Video 4 on Saddle points & duality

Recall:

DEFINITION. $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \in X \times U$ *is a* **saddle point** *of the Lagrange function L iff*

$$L(\bar{\boldsymbol{x}}; \boldsymbol{u}, \boldsymbol{v}) \le L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \le L(\boldsymbol{x}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \quad \forall (\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) \in X \times U.$$

LEMMA B. $\bar{\boldsymbol{x}}$ *is feasible and (CS)* $\implies$ $f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$.

LEMMA C. $\bar{\boldsymbol{x}}$ *is feasible and (CS)* $\iff$ $L(\bar{\boldsymbol{x}}; \boldsymbol{u}, \boldsymbol{v}) \le L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \quad \forall (\boldsymbol{u}, \boldsymbol{v}) \in U$.

The right inequality of the saddle-point property states that "$\bar{\boldsymbol{x}}$ is a minimizer of $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$".

THEOREM 1. $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \in X \times U$ *is a saddle point* $\implies$ $\bar{\boldsymbol{x}}$ *solves (P) and (KKT) holds.*

The **dual problem** to (P):

$$\text{(D)} \qquad \begin{aligned} &\text{maximize} && \theta(\boldsymbol{u}, \boldsymbol{v}) := \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) \\ &\text{subject to} && (\boldsymbol{u}, \boldsymbol{v}) \in U. \end{aligned}$$

For feasible points $\boldsymbol{x} \in S$ and $(\boldsymbol{u}, \boldsymbol{v}) \in U$, we have

$$(*) \qquad \theta(\boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{y} \in X} L(\boldsymbol{y}; \boldsymbol{u}, \boldsymbol{v}) \le L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) \underset{\substack{\uparrow \\ \text{Lemma A}}}{\le} f(\boldsymbol{x})$$

Now we continue:

THEOREM 2. *If (P) is (CP), then*

$\bar{\boldsymbol{x}}$ *is feasible and (KKT) holds* $\implies$ $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \in X \times U$ *is a saddle point.*

Then we can use Thm 1 and conclude that $\bar{\boldsymbol{x}}$ is a global solution, but this is nothing new!

THEOREM 3. $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \in X \times U$ *is a saddle point* $\iff$ $\bar{\boldsymbol{x}}$ *is feasible and* $\theta(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = f(\bar{\boldsymbol{x}})$.

PROOF: $\boxed{\Rightarrow}$ Lemma C gives that $\bar{\boldsymbol{x}}$ is feasible and (CS), and Lemma B that $L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = f(\bar{\boldsymbol{x}})$. This equality and the saddle-point property imply

$$\theta(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \min_{\boldsymbol{x} \in X} L(\boldsymbol{x}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = f(\bar{\boldsymbol{x}}).$$

$\boxed{\Leftarrow}$ The general inequalities $(*)$ gives

$$\theta(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \min_{\boldsymbol{x} \in X} L(\boldsymbol{x}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \le L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \le f(\bar{\boldsymbol{x}}).$$

By assumption, $f(\bar{\boldsymbol{x}}) = \theta(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$, so we have equality everywhere here. This implies partly that the right saddle-point inequality holds and partly that $\bar{\boldsymbol{u}}^{\mathsf{T}} \boldsymbol{g}(\bar{\boldsymbol{x}}) = 0$, i.e., (CS) holds. Since by assumption $\bar{\boldsymbol{x}}$ is feasible, Lemma C gives the left inequality of the saddle-point definition. $\square$

Usage

If (P) is difficult to solve, (D) may be easier; it has only the constraint $\boldsymbol{u} \geq \mathbf{0}$. Suppose $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ is the solution of (D) and that one finds a feasible $\bar{\boldsymbol{x}}$ for (P) that satisfies $f(\bar{\boldsymbol{x}}) = \theta(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$. Then Theorem 3 states that $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ is a saddle point and Theorem 1 that $\bar{\boldsymbol{x}}$ solves (P).

Otherwise one can use the following theorem.

THEOREM 4. *If (CQ) holds at a solution $\bar{\boldsymbol{x}}$ of (CP), then there is no duality gap between (CP) and its dual problem.*

PROOF: By the KKT-theorem, the point $(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ satisfies (KKT). Hence, (CS) holds and $\bar{\boldsymbol{x}}$ is feasible, so Lemma C gives the left inequality of the saddle-point property. (KKT) gives that $\bar{\boldsymbol{x}}$ is a stationary point of $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$; hence, a global minimizer of (CP). This gives the right inequality of the saddle-point property. Theorem 3 states that saddle point implies no duality gap. $\square$

This theorem is particularly useful for linear and quadratic programming problems where f is convex and $\boldsymbol{g}$ and $\boldsymbol{h}$ are affine functions, because then (CQ) is satisfied (if redundant constraints are removed).

Ex. Quadratic programming

(P) minimize $f(x) = \frac{1}{2} x^T H x + c^T x$
subject to $Ax = b$

where H pos. def. and A $m \times n$, rank $A = m < n$.
(CQ) is satisfied and (P) is CP.
Thm 3 gives no duality gap, if $\bar{x}$ is feasible.
We investigate the dual problem:

$$\begin{aligned} \mathcal{L}(x; v) &= f(x) + v^T(Ax - b) \\ &= \frac{1}{2} x^T H x + (c^T + (A^T v)^T) x - v^T b \\ &= -v^T b + \frac{1}{2} x^T H x + (c + A^T v)^T x \end{aligned}$$

The minimum of $\mathcal{L}(\cdot; v)$ occurs when

$$\nabla_x \mathcal{L}(x^*; v) = 0 \iff H x^* + c + A^T v = 0 \iff x^* = -H^{-1}(c + A^T v)$$

$$\Theta(v) = \min_{x \in \mathbb{R}^n} \mathcal{L}(x; v) = \mathcal{L}(x^*; v) =$$

$$= \frac{1}{2}(c^T + v^T A)\underbrace{H^{-1} H}_{I} H^{-1}(c + A^T v) - (c^T + A v^T) H^{-1}(c + A^T v) - v^T b$$

$$= -\frac{1}{2}(c^T + v^T A) H^{-1}(c + A^T v) - v^T b$$

$$= -\frac{1}{2} v^T \underbrace{A H^{-1} A^T}_{K} v - c^T H^{-1} A^T v - v^T b - \frac{1}{2} c^T H^{-1} c$$

$$= -\frac{1}{2} v^T K v - (A H^{-1} c + b)^T v - \frac{1}{2} c^T H^{-1} c = \Theta(v)$$

K is pos. def. because $z^T K z = z^T A H^{-1} A^T z$

$$= (A^T z)^T H^{-1} (\underbrace{A^T z}_{y}) = y \underset{\substack{\uparrow \\ \text{pos. def.}}}{H^{-1}} y > 0 \quad \forall\, y \neq 0 \iff$$

$$\forall\, A^T z \neq 0 \iff z \neq 0$$

(D) $\quad \underset{v \in \mathbb{R}^n}{\text{maximize}} \ \Theta(v) \qquad$ unconstrained

Since $\Theta(v)$ is strictly concave its unique solution $\bar{v}$ is given by

$$\nabla \Theta(\bar{v}) = 0 \iff -K\bar{v} - (A H^{-1} c + b) = 0 \iff$$

$$\bar{v} = K^{-1}(A H^{-1} c + b)$$

The duality theory gives $\Theta(\bar{v}) = \mathcal{L}(\bar{x}; \bar{v}) = f(\bar{x})$
where $\bar{x} = x^*(\bar{v}) = -H^{-1}\left(c - A^T K^{-1}(A H^{-1} c + b)\right)$
is the unique solution of (P), since $\bar{x}$ is feasible:

$$A\bar{x} = \ldots = b$$