LECTURE 6

LAGRANGIAN AND SECOND ORDER OPTIMALITY CONDITIONS

First order optimality conditions can be expressed in terms of the Lagrangian function

$$\mathcal{L}(x, \underline{y}) := f(x) - \sum_{j \in E \cup I} \lambda_j c_j(x).$$

THM

Suppose $x_* \in \mathbb{R}^n$ is a point where LICQ holds. If x_* is a local minimizer of (NLP), there exists a λ_* such that

(i) $\nabla_x \mathcal{L}(x_*, \lambda_*) = 0$

(ii) $(\lambda_*)_j \, c_j(x_*) = 0$ for each $j \in I \cup E$

(iii) $(\lambda_*)_j \geqslant 0$ for each $j \in I$

(iv) $c_j(x_*) = 0$ for each $j \in E$

(v) $c_j(x_*) \geqslant 0$ for each $j \in I$.

Ex

① Linear programs

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^T x$$

$$Ax = b$$

$$x \geq 0$$

$$\boxed{\begin{array}{l}\text{Lagrangian} \\ \mathcal{L}(x, \lambda, s) = \\ c^T x - (Ax - b)^T \lambda - x^T s\end{array}}$$

Suppose x_* is a local minimizer.
There exists λ_* and s_* such that

(i) $c = A^T \lambda_* + s_*$

(ii) $x_*^T s_* = 0$

(iii) $s_* \geq 0$

(iv) $A x_* = b$

(v) $x_* \geq 0$

② Quadratic programs

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} x^T \underset{\substack{\downarrow \\ \text{symmetric}}}{H} x + g^T x$$

$$Ax = b$$

$$\boxed{\begin{array}{l}\text{Lagrangian} \\ \mathcal{L}(x, \lambda) = \\ \left(\frac{1}{2} x^T H x + g^T x\right) - (Ax - b)^T \lambda\end{array}}$$

Suppose x_* is a local minimizer.

There exists λ_* such that

$$\text{(i)} \quad H x_* + g = A^T \lambda_*$$

$$\text{(ii)} \quad A x_* = b$$

equivalently

$$\begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x_* \\ \lambda_* \end{bmatrix} = \begin{bmatrix} -g \\ -b \end{bmatrix}$$

Second order optimality conditions

Suppose (x_*, λ_*) satisfy conditions (i) – (v) on page 1

$$\mathcal{L}(x_*, \lambda_*) = f(x_*) - \sum_{j \in \mathcal{I} \cup \mathcal{E}} (\lambda_*)_j \, c_j(x_*)$$

$$\overset{\text{(due to (ii))}}{=} f(x_*).$$

Moreover, assuming the active constraints remain the same at $x_* + d$ for all d small enough
feasible

$$\mathcal{L}(x_* + d, \lambda_*) = f(x_* + d)$$

Feasible directions (tangent cone) is a subset of

$$S_1 = \{ d \mid \nabla c_j(x_*)^T d = 0 \quad j \in E, \\ \nabla c_j(x_*)^T d \geq 0 \quad j \in I \cap A(x_*) \}$$

Let us also define $S_2 \subseteq S_1$ such that

$$S_2 = \{ d \in S_1 \mid \nabla c_j(x_*)^T d = 0 \text{ if } \\ j \in I \cap A(x_*) \text{ and } \lambda_j > 0 \}.$$

We have for each $d \in S_2$

$$\nabla f(x_*)^T d = \sum_{j \in A(x_*)} \lambda_j \nabla c_j(x_*)^T d$$

$$= \sum_{j \in E} \lambda_j \underbrace{\nabla c_j(x_*)^T d}_{0 \text{ since } d \in S_1}$$

$$+ \sum_{\substack{j \in I \cap A(x_*) \\ \lambda_j = 0}} \underbrace{\lambda_j \nabla c_j(x_*)^T d}_{0}$$

$$+ \sum_{\substack{j \in I \cap A(x_*) \\ \lambda_j > 0}} \lambda_j \underbrace{\nabla c_j(x_*)^T d}_{0 \text{ since } d \in S_2} = 0.$$

But for $d \in S_1 \setminus S_2$, we have $\nabla f(x_*)^T d > 0$.

THM (Second Order Necessary Conditions)

Suppose $x_* \in \mathbb{R}^n$ is a point where LICQ holds. If x_* is a local minimizer of (NLP), then the first order necessary conditions (KKT conditions) hold ~~and~~ for some λ_* s.t.

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d \geq 0$$

for all $d \in S_2$.

THM (Second Order Sufficient Conditions)

Suppose $x_* \in \mathbb{R}^n$ is a point where the first order conditions hold for some λ_* such that

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d > 0$$

for all $d \in S_2 \setminus \{0\}$.

Then x_* is a local minimizer of (NLP).

PROOF

To conclude x_* is a local minimizer, it suffices to show $f(z^{(k)}) \geq f(x_*)$ for all feasible sequences $\{z^{(k)}\}$ and all k sufficiently large.

For any feasible sequence $\{z^{(k)}\}$, we have

$$\mathcal{L}(z^{(k)}, \lambda_*) = f(z^{(k)}) - \sum_{j \in A(x_*)} [\lambda_*]_j \, c_j(z_k)$$

$$\leqslant f(z^{(k)}).$$

Furthermore, ~~for~~ by Taylor's thm,

$$\mathcal{L}(z^{(k)}, \lambda_*) = \mathcal{L}(x_*, \lambda_*) + \nabla_x \mathcal{L}(x_*, \lambda_*)^T (z^{(k)} - x_*) + \frac{1}{2}(z^{(k)} - x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*)(z^{(k)} - x_*) + o(\|z^{(k)} - x_*\|^2)$$

$$= f(x_*) + \frac{1}{2}(z^{(k)} - x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*)(z^{(k)} - x_*) + o(\|z^{(k)} - x_*\|^2)$$

Without loss of generality, assume $\{z^{(k)}\}$ is such that $\left\{(z^{(k)} - x_*)/\|z^{(k)} - x_*\|\right\}$ is convergent and $d = \lim_{k \to \infty} (z^{(k)} - x_*)/\|z^{(k)} - x_*\|$ (Why without loss of generality?)

Note that $(z^{(k)} - x_*) = d\,\|z^{(k)} - x_*\| + o(\|z^{(k)} - x_*\|)$.

We must have $d \in S_1$. There are two possibilities, $d \in S_2$ or $d \in S_1 \setminus S_2$.

First suppose $d \in S_2$.

$$f(z^{(k)}) \geq \mathcal{L}(z^{(k)}, \lambda_*)$$

$$= f(x_*) + \frac{1}{2}(z^{(k)} - x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*)(z^{(k)} - x_*) + o(\|z^{(k)} - x_*\|^2)$$

$$= f(x_*) + \frac{1}{2} \underbrace{d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d \, \|z^{(k)} - x_*\|^2}_{> 0} + o(\|z^{(k)} - x_*\|^2)$$

$$> f(x_*).$$

Next suppose $d \in S_1 \setminus S_2$. In this case, there is a $j \in \mathcal{I} \cap \mathcal{A}(x_*)$ with $\cancel{\lambda_* > 0}$ $[\lambda_*]_j > 0$ s.t.

$$\nabla c_j(x_*)^T d > 0.$$

By Taylor's thm

$$[\lambda_*]_j c_j(z_k) = [\lambda_*]_j c_j(x_*) + [\lambda_*]_j \nabla c_j(x_*)^T (z^{(k)} - x_*) + o(\|z^{(k)} - x_*\|)$$

$$= \underbrace{[\lambda_*]_j c_j(x_*)}_{0 \text{ since } c_j(x_*) = 0} + [\lambda_*]_j \nabla c_j(x_*)^T d \, \|z^{(k)} - x_*\| + o(\|z^{(k)} - x_*\|),$$

so

$$\mathcal{L}(z^{(k)}, \lambda_*) = f(z^{(k)}) - \sum_{l \in \mathcal{A}(x_*)} [\lambda_*]_l c_l(z^{(k)})$$

$$\leq f(z^{(k)}) - [\lambda_*]_j \, c_j(z^{(k)})$$

$$= f(z^{(k)}) - [\lambda_*]_j \nabla c_j(x_*)^T d \, \|z^{(k)} - x_*\| + o(\|z^{(k)} - x_*\|).$$

Consequently,

$$f(z^{(k)}) - [\lambda_*]_j \nabla c_j(x_*)^T d \, \|z^{(k)} - x_*\|$$

$$\geq \mathcal{L}(z^{(k)}, \lambda_*) \geq$$

$$f(x_*) + \frac{1}{2} (z^{(k)} - x_*)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) (z^{(k)} - x_*) + o(\|z^{(k)} - x_*\|^2)$$

$$\Longrightarrow f(z^{(k)}) \geq f(x_*) + \underbrace{[\lambda_*]_j \nabla c_j(x_*)^T d}_{>0} \, \|z^{(k)} - x_*\| + o(\|z^{(k)} - x_*\|)$$

$$\geq f(x_*).$$

$\square$

Examples

①
$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad -2x_1 + x_2$$

$$(1-x_1)^3 - x_2 \geq 0$$

$$x_2 + 0.25x_1^2 - 1 \geq 0$$

Let $x_* = (0,1)$

Notice $A(x_*) = \{1, 2\}$

$$\nabla f(x) = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \quad \nabla c_1(x) = \begin{bmatrix} -3(1-x_1)^2 \\ -1 \end{bmatrix} \quad \nabla c_2(x) = \begin{bmatrix} 0.5x_1 \\ 1 \end{bmatrix}$$

$$\nabla f(x_*) = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \quad \nabla c_1(x_*) = \begin{bmatrix} -3 \\ -1 \end{bmatrix} \quad \nabla c_2(x_*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Check KKT conditions

$$\nabla f(x_*) = \lambda_1 \nabla c_1(x_*) + \lambda_2 \nabla c_2(x_*)$$

$$\Longrightarrow \lambda_1 = 2/3 \qquad \lambda_2 = 5/3$$

Gradient of the Lagrangian

$$\nabla_x \mathcal{L}(x, \lambda_*) = \begin{bmatrix} -2 \\ 1 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} -3(1-x_1)^2 \\ -1 \end{bmatrix} - \frac{5}{3} \begin{bmatrix} 0.5x_1 \\ 1 \end{bmatrix}$$

Its Hessian

$$\nabla^2_{xx} \mathcal{L}(x, \lambda_*) = \begin{bmatrix} 4(x_1-1) - \frac{5}{6} & 0 \\ 0 & 0 \end{bmatrix}.$$

Sets

$$S_1 = \{ d \mid [-3 \ -1]\, d \geq 0 \text{ and } [0 \ 1]\, d \geq 0 \}$$

$$= \left\{ \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \middle| \ d_2 \geq 0 \text{ and } d_2 \leq -3d_1 \right\}$$

$$S_2 = \{ d \mid [-3 \ -1]\, d = 0 \text{ and } [0 \ 1]\, d = 0 \}$$

$$= \{0\}.$$

Second order condition holds trivially.

By second order sufficient conditions

$x_* = (0, 1)$

is a local minimizer.

② minimize $\quad -\frac{1}{2}(x_1-2)^2 + x_2^2$
$x \in \mathbb{R}^2$
$x_1^2 + x_2^2 - 1 \geq 0$

Let $\quad x_* = (1,0)$.

Notice $A(x_*) = \{1\}$.

$$\nabla f(x) = \begin{bmatrix} -(x_1-2) \\ 2x_2 \end{bmatrix} \qquad \nabla c_1(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla f(x_*) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \nabla c_1(x_*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Applying KKT condition

$$\nabla f(x_*) = \lambda_1 \nabla c_1(x_*)$$

$$\Longrightarrow \lambda_1 = 1/2.$$

Gradient of Lagrangian

$$\nabla_x \mathcal{L}(x, \lambda_*) = \begin{bmatrix} 2 - x_1 \\ 2x_2 \end{bmatrix} - \lambda_1 \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 - 2x_1 \\ x_2 \end{bmatrix}$$

Its Hessian

$$\nabla^2_{xx} \mathcal{L}(x, \lambda_*) = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$$

Sets

$$S_1 = \{ d \mid [2 \ \ 0]\, d \geq 0 \}$$

$$= \left\{ \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \,\middle|\, d_1 \geq 0 \right\}$$

$$S_2 = \{ d \mid [2 \ \ 0]\, d = 0 \}$$

$$= \left\{ \begin{bmatrix} 0 \\ d_2 \end{bmatrix} \,\middle|\, d_2 \in \mathbb{R} \right\}.$$

For each $d \in S_2 \setminus \{0\}$

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*)\, d = d_2^2 > 0.$$

By second order conditions $x_* = (1,0)$ is a local minimizer

Projected Hessians

Suppose LICQ holds at x_*.
In this case λ_* is unique. (Why?)

Suppose furthermore strict complementarity holds.

$$[\lambda_*]_j = 0 \quad \text{or} \quad c_j(x_*) = 0 \quad \left(\begin{smallmatrix}\text{but}\\ \text{not}\\ \text{both}\end{smallmatrix}\right)$$
$$\text{for each } j \in A(x_*)$$

In this case,

$$[\lambda_*]_j > 0 \quad \text{for } j \in A(x_*) \cap I$$

and

$$S_2 = \{ d \mid \nabla c_j(x_*)^T d = 0 \ \ j \in E$$
$$\nabla c_j(x_*)^T d = 0 \ \ j \in A(x_*) \cap I \}$$
$$= \Big\{ d \mid \underbrace{\begin{bmatrix} J_E(x_*) \\ J_{IA}(x_*) \end{bmatrix}}_{J_A(x_*) \text{ - Jacobian of active constraints}} d = 0 \Big\}$$
$$= \mathrm{Null}(J_{IA}(x_*)).$$

The second order condition

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d > 0$$
$$\text{for all } d \in S_2 \setminus \{0\}$$

can be expressed as

$$(Z\alpha)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) (Z\alpha) > 0$$
$$\text{for all } \alpha \neq 0$$

where Z is a matrix whose columns form an orthonormal basis for $\text{Null}(\perp_A(x_*))$. In other words,

$$Z^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z \text{ is PD.}$$

Conclusions

1. $$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d \geqslant 0$$
$$\text{for all } d \in S_2$$
$$\Longleftrightarrow$$
$$\underbrace{Z^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z}_{\text{projected Hessian}} \text{ is PSD}$$

(2)

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*)\, d > 0$$

for all $d \in S_2 \setminus \{0\}$

$$\Longleftrightarrow$$

$Z^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z$ is PD.

General case

$$S_2 = \{ d \mid \nabla c_j(x_*)^T d = 0 \quad j \in E,$$
$$\nabla c_j(x_*)^T d = 0 \quad j \in I \cap A(x_*) \text{ and } \lambda_j > 0,$$
$$\nabla c_j(x_*)^T d \geq 0 \quad j \in I \cap A(x_*) \text{ and } \lambda_j = 0 \}$$

$$\underline{S_2} = \{ d \mid \nabla c_j(x_*)^T d = 0 \quad j \in E$$
$$\nabla c_j(x_*)^T d = 0 \quad j \in I \cap A(x_*) \}$$
$$= \mathrm{Null}(J_{IA}(x_*))$$

$$\overline{S_2} = \{ d \mid \nabla c_j(x_*)^T d = 0 \quad j \in E,$$
$$\nabla c_j(x_*)^T d = 0 \quad j \in I \cap A(x_*) \text{ and } \lambda_j > 0 \}$$

Notice that

$$\underline{S_2} \subseteq S_2 \subseteq \overline{S_2}$$

Letting $\underline{Z}$ be a matrix whose columns form orthonormal basis for $\underline{S_2}$.

$\overline{Z}$ be a matrix whose columns form orthonormal basis for $\overline{S_2}$.

① $$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d \geq 0$$
for each $d \in S_2$

$$\Longrightarrow$$

$$(\underline{Z}\alpha)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) (\underline{Z}\alpha) \geq 0$$
for each α

$$\Longrightarrow$$

$$\underline{Z}^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) \underline{Z} \text{ is PSD}$$

② $$\overline{Z}^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) \overline{Z} \text{ is PD}$$

$$\Longrightarrow$$

$$(\overline{Z}\alpha)^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) (\overline{Z}\alpha) > 0$$
for each nonzero α

$$\Longrightarrow$$

$$d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d > 0$$
for each $d \in S_2 \setminus \{0\}$.

Summary

Necessary condition

x_* is a point where LICQ holds

x_* is a local minimizer

$$\Longrightarrow$$

(i) There exists λ_* s.t.

$$\nabla_x \mathcal{L}(x_*, \lambda_*) = 0;$$

(ii) $\underline{Z}^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) \underline{Z}$ is PSD.

Sufficient condition

(i) There exists λ_* s.t.

$$\nabla_x \mathcal{L}(x_*, \lambda_*) = 0$$

(ii) $\overline{Z}^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) \overline{Z}$ is PD

$$\Longrightarrow$$

x_* is a local minimizer.