

Example minimize $\frac{1}{2} (x_1^2 + x_2^2 + x_3^2)$

$$\text{s.t. } x_1 + x_2 + x_3 = 3$$

— (P)

$$f(x) = \frac{1}{2} (x_1^2 + x_2^2 + x_3^2) \quad \text{— strictly convex}$$

$$\mathcal{H} = \{x : x_1 + x_2 + x_3 = 3\} \quad \text{— convex, closed}$$

We showed in Lec 13 using LM's that

$x^* = (1, 1, 1)$ is the unique solution to (P)

We can also apply necessary and sufficient condition for minimizing convex function over closed convex set (lec 10) to verify:

$$\nabla f(x) = (x_1, x_2, x_3).$$

For all $x \in \mathcal{H}$,

$$\nabla f(x^*)^T (x - x^*) = \sum_{i=1}^3 x_i^* (x_i - x^*)$$

$$\begin{aligned} x^* = (1, 1, 1) &\rightarrow = \sum_{i=1}^3 (x_i - 1) \\ &= \sum_{i=1}^3 x_i - 3 = 3 - 3 = 0 \end{aligned}$$

Therefore $x^* = (1, 1, 1)$ is a global min

Uniqueness follows from strict convexity

Example Nonlinear equality constraint

$$\text{minimize} \quad x_1 + x_2$$

$$\text{s.t.} \quad x_1^2 + x_2^2 - 8 = 0$$

$$\nabla f(x) = (1, 1), \quad \nabla^2 f(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\nabla h(x) = (2x_1, 2x_2), \quad \nabla^2 h(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$\nabla h(x) \neq 0$, $\forall x \in \mathcal{H} \Rightarrow$ all $x \in \mathcal{H}$ are regular

x^*, λ satisfy :

$$(x_1^*)^2 + (x_2^*)^2 = 8 \quad - (1)$$

$$1 + 2\lambda x_1^* = 0 \quad - (2)$$

$$1 + 2\lambda x_2^* = 0 \quad - (3)$$

$$(2), (3) \Rightarrow x_1^* = x_2^* = -\frac{1}{2\lambda}$$

$$(1) \Rightarrow (x_1^*)^2 = (x_2^*)^2 = 4$$

Two possible solutions: $x_1^* = x_2^* = 2$

and $x_1^* = x_2^* = -2$

For $x^* = (2, 2)$, $\lambda = -\frac{1}{4}$

For $x^* = (-2, -2)$, $\lambda = \frac{1}{4}$

TWO possible candidates for local min:

$$x^* = (2, 2), \lambda = -\frac{1}{4} \text{ and } x^* = (-2, -2), \lambda = \frac{1}{4}$$

Check second order condition:

$$\begin{aligned}\nabla_{xx}^2 L(x^*, \lambda) &= \nabla^2 f(x^*) + \lambda \nabla^2 h(x^*) \\ &= 0 + 2\lambda I \\ &= \begin{cases} -\frac{1}{2} I \preceq 0 & \text{if } x^* = (2, 2) \\ +\frac{1}{2} I \succ 0 & \text{if } x^* = (-2, -2) \end{cases}\end{aligned}$$

i.e., $x^* = (-2, -2)$ is a (strict) local min
by second order suff. condition

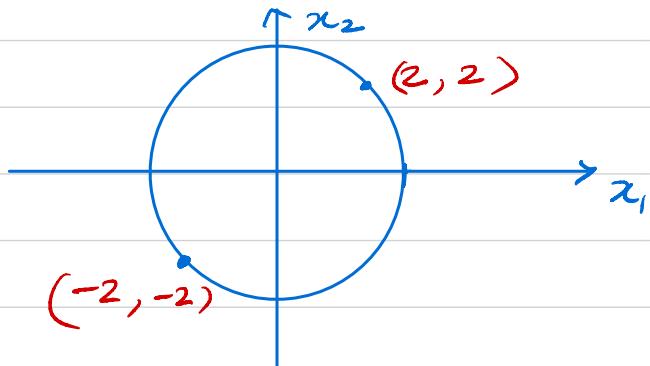
Note that $x^* = (2, 2)$ cannot be a local min
by second order nec. condition

So $x^* = (-2, -2)$ is the unique local min

Is $x^* = (-2, -2)$ also the global min?

Yes, by WT since H is compact.

By considering minimizing $-f(x)$ over H ,
can show that $(2, 2)$ is global max



INEQUALITY CONSTRAINTS

minimize $f(x)$

s.t. $h_i(x) = 0, i=1, \dots, m$ — (ICP)

$g_j(x) \leq 0, j=1, \dots, r$

constraints can be written compactly as:

$$\begin{array}{ll} h(x) = 0 & h : \mathbb{R}^n \rightarrow \mathbb{R}^m \\ g(x) \leq 0 & g : \mathbb{R}^n \rightarrow \mathbb{R}^r \end{array}$$

f, h_i, g_j
cont. diff.

Active v. Inactive Inequality constraints

The constraint $g_j(x) \leq 0$ is said to be active at x if $g_j(x) = 0$, and inactive if $g_j(x) < 0$

$$A(x) = \{j \in \{1, \dots, r\} : g_j(x) = 0\}$$

↳ set of active inequality constraints

Claim If x^* is a local min for (ICP), then x^* is also a local min for

minimize $f(x)$

s.t. $h_i(x) = 0, i=1, \dots, m$ — (ECP)

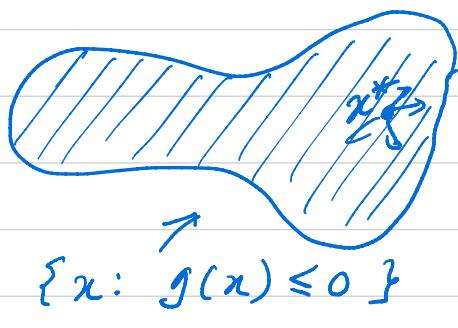
$g_j(x) = 0, \forall j \in A(x^*)$

i.e., we can discard inactive inequality constraints in (ICP) in search for local minima

Intuition for (CP) \rightarrow (ECP)

Consider minimize $f(x)$
 s.t. $g(x) \leq 0, g: \mathbb{R} \rightarrow \mathbb{R}$. — (1)

If x^* is a local min of (1) and $g(x^*) < 0$



By continuity of g ,

$\{x: g(x) < 0\}$ is open

\Rightarrow if $g(x^*) < 0$, then all directions are locally feasible at x^*

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

- constraint does not affect necessary condition.

This argument carries over to multiple $g_j(x^*) < 0$.

\Rightarrow only active inequality constraints at x^* matter

If x^* is regular for (ECP), i.e., if

$$\nabla h_1(x^*), \dots, \nabla h_m(x^*), \nabla g_j(x^*), j \in A(x^*)$$

are linearly independent, then \exists Lagrange multipliers $\lambda_1, \dots, \lambda_m, \mu_j, j \in A(x^*)$ s.t.

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0$$

- Follows from solution in lec 12 - 13.

It also turns out that $\mu_j \geq 0, j \in A(x^*)$.

Intuition for why $\mu_j \geq 0$ if $j \in A(x^*)$

minimize $f(x)$

$$\text{s.t. } h_i(x) = 0, i=1, \dots, m$$

$$g_j(x) \leq 0, j=1, \dots, r$$

If j^{th} constraint $g_j(x) \leq 0$ is relaxed to $g_j(x) \leq u_j$ for $u_j > 0$ (small), then

$$f(x^*(u_j)) \leq f(x^*(0)) = f(x^*)$$

Why? Optimization is over larger set

Now if $j \in A(x^*)$, then $g_j(x^*) = 0$.

By sensitivity interpretation of Lagrange multipliers for equality constraints,

$$f(x^*(u_j)) = f(x^*) - \mu_j u_j + o(u_j)$$

$$\Rightarrow -\mu_j u_j + o(u_j) = f(x^*(u_j)) - f(x^*) \leq 0$$

Dividing by u_j and letting $u_j \rightarrow 0 \Rightarrow \mu_j \geq 0$.

Complementary Slackness

$$\mu_j = 0 \quad \forall j \notin A(x^*)$$

$$\Leftrightarrow \mu_j = 0 \quad \text{whenever } g_j(x^*) < 0, j=1, \dots, r$$

$$\Leftrightarrow \mu_j g_j(x^*) = 0, j=1, \dots, r$$

Karush-Kuhn-Tucker (KKT) Necessary conditions

Lagrangian function for (ICP):

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x)$$

Proposition (KKT) Let x^* be a local min. of (ICP) and assume that x^* is regular for (ECP)

Then \exists unique Lagrange multipliers

$$\lambda = (\lambda_1, \dots, \lambda_m) \text{ and } \mu = (\mu_1, \dots, \mu_r) \text{ s.t.}$$

$$\nabla_x L(x^*, \lambda, \mu) = 0$$

$$\mu_j \geq 0, j = 1, \dots, r$$

$$\mu_j = 0, \forall j \notin A(x^*)$$

If f, h_i, g_j are twice cont. diff., then

$$y^T \nabla_{xx}^2 L(x^*, \lambda, \mu) y \geq 0$$

for all $y \in \mathbb{R}^n$ s.t.

$$\nabla h_i(x^*)^T y = 0, i = 1, \dots, m, \quad \nabla g_j(x^*)^T y = 0, \forall j \in A(x^*)$$

Proof Convert (ICP) to:

minimize $f(x)$

$$\text{s.t. } h_i(x) = 0, i = 1, \dots, m \quad - (P3)$$

$$g_j(x) + \bar{\gamma}_j^2 = 0, j = 1, \dots, r$$

Auxilliary variables $\bar{\gamma} = (\bar{\gamma}_1, \dots, \bar{\gamma}_r)$, $\bar{\gamma}_j \geq 0$

minimize $f(x)$

$$\text{s.t. } h_i(x) = 0, i=1, \dots, m \quad — (P3)$$

$$g_j(x) + \bar{z}_j^2 = 0, j=1, \dots, r$$

(P3) Can be considered as optimization problem over x and $\bar{z} = (\bar{z}_1, \dots, \bar{z}_r)$, with \bar{z} only appearing in constraint.

Let x^* be a local min. for (ICP), then (x^*, \bar{z}^*) is a local min. for (P3) with

$$\bar{z}_j^* = (-g_j(x^*))^{\frac{1}{2}}, j=1, \dots, r.$$

Define Lagrangian for (P3) :

$$\begin{aligned} L(x, \bar{z}, \lambda, \mu) &= f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x) \\ &\quad + \sum_{j=1}^r \mu_j \bar{z}_j^2 \end{aligned}$$

From first-order necessary condition (lec 12), assuming that (x^*, \bar{z}^*) is regular for (P3),

$$\nabla L(x^*, \bar{z}^*, \lambda, \mu) = 0 \quad (\text{Note: } \nabla \text{ w.r.t. both } x \text{ and } \bar{z})$$

$$\Rightarrow \nabla_x f(x^*) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x^*) + \sum_{j=1}^r \mu_j \nabla_x g_j(x^*) = 0$$

$$\text{and } \left. \sum_{j=1}^r \mu_j \nabla_{\bar{z}} (\bar{z}_j^2) \right|_{\bar{z}=\bar{z}^*} = 0$$

$$\nabla_x f(x) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x) + \sum_{j=1}^r \mu_j \nabla_x g_j(x) = 0 \quad -(1)$$

and $\left. \sum_{j=1}^r \mu_j \nabla_z (\bar{z}_j^2) \right|_{\bar{z}=\bar{z}^*} = 0 \quad -(2)$

Note: $\nabla_z (\bar{z}_j^2) = [0, \dots, 0, 2\bar{z}_j, 0 \dots 0]^T$
 $\uparrow j^{\text{th}} \text{ component}$

(1) $\Rightarrow \nabla_x L(x^*, \lambda, \mu) = 0$ with

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x)$$

(2) $\Rightarrow \begin{bmatrix} 2\mu_1 \bar{z}_1^* \\ \vdots \\ 2\mu_r \bar{z}_r^* \end{bmatrix} = 0 \Rightarrow \mu_j \bar{z}_j^* = 0 \text{ for } j = 1, \dots, r.$

But $\bar{z}_j^* = (-g_j(x^*))^{\frac{1}{2}} > 0 \quad \forall j \notin A(x^*)$

$$\Rightarrow \mu_j = 0 \quad \text{for } j \notin A(x^*)$$

Second-order necessary conditions for (P3):

$$\begin{bmatrix} y \\ w \end{bmatrix}^T \nabla^2 L(x^*, \bar{z}^*, \lambda, \mu) \begin{bmatrix} y \\ w \end{bmatrix} \geq 0$$

$\forall y \in \mathbb{R}^n, w \in \mathbb{R}^r$ satisfying

$$\nabla_x h_i(x^*)^T y = 0 \quad i = 1, \dots, m$$

and $\nabla_x g_j(x^*)^T y + 2\bar{z}_j^* w_j = 0, j = 1, \dots, r.$

$$L(x, \bar{z}, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j (g_j(x) + z_j^2)$$

$$\nabla^2 L(x, \bar{z}, \lambda, \mu) = \begin{bmatrix} \nabla_{xx}^2 L(x, \lambda, \mu) & \underbrace{0 \cdots 0}_{nxn} \\ 0 & \ddots & 2\mu_1 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2\mu_r \end{bmatrix}_{(n+r) \times (n+r)}$$

For every $j \in A(x^*)$ select (y, w) with
 $y = 0$, $w_j \neq 0$ and $w_k = 0 \quad \forall k \neq j$

Then (y, w) satisfies $\nabla h_i(x^*)^T y = 0, i = 1, \dots, m$

and $\nabla g_k(x^*)^T y + 2\bar{z}_k^* w_k = 0 \quad \forall k \neq j$

and $\nabla g_j(x^*)^T y + 2\bar{z}_j^* w_j = 0$

$\uparrow = 0$ since $j \in A(x^*)$

Thus for this choice of (y, w)

$$[y \quad w]^T \begin{bmatrix} \nabla_{xx}^2 L(x, \lambda, \mu) & 0 \cdots 0 \\ 0 & \ddots & 2\mu_1 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2\mu_r \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} \geq 0$$

$$\Rightarrow 2\mu_j w_j^2 \geq 0 \Rightarrow \mu_j \geq 0$$

Similarly we can show $\mu_j \geq 0 \quad \forall j \in A(x^*)$.

Now choose $y \in \mathbb{R}^n$ s.t.

$$\nabla h_i(x^*)^T y = 0, i=1, \dots, m, \quad \nabla g_j(x^*)^T y = 0, \quad \forall j \in A(x^*)$$

and w s.t.

$$w_j = \begin{cases} 0 & \text{if } j \in A(x^*) \\ -\frac{\nabla g_j(x^*)^T y}{2\beta_j^*} & \text{if } j \notin A(x^*) \end{cases}$$

$$\text{Then } \nabla g_j(x^*)^T y + 2\beta_j^* w_j = 0 \quad \forall j \notin A(x^*)$$

$$\text{and } \nabla g_j(x^*)^T y + 2\beta_j^* w_j = 0 \quad \forall j \in A(x^*)$$

$$\text{Thus } \nabla_x g_j(x^*)^T y + 2\beta_j^* w_j = 0 \quad \forall j=1, \dots, r$$

$$\begin{aligned} \text{Also, } w_j &= 0 \quad \forall j \in A(x^*) \text{ and } w_j = 0 \quad \forall j \notin A(x^*) \\ \Rightarrow \quad \mu_j w_j &= 0 \quad \forall j=1, \dots, r. \end{aligned}$$

Thus, for the above choice of (y, w) :

$$[y \quad w]^T \begin{bmatrix} \nabla_{xx}^2 L(x, \lambda, \mu) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 2\mu_r \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} \geq 0$$

$$\Rightarrow y^T \nabla_{xx}^2 L(x, \lambda, \mu) y \geq 0$$

QED