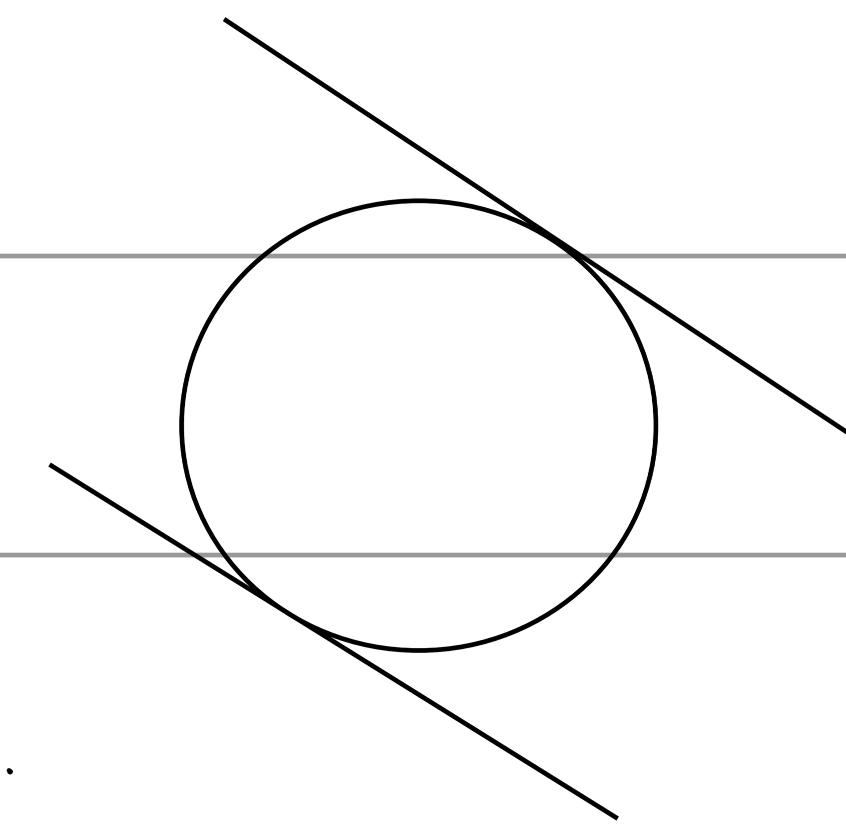
Lecture 19. Inequality constraints and KKT condition.

Consider the optimization: $\min f(x)$ s.t. $g_1(x)=0 \ldots g_m(x)=0$, $h_1(x)\le 0 \ldots h_\ell(x)\le 0$.

Question: How to certify optimal solutions?



Example: $\min x_1+x_2$ s.t. $x_1^2+x_2^2\le 2$

Is $\begin{pmatrix}\sqrt{2}\\0\end{pmatrix}$ optimal? $\begin{pmatrix}\sqrt{2}\\0\end{pmatrix}$ satisfies $x_1^2+x_2^2=2$.

If it is optimal under constraints $x_1^2+x_2^2\le 2$.

it must be optimal under constraints $x_1^2+x_2^2=2$.

$\begin{pmatrix}\sqrt{2}\\0\end{pmatrix}$ is regular but does not satisfy multiplier condition

Is $\begin{pmatrix}1\\1\end{pmatrix}$ optimal? It is regular and has multipliers. same as $\begin{pmatrix}-1\\-1\end{pmatrix}$.

Is $\begin{pmatrix}0\\0\end{pmatrix}$ optimal? It satisfies $x_1^2+x_2^2<2$. So $\exists\, \varepsilon>0$. $\forall \begin{pmatrix}x_1\\x_2\end{pmatrix}$

$\in \mathcal{B}\left(\begin{pmatrix}0\\0\end{pmatrix},\varepsilon\right)$. $x_1^2+x_2^2<2$. $\Rightarrow \begin{pmatrix}0\\0\end{pmatrix}$ must be a local minimum.

if no constraints. But it is not since $\nabla f(0,0)\ne\vec{0}$.

Given x. let $J(x)=\{j : h_j(x)=0\}$ called active constraints.

If x^* is an optimal solution to $\min f(x)$ s.t. $g_1(x)=0 \ldots g_m(x)=0$, $h_1(x)\le 0 \ldots h_m(x)\le 0$

then x^* is also optimal to $\min f(x)$. s.t. $g_i(x)=0 \quad i=1,2,\ldots,m$; $h_j(x)=0. \quad j\in J(x^*)$.

If x^* is further regular, then $\exists\, \lambda^*. \mu^*$. s.t. $\nabla f+\sum_{i=1}^m \lambda_i^*\nabla g_i+\sum_{j\in J}\mu_j^*\nabla h_j=0$.

Also, we can set $\mu_j = 0$ if h_j is inactive ($h_j(x) < 0$). Then

$\exists \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^l$ s.t. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla h_j(x^*) = 0$.

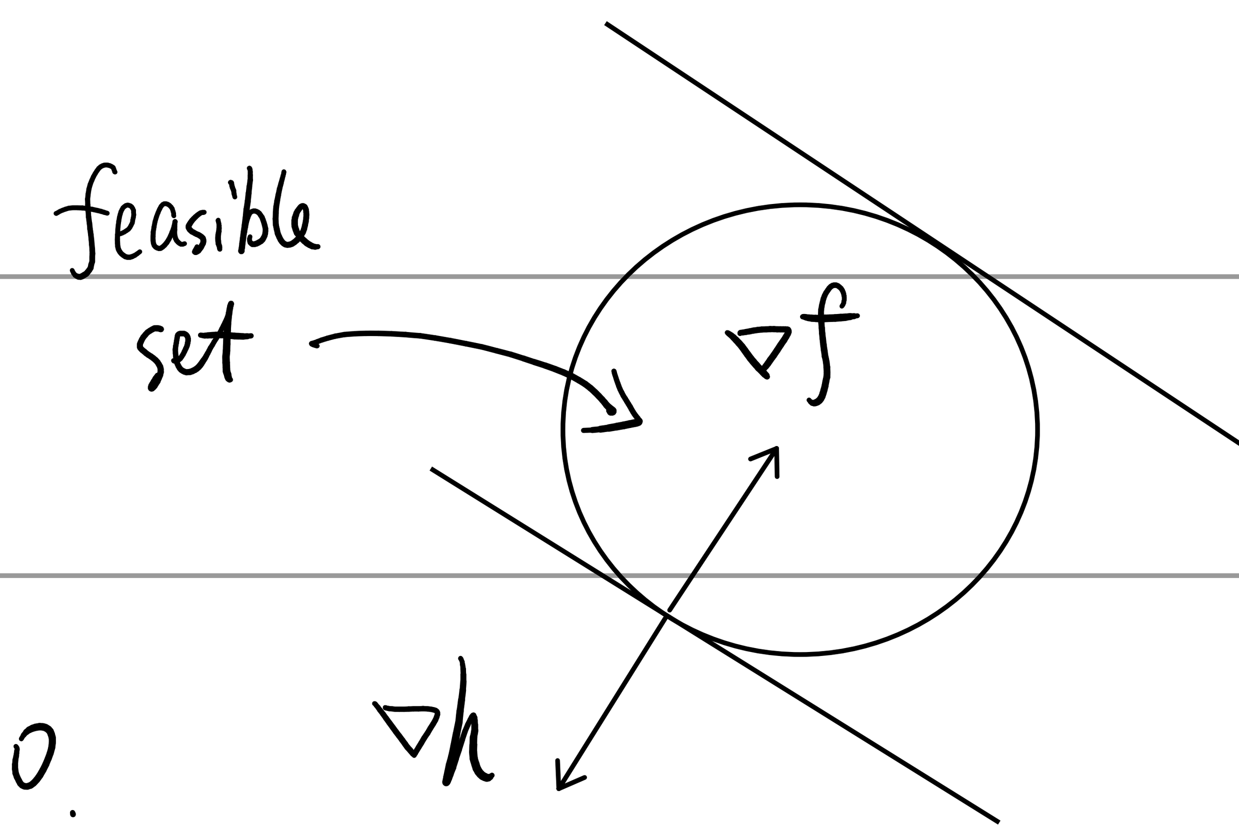
and $\forall j$, either $\mu_j^* = 0$, or $h_j(x^*) = 0$, namely $\mu_j^* h_j(x^*) = 0$.

Suffice? for previous example, we may want to exclude $\binom{1}{1}$.

$f(x_1, x_2) = x_1 + x_2$, $\nabla f = \binom{1}{1}$.

$h(x_1, x_2) = x_1^2 + x_2^2 - 2$, $\nabla h = \binom{2x_1}{2x_2}$.

$\binom{1}{1}$: $\nabla f - \frac{1}{2} \nabla h = 0$, $\binom{-1}{-1}$: $\nabla f + \frac{1}{2} \nabla h = 0$.



Note that ∇h should point out of Ω, while we hope ∇f point into it.

So $\mu \geq 0$ is reasonable. $\binom{1}{1}$ is excluded since $\mu = -\frac{1}{2} < 0$.

Theorem (KKT condition).

Suppose x^* is a local minimum point of an ICP (inequality constrained problem). If x^* is regular for equality and active inequality constraints, then there exist Lagrange / KKT multipliers $\lambda_1^* \dots \lambda_m^*$, $\mu_1^*, \dots \mu_l^*$ s.t.

1. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla h(x^*) = \vec{0}$
2. $\mu_j^* h_j(x^*) = 0$, $\forall j = 1, \dots, l$.
3. $\mu_j^* \geq 0$, $\forall j = 1, \dots l$.

4. $g_i(x^*) = 0 \quad \forall i = 1, \dots, m. \quad h_j(x^*) \le 0 \quad \forall j = 1, \dots, l.$

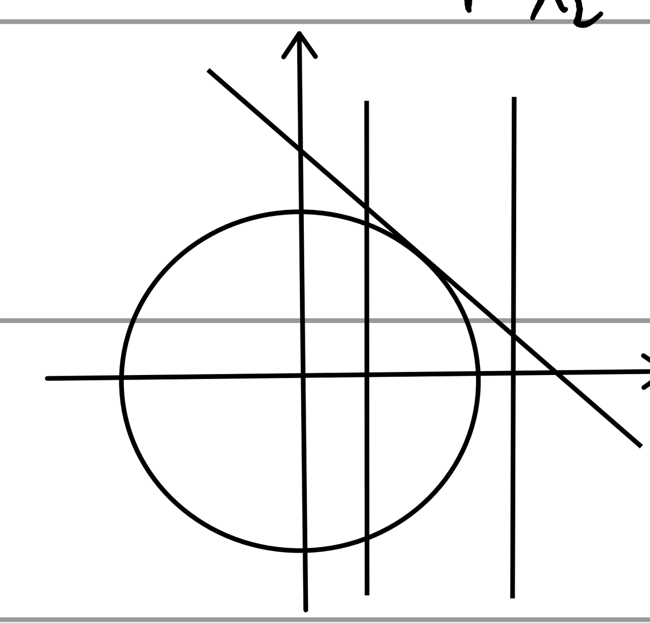
Example. $\min \ x_1^2 + x_2^2 \quad \text{s.t.} \ x_1 + x_2 = 1, \ x_2 \le \alpha.$

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

$2x_1^* + 2x_2^* + 2\lambda + \mu = 0 \Rightarrow 2\lambda + \mu = -2 \Rightarrow \begin{cases} x_1^* = 1/2 + \mu/4 \\ x_2^* = 1/2 - \mu/4 \end{cases}$

Case 1. $\alpha > \frac{1}{2}$. $\frac{1}{2} - \frac{\mu}{4} < \alpha \Rightarrow \mu = 0$ since $\mu(x_2^* - \alpha) = 0$. $\begin{cases} x_1^* = \frac{1}{2} \\ x_2^* = \frac{1}{2} \end{cases}$

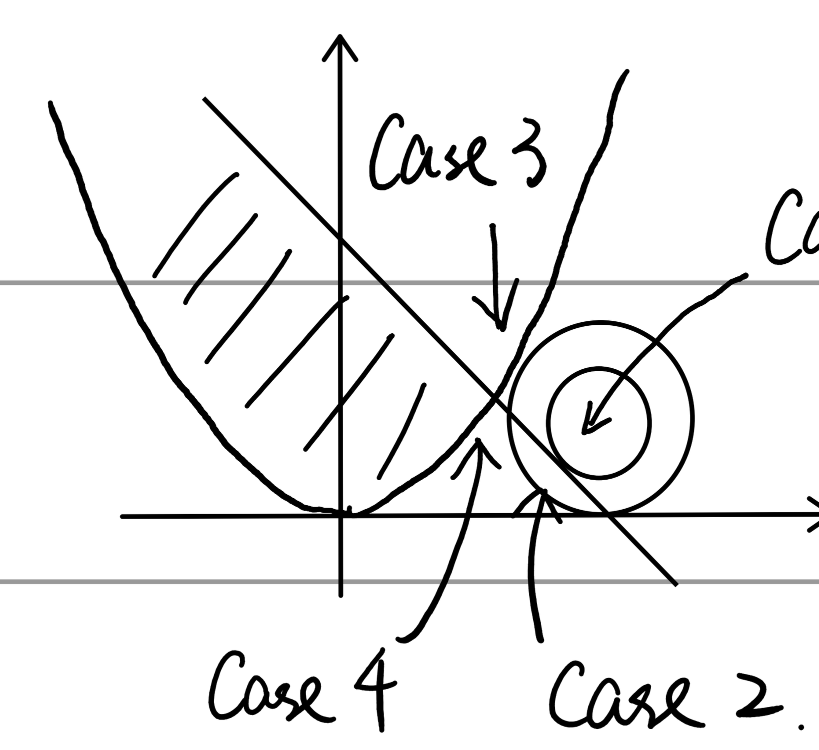
Case 2. $\alpha < \frac{1}{2}$. $\frac{1}{2} - \frac{\mu}{4} \le \alpha \Rightarrow \mu \ge 2 - 4\alpha > 0$



$\Rightarrow x_2^* = \alpha$ since $\mu(x_2^* - \alpha) = 0$. $\begin{cases} x_1^* = 1 - \alpha \\ x_2^* = \alpha \end{cases}$

Case 3. $\alpha = \frac{1}{2}$ $\quad \mu = 0$ or $x_2^* = \alpha = \frac{1}{2}$ since $\mu(x_2^* - \alpha) = 0$. $\begin{cases} x_1^* = \frac{1}{2} \\ x_2^* = \frac{1}{2} \end{cases}$

Example $\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$



subject to $\begin{matrix} g_1(x) = x_1^2 - x_2 \le 0 \\ g_2(x) = x_1 + x_2 - 2 \le 0. \end{matrix}$

KKT condition: $\begin{cases} 2(x_1 - 2) + 2\mu_1 x_1 + \mu_2 = 0 \\ 2(x_2 - 1) - \mu_1 + \mu_2 = 0 \end{cases}$

Case 1. both g_1, g_2 are inactive $\mu_1 = \mu_2 = 0$. $\begin{cases} x_1 = 2 \\ x_2 = 1 \end{cases}$ infeasible.

Case 2. g_1 inactive, g_2 active. $\begin{cases} \mu_1 = 0 \\ x_1 + x_2 = 2 \end{cases} \Rightarrow \begin{cases} \mu_2 = 1 \\ x_1 = 3/2 \\ x_2 = 1/2 \end{cases}$ infeasible.

Case 3. g_1 active, g_2 inactive. $\begin{cases} \mu_2 = 0 \\ x_1^2 = x_2 \end{cases} \Rightarrow \begin{cases} \mu_1 > 0 \\ x_1, x_2 > 1. \end{cases}$ infeasible.

Case 4. both g_1, g_2 are active. $\begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases} \Rightarrow \begin{cases} -2 + 2\mu_1 + \mu_2 = 0 \\ -\mu_1 + \mu_2 = 0 \end{cases}$

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

Example. (nonregularity). $\min\ x_1+x_2$ s.t. $(x_1-1)^2+x_2^2 \le 1$, $(x_1+1)^2+x_2^2 \le 1$.

KKT condition: $\begin{cases} 1+2\mu_1(x_1-1)+2\mu_2(x_1+1)=0 \\ 1+2\mu_1 x_2+2\mu_2 x_2=0 \end{cases}$ infeasible for $\begin{pmatrix}0\\0\end{pmatrix}$

Remark: KKT condition is possibly unsolved but critical optimal exists.

Example. (LP). $\min\ -c^T x$. s.t. $Ax \le b$, $x \ge 0$.

KKT condition: $\begin{cases} -c+A^T\mu_1-\mu_2=0 & \mu_1^T(Ax-b)=0 \\ \mu_1.\ \mu_2 \ge 0 & \mu_2^T x=0. \end{cases}$

Recall LP duality: $\min\ y^T b$. s.t. $y^T A \ge c^T$, $y \ge 0$.

Complementary slackness: x, y optimal for (P). (D). iff $\begin{cases} y^T(Ax-b)=0 \\ (y^TA-c^T)x=0. \end{cases}$

Remark: Multipliers are optimal for another optimization. not a coincidence.

Theorem. For a convex optimization problem $\min f(x)$. s.t. $g_i(x)=0$, $h_j(x)\le 0$.

if x^* is feasible, and there exists KKT multipliers λ^*. μ^*. such that

KKT condition holds. then x^* is an optimal solution.

Proof. It suffices to show $\forall$ feasible x. $\nabla f(x^*)^T(x-x^*) \ge 0$.

According to KKT. $\nabla f(x^*) = \sum_{i=1}^{m} -\lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{l} -\mu_j^* \nabla h_j(x^*)$.

We claim that $\nabla g_i(x^*)^T(x-x^*)=0$ and $\forall j \in J(x^*)$ $\nabla h_j(x^*)^T(x-x^*) \le 0$

$\forall i$. g_i is affine. $g_i(x)=g_i(x^*)=0 \Rightarrow \nabla g_i(x^*)^T(x-x^*)=0$.

$\forall j \in J(x^*)$. $h_j(x^*)=0$. $h_j(x)\le 0 \Rightarrow \nabla h_j(x^*)^T(x-x^*) \le 0$ by convexity. $\square$

Proof of $\mu_j^* \geq 0$ in KKT condition: Prove by contradiction.

Assume $\exists k \in J(x^*)$, $\mu_k^* < 0$. consider all other constraints.

$\hat{\Omega} = \{x : g_i(x) = 0,\ i = 1 \dots m,\ h_j(x) = 0,\ k \neq j \in J(x^*)\}$. $T = T_{x^*}\hat{\Omega}$

By regularity of x^*. $\hat{\Omega}$ is a submanifold and T is a Tangent space.

Note that $T = \ker \begin{pmatrix} \nabla g_i \\ \nabla h_j \end{pmatrix}_{\substack{1 \leq i \leq m \\ k \neq j \in J(x^*)}}$ so if $\forall v \in T$. $\nabla h_k(x^*)^T v = 0$

then $\nabla h_k(x^*) \in \text{span}\{\nabla g_i,\ 1 \leq i \leq m,\ \nabla h_j,\ k \neq j \in J(x^*)\}$. contradicts to

regularity of x^*. Thus $\exists v \in T$. $\nabla h_k(x^*)^T v \neq 0$. w.l.o.g assume < 0.

Consider the Lagrange condition (x^* is optimal under active constraints).

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

note that $\forall 1 \leq i \leq m$. $\nabla g_i(x^*)^T v = 0$, $\forall k \neq j \in J(x^*)$. $\nabla h_j(x^*)^T v = 0$.

So $\nabla f(x^*)^T v + \mu_k^* \nabla h_k(x^*)^T v = 0$. By assumption $\nabla f(x^*)^T v < 0$.

Since $v \in T$. $\exists \gamma : (-\varepsilon, \varepsilon) \to \hat{\Omega}$ such that $\gamma(0) = x^*$. $\gamma'(0) = v$.

$\frac{d}{dt}(f \circ \gamma)\big|_{t=0} = \nabla f(\gamma(0))^T \gamma'(0) = \nabla f(x^*)^T v$. $\frac{d}{dt}(h_k \circ \gamma)\big|_{t=0} = \nabla h_k(x^*)^T v$.

Thus. $\left.\begin{cases} \exists \varepsilon_0 > 0.\ \forall 0 < \varepsilon < \varepsilon_0.\ f(\gamma(\varepsilon)) < f(\gamma(0)) = f(x^*). \\ \exists \delta_0 > 0.\ \forall 0 < \delta < \delta_0.\ h_k(\gamma(\delta)) < h_k(\gamma(0)) = h_k(x^*). \\ \exists \xi_0 > 0.\ \forall 0 < \xi < \xi_0.\ h_j(\gamma(\xi)) \leq 0,\ \forall j \notin J(x^*) \end{cases}\right\} \Rightarrow$

$\exists x' = \gamma(\min(\varepsilon_0, \delta_0, \xi_0)/2) \in \hat{\Omega}$ feasible, and $f(x') < f(x^*)$. $\square$.