PART 1: Proof of the Separation Theorem and the Krein-Milman Theorem

Th. 3.8 [KRT, Th.2.23] (Separation Theorem)

Let $\emptyset \neq U \subset \mathbb{R}^n$ be convex and $w \notin U$. Then there exists a separating hyperplane $H = \{x \mid a^T x = \alpha\}$, $0 \neq a \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$ such that

$$a^T w \geq \alpha \geq a^T x \quad \forall x \in U$$

and $\alpha > a^T u_0$ for some $u_0 \in U$.

Proof: Case $w \notin \partial U$: ($\partial U$, boundary). Then by [FKS, Th.10.1] it holds for some $a \neq 0$, $\alpha$:

$$a^T w > \alpha \geq a^T x \quad \forall x \in U$$
Case \( w \in \partial U \): Then by [FKS, Th. 10.2] it holds for some \( a \neq 0, \alpha \):

\[
a^T w = \alpha \geq a^T x \quad \forall x \in U
\]

It remains to show: (in this case \( w \in \partial U \))

\[
(*) \quad \alpha > a^T u_0 \text{ for some } u_0 \in U.
\]

We reduce the construction to \( \text{aff } U \). Note that \( w \in \partial U \) implies \( w \in \text{aff } U \) (closed). The proof is in three steps:

1. Construction of \( u_0 \): With (fixed) \( \bar{u} \in U \) the affine hull of \( U \) is given by \( \text{aff } U = \bar{u} + V \) with

\[
V = \left\{ \sum_{j=1}^{m} \mu_j (x_j - \bar{u}), \mu_j \in \mathbb{R}, x_j \in U \right\}, \quad v_j := x_j - \bar{u} \text{ lin. indep.}
\]

So here, \( \dim V = m \).
Define $u_0 := \frac{1}{m+1} (\bar{u} + \sum_j x_j) \in U$. Then for any $k$ and any $0 < \varepsilon < \frac{1}{m+1}$ we have $u_0 \pm \varepsilon (x_k - \bar{u}) \in U$. This follows by (convex combination)

$$u_0 \pm \varepsilon (x_k - \bar{u}) = (\frac{1}{m+1} \pm \varepsilon)\bar{u} + (\frac{1}{m+1} \pm \varepsilon)x_k + \sum_{j\neq k} \frac{1}{m+1} x_j$$

Thus, there must be some $\delta > 0$ such that

$$(\star\star) \quad u_0 \pm v \in U \quad \forall v \in V, \|v\| < \delta$$

Note that $u_0$ is a point in the relative interior $U^0$ of $U$.

2. Transformation from “$\mathbb{R}^n$ to $\mathbb{R}^m$“: Consider the transformation $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $y \rightarrow x = f(y)$

$$(\star\star) \quad f(y) := \bar{u} + \sum_j y_j v_j = \bar{u} + By, \quad B := (v_1, \ldots, v_m)$$

This transformation defines a bijection between

$x \in U \leftrightarrow y \in U_y \subset \mathbb{R}^m$ with $U_y := \{y \in \mathbb{R}^m | x = f(y) \in U\}$
There is also a one to one correspondence: 
\[ w = f(y) \in \partial U \leftrightarrow y \in \partial U_y. \]
By using (\(~\star\star\~\)) one can show that \( U_y \) is convex.

Let us now consider \( y_w \in \partial U_y \) such that \( f(y_w) = w \in \partial U \)
and \( y_0 \) such that \( f(y_0) = u_0 \). By construction (see (\(~\star\star\~\)) in \( \mathbb{R}^m \)) \( y_0 \) is an interior point of \( U_y \). In \( \mathbb{R}^m \), \( y_0 \) and \( U_y \) can be separated, i.e. with \( b \neq 0, \beta \) (see [FKS, Th10.2]):

\[
    b^T y_w = \beta \geq b^T y \quad \forall y \in U_y.
\]

Using \( y_0 + \rho \, b \in U_y \) for some \( \rho > 0 \) we obtain

\[
    \beta \geq b^T (y_0 + \rho \, b) = b^T y_0 + \rho \|b\|^2 \quad \text{or} \quad \beta > b^T y_0.
\]

Then with \( a := B(B^T B)^{-1} b \in \mathbb{R}^n \) and \( \alpha := a^T u + \beta \) the relation (\( \star \)) is satisfied.  \( \square \)
Th.2.7 [Krein–Milman Theorem] ([KRT, Th.1.19])
Let $C \subset \mathbb{R}^n$ be a compact convex set. Then $C$ is the convex hull of its extreme points.

Proof By using
- [KRT, L.1.18] and Th. 3.8

Proof. By induction on $k := \dim C \ (= \dim \text{aff} \ (C))$

$k=0$: Then $C = \{c\}$, a singleton, and the result is true.
Define \( K := \text{conv} \{ x \mid x \text{ is extreme point of } C \} \subset C \). Then \( \emptyset \neq K \), convex. \underline{We show } \( K = C \) \underline{by contradiction.}

Assume that \( K \) is strictly smaller than \( C \), i.e., there exists a point \( y \in C \setminus K \). By the separation theorem, Th. 3.8, there exist \( H = \{ x \mid a^T x = \alpha \} \), \((0 \neq a)\), \( x_0 \in K \) such that

\[
(*) \quad a^T y \geq \alpha \geq a^T x \quad \forall x \in K, \quad \alpha > a^T x_0 .
\]

Now consider \( m := \max_{x \in C} a^T x \) (shift the hyperplane \( H \)) and define \( H' = \{ x \mid a^T x = m \} \),

\[
C' := C \cap H' .
\]

\( C' \) is nonempty, convex and compact. \( \text{(Note that for convex } \ C \text{ the dimension is defined by } \dim C := \dim \text{ aff } C. \)\)

We consider two cases.

\underline{Case, } \( a^T y = m \), i.e., \( y \in C' \): By

\[
a^T y \geq \alpha > a^T x_0 \text{ it follows } x_0 \in C \setminus H' \text{ and thus } \dim C' < \dim C .
\]
By applying the induction argument to $C'$ the element $y \in C'$ can be written as

$$y = \sum_{i=1}^{k} \lambda_i x^i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{k} \lambda_i = 1, \quad x^i \text{ extreme points in } C'.$$

But as we shall show below:

$$\star\star \quad \tilde{x} \text{ is extreme point of } C' \Rightarrow \tilde{x} \text{ is extreme point of } C.$$

Consequently, $y$ is a convex combination of extreme points in $C$ contradicting $y \notin K$.

Case, $a^T y < m$, i.e., $y \notin C'$: By [KRT,L 1.18], $C'$ has an extreme point $\tilde{x}$ and by $(\star\star)$ $\tilde{x} \in K$. But (see $(\star)$)

$$m = a^T \tilde{x} > a^T y \geq a^T x \quad \forall x \in K$$

yields a contradiction.

We finally show $(\star\star)$: Suppose, $\tilde{x} \in C'$ is not an extreme point of $C$. 
Then $\tilde{x}$ can be written as $\tilde{x} = \lambda x^1 + (1 - \lambda) x^2$, $0 < \lambda < 1$, $x^1 \neq x^2 \in C$. We find (using $a^T x^i \leq m$)

$$m = a^T \tilde{x} = \lambda a^T x^1 + (1 - \lambda) a^T x^2 \leq m$$

and then $a^T x^1 = a^T x^2 = m$, so that $x^1, x^2 \in C'$. Consequently $\tilde{x}$ cannot be an extreme point of $C'$.
Proofs of main results in Section 4.2: Saddle point theory

L.4.7 see [KRT;L.2.28,Cor.2.35]

A saddlepoint \((\bar{x}, \bar{y})\) of \(L(x, y)\) satisfies the strong duality relation

\[
\sup_{y \geq 0} \inf_{x \in C} L(x, y) = L(\bar{x}, \bar{y}) = \inf_{x \in C} \sup_{y \geq 0} L(x, y) .
\]

Moreover, if \((\bar{x}, \bar{y})\) is a saddle point of \(L\) then \(\bar{x}\) is a minimizer of \((CO)\) and \(\bar{y}\) is a maximizer of \((D)\).

**PROOF:** The saddle point relations for \((\bar{x}, \bar{y})\) yield:

\[
\inf_{x \in C} \sup_{y \geq 0} L(x, y) \leq \sup_{y \geq 0} L(\bar{x}, \bar{y}) = L(\bar{x}, \bar{y}) = \inf_{x \in C} L(x, \bar{y}) = \psi(\bar{y})
\]

\[
\leq \sup_{y \geq 0} \inf_{x \in C} L(x, y) \leq \inf_{x \in C} \sup_{y \geq 0} L(x, y) = v(Co)
\]

(last inequality by weak duality). So equality must hold everywhere.
In particular \( \psi(\bar{y}) = v(CO) \) and (by weak duality) \( \bar{y} \) is a solution of (D).

We now show: \((\bar{x}, \bar{y})\) saddle point \(\Rightarrow\) \(\bar{x}\) is a solution of (CO)

The saddle point relations read: for all \(x \in C, y \geq 0\) we have

\[
    f(\bar{x}) + \sum_j y_j g_j(\bar{x}) \leq f(\bar{x}) + \sum_j \bar{y}_j g_j(\bar{x}) \leq f(x) + \sum_j \bar{y}_j g_j(x) \quad (\star)
\]

Letting in the left hand sum \(y_j \rightarrow \infty\) yields \(g_j(\bar{x}) \leq 0\). So \(\bar{x} \in F\).

Choosing \(y = 0\), then \((\star)\) gives: for any \(x \in F\)

\[
    f(\bar{x}) \leq f(x) + \sum_j \bar{y}_j g_j(x) \leq f(x)
\]

So \(\bar{x}\) is a minimizer of (CO).

\(\diamond\)
Th.4.8  [KRT,Th.2.30]

Let the problem (CO) satisfy the Slater condition. Then $\bar{x}$ is a minimizer of (CO) if and only if there exists $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a saddle point of the Lagrange function $L$.

PROOF: ”$\iff$“ has been shown in Lem.4.7. (no Slater condition)
 ”$\Rightarrow$“ (similar to the proof of Strong Duality) Let $\bar{x}$ be a minimizer of (CO). Then the system

\[ (1) \quad f(x) < f(\bar{x}), \quad g_j(x) \leq 0 \quad \forall j, \quad x \in C \]

does not have a solution. By Farkas’s Lemma there exists $\bar{y} \in \mathbb{R}^m$ with

\[ \bar{y} \geq 0, \quad (L(x, \bar{y}) =) \quad f(x) + \sum_j \bar{y}_j \cdot g_j(x) \geq f(\bar{x}) \quad \forall x \in C . \]

Altogether using $\bar{x} \in F$ we find for all $x \in C, y \geq 0$:

\[ L(\bar{x}, y) = f(\bar{x}) + \sum_j y_j g_j(\bar{x}) \leq f(\bar{x}) \leq L(x, \bar{y}) . \]

Taking $x = \bar{x}$ to the right and $y = \bar{y}$ to the left gives $f(\bar{x}) = L(\bar{x}, \bar{y})$ and $(\bar{x}, \bar{y})$ is a saddle point. $\Diamond$
Proofs of main results in Section 4.3: Wolfe duality

**Th.4.11  [KRT, Th.3.9]  (weak duality)**

If $\hat{x}$ is feasible for (CO) and $(\bar{x}, \bar{y})$ is a feasible point of (WD) then

$$L(\bar{x}, \bar{y}) \leq f(\hat{x}).$$

If $L(\bar{x}, \bar{y}) = f(\hat{x})$ holds then $(\bar{x}, \bar{y})$ is a solution of (WD) and $\hat{x}$ is a minimizer of (CO)

Proof. Let $(\bar{x}, \bar{y})$ be feasible for (WD). Then $\nabla_x L(\bar{x}, \bar{y})$ holds and (since $L(x, \bar{y})$ is convex on $\mathbb{R}^n$) $\bar{x}$ is (global) minimizer with:

$$L(\bar{x}, \bar{y}) = \min_{x \in \mathbb{R}^n} L(x, \bar{y}) = \psi(\bar{y}).$$

By weak duality (Th.4.3) we obtain

$$\psi(\bar{y}) = L(\bar{x}, \bar{y}) \leq f(\hat{x}).$$

$\diamondsuit$
Th.4.12, [KRT,Th.3.10] (partial strong duality)

Let (CO) satisfy the Slater condition. If the feasible point \( \bar{x} \in \mathcal{F} \) is a minimizer of (CO) then there exists \( \bar{y} \geq 0 \) such that \((\bar{x}, \bar{y})\) is an optimal solution of (WD) and \( L(\bar{x}, \bar{y}) = f(\bar{x}) \).

**Proof.** Let \( \bar{x} \) be a minimizer of (CO). By Cor.4.9 with some \( \bar{y} \geq 0 \) the vector \((\bar{x}, \bar{y})\) satisfies the KKT conditions

\[
\nabla_x L(\bar{x}, \bar{y}) = 0, \quad \bar{y}_j g_j(\bar{x}) = 0 \quad \forall j.
\]

So (by first condition) \((\bar{x}, \bar{y})\) is feasible for (WD) and using \( \bar{y}_j g_j(\bar{x}) = 0 \) we find,

\[
L(\bar{x}, \bar{y}) = f(\bar{x}) + \sum_j \bar{y}_j g_j(\bar{x}) = f(\bar{x})
\]

and by Th.4.11, \((\bar{x}, \bar{y})\) solves (WD). \(\diamondsuit\)