true as long as $\mu \ge 0$. Since $\mu(x_2^* - \alpha) = 0$, we have $\mu = 0$, which gives that

$$
\begin{cases}
x_1^* = \frac{1}{2} \\
x_2^* = \frac{1}{2}
\end{cases}
$$

- Case 2. $\alpha = \frac{1}{2}$. $x_2^* = \frac{1}{2} - \frac{\mu}{4} \le \alpha$ is always true as long as $\mu \ge 0$. Then $\mu = 0$ or $x_2^* = \alpha = \frac{1}{2}$ since $\mu(x_2^* - \alpha) = 0$. Both of them imply that

$$
\begin{cases}
x_1^* = \frac{1}{2} \\
x_2^* = \frac{1}{2}
\end{cases}
$$

- Case 3. $\alpha < \frac{1}{2}$. $x_2^* = \frac{1}{2} - \frac{\mu}{4} \le \alpha \implies \mu \ge 2 - 4\alpha > 0 \implies x_2^* = \alpha$ since $\mu(x_2^* - \alpha) = 0$. Then

$$
\begin{cases}
x_1^* = 1 - \alpha \\
x_2^* = \alpha
\end{cases}
$$

Example 2

$$\begin{aligned} \min \quad & (x_1-2)^2+(x_2-1)^2 \\ \text{subject to} \quad & h_1(x)=x_1^2-x_2 \le 0 \\ & h_2(x)=x_1+x_2-2 \le 0 \end{aligned}$$

The KKT condition is

$$\begin{cases} 2(x_1-2)+2\mu_1 x_1+\mu_2=0 \\ 2(x_2-1)-\mu_1+\mu_2=0 \\ \mu_1 h_1(x)=0 \\ \mu_2 h_2(x)=0 \\ h_1(x), h_2(x) \le 0 \\ \mu_1, \mu_2 \ge 0 \end{cases}$$

- Case 1. Both h_1 and h_2 are inactive. Then $\mu_1=\mu_2=0$. So the solution is

$$\begin{cases} x_1=2 \\ x_2=1 \end{cases}$$

  However, the solution is infeasible.
- Case 2. h_1 is inactive and h_2 is active. Then

$$\begin{cases} \mu_1=0 \\ x_1+x_2-2=0 \end{cases} \implies \begin{cases} \mu_2=1 \\ x_1=\frac{3}{2} \\ x_2=\frac{1}{2} \end{cases}$$

  However, the solution is infeasible.
- Case 3. h_1 is active and h_2 is inactive. Then

$$\begin{cases} x_1^2-x_2=0 \\ \mu_2=0 \end{cases} \implies \begin{cases} \mu_1>0 \\ x_1>1 \\ x_2>1 \end{cases}$$

  However, the solution is infeasible.
- Case 4. Both h_1 and h_2 are active. Then we have the following two solutions

$$\begin{cases} x_1^2-x_2=0 \\ x_1+x_2=2 \end{cases} \implies \begin{cases} x_1=1 \\ x_2=1 \end{cases} \text{or} \begin{cases} x_1=-2 \\ x_2=4 \end{cases}$$

For the first solution,

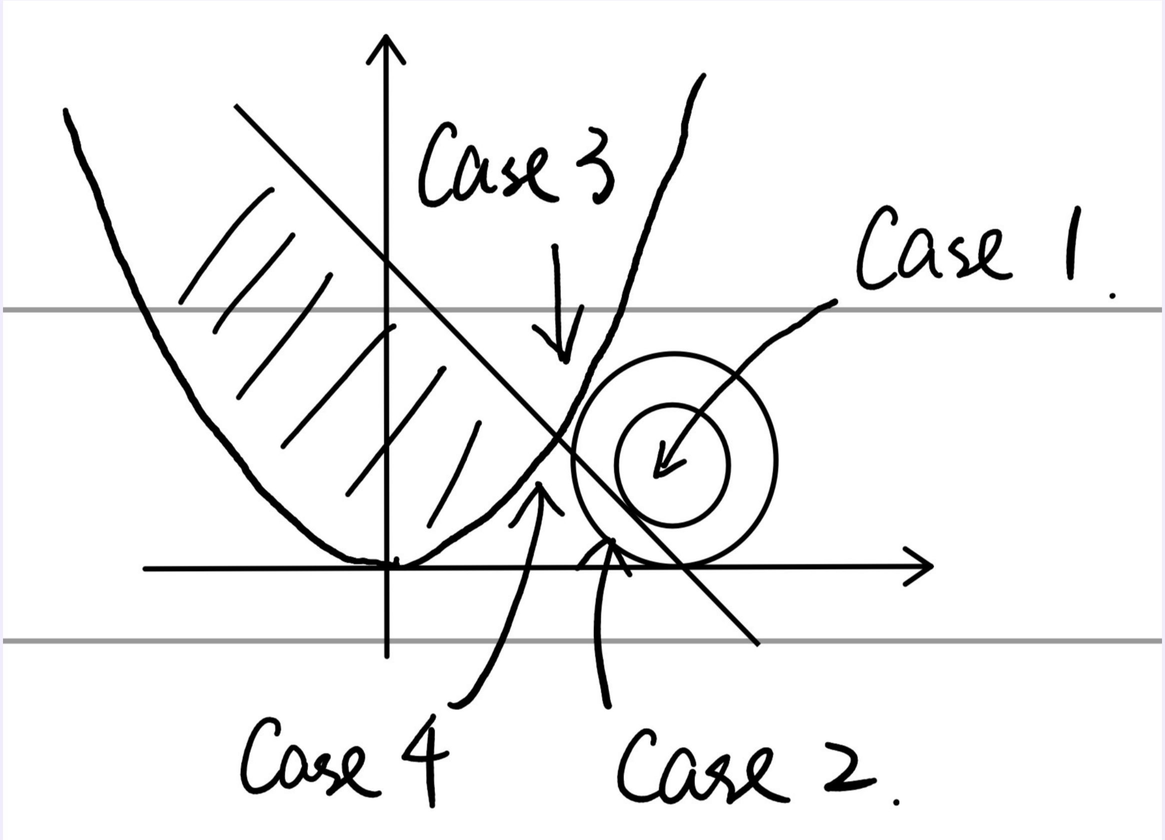
$$\begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases} \implies \begin{cases} -2 + 2\mu_1 + \mu_2 = 0 \\ -\mu_1 + \mu_2 = 0 \end{cases} \implies \begin{cases} \mu_1 = \frac{2}{3} \\ \mu_2 = \frac{2}{3} \end{cases}$$

The solution satisfies the KKT condition.

For the second solution

$$\begin{cases} x_1 = -2 \\ x_2 = 4 \end{cases} \implies \begin{cases} -8 - 4\mu_1 + \mu_2 = 0 \\ 6 - \mu_1 + \mu_2 = 0 \end{cases} \implies \begin{cases} \mu_1 = -\frac{14}{3} \\ \mu_2 = -\frac{32}{3} \end{cases}$$

The solution is invalid.



Remark

KKT condition is possibly unsolved but a critical optimal point exists.

Example 4 (*Linear program*)

$$\begin{aligned} \min \quad & -\boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

The KKT condition is

$$\begin{cases} -\boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{0} \\ \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \geq \boldsymbol{0} \\ \boldsymbol{\mu}_1^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = 0 \\ \boldsymbol{\mu}_2^\top \boldsymbol{x} = 0 \\ \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0} \end{cases}$$

Recall LP duality and complementary slackness:

$$\begin{aligned} \min \quad & \boldsymbol{y}^\top \boldsymbol{b} \\ \text{subject to} \quad & \boldsymbol{y}^\top \boldsymbol{A} \geq \boldsymbol{c}^\top \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

and

$$\begin{cases} (\boldsymbol{y}^*)^\top (\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}) = 0 \\ (\boldsymbol{A}\boldsymbol{y}^* - \boldsymbol{c})^\top \boldsymbol{x}^* = 0 \end{cases}$$

for primal optimal solution $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$. It is easy to see that

$$\boldsymbol{\mu}_1 = \boldsymbol{y}^*, \qquad \boldsymbol{\mu}_2 = \boldsymbol{A}\boldsymbol{y}^* - \boldsymbol{c}$$

are KKT multipliers of $\boldsymbol{x}^*$.

As we mentioned before, if we define the Lagrangian as follows

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\mu}^\top \boldsymbol{h}(\boldsymbol{x})$$

where $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x}))^\top$ and $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_\ell(\boldsymbol{x}))^\top$, then the domain of $\mathcal{L}$ is given by

$$\boldsymbol{x} \in D \triangleq \operatorname{dom} f \cap \operatorname{dom} g_1 \cap \cdots \cap \operatorname{dom} g_m \cap \operatorname{dom} h_1 \cap \cdots \cap \operatorname{dom} h_\ell, \quad \boldsymbol{\lambda} \in \mathbb{R}^m, \quad \boldsymbol{\mu} \in \mathbb{R}^\ell_{\geq 0},$$

and the KKT condition can be expressed as

$$\nabla_{\boldsymbol{x},\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{0}, \quad \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \boldsymbol{0}, \quad (\boldsymbol{\mu}^*)^\top \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$$

for some KKT multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^\ell_{\geq 0}$.

16.3 Necessity and sufficiency of KKT conditions

Now we prove the necessity of KKT conditions. Cleary if $\boldsymbol{x}^*$ is an optimal solution then it must be a local minimum. Consider the following set

$$\tilde{\Omega} \triangleq \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) = 0 \text{ for all } i,\, h_j(\boldsymbol{x}) = 0 \text{ for all } j \in J(\boldsymbol{x}^*), \text{ and } h_j(\boldsymbol{x}) < 0 \text{ for all } j \notin J(\boldsymbol{x}^*)\}\,.$$

It is a subset of the feasible set Ω, and thus $\boldsymbol{x}^*$ must be a local minimum on $\tilde{\Omega}$. If we assume that h_j is continuous for all j, then there exists $\varepsilon > 0$ such that for all $\boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}^*, \varepsilon)$, $h_j(\boldsymbol{x}) < 0$ for all j. So locally we have

$$\tilde{\Omega} \cap \mathcal{B}(\boldsymbol{x}^*, \varepsilon) = \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) = 0 \text{ for all } i, \text{ and } h_j(\boldsymbol{x}) = 0 \text{ for all } j \in J(\boldsymbol{x}^*)\,.$$

Hence, $\boldsymbol{x}^*$ should be a local minimum on the is set. There are only equality constraints. Lagrange condition applies. So there exists KKT multipliers $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_\ell^*$ such that $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^\ell \mu_j^* \nabla h_j(x^*) = \mathbf{0}$ and $\mu_j^* h_j(x^*) = 0$ for all $j = 1, \ldots, \ell$. The remaining part is to show that $\mu_j^* \geq 0$.

Proof of $\mu_j^* \geq 0$ for all $j \in J(x^*)$

We prove this by contradiction. Assume there exists an active $k \in J(x^*)$ and $\mu_k^* < 0$. Then, we consider the set containing all other active constraints

$$\widehat{\Omega} = \{x \mid g_i(x) = 0, i = 1, \cdots, m;\ h_j(x) = 0, j \neq k, j \in J(x^*)\}.$$

If x^* is regular, $T = T_{x^*}\widehat{\Omega}$ is a linear space, where

$$T = \ker \begin{pmatrix} \nabla g_i, & 1 \leq i \leq m \\ \nabla h_j, & k \neq j \in J(x^*) \end{pmatrix}$$

By regularity of x^*, $\nabla h_k(x^*) \notin \operatorname{span}\{\nabla g_i(x^*), \nabla h_j(x^*)\}$ where $i = 1, 2, \ldots, m$ and $j \in J(x^*)$, $j \neq k$. So there exists $v \in T$ such that $\nabla h_k(x^*)^\mathsf{T} v \neq 0$, otherwise above fact does not hold. Without loss of generality, assume $\nabla h_k(x^*)^\mathsf{T} v < 0$. Now we consider the Lagrange condition

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J(x^*)} \mu_j^* \nabla h_j(x^*) = 0.$$

Multiplying by v, we have

$$\Big(\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J(x^*)} \mu_j^* \nabla h_j(x^*)\Big)^\mathsf{T} v = 0.$$

Note that $\nabla g_i(x^*)^\mathsf{T} v = 0$ and $\nabla h_k(x^*)^\mathsf{T} v = 0$ if $j \neq k$. Then,

$$\nabla f(x^*)^\mathsf{T} v + \mu_k^* \nabla h_k(x^*)^\mathsf{T} v = 0 \implies \nabla f(x^*)^\mathsf{T} v < 0.$$

Since $v \in T$, then there exists $\gamma : (-\varepsilon, \varepsilon) \to \widehat{\Omega}$ such that $\gamma(0) = x^*$ and

$\gamma'(0) = v$. Then,

$$\begin{cases} f'(\gamma(t))|_{t=0} = \nabla f(\gamma(0))^\mathsf{T} \gamma'(0) = \nabla f(x^*)^\mathsf{T} v < 0 \\ h_k'(\gamma(t))|_{t=0} = \nabla h_k(\gamma(0))^\mathsf{T} \gamma(0) = \nabla h_k(x^*)^\mathsf{T} v < 0 \end{cases}$$

which leads to

$$\begin{cases} \exists\, \varepsilon_0 > 0,\ \forall\, 0 < \varepsilon \le \varepsilon_0,\ f(\gamma(\varepsilon)) < f(\gamma(0)) = f(x^*) \\ \exists\, \delta_0 > 0,\ \forall\, 0 < \delta \le \delta_0,\ h_k(\gamma(\delta)) < h_k(\gamma(0)) = h_k(x^*) \\ \exists\, \xi_0 > 0,\ \forall\, 0 < \xi \le \xi_0,\ h_j(\gamma(\xi)) \le 0 \text{ for any } j \notin J(x^*) \end{cases}.$$

Now we obtain that for $x' \in \gamma(\min\{\varepsilon_0, \delta_0, \xi_0\})$,

$$\begin{cases} h_k(x') < h_k(x^*) \le 0 \\ f(x') < f(x^*) \\ x' \in \widehat{\Omega} \\ h_j(x') \le 0 \text{ for any } j \notin J(x^*) \end{cases},$$

which contradicts to that x^* is optimal. Thus we conclude $\mu_j^* \ge 0$.

KKT condition is a necessary condition for optimization problems. For convex optimization problems, as we showed for equality constrained problems, it is also sufficient.

Theorem

For a convex optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0, 1 \le i \le m \\ & h_j(x) \le 0, 1 \le j \le \ell \end{aligned}$$

If x^* is feasible and there exist KKT multipliers λ^*, μ^* such that KKT condition holds, then x^* is an optimal solution.

Proof

It suffices to show that for any feasible x, $\nabla f(x^*)^\mathsf{T}(x - x^*) \ge 0$ since $f(x) \ge f(x^*) + \nabla f(x^*)^\mathsf{T}(x - x^*)$.

By KKT condition, $\nabla f(x^*) = \sum_{i=1}^{m} -\lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{\ell} -\mu_j^* \nabla h_j(x^*)$.

We claim that $\nabla g_i(x^*)^\mathsf{T}(x - x^*) = 0$ for all i and $\nabla h_j(x^*)^\mathsf{T}(x - x^*) \le 0$ for all

j. Note that

$$\begin{cases} \forall\, i,\ g_i \text{ is affine, so } \ g_i(x) = g_i(x^*) = 0 \implies \nabla g_i(x^*)^\mathsf{T}(x - x^*) = 0; \\ \forall\, j \notin J(x^*),\ \mu_j^* = 0; \\ \forall\, j \in J(x^*),\ h_j(x^*) = 0, h_j(x) \leq 0 \implies \nabla h_j(x^*)^\mathsf{T}(x - x^*) \leq h_j(x) - h_j(x^*) \leq \end{cases}$$

Hence, we conclude that $\nabla f(x^*)^\mathsf{T}(x - x^*) \geq 0$.