8. Saddle points and duality

Idea: Squeeze more information out of the Lagrange function and define a dual optimization problem, which can help in finding the solution to the original one.

Let $X \subseteq \mathbb{R}^n$ and assume that $f, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{C}^1(X)$ if X is open, otherwise $\mathcal{C}^1$ in a neighbourhood of X. The **primal problem** is

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in S = \{\boldsymbol{x} \in X;\ \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0},\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}. \end{array}$$

We define the *Lagrange function*

$$L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) := f(\boldsymbol{x}) + \boldsymbol{u}^{\mathsf{T}}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{v}^{\mathsf{T}}\boldsymbol{h}(\boldsymbol{x}), \quad \boldsymbol{x} \in X, \quad (\boldsymbol{u}, \boldsymbol{v}) \in U := \{(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u} \geq \boldsymbol{0}\},$$

and use the terminology:

$$\begin{array}{lll} \bar{\boldsymbol{x}} \text{ is } \textit{feasible} & \iff & \bar{\boldsymbol{x}} \in S, \\ \bar{\boldsymbol{x}} \textit{ solves } \text{(P)} & \iff & \bar{\boldsymbol{x}} \text{ is a } \textit{global solution} \text{ of (P)} \quad \iff \quad f(\bar{\boldsymbol{x}}) \leq f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in S. \end{array}$$

The *Karush-Kuhn-Tucker* conditions are

$$\text{(KKT)} \qquad \begin{cases} \nabla_x L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \boldsymbol{0}, \\ \bar{\boldsymbol{u}}^{\mathsf{T}}\boldsymbol{g}(\bar{\boldsymbol{x}}) = 0, \quad \text{(complementary slackness; CS)} \\ (\bar{\boldsymbol{u}} \geq \boldsymbol{0}). \end{cases}$$

A **convex problem** is defined as

(CP) (P) with X convex set and the functions $f, \boldsymbol{g}$ convex and $\boldsymbol{h}$ affine.

LEMMA A. *$\boldsymbol{x}$ is feasible* $\implies$ $L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) \leq f(\boldsymbol{x})$.

PROOF: $L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \underbrace{\boldsymbol{u}}_{\geq 0}{}^{\mathsf{T}}\underbrace{\boldsymbol{g}(\boldsymbol{x})}_{\leq 0} + \boldsymbol{v}^{\mathsf{T}}\underbrace{\boldsymbol{h}(\boldsymbol{x})}_{=0} \leq f(\boldsymbol{x}) \quad \forall (\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) \in S \times U.$

THEOREM (FIRST-ORDER SUFFICIENT CONDITIONS) (Ch. 7, pp 264–265).

$$\bar{\boldsymbol{x}} \textit{ feasible for (CP) and (KKT)} \quad \implies \quad \bar{\boldsymbol{x}} \text{ solves } \textit{(CP)}$$

PROOF: Conditions (KKT) imply that $\bar{\boldsymbol{x}}$ is stationary point of the convex function $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ (positive linear combination of convex functions); hence, $\bar{\boldsymbol{x}}$ is a global minimizer of $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ and we get

$$(2) \quad f(\bar{\boldsymbol{x}}) = f(\bar{\boldsymbol{x}}) + \underbrace{\bar{\boldsymbol{u}}^{\mathsf{T}}\boldsymbol{g}(\bar{\boldsymbol{x}})}_{=0\ \text{(CS)}} + \bar{\boldsymbol{v}}^{\mathsf{T}}\underbrace{\boldsymbol{h}(\bar{\boldsymbol{x}})}_{=0} = L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \leq L(\boldsymbol{x}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \underset{\substack{\uparrow \\ \text{Lemma A}}}{\leq} f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in S.$$

$\square$

We note that (CQ) is not needed. The proof gives that ~~we~~ the convexity assumption can be replaced by the weaker blue inequality, which says "$\bar{\boldsymbol{x}}$ is a minimizer of $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$". Then we need not the condition $\nabla_x L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \boldsymbol{0}$ of (KKT), but only (CS) and that $\bar{\boldsymbol{x}}$ is feasible. The latter two conditions are important and are in fact equivalent to another inequality of the Lagrange function.

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

PROOF: See (2).

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

PROOF: $\boxed{\Rightarrow}$

$$L(\bar{\boldsymbol{x}}; \boldsymbol{u}, \boldsymbol{v}) = f(\bar{\boldsymbol{x}}) + \boldsymbol{u}^\top \underbrace{\boldsymbol{g}(\bar{\boldsymbol{x}})}_{\le 0} + \boldsymbol{v}^\top \underbrace{\boldsymbol{h}(\bar{\boldsymbol{x}})}_{=0} \le f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}).$$

$\boxed{\Leftarrow}$

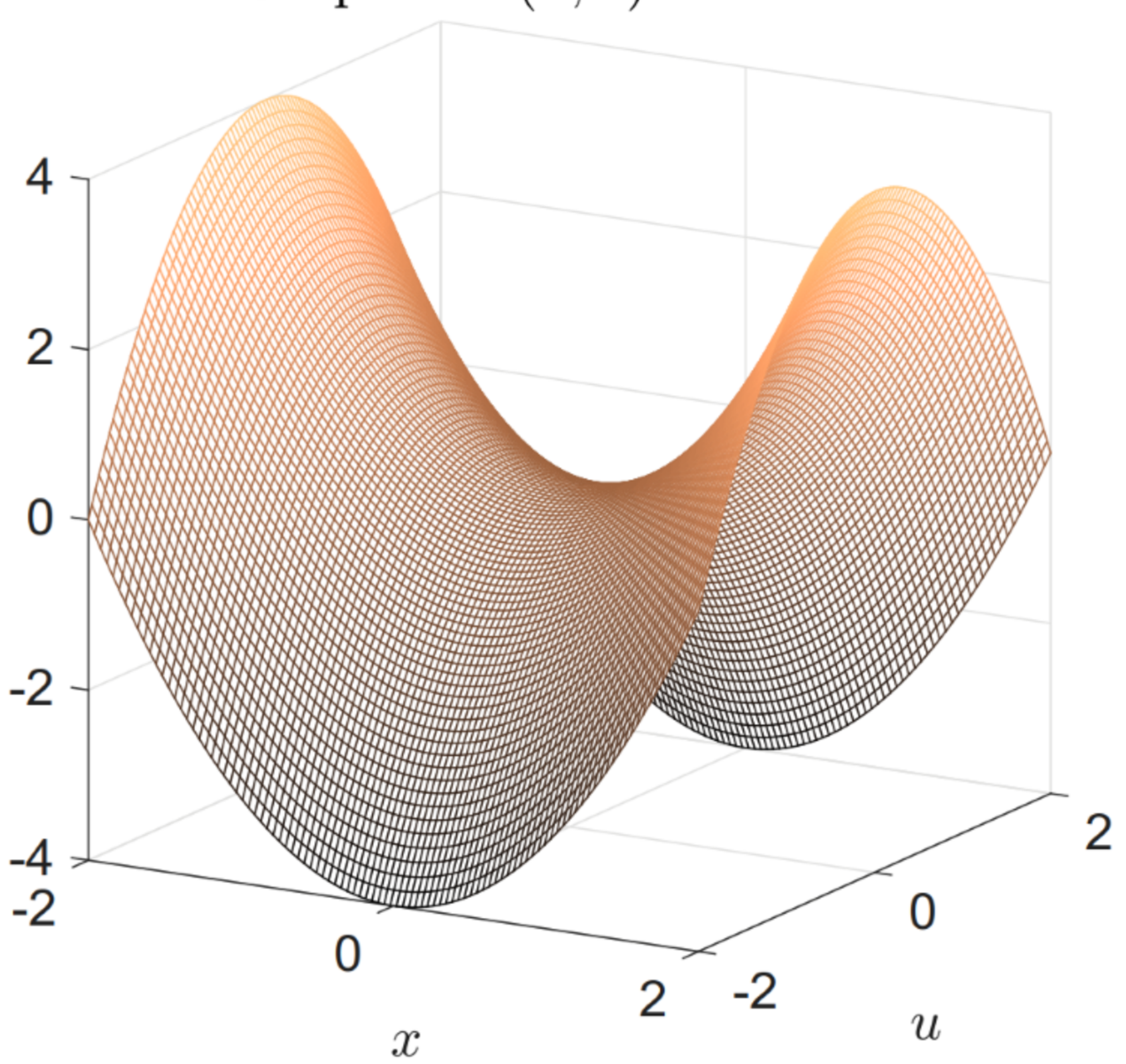
$$L(\bar{\boldsymbol{x}}; \boldsymbol{u}, \boldsymbol{v}) - L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = (\boldsymbol{u} - \bar{\boldsymbol{u}})^\top \boldsymbol{g}(\bar{\boldsymbol{x}}) + (\boldsymbol{v} - \bar{\boldsymbol{v}})^\top \boldsymbol{h}(\bar{\boldsymbol{x}}) \le 0 \quad \forall (\boldsymbol{u}, \boldsymbol{v}) \in U.$$

Firstly, we choose $\boldsymbol{u} = \bar{\boldsymbol{u}}$ and $\boldsymbol{v} = \bar{\boldsymbol{v}} + \boldsymbol{h}(\bar{\boldsymbol{x}})$ and get $\|\boldsymbol{h}(\bar{\boldsymbol{x}})\|^2 = \boldsymbol{h}(\bar{\boldsymbol{x}})^\top \boldsymbol{h}(\bar{\boldsymbol{x}}) \le 0 \Leftrightarrow \boldsymbol{h}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$.
Secondly, we choose $\boldsymbol{v} = \bar{\boldsymbol{v}}$ and $\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{e}_1 \ge 0$, where $\boldsymbol{e}_1 = (1, 0 \ldots, 0)^\top$ to obtain $g_1(\bar{\boldsymbol{x}}) \le 0$, and repeat the same for $\boldsymbol{e}_i$, $i = 2, \ldots, m$; hence, $\boldsymbol{g}(\bar{\boldsymbol{x}}) \le \boldsymbol{0}$.
Thirdly, with $\boldsymbol{v} = \bar{\boldsymbol{v}}$, the choice $\boldsymbol{u} = \boldsymbol{0}$ implies $-\bar{\boldsymbol{u}}^\top \boldsymbol{g}(\bar{\boldsymbol{x}}) \le 0$ and the choice $\boldsymbol{u} = 2\bar{\boldsymbol{u}} \ge 0$ implies $\bar{\boldsymbol{u}}^\top \boldsymbol{g}(\bar{\boldsymbol{x}}) \le 0$; hence, $\bar{\boldsymbol{u}}^\top \boldsymbol{g}(\bar{\boldsymbol{x}}) = 0$. $\square$

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

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

Graph of $L(x; u) = x^2 - u^2$



Now we can reformulate the theorem on first-order sufficient conditions without convexity.

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.*

PROOF: Lemma C gives that $\bar{\boldsymbol{x}}$ is feasible and (CS) holds. The remaining KKT condition is $\nabla_{\boldsymbol{x}} L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \mathbf{0}$, which is implied by the fact that $\bar{\boldsymbol{x}}$ is a minimizer of $L(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$. Lemmas B and A now give

$$f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \le L(\boldsymbol{x}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \le f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in S,$$

i.e., $\bar{\boldsymbol{x}}$ solves (P). $\square$

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

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

DEFINITION. *We define the function*

$$\theta(\boldsymbol{u}, \boldsymbol{v}) := \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}), \quad (\boldsymbol{u}, \boldsymbol{v}) \in U,$$

which may attain $-\infty$ if $L(\cdot; \boldsymbol{u}, \boldsymbol{v})$ is unbounded, and the (Lagrange) ***dual problem*** *to (P):*

$$\text{(D)} \qquad \begin{array}{ll} \text{maximize} & \theta(\boldsymbol{u}, \boldsymbol{v}) \\ \text{subject to} & (\boldsymbol{u}, \boldsymbol{v}) \in U. \end{array}$$

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

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

For any fixed $(\boldsymbol{u}, \boldsymbol{v})$, we can take the infimum over $\boldsymbol{x} \in S \ne \emptyset$ and get

$$\theta(\boldsymbol{u}, \boldsymbol{v}) \le \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$

Thus, any feasible point for the dual problem gives a lower bound of the primal one. If (P) is unbounded (below), then (D) has no solution. If (P) has a solution, we take the supremum of the left-hand side over $(\boldsymbol{u}, \boldsymbol{v}) \in U$ to get the duality inequality

$$\text{(DI)} \qquad \sup_{\boldsymbol{u} \ge \mathbf{0}} \theta(\boldsymbol{u}, \boldsymbol{v}) \le \inf_{\boldsymbol{x} \in S} f(\boldsymbol{x}).$$

(If we analogously start from $(*)$ and see that the supremum of $\theta(\boldsymbol{u}, \boldsymbol{v})$ is infinity, then the the primal problem has no solution.) A positive difference between the right- and left-hand sides in (DI) is called a *duality gap*. If this is zero, we have *strong duality*.

Ex. in 1D.

$$(P) \quad \begin{array}{l} \text{minimize } f(x) = 4 - x^2 \\ \text{subject to } g(x) = x^2 - 1 \le 0 \end{array}$$

Solution: $f(x) = 4 - x^2 \ge 4 - 1 = \boxed{3}$ with equality for $x = \pm 1$.

$$\mathcal{L}(x;u) = 4 - x^2 + u(x^2-1) = -u + 4 + (u-1)x^2, \quad u \ge 0$$

$$\theta(u) = \inf_{x \in \mathbb{R}} \mathcal{L}(x;u) = \begin{cases} -u+4 & \text{if } u \ge 1 \\ -\infty & \text{if } 0 \le u < 1 \end{cases}$$

$$(D) \quad \underset{u \ge 1}{\text{maximize}} \; \theta(u) = -1 + 4 = \boxed{3}$$

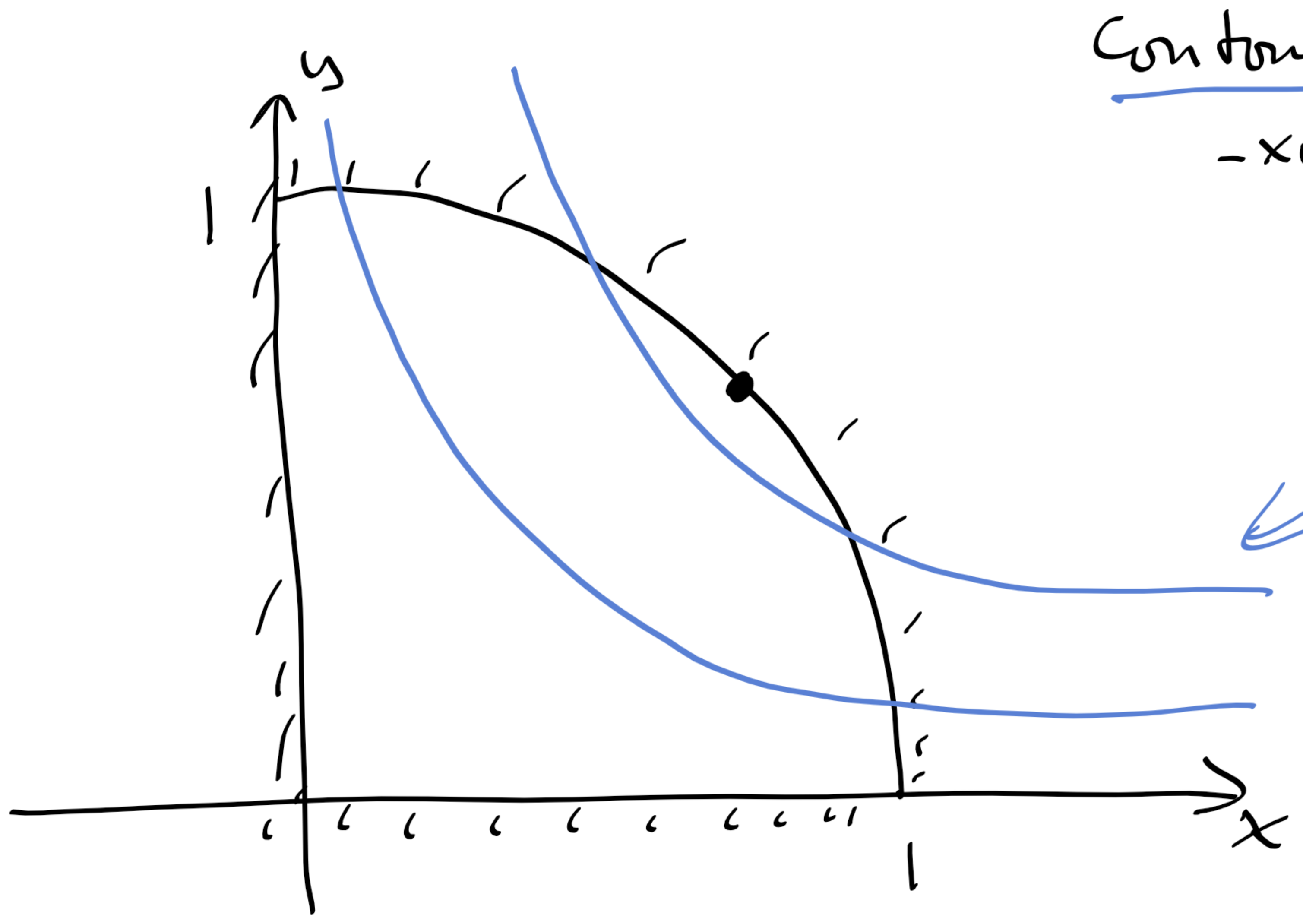
No duality gap (strong duality)

Read Ex. 6 on p. 303: similar but with duality gap

Ex. in 2D.

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x,y) = -xy \\ \text{subject to} & x^2+y^2 \le 1 \\ & x \ge 0 \\ & y \ge 0 \end{array}$$

Contours of f: $-xy = C \iff y = -\frac{c}{x}$



Compact region, so Weirstraß thm gives min. exists

Solution of (P): Set $X = \{(x,y) : x \ge 0, y \ge 0\}$

and $\mathcal{L}(x,y;u) = -xy + u(\underbrace{x^2+y^2-1}_{g(x,y)})$, $(x,y,u) \in X \times \{u \ge 0\}$

$\nabla g = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$ so (CQ) is satisfied except at $(0,0)$ where $f(0,0) = 0$

$$(KKT) \left\{ \begin{array}{l} \nabla_{(x,y)} \mathcal{L} = \begin{pmatrix} -y + u2x \\ -x + u2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ u(x^2+y^2-1) = 0 \\ u \ge 0 \end{array} \right.$$

$\underline{u=0} \Rightarrow \quad (x,y)=(0,0) \quad$ and $\quad f=0$

$$\underline{u>0} \Rightarrow \quad \begin{cases} \underbrace{\begin{pmatrix} 2u & -1 \\ -1 & 2u \end{pmatrix}}_{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ x^2+y^2=1 \end{cases}$$

Non-trivial solution iff $\det A = 0 \Leftrightarrow$
$4u^2-1=0 \Leftrightarrow u=\frac{1}{2}$. Then

$$\begin{cases} x-y=0 \\ -x+y=0 \\ x^2+y^2=1 \end{cases} \Leftrightarrow \begin{cases} x=y \\ 2x^2=1 \end{cases} \Leftrightarrow x=y=\frac{1}{\sqrt{2}}$$

$$f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \boxed{-\frac{1}{2}}$$

Dual problem:

$$\mathcal{L}(x,y;u) = -xy + u(x^2+y^2-1)$$

If $u=0$, then $\mathcal{L}=-xy$ is unbounded below.
Otherwise $u>0$:

$$\mathcal{L} = -xy + ux^2+uy^2-u = -u + u\left(x^2 - \frac{1}{u}xy + y^2\right)$$

$$= -u + u\left(\left(x-\frac{y}{2u}\right)^2 - \frac{y^2}{4u^2} + y^2\right)$$

$$= -u + u\left(x-\frac{y}{2u}\right)^2 + u\left(1-\frac{1}{4u^2}\right)y^2$$

$$\theta(u) = \begin{cases} -u & \text{if } 1-\frac{1}{4u^2} \geq 0 \Leftrightarrow u \geq \frac{1}{2} \\ -\infty & \text{if } 0 \leq u < \frac{1}{2} \end{cases}$$ (by choosing $x=\frac{y}{2u}$ and $y\to\infty$

$$\theta(u) = \inf_{(x,y)\in X} \mathcal{L}(x,y;u) = \begin{cases} -u & \text{if } u \geq \frac{1}{2} \\ -\infty & \text{if } 0 \leq u < \frac{1}{2} \end{cases}$$

(D) $\quad \underset{u\geq 0}{\text{maximize}} \ \theta(u)$

has the solution $u=\frac{1}{2}$ and $\theta(\frac{1}{2}) = \boxed{-\frac{1}{2}}$

so no duality gap